Doctoral theses at NTNU, 2023:90

Caitlin Mandeville

Applications of participatory monitoring in biodiversity science and conservation

Doctoral thesis

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of
Philosophiae Doctor
Faculty of Natural Sciences
Department of Natural History

NTNU
Norwegian University of
Science and Technology

Caitlin Mandeville

Applications of participatory monitoring in biodiversity science and conservation

Thesis for the Degree of Philosophiae Doctor

Trondheim, March 2023

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Natural History

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Natural Sciences
Department of Natural History

ISBN 978-82-326-6039-1 (printed ver.)
ISBN 978-82-326-6685-0 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2023:90

Printed by NTNU Grafisk senter

Abstract

Meeting the challenge of the growing biodiversity crisis requires high quality biodiversity knowledge integrated across scales from local to international. But despite rapid developments in the collection, integration, and mobilization of biodiversity data, the current extent of available species occurrence data is insufficient to develop conservation strategies for the majority of species worldwide. Participatory monitoring offers a way to increase the spatial, temporal, and taxonomic resolution of biodiversity data while integrating local knowledge into international conservation strategy. Recent decades have seen the rapid integration of participatory biodiversity monitoring into the mainstream of biodiversity science, though there remain challenges, including the analysis of unstructured data, integration of data across scales, and inclusion of underrepresented regions and communities. Research that seeks to characterize the current role of participatory biodiversity monitoring and the conditions that enable its contribution across varying contexts will be instrumental for guiding its continued development. The central aim of this thesis is to contribute to better understanding the current role of participatory biodiversity monitoring and to strengthening its future impact.

The thesis contains four articles. The first two address the role of participatory monitoring in protected areas. The first takes a global perspective, characterizing the contributions of participatory monitoring across a variety of protected area contexts worldwide. The second takes a local perspective, modeling participatory monitoring observations at a fine spatial scale within a small natural area with the aim of improving the utility of unstructured monitoring data in local applications. The third article directly applies participatory monitoring data in the context of conservation-relevant research, using multi-species occupancy modeling to investigate how competition may affect the range limits of willow ptarmigan and rock ptarmigan, two alpine bird species that are expected to face climate-driven habitat loss and range shifts throughout their ranges in Norway. The fourth article explores the relationship between participatory biodiversity monitoring and open data sharing, finding that participatory monitoring is paving the way in open sharing of biodiversity data but identifying several areas for potential improvement.

The results of this thesis make it clear that participatory monitoring drives a growing proportion of the world's biodiversity knowledge and highlight some of the developments that support its increasingly central role in biodiversity science and conservation. The thesis further contributes to these developments, advancing an improved understanding of the participatory monitoring observation process at a fine spatial scale and identifying opportunities to expand the sharing and integration of data from participatory monitoring. Finally, the thesis directly applies participatory monitoring data to further the understanding of climate threats faced by two Norwegian species widely considered to be sentinels of climate change. Overall, this thesis suggests great capacity for the contribution of participatory monitoring to continue to increase. With growing awareness that bending the curve of biodiversity loss will require integrated action across scales from local to international, participatory monitoring is poised to have a central role.

Contents

Acknowledgements

Foremost, I am grateful for the support of a fantastic team of supervisors. Anders Finstad and Erlend Nilsen have been encouraging, enthusiastic, and supportive from the beginning, striking a balance between trusting in my ideas and pushing me to refine them that has without doubt helped me grow as a researcher and as a person. Bob O'Hara has been a formidable leader at the helm of the Transforming Citizen Science for Biodiversity research group, and welcomed me into his statistical research group meetings with much patience for my background that lies firmly on the 'ecological' side of ecological statistics. I also appreciate the support given by Francesca Verones throughout my thesis work.

I have been fortunate to have many institutional homes across NTNU, each filled with welcoming and kind colleagues. To my communities at the University Museum, the Department of Biology, and the Centre for Biodiversity Dynamics, thank you for many conversations, coffee breaks, and fun times. I have especially appreciated being part of the Transforming Citizen Science group: special thanks to Ben, Jan, Jorge, Kwaku, Philip, and Wouter for being there through the whole PhD journey with all the fun that has entailed, from doing citizen science in Bymarka to travels beyond Norway to Iceland and Denmark.

I am further grateful to the many staff members in various roles across NTNU who keep everything running smoothly and have been quick to respond when I have needed administrative or technological help—special thanks to Renate Kvernberg (Department of Natural History Head of Office), Maja Haaker (Department of Biology contact for PhD students), and Roger Alterskjær (University Museum IT Team Leader, who has saved me from many a last-minute IT emergency).

My research was funded by the NTNU Digital Transformation Initiative. My employment and duty work were supported by the NTNU University Museum, for which I am immensely grateful.

Among the most rewarding parts of my time at NTNU was the opportunity to teach in a wide range of courses. I am grateful to everyone that I taught alongside and to the many students who contributed to making these courses collaborative, enriching experiences. In particular, I thank Bente Graae, Gunnar Austrheim, and Mike Martin for entrusting me with a great deal of teaching responsibility and providing guidance and collaboration in the teaching process. I am also thankful to María Azucena Gutiérrez González for shepherding the Experts in Teamwork Community of Practice into existence and turning it into such a rewarding community to be part of.

Many individuals have contributed in large and small ways to assembling the data and assisting with the research in this thesis. Many of these are listed in the acknowledgements associated with the individual manuscripts in this thesis, and I reiterate my thanks to them here. I wish to further recognize that, being a thesis about participatory monitoring, my work relies on the voluntary contributions of countless individual participants who have gathered and shared biodiversity data. I am grateful for their contributions.

I have had many mentors who influenced the path that led me to this PhD. First, enormous thanks go to Alyson Eberhardt, who exemplifies a high standard for how community science can be done right. I have worked my way through many professional situations by asking myself "what would Alyson do here?" and this metric has never yet steered me wrong. My MSc advisors at the

University of Wyoming, Annika Walters and Frank Rahel, were welcoming and supportive mentors throughout my first graduate research experience. And although my undergraduate education was already a decade ago, I would like to name two individuals whose relatively small role in my education had, in retrospect, an outsized influence: Gretchen Hansen mentored my first-ever research experience and, along with other researchers at the University of Wisconsin-Madison Center for Limnology, modeled a kind and collegial approach to research that instilled in me a formative impression of science as a team activity; and Kris Stepenuck graciously mentored me in my first experience with citizen science, which was the first step on a path that directly led to many of the professional experiences that followed.

Moving to a new country for a PhD has been enriching, humbling, and very fun! Thanks to all who patiently put up with my faltering attempts to speak Norwegian and made me feel very welcome. Jeg setter stor pris på det. Included in this are my lovely landlords/neighbors, Lars and Sylvia, who set us off on the right foot with a lutefisk dinner soon after our arrival to Trondheim near the winter solstice in 2018 and have been an extraordinarily welcoming presence ever since.

Can I express gratitude simply to the environment here in Norway? Bymarka feels like a second home to me in Trondheim and I appreciate all who play a role in its stewardship and protection. Norway's public transportation and excellent cycling roads have led me to countless remote corners of Norwegian nature and I have squinted through fog and rain and snow to follow many a red 'T' along DNT trails near and far. It is a joy to live in a place where *friluftsliv* is, it seems, a core cultural value. I am thankful to everyone with whom I have shared time in the outdoors and look forward to more of the same.

Massive thanks go to my family, the Petersons, as well as the family I've happily been welcomed into, the Mandevilles. First, for reacting with enthusiasm rather than alarm when we first announced the plan to move across an ocean for this PhD; and then for being a constant source of support and fun that has made the distance feel not so far at all.

Finally, the biggest of thanks to Willie, my partner in life and on this grand Norwegian adventure. The academic life has led us to call many different cities home; the adventure life means that home is often a tent stuffed in our packs or stashed on our bikes; but for me, home is wherever we go together.

List of articles

Articles included in this thesis

I. **Mandeville CP**, Nilsen EB, Herfindal I, Finstad AG. Participatory monitoring drives biodiversity knowledge in global protected areas. *Manuscript.*

II. **Mandeville CP**, Nilsen EB, Finstad AG. 2022. Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use. *Ecological Solutions and Evidence* 3(4): e12185. **https://doi.org/10.1002/2688-8319.12185**

III. **Mandeville CP**, Finstad AG, Nilsen EB. Interspecific competition impacts the occupancy and range limits of two ptarmigan species along the elevation gradient in Norway. *Manuscript.*

IV. **Mandeville CP**, Koch W, Nilsen EB, Finstad AG. 2021. Open data practices among users of primary biodiversity data. *BioScience* 71(11): 1128– 1147. **https://doi.org/10.1093/biosci/biab072**

Declaration of author contributions

I. *Participatory monitoring drives biodiversity knowledge in global protected areas.* AGF and IH led the biodiversity data acquisition and management. **CPM** led the conceptual development, analysis, and writing. All authors participated in conceptual development and revision.

II. *Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use.* **CPM** conceived of the idea, analyzed the data, and led the writing of the manuscript. EBN and AGF supported the conceptual development and writing of the manuscript. All authors contributed to the drafts and gave approval for publication.

III. *Interspecific competition impacts the occupancy and range limits of two ptarmigan species along the elevation gradient in Norway.* EBN and **CPM** jointly led the conceptual development. **CPM** led the analysis and writing. All author participated in conceptual development and revision.

IV. *Open data practices among users of primary biodiversity data.* **CPM** conducted the literature search, data collection, analyses, and wrote the first draft of the manuscript. All authors participated in conceptual development, tested a pilot data collection process, and revised the manuscript.

Funding

This thesis was funded by the NTNU Digital Transformation Initiative and by the NTNU University Museum Department of Natural History.

Preface

"Building community is a requisite foundation for building a better world."

— Dr. Ayana Elizabeth Johnson[4]

Aldo Leopold famously wrote in 1949 that "one of the penalties of an ecological education is that one lives alone in a world of wounds[5]." Threats to nature have not lessened in the seven decades since those words were written, but I think that we who take note of them are no longer—if we ever really were—alone.

My work with participatory monitoring has given me the great privilege of connecting with countless individuals within and outside of academic settings who are committed to bending the curve of biodiversity loss and building a world that lives in greater harmony with nature. I am immensely grateful to the many individuals who contribute, often quietly and with little fanfare, to the coordination of participatory monitoring around the world. And most of all I am grateful to the innumerable individuals who volunteer their time, effort, and expertise to advancing our knowledge and informing our care for the natural world as participants in voluntary monitoring. Doing research with data that were gathered by the collective effort of quite literally millions of individuals is a joy, and makes me feel hopeful for the future of biodiversity.

We may together witness a world of wounds, but it is equally true that the study of nature invites us ever deeper into a world of wonder, awe, and appreciation. The communities that we build through these shared experiences are no less wonderful. It has been my great pleasure to work and learn alongside so many inspiring individuals during and beyond my PhD.

Caitlin Mandeville
Trondheim, December 2022

1 Introduction

1.1 The growing biodiversity crisis

We face a growing biodiversity crisis. Human-caused drivers of global change, including climate change, habitat loss and fragmentation, overexploitation of natural resources, the spread of invasive species, and environmental pollution, threaten the function and stability of the world's ecosystems[6–8]. As a result, an estimated one million species face extinction within decades[6,9]. Furthermore, the abundance and diversity of species are in decline worldwide; it is estimated that the populations of many monitored animal species have declined substantially since 1970[9,10], with ripple effects including reduced diversity among ecological communities and alteration of biotic interactions[6,11].

This crisis threatens humanity along with the rest of nature. Biodiversity is foundational to nature's contributions to people, from direct provisioning of food and other resources to regulation of natural processes and cycles to carbon sequestration and other nature-based forms of resilience to global change threats[10,12–14]. Further, biodiversity is inextricably linked to the profound cultural connections to nature that have been central to human society for millennia[6,15,16].

Despite these dire threats, it is still possible for humanity to 'bend the curve' of biodiversity loss and set a new course of sustainable conservation, management, and use of nature[6,17]. Reversing the current biodiversity crisis will require transformational change that cuts across all sectors of society at scales from international to local[18]. International targets and agreements establish shared priorities and align the world in pursuit of evidence-based conservation goals[19]. However, without specific, measurable, and science-based indicators, targets are unlikely to be met[20]. In the recently completed negotiations to define the post-2020 Global Biodiversity Framework, quantifiable indicators were a key component of discussions[20–22]. The completed Framework entrusts parties to the Convention on Biological Diversity with responsibility for monitoring and reporting progress on a large number of quantitative indicators measuring the conservation and sustainable use of biodiversity[23].

1.1.1 Challenges in biodiversity conservation: biodiversity data

Strategies for achieving biodiversity conservation targets are planned and evaluated using indicators that represent specific dimensions of biodiversity status and trends[24,25]. This requires high-resolution, long-term, multidimensional biodiversity data[25–27]. In some ways, biodiversity science has access to data at a scale never seen before. Technological advances in areas ranging from remote sensing to genomic analysis have vastly expanded the types of data available to monitor biodiversity; digital infrastructure supports data storage, management, and sharing at large scales; and the collection and management of data is increasingly being incentivized, as seen in the growing publication of data papers and public datasets[26,28,29]. A recent call even proposed establishing UNESCO World Heritage status for long-term datasets that are particularly integral to documenting changes in the environment[30]. In short, the past decades have seen the transformation of biodiversity research into a big data science[31].

Nevertheless, the current extent of available species occurrence data is insufficient to assess the conservation status of and develop conservation strategies for the majority of species worldwide[25,32,33]. For instance, fewer than five percent of named species have been assessed for the International Union for Conservation of Nature (IUCN) Red List of Threatened Species (hereafter, 'Red List'), and of these, nearly 15% are classified as data deficient[25,34]. Differences in data availability between species are driven by steep taxonomic biases. A small number of taxa are heavily overrepresented while the majority are neglected[25,35,36]. The biodiversity evidence base is most limited in the Global South, where biodiversity is highest and conservation stakes often greatest[32,36–38]. In general, the collection of data to monitor species' status and trends is limited by the cost-, time-, and labor-intensive nature of such monitoring[26,32,39]. Furthermore, there are direct trade-offs between resources expended on the monitoring of biodiversity and resources available to implement conservation actions[40,41]. Therefore, finding new ways to fill gaps in the biodiversity evidence base and leveraging existing data to the greatest extent possible are major priorities for making progress towards targets for the conservation and sustainable use of biodiversity[42,43].

1.1.2 Challenges in biodiversity conservation: integrating conservation across scales

Progress towards biodiversity targets relies on integrated action across scales from international to local[44,45]. Because biodiversity threats are caused by multiple socio-ecological drivers that interact across spatial scales, engagement with the local context is critical for implementing effective science-based conservation strategies[6,46–51]. In light of this, there are growing calls for localization of conservation efforts, reliance on local leadership, and furtherance of community engagement in biodiversity conservation[48,52,53]. A core component of this is the growing awareness that the majority of lands with high conservation value have long been actively managed by Indigenous peoples and local communities[54,55]. Partnerships at a local level are therefore essential for successful implementation of international targets[56,57].

A further challenge lies in ensuring that conservation actions support the long-term needs of people around the world. Many parts of the world experiencing the most severe global change threats are home to Indigenous peoples and marginalized communities[58–60]. Past conservation actions initiated by the global community in regions facing severe global change threats have all too often been detrimental to the people living in these regions[56,60,61]. As the international community works together to enact the post-2020 Global Biodiversity Framework, there is a responsibility to ensure that the needs of marginalized communities are prioritized and that conservation strategies seek to advance and sustain positive relationships between people and nature.

1.2 Participatory biodiversity monitoring

Participatory monitoring of biodiversity is increasingly highlighted as a way to expand data collection while engaging the public in biodiversity science and conservation[56,62–65]. Here, participatory biodiversity monitoring refers to any form of voluntary participation of members of the public, acting outside of their typical professional capacity, in the collection of biodiversity data[†]. Participatory monitoring has a long history—in fact, prior to the professionalization of science, nearly all research on the natural world was conducted by amateurs and could therefore be considered a form of participatory monitoring[66]. In the past few decades, however, it has entered an unprecedented era of popularity, relevance, and mainstream acceptance in scientific research[62,67,68].

One way to conceptualize the wide range of forms that participatory monitoring can take is by differentiating among types of participant engagement[69]. Bonney and colleagues[69] defined a spectrum ranging from contributory (where the main role played by participants is that of data collectors) to collaborative (where participants are more deeply involved in formulating research questions or analyzing and interpreting data) to co-created (where participants hold significant or sole leadership over monitoring). Co-created biodiversity monitoring often focuses on local priorities or concerns, tends to be more deeply situated in a specific place, and often has more explicit links to local governance or decision-making[70,71]. In contrast, the aims of contributory monitoring programs are more often set from the top down by researchers or monitoring program coordinators[70]. Contributory monitoring programs, often administered digitally through apps or websites (**Figure 1**), can be immense and can reach a large, geographically dispersed participant base[70,72].

Participatory biodiversity monitoring programs also vary widely in the type of data that they collect; an assessment of all indicators proposed for the post-2020 Global Biodiversity Framework found that 45% could be monitored by citizens[21]. Participatory monitoring data can be highly structured, collected by participants following a strict sampling protocol. It can also be technically nuanced, requiring specialized training, expertise, or equipment; for instance, participatory monitoring has been used to monitor physiological traits, species interactions, and measures of ecosystem function[56,65,71,73,74]. On the other hand, participatory monitoring can also be opportunistic, spontaneous, and unstructured. Consider the example of iNaturalist (**Figure 1**), a contributory platform for gathering species occurrence data. Although iNaturalist can be used in a structured way, it is also possible (and common) for a participant to spontaneously upload any number of species observations with no training or protocol—or perhaps in some cases, without even being aware of having contributed data that could be used in research. A great deal of participatory monitoring data can be considered semi-structured, existing somewhere on a spectrum between structured and unstructured; commonly, semi-structured data do not follow a set sampling protocol but do include metadata about the observation process[42,75].

In this thesis, I focus on participatory biodiversity monitoring of species occurrence data, including both structured and unstructured data.

---

† Participatory biodiversity monitoring is diverse. Here, I use it as an umbrella term for many distinct, though related, concepts, including the collection of biodiversity data through both citizen science and community-based monitoring. For discussion of differences between these approaches, I refer the reader to Danielsen et al. 2022[56], Shirk et al. 2012[70], and Conrad & Hilchey 2011[76]. Participatory monitoring also has much in common with community science, though the latter has a stronger emphasis on community leadership and centering of community priorities[77]. In this thesis, I use 'participatory monitoring' as a general umbrella term to refer to all related forms of public inclusion in biodiversity monitoring. When it is relevant to refer to a specific form of participatory monitoring, I define the term in context.

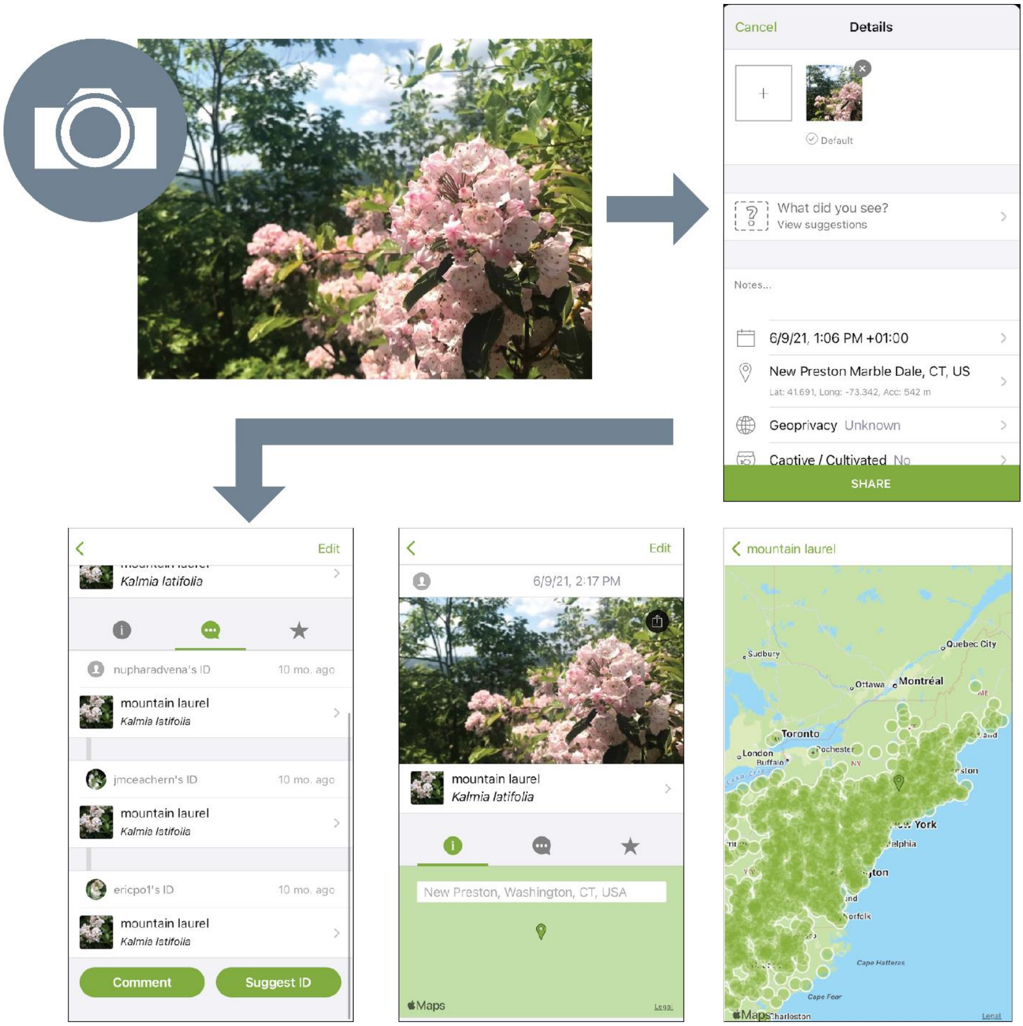


FIGURE 1. An illustration of the reporting of a species observation with the iNaturalist app. Top left: A participant observes a species they'd like to report. Top right: The participant can report a taxonomic identification (if known) and any relevant metadata. If suitable settings are enabled, the date, time, and location can be recorded automatically. Bottom row: The app allows for community interaction that allows other participants to both share additional information and learn from each observation. Bottom left: Other participants have confirmed the identification suggested by the original observer. Bottom middle: The observation can be viewed on a map. Bottom right: The observation is viewed alongside other observations of the same species.

1.2.1 The promise of participatory biodiversity monitoring

Participatory monitoring has grown tremendously in recent years. Much of this growth has been driven by technological developments that facilitate the collection, sharing, and accessibility of biodiversity data[68,78,79]. App- and website-based platforms have made it fun and easy for participants to contribute data with very little effort, and digital infrastructures for aggregating and sharing data provide a seamless pipeline from digital monitoring platforms to open databases[79]. At the same time, technological and community-based solutions for data standardization and quality control have made these data more useful than ever[71,79,80].

Co-created and community-based monitoring programs have also seen rapid growth. Increased awareness of global change threats has heightened the urgency of leveraging local biodiversity knowledge and cross-scale partnerships to inform science-based conservation strategies[44,45,48,53]. As a result, a growing number of professional societies, funding entities, and governing agencies have begun to formally recognize and call for greater engagement with the public through participatory monitoring[81–85]. And though technology does not play as central a role in co-created participatory monitoring as in digital contributory monitoring platforms, technological advances have also contributed to its growth and widespread accessibility[78].

The expansion of participatory monitoring has resulted in an incredible amount of biodiversity data. It has been estimated that over 60% of the biodiversity data shared on the world's largest biodiversity data repository derive from citizen science[86]. One of the largest contributory monitoring programs, eBird, has, on its own, contributed over one billion of these species observations[87]. These data have the potential to fuel state-of-the-art analyses and visualizations of species' status and trends at spatial and temporal resolutions that would be unimaginable without the contribution of participatory monitoring[80,88–90]. Beyond the sheer immensity of data from participatory monitoring, it can also expand the scope of biodiversity data collection to places, times, and topics that might not otherwise receive attention from non-participatory monitoring approaches[68,91].

The use of participatory monitoring data has become mainstream in biodiversity research and conservation[67,92]. The early years of its growth saw extensive discussions of the rigor and quality of participatory monitoring data, with many studies comparing data collected by volunteers with data collected by professional researchers[80,93]. For the most part, these debates have been laid to rest[62,71,80,94]. Well designed and fit-for-purpose participatory monitoring is widely accepted as an effective data collection approach, and it has been used as the basis of thousands of peer-reviewed articles as well as reports, assessments, and conservation decisions[63,67,95,96]. Its growing use is supported by recent advancements in analytical approaches for participatory monitoring data, including novel methods for analyzing unstructured data and for integrating it with structured data[97–100].

1.2.2 Beyond the data: the participants in participatory monitoring

Of course, participatory monitoring would be nothing without its participants. The impact of the data collected by participatory biodiversity monitoring is equaled, if not surpassed, by its role in engaging the public in biodiversity research and conservation (**Figure 2**). This engagement has effects both on individual participants and on society more broadly. Effects on individual participants include learning gains, enhanced science literacy, deeper connections with nature, heightened conservation attitudes and behaviors, and a sense of community with other participants[15,101–104]. In community-based monitoring, these outcomes are additionally rooted in a sense of place and situated in a local socio-ecological context[105–107].

At a societal level, participatory monitoring provides a forum for the public to engage with biodiversity research and conservation[74]. It can facilitate two-way communication and build partnerships between the public and research institutions, governing agencies, and conservation decision-makers. This can lead to increased research and management focus on public priorities as well as greater public support for science-based conservation strategies[108–110]. When monitoring is primarily led by communities, it can be part of a frontline response to local environmental challenges and injustices, serving as a mechanism for communities to leverage local knowledge and gather data in support of science-based advocacy and action[56,111].

Finally, participant and societal outcomes can strengthen the conservation impact of participatory monitoring data. Engagement with the priorities of local communities promotes a stronger fit between research and local decision-making contexts and can strengthen public support for science-based conservation actions[95,106]. As a result, inclusion of the public in biodiversity monitoring makes it more likely that monitoring data will be mobilized to inform conservation actions[112,113]. Additionally, participatory monitoring stimulates greater civic engagement around biodiversity conservation[114]. Participants report increased engagement in conservation behaviors like contribution to local decision-making boards and committees, political advocacy, and engagement with local policymakers[106,115,116].

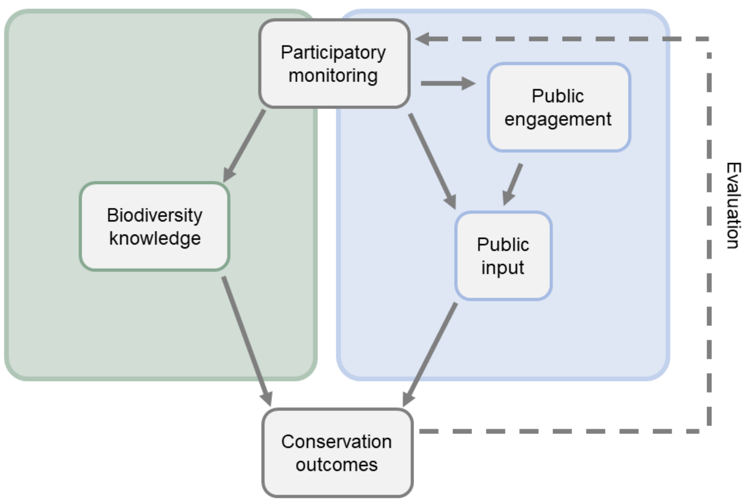


FIGURE 2. Participatory monitoring impacts biodiversity science and conservation through the direct contribution of data as well as through public engagement in the conservation process. Figure is adapted from McKinley et al. 2017[74].

1.2.3 Challenges of participatory biodiversity monitoring

Data from participatory monitoring are widely accepted as valid for research and conservation when paired with appropriate analyses[62,80]. Like all data, however, participatory monitoring data contain biases, error, and uncertainties that must be accounted for in data analysis and interpretation[97,117]. The highly heterogeneous and often unstructured nature of participatory monitoring data mean that quantifying and accounting for these analytical challenges can be substantially more difficult than for other types of data[89,118].

The major challenge in interpreting unstructured and semi-structured participatory monitoring data relates to observer behavior[118]. The same characteristic that enables unstructured participatory monitoring to amass immense datasets—the opportunistic, flexible nature of participation—also means that data analysts have minimal understanding of the observation process behind the data. This results in spatial biases that can be difficult to quantify (**Figure 3**). Species occurrence data do not strictly represent the locations where a species is found, but rather the locations where an observer has encountered, noticed, identified, and chosen to record an observation of that species. In a structured observation process, it is easier to account for the bias and uncertainty introduced by each of these hidden steps in the observation process. In unstructured sampling, however, they can result in unknown spatial biases that pose much greater challenges for inference. It is even more challenging, but critical for informing conservation strategies, to disentangle the effects of spatial bias when it may be correlated with spatial environmental drivers[119].

There is a rapidly expanding research focus on approaches for addressing the challenge of spatial bias in unstructured monitoring data. First, several novel analytical methods have been advanced that incorporate a model of the observation process into the analysis[89,98,118]. A second approach aims to shift the observation process during data collection rather than account for it during analysis, using nudges, incentives, and gamification to encourage sampling of high-value locations and taxa[120]. Both approaches are most effective when informed by the best available current understanding of the observation process. This is supported by a third line of inquiry, which investigates drivers of the participatory monitoring observation process in various contexts. This research has generally found that, at broad spatial scales, participatory monitoring oversamples areas with high population density and road accessibility, areas of perceived natural value, and areas facing perceived conservation threats[121–124]. However, variation in the observation process of participatory monitoring at finer spatial scales remains poorly understood.

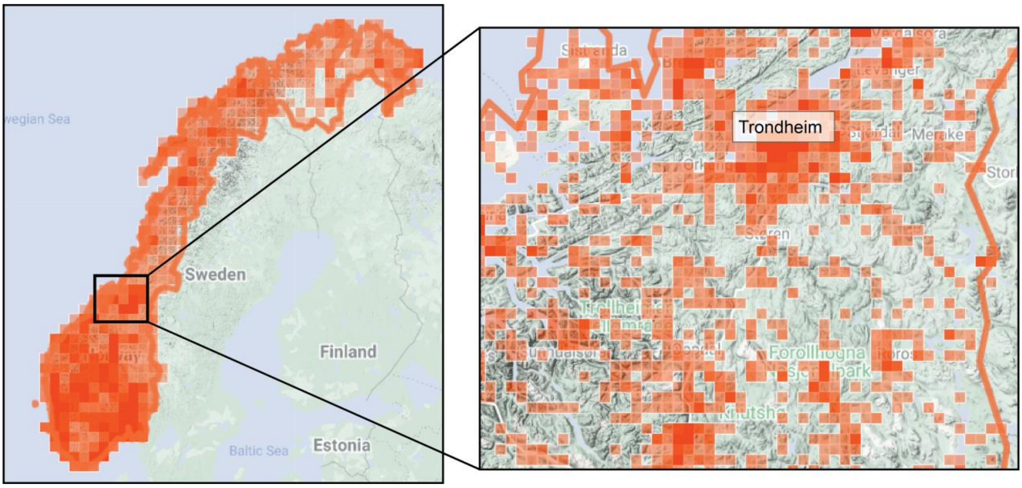


FIGURE 3. A map of all iNaturalist observations reported in Norway illustrates some of the most easily detected spatial biases that are common in unstructured biodiversity data from participatory monitoring. At left, more densely populated areas in Southern and Central Norway have many more observations than more sparsely populated areas. The inset at right shows observations in the region around Trondheim, Norway. Observations are most dense near populated areas and can be seen to follow linear features such as roads and railways. Maps and data are from iNaturalist.

In addition to spatial bias, observer behavior also introduces taxonomic bias into participatory monitoring data. Taxonomic bias is not unique to participatory monitoring; to the contrary, steep taxonomic biases are well known in global biodiversity data[25,35]. Still, participatory monitoring is expected to be characterized by greater biases towards charismatic species and those that are easily noticed and identified by casual observers[125,126]. Adjusting for this bias in analysis is complicated by the fact that the definitions of 'charismatic' and 'easily noticed and identified' are context dependent. In springtime in Central Norway, for instance, a birdwatcher will probably not report every observation of a fieldfare (*Turdus pilaris*), lest they spend all day recording bird sightings. But when a vagrant fieldfare was observed in Maine, USA, in April 2017, it drew birders from several states who collectively reported it to eBird dozens of times during its short stay in the area[127,128]. Observers' general bias towards rare and noteworthy species can pose challenges for monitoring trends in species' status across space and time. On the other hand, many monitoring programs leverage this bias very effectively to solicit data on specific species of interest, including rare, threatened, or invasive species[96,129].

Structured and community-based monitoring programs, especially those that directly target a specific research question or conservation issue, often face a very different set of challenges. Many such programs collect data for direct application in a specific decision or management context and do not publish the data externally[65,96,130]. Direct application of data can result in rapid conservation outcomes, and these programs often have a significant impact in their target area[96]. However, a lack of external publishing may mean that such programs can struggle to receive recognition from the scientific community, which may be necessary to sustain funding or to enable participation in in relevant professional communities[96,130,131]. Furthermore, data that are not shared more widely are likely to be excluded from high-level assessments and measurements of progress against biodiversity indicators[132]. The contribution of these programs could be expanded by facilitating increased data sharing through continued infrastructure development, incentivization, awareness raising, and normalization of frameworks that consider data sharing in context of data rights and data justice[57,78,95,132,133].

A further challenge, particularly for top-down contributory monitoring, is to expand the reach and relevance of participatory monitoring to better reflect the full diversity of society. Participatory monitoring is rooted in ideals of democratizing science and enabling egalitarian participation in knowledge generation[77]. Yet marginalized communities are underrepresented in many participatory monitoring programs across scales[77,134,135]. Globally, the majority of large contributory monitoring programs are headquartered in the Global North, a fact which risks exacerbating existing geographical biases in the global biodiversity evidence base[36,63,136]. Conversely, community-based monitoring has long played a strong role in generating biodiversity knowledge on a local scale in otherwise underrepresented regions, but practices for integrating this knowledge across scales to support regional and national assessments are still developing[78,113,132,137]. Research that seeks to characterize the current contributions of and enabling conditions for participatory monitoring data across varying contexts will be instrumental for guiding its continued development.

2 Aims of thesis

The overall aim of this thesis is to contribute to a better understanding of the contribution of participatory monitoring to biodiversity science and conservation, with an emphasis on generating knowledge that might illuminate pathways for strengthening the future impact of participatory monitoring data. A secondary aim is to apply participatory monitoring data, demonstrating its contribution while directly contributing to conservation research.

These aims are split into four objectives, each corresponding to one manuscript.

The first two objectives relate to the current and potential contributions of citizen science and other forms of participatory monitoring in protected areas:

I: Characterize the relative contribution of participatory monitoring in global terrestrial protected areas, focusing on the heterogeneity of participatory monitoring across protected area contexts.

II: Characterize the spatial distribution of opportunistic participatory monitoring activity relative to other visitor activity in a small, recreationally popular natural area.

The third objective demonstrates the potential of structured participatory monitoring data to address a data-hungry ecological question with conservation relevance:

III: Apply structured participatory monitoring data to investigate whether interspecific competition plays a role in setting the range limits of two closely related alpine species in Norway, willow ptarmigan (*Lagopus lagopus*) and rock ptarmigan (*Lagopus muta*).

Finally, the fourth objective relates to the open data infrastructure that supports all of the biodiversity data used in my thesis:

IV: Examine the extent to which open data, including data from participatory monitoring, underlie the unstructured biodiversity data reported in the peer-reviewed literature, and the extent to which any new open biodiversity data are shared.

3 Methodological framing

3.1 Open data

All biodiversity data used in **Articles I**, **II**, and **III** have been accessed from open data repositories, primarily the Global Biodiversity Information Facility (GBIF; https://www.gbif.org/). Therefore, the scope of this thesis is largely restricted to species occurrence data that are openly shared, though **Article IV** looks beyond this scope to characterize openly shared monitoring data in comparison with data that are unshared.

In many ways, the development of digital infrastructures that facilitate the sharing, standardization, and open access of biodiversity data has led to the participatory monitoring movement as we know it today[68,79]. An incredible number of monitoring data are 'born digital'; that is, they are collected on digital platforms with an automated pipeline often including data standardization, quality control, and aggregation[79]. Other participatory monitoring data are added directly to online data repositories in a similar manner to the digitization of historical data, museum data, and non-participatory monitoring datasets[86]. The gathering of data from diverse sources into open repositories is made possible through adherence to common standards, such as the FAIR (Findable, Accessible, Interoperable, Replicable) data principles, which further increases their usability[138].

The availability of open biodiversity data has been transformational for biodiversity science, enabling the aggregation of massive datasets to inform research that was previously unattainable[86,139]. The integration of large numbers of modern and historical datasets supports new applications, including evidence synthesis and high-resolution analysis of long-term trends[86,140]. This repurposing of data means that even small datasets can be used multiple times in various combinations[141]. Open data sharing further supports research reproducibility[142] and reduces duplication of research effort[143].

There are many synergies between participatory biodiversity monitoring and open data sharing. Both are widely considered to be components of the broader open science movement, with parallel roles in making science more open to all[2,144]. Studies suggest that many monitoring participants want their data to be made available for research and conservation and are supportive of open sharing to an extent that facilitates this use[145,146]. Nevertheless, there are still diverse barriers and limitations to the sharing of participatory monitoring data, including lack of incentives or awareness, concerns about data privacy or data custody, and technical barriers[78,147,148]. As such, a large number of participatory monitoring data remain outside of the open data infrastructure. Continued efforts to integrate open science practices into participatory monitoring will further mobilize participatory monitoring data to have greater impact in biodiversity research and conservation and will support their integration into regional and international syntheses and assessments[132,149,150,151].

3.2 Protected areas

Article I and **Article II** consider participatory monitoring in the context of terrestrial protected areas. Protected areas are central in international strategies to conserve biodiversity, but the specific aims and approaches of area-based conservation remain a focus of fierce debate[152,153]. The recently adopted post-2020 Global Biodiversity Framework formally established an international target to preserve 30% of the Earth's terrestrial, inland freshwater, and marine habitat by the year 2030[154]. Despite broad convergence on area-based conservation as a hallmark of international conservation strategy, there remain serious concerns about how to define and manage protected areas to meet the needs of the billions of people living near and relying on conservation lands[58,60]. There is also debate about the most effective indicators for evaluating the biodiversity outcomes of protected area management strategies[155–158]. Indicators will rely on high-quality biodiversity monitoring data, though such data are sparse throughout much of the world[158].

Participatory monitoring has strong potential to contribute to filling gaps in biodiversity monitoring in protected areas. Protected areas are known to be hotspots for participatory monitoring[159,160]. Drawn by both natural interest and accessibility for recreation, monitoring participants often opportunistically collect data in protected areas that support visitor access[123,159]. In other cases, strong pre-existing relationships between communities and nearby protected areas motivate participants to contribute to relevant community-based monitoring programs[105]. It is increasingly recognized that area management can only succeed if management strategies involve local knowledge and local community members[56,161,162]. Importantly, participatory monitoring provides a forum that conservation practitioners or decision-makers can use to build relationships and partner with local communities.

Although participatory monitoring is common in protected areas, it is underrepresented in the literature on protected area resilience[131]. This may be partly attributed to the heterogeneous, place-based nature of much community-based participatory monitoring; the local focus and direct applications of data mean that it can be hard to summarize general characteristics about these programs[62,96,130]. Further examination of participatory monitoring in different protected area contexts will help identify areas of high impact and possible transferability between contexts. Therefore, I use protected areas as a framework for a portion of this thesis both because they are sites where participatory monitoring offers strong potential for conservation impact and because there remain many unknowns about its current role in this area.

3.3 Interdisciplinarity

As a growing phenomenon that is situated in complex and varied socio-ecological contexts, a study of participatory monitoring benefits from the integration of diverse approaches drawn from across academic disciplines. My thesis research was envisioned from the beginning as situated in an interdisciplinary research group, and I have strived to maintain that standard through both my thesis research and additional collaborative efforts. I aim in this thesis to apply a variety of approaches to examine a broad central question from multiple angles and perspectives. In the following summary of the manuscripts that comprise this thesis, I aim to synthesize my results so that these distinct lines of inquiry converge into a broader understanding that sheds new light on the current contribution and future potential of participatory biodiversity monitoring.

Finally, I wish to highlight one additional contribution of participatory monitoring as it relates to interdisciplinarity. Beyond its own inherent value and contributions to biodiversity science and conservation, participatory monitoring is poised to play a key role as a boundary object—in other words, it brings together researchers, practitioners, and community members from varied backgrounds and disciplines who each relate to the concept through the lens of their own experience and expertise. This process of engaging with a shared concept through multiple intertwined perspectives refines and deepens communication across disciplinary boundaries. I have observed this to be true in my own research, as my own understanding of participatory monitoring has repeatedly been challenged (in the most positive sense of the term) through conversations and collaborations with colleagues. I believe that I emerged with a richer and more nuanced understanding of the many roles played by participatory monitoring and the many contexts in which it is situated.

Collaboration across academic disciplines and beyond academia, grounded in effective interdisciplinary practices[163], will be critical to address the biodiversity crisis, climate change, and other challenges facing society. Participatory monitoring sits squarely at the intersection of academic disciplines and real-world challenges. As a result, researchers, practitioners, and participants of participatory monitoring are poised to pave the way forward in strengthening partnerships between disciplines and between science and society. I hope to rise to this challenge in my own work going forward.

4 Summary of articles

4.1 Article I

Objective I: Characterize the relative contribution of participatory monitoring in global terrestrial protected areas, focusing on the heterogeneity of participatory monitoring across protected area contexts.

Participatory monitoring, including citizen science and community-based monitoring, is increasingly highlighted as a way to increase the data available to inform conservation strategies and evaluate conservation outcomes in protected areas while engaging local communities[56,105,161]. However, it risks replicating the taxonomic and spatial biases that are currently associated with biodiversity data, or introducing new ones[27,35]. Participatory monitoring varies greatly across ecological and social contexts, so effective monitoring program design cannot be defined in a one-size-fits-all approach. It is therefore critical to understand which practices in participatory monitoring are transferable across protected area contexts in order to guide its improvement, expansion, and increased application in science-based conservation strategies.

In **Article I**, we assessed the varying contribution of participatory monitoring to the biodiversity data available on the Global Biodiversity Information Facility (GBIF) for global protected areas (**Figure 4**), focusing on the extent to which patterns in participatory monitoring data differ from or replicate spatial and taxonomic patterns associated with non-participatory monitoring. We further examined a small number of characteristics of both protected areas and monitoring programs to identify protected area contexts associated with high relative contribution from participatory monitoring.

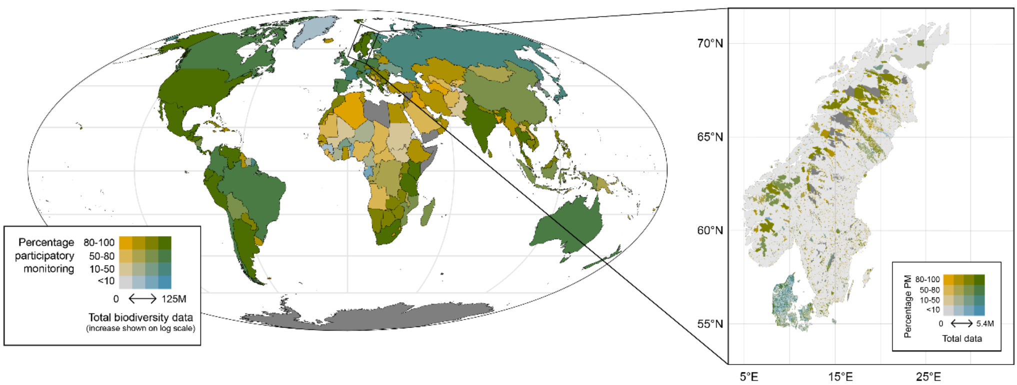


FIGURE 4. A bivariate scale illustrates the total quantity of biodiversity data available on GBIF and the ratio of this data that derives from participatory monitoring summarized at a national level. The inset map illustrates variation in these metrics among protected areas within Scandinavia.

Our findings indicate a rapidly changing landscape of biodiversity monitoring (**Figure 5**). Participatory monitoring contributed the majority of all data, and it was the sole data source for 25% of all terrestrial protected areas. Patterns in geographic, taxonomic, and threatened species coverage by participatory monitoring differed from non-participatory monitoring, suggesting its strong potential to complement other monitoring approaches. The taxonomic distribution of participatory monitoring data is commonly expected to emphasize charismatic and easily identified species[125,126,164], resulting in highly skewed coverage of species within broad taxonomic groups, but we showed that participatory monitoring in protected areas achieves a similar or less skewed distribution of observations per species than non-participatory monitoring for birds, reptiles, and amphibians. Participatory monitoring was generally less likely to record data on threatened species than non-participatory monitoring. Nevertheless, its contribution to threatened species monitoring is noteworthy: 44% of the IUCN Red List species reported through participatory monitoring were not reported through any other means.

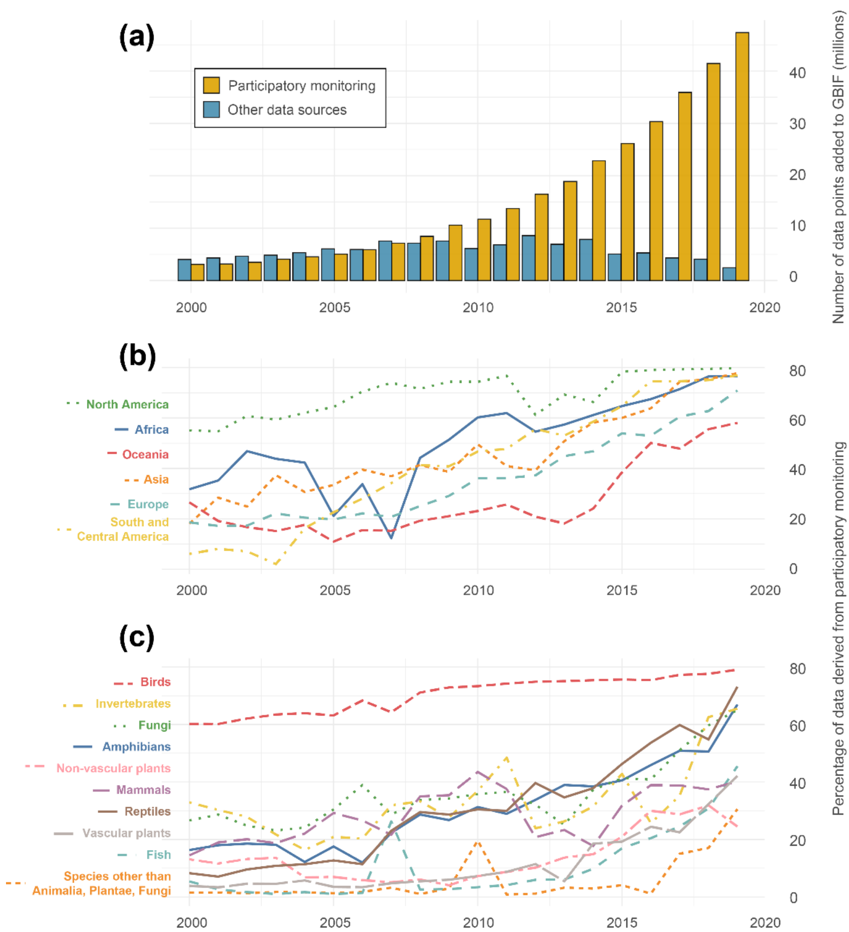


FIGURE 5. a) The amount of data contributed to GBIF annually from (yellow) participatory monitoring and (blue) other monitoring approaches for global protected areas; (b) the annual ratio of data derived from participatory monitoring across global regions; (c) the annual ratio of data derived from participatory monitoring across broad taxonomic groups.

Large areas, and those with the strictest protection criteria, were associated with the largest participatory monitoring datasets. However, small and less strictly protected areas were associated with a high reliance on participatory monitoring, which often served as the primary or only data source available for these areas on GBIF. We showed that the growing reliance on participatory monitoring may be exacerbated by a parallel decline in sharing of biodiversity data from other sources, which could indicate either a true decline in data collection or a decline or time lag in data sharing. Finally, we showed that small and taxonomically focused monitoring programs made the greatest contribution to threatened and data deficient species, and that these were most associated with large and more strictly protected areas.

Our results suggest opportunities to advance the contribution of participatory monitoring by sharing best practices at multiple scales. We showed that there are many protected areas with no data on GBIF, even among areas that share characteristics and proximity to areas with robust participatory monitoring; this underscores the heterogeneity of participatory monitoring. At the same time, this suggests that areas with robust monitoring could be used as models to extend participatory monitoring into nearby areas. Further, our results reveal substantial variation across national boundaries. Nations with low contribution from participatory monitoring can follow national-level guidance to create an enabling policy environment for participatory monitoring[3]. Finally, our results also suggest the need to continue expanding the practice of open data sharing. Among other reasons, the decline in contribution of non-participatory monitoring sources to GBIF is of concern because unstructured biodiversity data are often most useful when integrated with structured data from non-participatory sampling[42,98,100]. Like participatory monitoring, open data sharing practices can be normalized and incentivized through both local networks of protected areas and monitoring programs as well as through national-level support and infrastructure[3,78,132].

4.2 Article II

Objective II: Characterize the spatial distribution of opportunistic participatory monitoring activity relative to other visitor activity in a small, recreationally popular natural area.

Article II builds upon findings from **Article I** about the unique relationship between participatory monitoring and small protected areas. **Article I** revealed that participatory monitoring provides an exceptionally high proportion of the biodiversity data available on GBIF for small areas, providing empirical support for previous research that indicated a strong relationship between participatory monitoring and protected areas[105,131,159,165]. Protected areas and other natural areas are known to be hotspots for participatory monitoring, driven by participants' personal relationships with local natural areas and conservation motivations[159,166]. However, openly available data from opportunistic monitoring is not commonly used to inform local-scale conservation and management in small natural areas[86,130,131,167,168]. This may be partially due to a mismatch in scale between studies that investigate spatial bias in participatory monitoring data and the scale at which local conservation actions are implemented. Most studies that have modeled the participatory monitoring observation process in order to characterize spatial bias have been conducted at a regional or national scale. Analyses at this scale generally recognize small natural areas as hotspots of participatory monitoring but do not have sufficient spatial resolution to characterize the distribution of monitoring activity within these hotspots.

In **Article II**, we responded to this gap by investigating spatial bias in the distribution of participatory monitoring activity within a small natural area. We explored this question in the context of Bymarka, a popular natural area that is adjacent to the city of Trondheim and is a regional hotspot for participatory monitoring[169]. Our analysis adapted methods that have been used to investigate the distribution of participatory monitoring activity at broader spatial scales. Participatory monitoring is generally associated with two categories of landscape variables: accessibility (e.g., population density, road density, trail density) and natural interest (e.g., area protection status, conservation threats, species richness)[121–123,126,159,170,171]. Therefore, we derived ten spatial variables that we expected to represent accessibility and natural interest at a fine spatial resolution in our study area. We fit negative binomial generalized linear models in a multi-model inference approach to explore associations between these variables and participatory monitoring activity. This approach revealed that variables related to accessibility were more important in explaining the distribution of participatory monitoring activity than variables related to the perceived 'naturalness' of the area. In fact, citizen science was positively associated with developed land cover and negatively associated with land cover types that may be perceived as more 'natural', including forest and wetlands. These results differ somewhat from the many studies that have identified a strong relationship between participatory monitoring and natural interest at broader scales.

Trails have generally been found to be positively associated with citizen science, and our findings in **Article II** corroborate this. However, some studies have indicated that the relationship between trail access and citizen science activity may be more nuanced[170]. Therefore, we repeated our modeling process using general recreational trail activity, represented by activity tracking data from Strava Metro, as a response variable in place of participatory monitoring activity. This allowed us to compare patterns of trail use between monitoring participants and users of the Strava activity tracking app (**Figure 6**). We found that the two forms of trail use were not correlated (Pearson correlation test, $r = -0.01$, $p = 0.414$). Strava activity was much more strongly associated with well-established trails, while participatory monitoring activity spread more evenly over a wider range of trail conditions. Additionally, Strava activity had a positive association with 'natural' land cover types (forest and wetland land cover) and a negative relationship with developed land cover, while citizen science activity was characterized by the opposite associations.

These results suggest potential for expanded application of opportunistic participatory monitoring to the management of small natural areas. The predominant importance of a small number of variables for explaining monitoring activity indicates that it may be possible to coarsely model the observation process at a local scale, as has been done previously at broader scales. This could allow for improved use of opportunistic participatory monitoring data for local-scale inference. Small protected areas, green spaces and other multiple-use areas that contribute to other effective area-based conservation measures (OECMs) are increasingly recognized as crucial for meeting biodiversity conservation targets[172–174]. However, they often struggle with limited resources for biodiversity management and monitoring[153,175], a fact that is illustrated by the large number of small protected areas revealed by **Article I** to have no available open biodiversity data. Therefore, in lieu of structured monitoring programs, making effective use of the available data from opportunistic monitoring is essential.

Additionally, our results may suggest a way for area managers to leverage the infrastructure of opportunistic monitoring platforms to increase the structure of participatory monitoring data collected within areas of interest. Many participatory monitoring platforms offer options for customization, such as the 'Projects' feature of iNaturalist. These can be used to communicate with participants and guide the observation process to address topics of local relevance[120,168,176]. Currently, area managers may be hampered from using these tools by a lack of awareness about the current participatory monitoring activity within the areas they manage. Our results may contribute to addressing this need and informing guidance of future participatory monitoring activity. For example, managers could post signage promoting specific sampling requests in areas known to be frequented by monitoring participants, could identify areas of low monitoring activity for recruiting new participants, or could use knowledge of highly active monitoring areas to prioritize other areas for non-participatory monitoring. In these ways, area managers could use information about local opportunistic monitoring to replicate some of the benefits of structured monitoring programs.

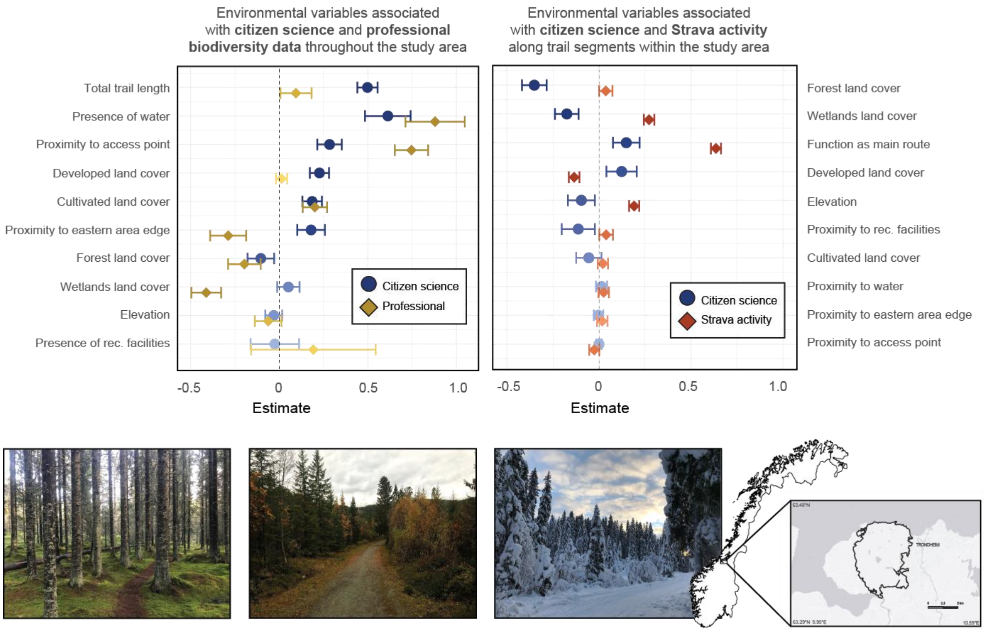


FIGURE 6. Model-averaged effect of all covariates on the response variable of (left) participatory monitoring and professional data collection, modeled in 150 x 150 m² grid cells, and (right) participatory monitoring activity and Strava activity, modeled along trail segments standardized by segment length. Decreasing color intensity indicates decreasing variable importance. All continuous variables have been centered and scaled. Photos indicate some of the varied conditions present in the natural area used as a study system in Article II. Inset map indicates the boundaries of the natural area and its position in Norway.

4.3 Article III

Objective III: Apply structured participatory monitoring data to investigate whether interspecific competition plays a role in setting the range limits of two closely related alpine species in Norway, willow ptarmigan and rock ptarmigan.

Participatory monitoring produces datasets at a size and spatial extent that would be impossible with conventional monitoring approaches, as illustrated by **Article I**. These data have opened new opportunities for analyses that require large amounts of data, including studies in macroecology, biogeography, and community dynamics[86]. Increasingly sophisticated inference is possible with even unstructured monitoring data, thanks to improvements in data quality control, statistical methods, and a growing understanding of the observation process associated with opportunistic monitoring data[89,98,118]. However, participatory monitoring programs that collect structured, complex data offer even greater potential for inference[42].

In **Article III**, we leveraged data from a structured participatory monitoring program to address a data-hungry ecological question with strong conservation implications. We applied survey data from the long-running Norwegian breeding bird monitoring program (https://tov-e.nina.no/) to model the conditional occupancy probability of two congeneric alpine bird species, willow ptarmigan (*Lagopus lagopus*) and rock ptarmigan (*Lagopus muta*). The Norwegian breeding bird monitoring surveys are coordinated annually by the Norwegian Institute for Nature Research, BirdLife Norway, and the Norwegian Environmental Agency[177]. Surveys are conducted by volunteer birdwatchers and ornithologists who are members of BirdLife Norway. Survey participants conduct annual point counts at fixed sites distributed in a grid across Norway. Each site consists of between 12 and 20 observation points that trace the outline of a 1.5 x 1.5 km^2 square. We extracted detection and non-detection data for willow ptarmigan and rock ptarmigan at all sampling sites within the species' potential range, using the spatial replicates within each site to establish a sampling history for each combination of site and year (hereafter, 'site'; n = 2560). At each site, we also derived data on two spatial variables: distance above or below the treeline[178] and elevation of the nearest treeline.

This dataset allowed us to address a little-considered aspect of ptarmigan ecology: Does interspecific competition between ptarmigan species affect their range limits along the elevation gradient? The question is of urgent conservation importance because ptarmigan, as high-latitude alpine species, are expected to shift their ranges upslope in response to changing climate conditions[179,180]. However, competition between the two species may interfere with their ability to adapt in this way[181,182]. Currently, the species share overlapping ranges during breeding season that are loosely differentiated by the treeline ecotone, with willow ptarmigan primarily occupying lower elevation sites with dense thicket or treeline vegetation and rock ptarmigan primarily occupying higher elevation habitat above treeline[183–186]. Long-standing theory suggests that asymmetric competition will favor the lower elevation species within a pair of species that share habitat along an elevation gradient, which will restrict the lower range limit of the higher elevation species[181,182,187,188]. Under this theory, upward range expansion of willow ptarmigan would be expected to reduce the space available for rock ptarmigan in what has been termed an 'escalator to extinction'[182,189,190]. However, recent studies have revealed a variety of alternative outcomes for competition along elevation gradients[181,182,191,192]. We drew upon this literature to formulate four competing hypotheses (below) for how competition might affect the range limits of willow ptarmigan and rock ptarmigan, and we used a model selection approach to test these hypotheses.

H1: Asymmetric competition favors willow ptarmigan, restricting the lower range limit of rock ptarmigan.

H2: Asymmetric competition favors rock ptarmigan, restricting the upper range limit of willow ptarmigan.

H3: Condition-dependent competition, where species' competitive dominance depends upon the environmental conditions, drives range limits such that the lower range limit of rock ptarmigan and the upper range limit of willow ptarmigan are both restricted.

H4: Competition does not affect the occupancy probability or range limits of either species.

We found that the best fitting models offered partial support for both hypothesis H2 (asymmetric competition that favors rock ptarmigan) and hypothesis H3 (condition-dependent competition) (**Figure 7**). Across most of the elevation gradient, occupancy probabilities of the two species were negatively associated. This negative association showed signs of condition-dependency: the strong positive relationship between rock ptarmigan occupancy and distance above the treeline meant that rock ptarmigan had a greater negative impact on willow ptarmigan at higher distances from the treeline, while willow ptarmigan had a greater negative impact on rock ptarmigan closer to the treeline.

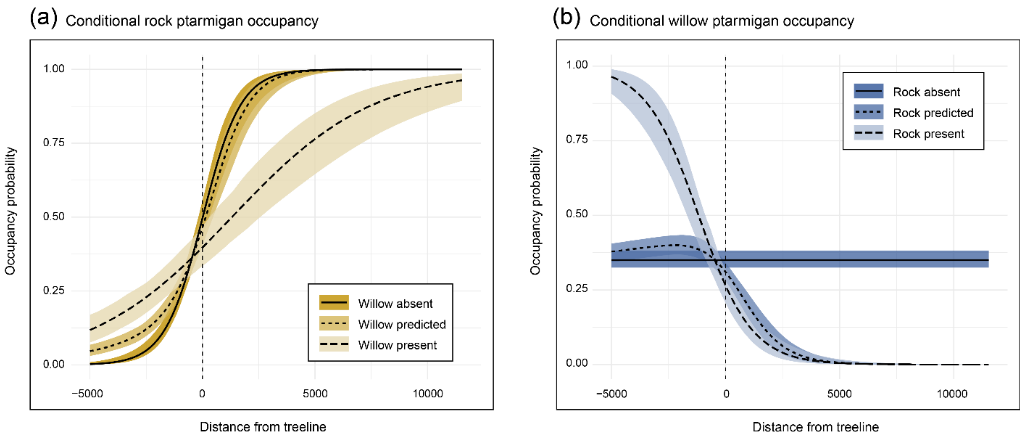


FIGURE 7. Conditional occupancy probability of (a) rock ptarmigan and (b) willow ptarmigan under three distinct modeled conditions. For both species, the solid line indicates occupancy probability given the absence of the other species; the dotted line indicates occupancy probability given the modeled occupancy of the other species; and the dashed line indicates occupancy probability given the absence of the other species. Ribbons indicate 95% confidence intervals.

However, these competitive interactions were predicted to occur entirely in the above-treeline habitat known to be preferred by rock ptarmigan and avoided by willow ptarmigan. As a result, the overall negative impact on willow ptarmigan is expected to be much greater than the negative impact on rock ptarmigan. This results in a highly asymmetric impact, where the upper range limit of willow ptarmigan is restricted but rock ptarmigan occur fairly independently of willow ptarmigan (**Figure 7**). This outcome is opposite to the traditional expectation that lower elevation species will be dominant competitors, and it casts doubt upon the relevance of the 'escalator to extinction' scenario for ptarmigan species in Norway. Rather, the competitive dominance of rock ptarmigan above the treeline may under some conditions prevent upslope range expansion of willow ptarmigan, a scenario that has been termed 'kings of the mountain'[182].

Thus, our results demonstrate that interspecific competition introduces new complexity into the question of predicting ptarmigan species' responses to climate change. This further reinforces the general importance of considering biotic interactions at both the upper and lower range limits of species' elevational distributions. The general role of biotic interactions in setting species' range limits along gradients has been debated since Darwin and is still an open question, with growing urgency as climate change reshuffles species distributions[181,182,188]. Exploration of this question has, to some extent, been limited in the past by the size of datasets required to identify signals of interspecific interactions from observational data[181]. The rise of participatory monitoring opens promising new opportunities to collect large datasets that can offer insights in this area.

4.4 Article IV

Objective IV: Examine the extent to which open data, including data from participatory monitoring, underlie the unstructured biodiversity data reported in the peer-reviewed literature, and the extent to which any new open biodiversity data are shared.

Articles I, **II**, and **III** illustrated the strong potential of participatory monitoring to contribute to biodiversity science and conservation. In all articles in this thesis, these contributions were made possible through the open sharing of participatory monitoring data. Although open sharing of biodiversity data has quickly become normalized in recent years, it is nevertheless expected that many biodiversity data remain unshared[193]. In **Article IV**, we aimed to investigate the extent to which the open data infrastructure relied upon for this thesis might be further leveraged to share additional data. We narrowed this aim into two related questions. First: Are open databases the primary sources of unstructured biodiversity data analyzed in the peer-reviewed literature, or are new unstructured biodiversity data being introduced from other sources? And second: If new unstructured biodiversity data are introduced in the literature, are they subsequently shared in an openly accessible format for further reuse?

We used a systematic approach to review a broad subset of the biodiversity literature, filtered from articles accessed using Web of Science. The criterion for inclusion in our review was simply that an article present or analyze unstructured biodiversity data; this included articles that collected structured data but reduced it to unstructured presence-only occurrence data for analysis.

Question 1: Are open databases the primary sources of unstructured biodiversity data analyzed in the peer-reviewed literature, or are new unstructured biodiversity data being introduced from other sources?

Because unstructured biodiversity data have many well-known shortcomings that create challenges for inference, we expected that most uses of unstructured data would consist of analyses that make use of openly available unstructured datasets rather than analyses of novel unstructured data. However, we found that only 19% of the articles that we reviewed relied upon exclusively open data (**Figure 8**). Others collected existing data from other sources. A large number, however, reported novel data gathered by researchers, governance agencies, and private organizations. Our results indicated that the number of articles relying on unstructured biodiversity data from all sources continues to increase annually. This suggests that the rapid innovations in analysis approaches for unstructured data, commonly associated with participatory monitoring, may also facilitate the analysis of data from a much wider range of sources. In fact, only 33% of the articles that we reviewed indicated that any of their data derived from participatory monitoring. These articles obtained participatory monitoring data from diverse direct sources, including open databases, government and independent organizations, participatory monitoring programs coordinated by the study authors.

Question 2: If new unstructured biodiversity data are introduced in the literature, are they subsequently shared in an openly accessible format for further reuse?

Having established that the peer-reviewed literature contains extensive unstructured biodiversity data from sources other than open data repositories, we next investigated whether these data go on to be shared openly after being reported in the literature. In other words: Is the literature a source or sink for potential open biodiversity data?

We found surprisingly low rates of data sharing for unstructured biodiversity data reported in the peer-reviewed literature (**Figure 8**). This was true even for articles whose authors demonstrated familiarity with open data by integrating it with data from other sources. Researchers refrain from sharing data for many reasons, including lack of time and incentives, concerns about data custody or privacy, lack of familiarity with open data infrastructure, technological challenges, and uncertainty about data ownership[133,194–196]. Our results indicate that the peer-reviewed literature is not yet a main pathway by which new open biodiversity data are generated. This suggests that the literature represents a significant new opportunity to target for further incentivization and support for open data sharing.

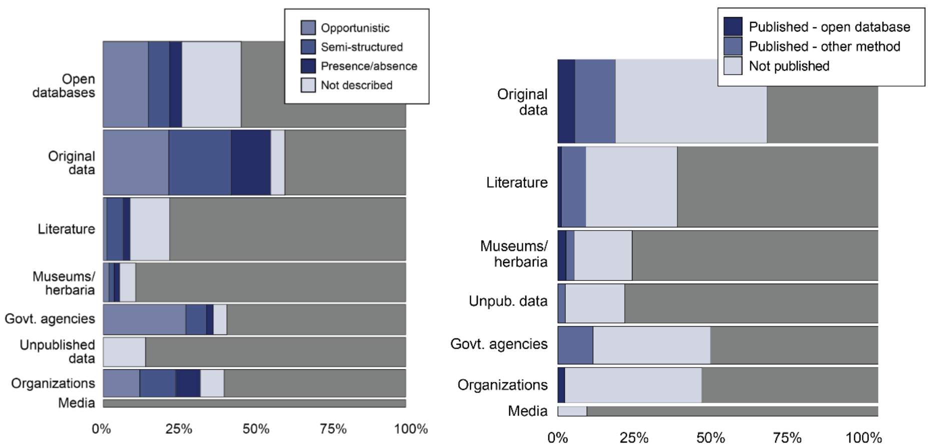


FIGURE 8. The left panel indicates the percentage of articles reviewed in this study that obtained data from each type of data source, shown according to the original structure of the data accessed from each source. The right panel indicates the percentage of articles, again shown by original data source, that subsequently shared the novel data in an open format. The bar widths indicate the number of articles within each category. The gray portions of the bars represent articles that integrated data from the indicated source with data from other sources; because of the confounding effect of data integration, trends are not reported for these articles.

Finally, it is noteworthy that the sharing rate of original data collected through participatory monitoring was somewhat higher than average, though still lower than 50%. This result corroborates the observation, mentioned in the discussion of **Articles I** and **II**, that many participatory monitoring programs apply data directly to research or management but do not share the data openly. Continuing to reduce barriers to open data sharing will make it possible to expand the already-substantial impact of such monitoring programs through opening data for reuse in new contexts, such as regional and international assessments and syntheses.

5 Discussion

In this thesis, I aimed to contribute to a better understanding of the role played by participatory monitoring in biodiversity science and conservation and to use these findings to illuminate possible pathways for strengthening the future impact of participatory monitoring data. A secondary aim was to directly apply participatory monitoring data in conservation research.

First, this thesis makes it clear that participatory monitoring drives our current knowledge of biodiversity in global protected areas. In 2021 it was estimated that participatory monitoring contributes 65% of the data on the Global Biodiversity Information Facility (GBIF)[86], but **Article I** revealed that for the past twenty years, within protected areas, its contribution has been even higher. **Article I** also revealed for the first time that participatory monitoring is the sole source of open biodiversity data for tens of thousands of protected areas. These results empirically support recent research indicating that monitoring participants are motivated by personal attachment to protected areas[105] and that monitoring activity is positively associated with protected and natural areas[159]. The results further reveal that participatory monitoring achieves more even coverage of species within some taxonomic groups than it is generally expected to[164].

Article I demonstrates the heterogeneous nature of participatory monitoring, identifying nearly one thousand distinct participatory monitoring programs that have collected data in protected areas around the world. The datasets collected by these programs range in size from less than one hundred to over one billion. The true number of active programs is surely higher, as the scope of this research excluded any programs that do not share data to GBIF, a group that includes many community-based and thematically focused monitoring programs[95,96,118]. Nevertheless, **Article I** identified a greater number of participatory monitoring programs on GBIF than similar efforts to do so in 2017[63] and 2018[197], suggesting that the number of participatory monitoring programs may either be growing or may have been historically underestimated.

One of the values of global mapping analyses is to identify priorities for further research to fill knowledge gaps[198]. **Article I** newly revealed that small and less strictly protected areas are most likely to rely on participatory monitoring as the primary source of open biodiversity data; however the literature suggests that open biodiversity data are not commonly applied in local-scale analyses, which may limit the usability of these data for area management[86,130]. **Article II** aimed to address this gap by modeling the observation process of participatory monitoring in a small natural area. Such modeling has greatly increased the usability of unstructured participatory monitoring data at broader spatial scales, so in **Article II** I adapted this approach to a finer spatial scale. This revealed that, within a small natural area that is a participatory monitoring hotspot, monitoring activity was mainly driven by accessibility and less influenced by differences in perceived 'naturalness' throughout the area. This is among the first times that these potential drivers, which both tend to be important at broader scales[122,123,159], have been evaluated within a natural area at a local scale. Small, fragmented, and multi-use areas are expected to play an important role in meeting area protection targets for the post-2020 Global Biodiversity Framework[153,199], yet **Article I** revealed that these areas tend to be deficient in conventional monitoring data. Participatory monitoring is therefore poised to fill a key gap, so efforts to make it more usable at a local scale will be critical.

Structured biodiversity data, which are often collected through locally led participatory monitoring programs, can open more opportunities for analysis. **Article III** demonstrated one such data application with a structured participatory monitoring dataset from Norway. Data from the Norwegian breeding bird monitoring program were used to address a data-hungry research question with both ecological and conservation relevance. The results from this analysis are among the first to indicate that interspecific competition may play a role in setting the elevation range limits between two ptarmigan species. This result has strong conservation implications; the range limits of ptarmigan species are expected to shift upslope in response to the projected loss of over half of their current habitat in coming decades, and such shifts can be complicated by competitive interactions[179,180]. **Article III** revealed that competition may reinforce the upper range limits of willow ptarmigan, suggesting that the species may face greater challenges in adapting to climate threats than previously expected. These results further contribute to a growing literature evaluating the general role of biotic interactions in setting range limits along gradients[181]. The analyses in **Article III** required a substantial amount of structured biodiversity data that may not have been feasible to collect without the contribution of participatory monitoring.

None of the results in this thesis would have been possible without access to open biodiversity data. **Article IV** therefore aimed to better understand the contribution of unstructured participatory monitoring data to open data repositories. The results indicate that participatory monitoring is paving the way for biodiversity data sharing, echoing the finding from **Article I** that revealed a high proportion of participatory data on GBIF. **Article IV** further revealed that data collected through participatory monitoring and reported in the peer-reviewed literature were made openly available more often than data from any other source. Nevertheless, there is room for improvement; over 50% of articles based on participatory monitoring data still did not share the data openly. Barriers keeping many participatory monitoring programs, especially those that are community-based, from sharing data are well known, and there are ongoing efforts to improve the sharing and integration of these data[118,132]. The results of **Article IV** suggest a specific category of participatory monitoring programs that may be relevant to address for increased data sharing: those that make data available for publication but not yet for open data sharing.

Article IV also showed that new unstructured biodiversity data are generated from diverse sources beyond participatory monitoring and that, in general, sharing rates for these data are low. This aligns with the finding from **Article I** that the rate at which non-participatory monitoring data are added to GBIF is declining, and suggests that ongoing efforts to normalize the open sharing of biodiversity data remain important for strengthening the impact of participatory and non-participatory data alike[141].

Despite some continued barriers, the open sharing of biodiversity data has become normalized to such an extent that it can be easy to forget that this is a relatively new development. Given the central role of GBIF in this thesis, it seems fitting to briefly take a look back and see how the central themes of this thesis have developed since GBIF was first established in 2001[200]. Its initiation was closely followed in 2002 by the launch of eBird, today the world's largest digital platform for participatory monitoring. The two grew in parallel. At first, growth was slow—ten years after its establishment, GBIF held around 300 million data points, a far cry from today's 2.3 billion[201]. By 2009, eBird had accumulated a total of 1.6 million participant checklists[202]; in contrast, more than 1 million checklists were submitted each month of the year in 2021[203]. The pace of growth for both organizations increased in the 2010s, which also saw the establishment of several new professional societies related to participatory monitoring[85]. In parallel, a new strand of literature illuminating the contributions of community-based monitoring began to gain speed[113,204].

The following decade saw a meteoric increase in both participatory monitoring and open data sharing that continues to this day. The degree to which participatory monitoring drives available knowledge about biodiversity is especially remarkable when viewed in light of its rapid growth. Its swift integration into the mainstream of biodiversity science has been facilitated by immense efforts to refine approaches at every stage of the participatory monitoring process, from data analysis to platform development to program coordination. Diverse challenges still remain, including analysis of unstructured data, integration of data across scales, and inclusion of underrepresented regions and communities. However, the rapid advancements that have brought participatory monitoring to its current central position in biodiversity research suggest great capacity for the field to continue growing in response to these challenges.

Approaches for conserving biodiversity, too, have changed since the early 2000s. In 2005, around 12% of the Earth's land was protected[205]. By 2020, this number had grown to 17%[206]. Meeting the post-2020 Global Biodiversity Framework target for 30% protection, then, will require a heightened pace of area protection[154] . Rising to this challenge in a way that is effective for conservation and meets the needs of people around the world will require a large amount of high-quality data to inform conservation actions. Participatory monitoring has already changed the landscape of biodiversity monitoring, and it is likely to have an even greater role moving forward. Recent years have seen new emergence of formal calls for the inclusion of participatory monitoring in conservation strategies, including an emphasis on local knowledge in the 2019 Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services global assessment[6], the identification of 'open engagement of societal actors' as a core pillar in the 2021 UNESCO Recommendation on Open Science[2], and a growing focus on participatory monitoring in government policy[3]. With growing awareness that bending the curve of biodiversity loss will require integrated action across scales from local to international, participatory monitoring is poised to have a central role.

References

1. Schmidt, R. *Rethinking Research Assessment: Building Blocks for Impact*. https://doi.org/10.5281/zenodo.7249187 (2022).
2. UNESCO. UNESCO Recommendation on Open Science. (2021).
3. Wehn, U. & Hepburn, L. *Guidance for the implementation of the UNESCO Open Science Recommendation re. 'Opening science to society' (FINAL)*. https://zenodo.org/record/7472827 (2022) doi:10.5281/zenodo.7472827.
4. *All We Can Save: Truth, Courage, and Solutions for the Climate Crisis*. (One World, 2020).
5. Leopold, A. *A Sand County Almanac. With Other Essays on Conservation from Round River*. (Oxford University Press, 1966).
6. IPBES. Global Assessment Report on Biodiversity and Ecosystem Services. *IPBES secretariat* https://ipbes.net/node/35274 (2019).
7. Johnson, C. N. *et al.* Biodiversity losses and conservation responses in the Anthropocene. *Science* **356**, 270–275 (2017).
8. Jaureguiberry, P. *et al.* The direct drivers of recent global anthropogenic biodiversity loss. *Sci. Adv.* **8**, eabm9982 (2022).
9. WWF. *Living Planet Report 2022 – Building a nature-positive society*. https://livingplanet.panda.org/en-US/ (2022).
10. Dirzo, R. *et al.* Defaunation in the Anthropocene. *Science* **345**, 401–406 (2014).
11. Valiente-Banuet, A. *et al.* Beyond species loss: the extinction of ecological interactions in a changing world. *Funct. Ecol.* **29**, 299–307 (2015).
12. Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
13. Hooper, D. U. *et al.* A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* **486**, 105–108 (2012).
14. Mace, G. M., Norris, K. & Fitter, A. H. Biodiversity and ecosystem services: a multilayered relationship. *Trends Ecol. Evol.* **27**, 19–26 (2012).
15. Schuttler, S. G., Sorensen, A. E., Jordan, R. C., Cooper, C. & Shwartz, A. Bridging the nature gap: can citizen science reverse the extinction of experience? *Front. Ecol. Environ.* **16**, 405–411 (2018).
16. Bridgewater, P. & Rotherham, I. D. A critical perspective on the concept of biocultural diversity and its emerging role in nature and heritage conservation. *People Nat.* **1**, 291–304 (2019).
17. Mace, G. M. *et al.* Aiming higher to bend the curve of biodiversity loss. *Nat. Sustain.* **1**, 448–451 (2018).
18. Díaz, S. *et al.* Pervasive human-driven decline of life on Earth points to the need for transformative change. *Science* **366**, eaax3100 (2019).
19. Doherty, T. S. *et al.* Expanding the Role of Targets in Conservation Policy. *Trends Ecol. Evol.* **33**, 809–812 (2018).
20. Xu, H. *et al.* Ensuring effective implementation of the post-2020 global biodiversity targets. *Nat. Ecol. Evol.* **5**, 411–418 (2021).
21. Danielsen, F. Prospects for public involvement in monitoring the post-2020 global biodiversity framework. *Nat. Portf. Rev. Prepr.* (2022) doi:10.21203/rs.3.rs-1844582/v1.

22. Convention on Biological Diversity. Monitoring framework for the Kunming-Montreal global biodiversity framework. (2022).
23. Convention on Biological Diversity. Mechanisms for planning, monitoring, reporting and review. (2022).
24. Navarro, L. M. *et al.* Monitoring biodiversity change through effective global coordination. *Curr. Opin. Environ. Sustain.* **29**, 158–169 (2017).
25. Hochkirch, A. *et al.* A strategy for the next decade to address data deficiency in neglected biodiversity. *Conserv. Biol.* **35**, 502–509 (2021).
26. Besson, M. *et al.* Towards the fully automated monitoring of ecological communities. *Ecol. Lett.* **25**, 2753–2775 (2022).
27. Bayraktarov, E. *et al.* Do Big Unstructured Biodiversity Data Mean More Knowledge? *Front. Ecol. Evol.* **6**, (2019).
28. Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z.-Q. & Bourne, P. E. Biodiversity data should be published, cited, and peer reviewed. *Trends Ecol. Evol.* **28**, 454–461 (2013).
29. Chavan, V. & Penev, L. The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics* **12**, S2 (2011).
30. Rosi, E. J. *et al.* Give long-term datasets World Heritage status. *Science* **378**, 1180–1181 (2022).
31. Farley, S. S., Dawson, A., Goring, S. J. & Williams, J. W. Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *BioScience* **68**, 563–576 (2018).
32. Schmeller, D. S. *et al.* Building capacity in biodiversity monitoring at the global scale. *Biodivers. Conserv.* **26**, 2765–2790 (2017).
33. Hortal, J. *et al.* Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. *Annu. Rev. Ecol. Evol. Syst.* **46**, 523–549 (2015).
34. Cazalis, V. *et al.* Bridging the research-implementation gap in IUCN Red List assessments. *Trends Ecol. Evol.* **37**, 359–370 (2022).
35. Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R. & Legendre, F. Taxonomic bias in biodiversity data and societal preferences. *Sci. Rep.* **7**, 9132 (2017).
36. Hughes, A. C. *et al.* Sampling biases shape our view of the natural world. *Ecography* **44**, 1259–1269 (2021).
37. Tydecks, L., Jeschke, J. M., Wolf, M., Singer, G. & Tockner, K. Spatial and topical imbalances in biodiversity research. *PLOS ONE* **13**, e0199327 (2018).
38. Oliver, R. Y., Meyer, C., Ranipeta, A., Winner, K. & Jetz, W. Global and national trends, gaps, and opportunities in documenting and monitoring species distributions. *PLOS Biol.* **19**, e3001336 (2021).
39. Stephenson, P. Technological advances in biodiversity monitoring: applicability, opportunities and challenges. *Curr. Opin. Environ. Sustain.* **45**, 36–41 (2020).
40. Buxton, R. T. *et al.* Half of resources in threatened species conservation plans are allocated to research and monitoring. *Nat. Commun.* **11**, 4668 (2020).
41. McDonald-Madden, E. *et al.* Monitoring does not always count. *Trends Ecol. Evol.* **25**, 547–550 (2010).
42. Dobson, A. D. M. *et al.* Making Messy Data Work for Conservation. *One Earth* **2**, 455–465 (2020).
43. Nichols, J. D. & Williams, B. K. Monitoring for conservation. *Trends Ecol. Evol.* **21**, 668–673 (2006).

44. Perino, A. *et al.* Biodiversity post-2020: Closing the gap between global targets and national-level implementation. *Conserv. Lett.* **15**, e12848 (2022).
45. Dreiss, L. M. & Malcom, J. W. Identifying key federal, state, and private lands strategies for achieving 30 × 30 in the United States. *Conserv. Lett.* **15**, e12849 (2022).
46. Spake, R. *et al.* An analytical framework for spatially targeted management of natural capital. *Nat. Sustain.* **2**, 90–97 (2019).
47. Darling, E. S. & Côté, I. M. Quantifying the evidence for ecological synergies. *Ecol. Lett.* **11**, 1278–1286 (2008).
48. Leadley, P. *et al.* Achieving global biodiversity goals by 2050 requires urgent and integrated actions. *One Earth* **5**, 597–603 (2022).
49. Chaplin-Kramer, R. *et al.* Mapping the planet's critical natural assets. *Nat. Ecol. Evol.* 1–11 (2022) doi:10.1038/s41559-022-01934-5.
50. Game, E. T., Meijaard, E., Sheil, D. & McDonald-Madden, E. Conservation in a Wicked Complex World; Challenges and Solutions. *Conserv. Lett.* **7**, 271–277 (2014).
51. Jansujwicz, J. S. *et al.* Localism "Reimagined": Building a Robust Localist Paradigm for Overcoming Emerging Conservation Challenges. *Environ. Manage.* **67**, 91–108 (2021).
52. Wyborn, C. & Evans, M. C. Conservation needs to break free from global priority mapping. *Nat. Ecol. Evol.* **5**, 1322–1324 (2021).
53. Bhola, N. *et al.* Perspectives on area-based conservation and its meaning for future biodiversity policy. *Conserv. Biol.* **35**, 168–178 (2021).
54. Ellis, E. C. *et al.* People have shaped most of terrestrial nature for at least 12,000 years. *Proc. Natl. Acad. Sci.* **118**, e2023483118 (2021).
55. Brondízio, E. S. *et al.* Locally Based, Regionally Manifested, and Globally Relevant: Indigenous and Local Knowledge, Values, and Practices for Nature. *Annu. Rev. Environ. Resour.* **46**, 481–509 (2021).
56. Danielsen, F. *et al.* Community Monitoring of Natural Resource Systems and the Environment. *Annu. Rev. Environ. Resour.* **47**, 637–670 (2022).
57. Reyes-García, V. *et al.* Recognizing Indigenous peoples' and local communities' rights and agency in the post-2020 Biodiversity Agenda. *Ambio* **51**, 84–92 (2022).
58. Mehrabi, Z., Ellis, E. C. & Ramankutty, N. The challenge of feeding the world while conserving half the planet. *Nat. Sustain.* **1**, 409–412 (2018).
59. Schleicher, J. *et al.* Protecting half of the planet could directly affect over one billion people. *Nat. Sustain.* **2**, 1094–1096 (2019).
60. Obura, D. O. *et al.* Integrate biodiversity targets from local to global levels. *Science* **373**, 746–748 (2021).
61. Stefanoudis, P. V. *et al.* Turning the tide of parachute science. *Curr. Biol.* **31**, R184–R185 (2021).
62. Bonney, R. Expanding the Impact of Citizen Science. *BioScience* **71**, 448–451 (2021).
63. Chandler, M. *et al.* Contribution of citizen science towards international biodiversity monitoring. *Biol. Conserv.* **213**, 280–294 (2017).
64. Pocock, M. J. O. *et al.* Chapter Six - A Vision for Global Biodiversity Monitoring With Citizen Science. in *Advances in Ecological Research* (eds. Bohan, D. A., Dumbrell, A. J., Woodward, G. & Jackson, M.) vol. 59 169–223 (Academic Press, 2018).
65. Theobald, E. J. *et al.* Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biol. Conserv.* **181**, 236–244 (2015).

66. Miller-Rushing, A., Primack, R. & Bonney, R. The history of public participation in ecological research. *Front. Ecol. Environ.* **10**, 285–290 (2012).
67. Cooper, C. B. The Invisible Prevalence of Citizen Science in Global Research: Migratory Birds and Climate Change. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0106508 (2014).
68. MacPhail, V. J. & Colla, S. R. Power of the people: A review of citizen science programs for conservation. *Biol. Conserv.* **249**, 108739 (2020).
69. Bonney, R. *et al. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. Online Submission* https://eric.ed.gov/?id=ED519688 (2009).
70. Shirk, J. *et al.* Public Participation in Scientific Research: a Framework for Deliberate Design. *Ecol. Soc.* **17**, (2012).
71. Parrish, J. K. *et al.* Exposing the Science in Citizen Science: Fitness to Purpose and Intentional Design. *Integr. Comp. Biol.* **58**, 150–160 (2018).
72. Di Cecco, G. J. *et al.* Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science. *BioScience* **71**, 1179–1188 (2021).
73. Fraisl, D. *et al.* Mapping citizen science contributions to the UN sustainable development goals. *Sustain. Sci.* **15**, 1735–1751 (2020).
74. McKinley, D. C. *et al.* Citizen science can improve conservation science, natural resource management, and environmental protection. *Biol. Conserv.* **208**, 15–28 (2017).
75. Kelling, S. *et al.* Using Semistructured Surveys to Improve Citizen Science Data for Monitoring Biodiversity. *BioScience* **69**, 170–179 (2019).
76. Conrad, C. C. & Hilchey, K. G. A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environ. Monit. Assess.* **176**, 273–291 (2011).
77. Cooper, C. B. *et al.* Inclusion in citizen science: The conundrum of rebranding. *Science* **372**, 1386–1388 (2021).
78. Johnson, N., Druckenmiller, M. L., Danielsen, F. & Pulsifer, P. L. The Use of Digital Platforms for Community-Based Monitoring. *BioScience* **71**, 452–466 (2021).
79. Kays, R., McShea, W. J. & Wikelski, M. Born-digital biodiversity data: Millions and billions. *Divers. Distrib.* **26**, 644–648 (2020).
80. Kosmala, M., Wiggins, A., Swanson, A. & Simmons, B. Assessing data quality in citizen science. *Front. Ecol. Environ.* **14**, 551–560 (2016).
81. Notermans, I. *et al.* Recommendations to mainstream citizen science in policy. (2022) doi:10.5281/zenodo.5772237.
82. Haklay, M. *Citizen Science and Policy: A European Perspective.* (2015).
83. Tyson, E. & Rejeski, D. *Strategic Recommendations for Federal Citizen Science and Crowdsourcing.* (2016).
84. Noone, K. J., Schensul, J. J. & Buchanan, P. R. Augmenting the Grammar of Science—The Community Science Exchange. *Community Sci.* **1**, e2022CSJ000012 (2022).
85. Storksdieck, M. *et al.* Associations for Citizen Science: Regional Knowledge, Global Collaboration. *Citiz. Sci. Theory Pract.* **1**, 10 (2016).
86. Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B. & Schigel, D. Data integration enables global biodiversity synthesis. *Proc. Natl. Acad. Sci.* **118**, e2018093118 (2021).
87. Auer, T. *et al.* EOD – eBird Observation Dataset. (2022) doi:10.15468/aomfnb.

88. Callaghan, C. T., Nakagawa, S. & Cornwell, W. K. Global abundance estimates for 9,700 bird species. *Proc. Natl. Acad. Sci.* **118**, e2023170118 (2021).
89. Altwegg, R. & Nichols, J. D. Occupancy models for citizen-science data. *Methods Ecol. Evol.* **10**, 8–21 (2019).
90. Taylor, S. D., Meiners, J. M., Riemer, K., Orr, M. C. & White, E. P. Comparison of large-scale citizen science data and long-term study data for phenology modeling. *Ecology* **100**, e02568 (2019).
91. Knape, J., Coulson, S. J., van der Wal, R. & Arlt, D. Temporal trends in opportunistic citizen science reports across multiple taxa. *Ambio* **51**, 183–198 (2022).
92. Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B. & Major, R. E. Improving big citizen science data: Moving beyond haphazard sampling. *PLOS Biol.* **17**, e3000357 (2019).
93. Dickinson, J. L. *et al.* The current state of citizen science as a tool for ecological research and public engagement. *Front. Ecol. Environ.* **10**, 291–297 (2012).
94. Brown, E. D. & Williams, B. K. The potential for citizen science to produce reliable and useful information in ecology. *Conserv. Biol.* **33**, 561–569 (2019).
95. Newman, G. *et al.* Leveraging the power of place in citizen science for effective conservation decision making. *Biol. Conserv.* **208**, 55–64 (2017).
96. Fontaine, A., Simard, A., Brunet, N. & Elliott, K. H. Scientific contributions of citizen science applied to rare or threatened animals. *Conserv. Biol.* **n/a**, e13976.
97. Johnston, A., Matechou, E. & Dennis, E. B. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods Ecol. Evol.* **n/a**,.
98. Rapacciuolo, G., Young, A. & Johnson, R. Deriving indicators of biodiversity change from unstructured community-contributed data. *Oikos* **130**, 1225–1239 (2021).
99. Isaac, N. J. B. *et al.* Data Integration for Large-Scale Models of Species Distributions. *Trends Ecol. Evol.* **35**, 56–67 (2020).
100. Zipkin, E. F. *et al.* Addressing data integration challenges to link ecological processes across scales. *Front. Ecol. Environ.* **19**, 30–38 (2021).
101. Larson, L. R. *et al.* The diverse motivations of citizen scientists: Does conservation emphasis grow as volunteer participation progresses? *Biol. Conserv.* **242**, 108428 (2020).
102. Bonney, R. *et al.* Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience* **59**, 977–984 (2009).
103. Sharma, N. *et al.* From citizen science to citizen action: analysing the potential for a digital platform to cultivate attachments to nature. *J. Sci. Commun.* **18**, 1–35 (2019).
104. Bonney, R., Phillips, T. B., Ballard, H. L. & Enck, J. W. Can citizen science enhance public understanding of science? *Public Underst. Sci.* **25**, 2–16 (2016).
105. Halliwell, P. *et al.* The power and potential of citizen science for park bonding, advocacy, and stewardship. *Parks Steward. Forum* **38**, (2022).
106. Haywood, B. K., Parrish, J. K. & Dolliver, J. Place-based and data-rich citizen science as a precursor for conservation action. *Conserv. Biol.* **30**, 476–486 (2016).
107. Toomey, A. H., Strehlau-Howay, L., Manzolillo, B. & Thomas, C. The place-making potential of citizen science: Creating social-ecological connections in an urbanized world. *Landsc. Urban Plan.* **200**, 103824 (2020).
108. Pandya, R. E. A framework for engaging diverse communities in citizen science in the US. *Front. Ecol. Environ.* **10**, 314–317 (2012).

109. Eicken, H. *et al.* Connecting Top-Down and Bottom-Up Approaches in Environmental Observing. *BioScience* **71**, 467–483 (2021).
110. Buschman, V. Q. Framing co-productive conservation in partnership with Arctic Indigenous peoples. *Conserv. Biol.* **36**, e13972 (2022).
111. Vance-Chalcraft, H. D. & Jelks, N. O. Community-engaged learning to broaden the impact of applied ecology: A case study. *Ecol. Appl.* **n/a**, e2768.
112. Danielsen, F., Burgess, N. D., Jensen, P. M. & Pirhofer-Walzl, K. Environmental monitoring: the scale and speed of implementation varies according to the degree of peoples involvement. *J. Appl. Ecol.* **47**, 1166–1168 (2010).
113. Danielsen, F. *et al.* The Concept, Practice, Application, and Results of Locally Based Monitoring of the Environment. *BioScience* **71**, 484–502 (2021).
114. Stepenuck, K. & Green, L. Individual- and community-level impacts of volunteer environmental monitoring: a synthesis of peer-reviewed literature. *Ecol. Soc.* **20**, (2015).
115. Stepenuck, K. F. & Genskow, K. D. Traits of Volunteer Water Monitoring Programs that Influence Natural Resource Management and Policy Impacts. *Soc. Nat. Resour.* **32**, 275–291 (2019).
116. Day, G., Fuller, R. A., Nichols, C. & Dean, A. J. Characteristics of immersive citizen science experiences that drive conservation engagement. *People Nat.* **4**, 983–995 (2022).
117. Kelling, S. *et al.* Taking a 'Big Data' approach to data quality in a citizen science project. *Ambio* **44**, 601–611 (2015).
118. Johnston, A. *et al.* Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. *Divers. Distrib.* **27**, 1265–1277 (2021).
119. Bullion, C. M. & Bahlai, C. A. Data Gap or Biodiversity Gap? Evaluating apparent spatial biases in community science observations of Odonata in the east-central United States. 2022.11.29.518107 Preprint at https://doi.org/10.1101/2022.11.29.518107 (2022).
120. Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B. & Major, R. E. Improving big citizen science data: Moving beyond haphazard sampling. *PLOS Biol.* **17**, e3000357 (2019).
121. Boakes, E. H. *et al.* Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. *Sci. Rep.* **6**, 33051 (2016).
122. Geldmann, J. *et al.* What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Divers. Distrib.* **22**, 1139–1149 (2016).
123. Tiago, P., Ceia-Hasse, A., Marques, T. A., Capinha, C. & Pereira, H. M. Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Sci. Rep.* **7**, 12832 (2017).
124. Tulloch, A. I. T., Possingham, H. P., Joseph, L. N., Szabo, J. & Martin, T. G. Realising the full potential of citizen science monitoring programs. *Biol. Conserv.* **165**, 128–138 (2013).
125. Callaghan, C. T., Poore, A. G. B., Hofmann, M., Roberts, C. J. & Pereira, H. M. Large-bodied birds are over-represented in unstructured citizen science data. *Sci. Rep.* **11**, 19073 (2021).
126. Petersen, T. K., Speed, J. D. M., Grøtan, V. & Austrheim, G. Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. *Ecol. Solut. Evid.* **2**, e12048 (2021).

127. When the Fieldfare Came to Maine. *Boothbay Register* https://www.boothbayregister.com/node/85273.
128. eBird. eBird: An online database of bird distribution and abundance [web application].
129. Howard, L., Rees, C. B. van, Dahlquist, Z., Luikart, G. & Hand, B. K. A review of invasive species reporting apps for citizen science and opportunities for innovation. *NeoBiota* **71**, 165–188 (2022).
130. Mandeville, C. P. & Finstad, A. G. Community science supports research on protected area resilience. *Conserv. Sci. Pract.* **3**, e442 (2021).
131. Binley, A. D., Proctor, C. A., Pither, R., Davis, S. A. & Bennett, J. R. The unrealized potential of community science to support research on the resilience of protected areas. *Conserv. Sci. Pract.* **3**, e376 (2021).
132. Kühl, H. S. *et al.* Effective Biodiversity Monitoring Needs a Culture of Integration. *One Earth* **3**, 462–474 (2020).
133. Pritchard, R., Sauls, L. A., Oldekop, J. A., Kiwango, W. A. & Brockington, D. Data justice and biodiversity conservation. *Conserv. Biol.* **36**, e13919 (2022).
134. Allf, B. C. *et al.* Citizen Science as an Ecosystem of Engagement: Implications for Learning and Broadening Participation. *BioScience* **72**, 651–663 (2022).
135. Pateman, R. M., Dyke, A. & West, S. E. The Diversity of Participants in Environmental Citizen Science. *Citiz. Sci. Theory Pract.* (2021).
136. Amano, T., Lamming, J. D. L. & Sutherland, W. J. Spatial Gaps in Global Biodiversity Information and the Role of Citizen Science. *BioScience* **66**, 393–400 (2016).
137. Evans, K., Guariguata, M. R. & Brancalion, P. H. S. Participatory monitoring to connect local and global priorities for forest restoration. *Conserv. Biol.* **32**, 525–534 (2018).
138. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
139. Ball-Damerow, J. E. *et al.* Research applications of primary biodiversity databases in the digital age. *PLOS ONE* **14**, e0215794 (2019).
140. Nakagawa, S. *et al.* A new ecosystem for evidence synthesis. *Nat. Ecol. Evol.* **4**, 498–501 (2020).
141. Poisot, T., Bruneau, A., Gonzalez, A., Gravel, D. & Peres-Neto, P. Ecological Data Should Not Be So Hard to Find and Reuse. *Trends Ecol. Evol.* **34**, 494–496 (2019).
142. Powers, S. M. & Hampton, S. E. Open science, reproducibility, and transparency in ecology. *Ecol. Appl.* **29**, e01822 (2019).
143. Grainger, M. J., Bolam, F. C., Stewart, G. B. & Nilsen, E. B. Evidence synthesis for tackling research waste. *Nat. Ecol. Evol.* **4**, 495–497 (2020).
144. Wagenknecht, K. *et al.* EU-Citizen.Science: A Platform for Mainstreaming Citizen Science and Open Science in Europe. *Data Intell.* **3**, 136–149 (2021).
145. Ganzevoort, W., van den Born, R. J. G., Halffman, W. & Turnhout, S. Sharing biodiversity data: citizen scientists' concerns and motivations. *Biodivers. Conserv.* **26**, 2821–2837 (2017).
146. Fox, R. *et al.* Opinions of citizen scientists on open access to UK butterfly and moth occurrence data. *Biodivers. Conserv.* **28**, 3321–3341 (2019).
147. Anhalt-Depies, C., Stenglein, J. L., Zuckerberg, B., Townsend, P. A. & Rissman, A. R. Tradeoffs and tools for data quality, privacy, transparency, and trust in citizen science. *Biol. Conserv.* **238**, 108195 (2019).

148. Groom, Q., Weatherdon, L. & Geijzendorffer, I. R. Is citizen science an open science in the case of biodiversity observations? *J. Appl. Ecol.* **54**, 612–617 (2017).
149. MacFadyen, S. *et al.* Drowning in data, thirsty for information and starved for understanding: A biodiversity information hub for cooperative environmental monitoring in South Africa. *Biol. Conserv.* **274**, 109736 (2022).
150. Wetzel, F. T. *et al.* Unlocking biodiversity data: Prioritization and filling the gaps in biodiversity observation data in Europe. *Biol. Conserv.* **221**, 78–85 (2018).
151. Wehn, U. Global Citizen Science perspectives on Open Science. (2020).
152. Watson, J. E. M., Dudley, N., Segan, D. B. & Hockings, M. The performance and potential of protected areas. *Nature* **515**, 67–73 (2014).
153. Maxwell, S. L. *et al.* Area-based conservation in the twenty-first century. *Nature* **586**, 217–227 (2020).
154. Convention on Biological Diversity. Kunming-Montreal Global biodiversity framework. (2022).
155. Rodrigues, A. S. L. & Cazalis, V. The multifaceted challenge of evaluating protected area effectiveness. *Nat. Commun.* **11**, 5147 (2020).
156. Jetz, W. *et al.* Include biodiversity representation indicators in area-based conservation targets. *Nat. Ecol. Evol.* **6**, 123–126 (2022).
157. Ghoddousi, A., Loos, J. & Kuemmerle, T. An Outcome-Oriented, Social–Ecological Framework for Assessing Protected Area Effectiveness. *BioScience* **72**, 201–212 (2022).
158. Carroll, C. & Noss, R. F. How percentage-protected targets can support positive biodiversity outcomes. *Conserv. Biol.* **36**, e13869 (2022).
159. Tulloch, A. I. T., Mustin, K., Possingham, H. P., Szabo, J. K. & Wilson, K. A. To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale. *Divers. Distrib.* **19**, 465–480 (2013).
160. Cambria, V. E. *et al.* Citizen Science Data to Measure Human Use of Green Areas and Forests in European Cities. *Forests* **12**, 779 (2021).
161. Andrade, G. & Rhodes, J. Protected Areas and Local Communities: an Inevitable Partnership toward Successful Conservation Strategies? *Ecol. Soc.* **17**, (2012).
162. Gatiso, T. T. *et al.* Sustainable protected areas: Synergies between biodiversity conservation and socioeconomic development. *People Nat.* **4**, 893–903 (2022).
163. Kelly, R. *et al.* Ten tips for developing interdisciplinary socio-ecological researchers. *Socio-Ecol. Pract. Res.* **1**, 149–161 (2019).
164. Koch, W., Hogeweg, L., Nilsen, E. B. & Finstad, A. G. Maximizing citizen scientists' contribution to automated species recognition. 2022.02.17.480847 Preprint at https://doi.org/10.1101/2022.02.17.480847 (2022).
165. Rosemartin, A., Watkins, T. & Miller-Rushing, A. J. Monitoring phenology in US national parks through citizen science: Some preliminary lessons and prospects for protected areas. *Parks Steward. Forum* **37**, (2021).
166. Halliwell, P., Whipple, S. & Bowser, G. Learning to love protected areas: Citizen science projects inspire place attachment for diverse students in United States National Parks. *J. Geosci. Educ.* **0**, 1–9 (2021).
167. Callaghan, C. & Gawlik, D. Efficacy of eBird data as an aid in conservation planning and monitoring. *J. Field Ornithol.* **86**, 298–304 (2015).

168. Salmon, R. A., Rammell, S., Emeny, M. T. & Hartley, S. Citizens, Scientists, and Enablers: A Tripartite Model for Citizen Science Projects. *Diversity* **13**, 309 (2021).
169. Hagen, D., Vistad, O. I. & Rød-Eriksen, L. *Sårbarhetsvurdering av utvalgte ferdselslokaliteter i Bymarka naturreservat. Kvistingstien, Gråkallen rundt, St. Olavsspranget. 38* https://brage.nina.no/nina-xmlui/handle/11250/2602455 (2019).
170. Mair, L. & Ruete, A. Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. *PLOS ONE* **11**, e0147796 (2016).
171. Millar, E. E., Hazell, E. C. & Melles, S. J. The 'cottage effect' in citizen science? Spatial bias in aquatic monitoring programs. *Int. J. Geogr. Inf. Sci.* **33**, 1612–1632 (2019).
172. Baldwin, R. F. & Fouch, N. T. Understanding the Biodiversity Contributions of Small Protected Areas Presents Many Challenges. *Land* **7**, 123 (2018).
173. Kendal, D. *et al.* The importance of small urban reserves for plant conservation. *Biol. Conserv.* **213**, 146–153 (2017).
174. Rodríguez-Rodríguez, D., Sánchez-Espinosa, A. & Abdul Malak, D. Potential contribution of OECMs to international area-based conservation targets in a biodiversity rich country, Spain. *J. Nat. Conserv.* **62**, 126019 (2021).
175. Armsworth, P. R., Cantú-Salazar, L., Parnell, M., Davies, Z. G. & Stoneman, R. Management costs for small protected areas and economies of scale in habitat conservation. *Biol. Conserv.* **144**, 423–429 (2011).
176. Koen, E. L. & Newton, E. J. Outreach increases detections of an invasive species in a crowdsourced monitoring program. *Biol. Invasions* **23**, 2611–2620 (2021).
177. Program for Terrestrial Monitoring. TOV-Extensive monitoring of breeding birds. Method booklet – Field work. Version 4. (2021).
178. Bryn, A. & Potthoff, K. Elevational treeline and forest line dynamics in Norwegian mountain areas – a review. *Landsc. Ecol.* **33**, 1225–1245 (2018).
179. Scridel, D. *et al.* A genus at risk: Predicted current and future distribution of all three Lagopus species reveal sensitivity to climate change and efficacy of protected areas. *Divers. Distrib.* **27**, 1759–1774 (2021).
180. Brambilla, M. *et al.* Identifying climate refugia for high-elevation Alpine birds under current climate warming predictions. *Glob. Change Biol.* **28**, 4276–4291 (2022).
181. Sirén, A. P. K. & Morelli, T. L. Interactive range-limit theory (iRLT): An extension for predicting range shifts. *J. Anim. Ecol.* **89**, 940–954 (2020).
182. Freeman, B. G. Lower elevation animal species do not tend to be better competitors than their higher elevation relatives. *Glob. Ecol. Biogeogr.* **29**, 171–181 (2020).
183. Fuglei, E. *et al.* Circumpolar status of Arctic ptarmigan: Population dynamics and trends. *Ambio* **49**, 749–761 (2020).
184. Martin, K. & Wilson, S. Ptarmigan in North America: Influence of life history and environmental conditions on population persistence. *Gyrfalcons Ptarmigan Chang. World* **1**, 45–54 (2011).
185. Pedersen, Å. Ø. *et al.* Rock Ptarmigan (Lagopus muta) breeding habitat use in northern Sweden. *J. Ornithol.* **155**, 195–209 (2014).
186. Kvasnes, M. A. J., Pedersen, H. C. & Nilsen, E. B. Quantifying suitable late summer brood habitats for willow ptarmigan in Norway. *BMC Ecol.* **18**, 41 (2018).
187. Darwin, C. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life.* (John Murray, 1859).

188. Louthan, A. M., Doak, D. F. & Angert, A. L. Where and When do Species Interactions Set Range Limits? *Trends Ecol. Evol.* **30**, 780–792 (2015).
189. Urban, M. C. Escalator to extinction. *Proc. Natl. Acad. Sci.* **115**, 11871–11873 (2018).
190. Paquette, A. & Hargreaves, A. L. Biotic interactions are more often important at species' warm versus cool range edges. *Ecol. Lett.* **24**, 2427–2438 (2021).
191. Lyu, S. & Alexander, J. M. Competition contributes to both warm and cool range edges. *Nat. Commun.* **13**, 2502 (2022).
192. Twomey, E., Morales, V. & Summers, K. Evaluating condition-specific and asymmetric competition in a species-distribution context. *Oikos* **117**, 1175–1184 (2008).
193. Peterson, A. T., Asase, A., Canhos, D. A. L., de Souza, S. & Wieczorek, J. Data Leakage and Loss in Biodiversity Informatics. *Biodivers. Data J.* e26826 (2018) doi:10.3897/BDJ.6.e26826.
194. Tenopir, C. *et al.* Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLOS ONE* **15**, e0229003 (2020).
195. Escribano, N., Galicia, D. & Ariño, A. H. The tragedy of the biodiversity data commons: a data impediment creeping nigher? *Database* **2018**, bay033 (2018).
196. Huang, X. *et al.* Willing or unwilling to share primary biodiversity data: results and implications of an international survey. *Conserv. Lett.* **5**, 399–406 (2012).
197. Grosjean, M. Finding citizen science datasets on GBIF. https://data-blog.gbif.org/post/gbif-citizen-science/ (2018).
198. Chaplin-Kramer, R. *et al.* Conservation needs to integrate knowledge across scales. *Nat. Ecol. Evol.* **6**, 118–119 (2022).
199. Gurney, G. G. *et al.* Biodiversity needs every tool in the box: use OECMs. *Nature* **595**, 646–649 (2021).
200. GBIF. GBIF: The Global Biodiversity Information Facility. (2021).
201. GBIF. *GBIF Annual Report 2011*. http://links.gbif.org/ar2011.pdf (2011).
202. Sullivan, B. L. *et al.* eBird: A citizen-based bird observation network in the biological sciences. *Biol. Conserv.* **142**, 2282–2292 (2009).
203. eBird. 2021 Year in Review: eBird, Merlin, Macaulay Library, and Birds of the World - eBird. https://ebird.org/ebird/news/2021-year-in-review (2021).
204. Danielsen, F., Burgess, N. D. & Balmford, A. Monitoring Matters: Examining the Potential of Locally-based Approaches. *Biodivers. Conserv.* **14**, 2507–2542 (2005).
205. Chape, S., Harrison, J., Spalding, M. & Lysenko, I. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 443–455 (2005).
206. UNEP-WCMC, IUCN and NGS. *Protected Planet Report 2020*. (2020).

Article 1

Participatory monitoring drives biodiversity knowledge in global protected areas

Caitlin P. Mandeville[1,2]*, Erlend B. Nilsen[3,4], Ivar Herfindal[2], Anders G. Finstad[1,2]

[1] Department of Natural History, Norwegian University of Science and Technology
[2] Centre for Biodiversity Dynamics, Norwegian University of Science and Technology
[3] Department of Terrestrial Biodiversity, Norwegian Institute for Nature Research
[4] Faculty of Biosciences and Aquaculture, Nord University, Steinkjer, Norway
* Corresponding author

Abstract

Protected areas are central in international strategies to conserve biodiversity. However, the biodiversity data available to monitor protected areas are insufficient and limited by geographic and taxonomic biases. Participatory biodiversity monitoring vastly increases data collection while engaging the public, but it risks replicating existing biases or introducing new ones. Better understanding of participatory monitoring in protected areas will help guide its ongoing expansion. We explored the contribution of participatory monitoring, across various protected area contexts, to the world's largest biodiversity data repository. Participatory monitoring contributed the majority of all data and was the sole data source for 25% of global protected areas. Patterns in geographic, taxonomic, and threatened species coverage by participatory monitoring differed from non-participatory monitoring, suggesting strong potential to complement other monitoring approaches. Our findings indicate a changing landscape of biodiversity monitoring and suggest directions for the ongoing growth of participatory monitoring across protected area contexts.

This paper is awaiting publication and is not included in NTNU Open

Article 2

Received: 1 April 2022 | Accepted: 21 September 2022
DOI: 10.1002/2688-8319.12185

Ecological Solutions and Evidence | BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use

Caitlin P. Mandeville[1] | **Erlend B. Nilsen**[2,3] | **Anders G. Finstad**[1]

[1]Centre for Biodiversity Dynamics, Department of Natural History, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Terrestrial Biodiversity, Norwegian Institute for Nature Research, Trondheim, Norway

[3]Faculty of Biosciences and Aquaculture, Nord University, Steinkjer, Norway

Correspondence
Centre for Biodiversity Dynamics, Department of Natural History, Norwegian University of Science and Technology, Erling Skakkes Gate 47B, 7021 Trondheim, Norway.
Email: caitlin.mandeville@ntnu.no

Handling Editor: Daniel Bebber

Abstract

1. Opportunistic citizen science produces large amounts of primary biodiversity data but is underutilized in the conservation and management of protected areas despite these areas' status as citizen science hotspots. Application of these data may be limited by the challenge of understanding sampling patterns associated with opportunistic data at a scale relevant to local area management. An improved understanding of citizen science activity patterns within protected areas could strengthen both data analysis and the local promotion and guidance of citizen science activity.
2. We investigated local-scale patterns of citizen science activity, using a case study approach to examine citizen science activity in a recreationally popular natural area that serves as a regional citizen science hotspot. We modelled the relationship between local citizen science activity and 10 spatial covariates broadly related to ease of access and natural interest, factors which have been shown to drive citizen science activity at regional scales. We further compared the distribution of citizen science activity with that of professional data collection and recreational visitor activity in the study area.
3. We found that citizen science data largely complement rather than replicate openly available professional data. Citizen science participation was primarily driven by ease of access, especially the presence of trails. However, citizen science use of the trail network differed from other types of recreational trail use, including a weaker preference for well-established trails and a stronger association with developed areas.
4. This improved understanding of patterns in citizen science participation may be used to better account for spatial biases in citizen science data and to manage natural areas in a way that supports and guides future citizen science activity.

KEYWORDS
biodiversity, biodiversity data, citizen science, community science, green space, natural areas, protected areas, recreation ecology, trails, visitor management

1 | INTRODUCTION

Public participation in biodiversity research, often termed biodiversity citizen science, produces massive amounts of data and contributes extensively to research in biodiversity, conservation and related fields (Bonney, 2021; Cooper et al., 2014; Callaghan et al., 2021; Kays et al., 2020). Much of this contribution comes from mass participation citizen science, in which participants opportunistically upload species observations to digital platforms that are often national to international in scope, due largely to the accessibility of these data in open digital repositories (Ball-Damerow et al., 2019; Callaghan et al., 2021; Mandeville et al., 2021). But despite the mainstream recognition and application of mass participation citizen science in biodiversity science at broader spatial scales, it is generally underutilized in the conservation and management of protected areas and other natural areas on a local scale (Binley et al., 2021; Callaghan & Gawlik, 2015; Cheung et al., 2022; Mandeville & Finstad, 2021; Rapacciuolo et al., 2021; Salmon et al., 2021).

Biodiversity data from mass participation citizen science could play a greater role in filling a critical data gap for small protected areas, green spaces and other multiple-use areas that contribute to other effective area-based conservation measures (OECMS [IUCN WCPA Task Force on OECMs, 2019]) (Adams et al., 2021; Maxwell et al., 2020; Schmeller et al., 2017), which are increasingly recognized as crucial for meeting biodiversity conservation targets (Baldwin & Fouch, 2018; Bonnet et al., 2020; Häkkilä et al., 2021; Kendal et al., 2017; Rodríguez-Rodríguez et al., 2021). Such areas enhance connectivity, support ecosystem services and play a key role in addressing environmental threats that manifest at a local scale (Dreiss & Malcom, 2022; Gaget et al., 2022; Hlásny et al., 2021; Oldekop et al., 2016; Volenec & Dobson, 2020; Wintle et al., 2019). Still, small natural areas often have limited resources for biodiversity conservation, management and monitoring, despite their high conservation value (Armsworth et al., 2011; Jansujwicz et al., 2021; Maxwell et al., 2020).

Mass participation citizen science data are already regularly collected in protected areas and other natural areas and green spaces, which tend to be hotspots for citizen science activity (Tulloch et al., 2013). At broad spatial resolutions, citizen science activity is largely associated with two main types of predictors: accessibility (e.g. population density, road access, regional trail availability) and natural interest (e.g. aesthetic and recreational value, high biodiversity and threatened ecosystems) (Boakes et al., 2016; Geldmann et al., 2016; Mair & Ruete, 2016; Millar et al., 2019; Petersen et al., 2021; Tiago et al., 2017; Tulloch et al., 2013). As accessible areas of local natural interest, small natural areas within or near population centres are popular destinations for citizen science participants.

An improved understanding of spatial sampling patterns within these citizen science hotspots may enhance the utility of opportunistic citizen science data for informing local area management (Callaghan & Gawlik, 2015; Dobson et al., 2020). First, an understanding of sampling patterns might open the door for a wider range of analysis approaches and allow for greater statistical inference (Johnston et al., 2022; Mandeville et al., 2021). At broader spatial scales, information about sampling has been used to overcome analysis challenges related to the spatial and temporal biases and lack of non-detection data that are typical of citizen science data (Cretois et al., 2021; Di Cecco et al., 2021; Johnston et al., 2020; Mueller et al., 2019; Sicacha-Parada et al., 2021; Zulian et al., 2021). But covariates commonly used to model the citizen science sampling process at broader spatial scales are often not well suited to characterize the sampling process at scales relevant to local management. As such, little is known about how the fine-scale distribution of citizen science activity varies within regional citizen science hotspots (Callaghan & Gawlik, 2015; Dobson et al., 2020).

Second, a better understanding of citizen science activity within natural areas can help managers utilize citizen science more effectively (Feldman et al., 2021). Accessible natural areas are commonly managed for both conservation and recreation objectives, both of which can be furthered by citizen science (Buta et al., 2014; Gurney et al., 2021; Halliwell et al., 2022; Newman et al., 2017; Vimal et al., 2021). Citizen science is increasingly recognized by protected area managers as a desirable activity for many area visitors (Weaver & Lawton, 2017), and a better understanding of spatial patterns in their activity would allow managers to actively promote and direct citizen science to meet local objectives. Such direction (e.g. interpretive signage, the use of customized settings on citizen science platforms and promotional events such as bioblitzes) can effectively guide mass participation citizen science data collection (Callaghan et al., 2019; Kays et al., 2020; Knape et al., 2022; Koen & Newton, 2021; Salmon et al., 2021). For these reasons, researchers and managers of protected areas have called for greater research into trends in citizen science participation within protected areas (Binley et al., 2021; Gosal et al., 2021; Leung et al., 2018; Weaver & Lawton, 2017).

We aimed to respond to this call by investigating the spatial distribution of citizen science participation at a scale relevant to local area management. We took a case study approach, characterizing citizen science activity within a small, recreationally popular natural area in Central Norway. The site was selected because it is a regional citizen science hotspot. Our objectives were to (1) test the hypothesis that the main predictors of citizen science activity at a broad spatial resolution—accessibility and natural interest—also drive citizen science at a local scale; (2) compare the distribution of citizen science activity throughout the study area with that of professional

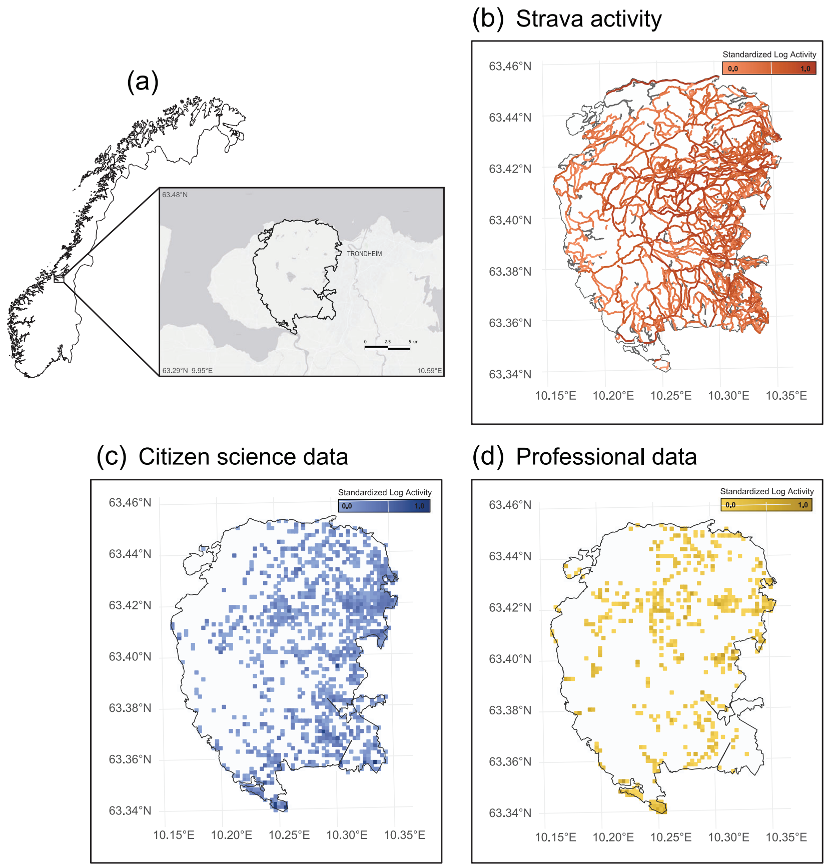


FIGURE 1 Map of study area in Trondheim, Central Norway. Panel (a) indicates the position of the study area in relation to the population centre of Trondheim. Panel (b) indicates the density of reported Strava activities per trail segment and panels (c) and (d) indicate the density of citizen science data and professional biodiversity data available on Global Biodiversity Information Facility (GBIF) per grid cell, respectively.

biodiversity data collection; (3) test the hypothesis that citizen science activity would primarily occur within a short distance of trails and roads and (4) compare the distribution of citizen science activity along the area's trail network with that of other recreational trail use, represented by activity tracking data from Strava Metro.

2 | MATERIALS AND METHODS

2.1 | Study site

Our study site is an 86-km^2 natural area located on the periphery of Trondheim, Central Norway, a regionally dominant city with a population of around 190,000 (Figure 1; https://www.trondheim.kommune.no/). The area consists of a diverse range of southern-boreal habitat types, including mires, mixed forest, lakes and coastline (Moen, 1998). Land management objectives vary within the study area; the entire area is designated as a natural space for public use, while three smaller subsets of the area comprising a total of 12 km^2 are designated as nature reserves with greater conservation protections. The area contains an extensive trail network that is used throughout the year for a range of activities including hiking, running, cycling and skiing, as well as a small number of access roads. There are also a small number of private homes within the area, primarily concentrated near the access roads. The area is recognized as highly important for recreation, but visitor activity patterns are not well studied (Hagen et al., 2019).

2.2 | Data

2.2.1 | Citizen science and professional biodiversity data

All biodiversity data available on the Global Biodiversity Information Facility (GBIF) for the study area were downloaded on 3 August 2021 (GBIF, 2021). The descriptions on GBIF of contributing data providers were used to classify all data as either opportunistic citizen science, structured citizen science or professionally collected data (Table S1.1). If a dataset was attributed to a professional research or management institution with no mention of volunteer participation in the dataset description, the dataset was classified as professional. The single dataset classified as structured citizen science, which was collected by a school-based program, was excluded from analysis because its data derive from a different sampling process than opportunistic citizen science data. Data from before 2000 were excluded, as digital platforms for opportunistic citizen science largely grew in popularity after that year (Figure S1.2). Bacteria and freshwater-obligate species, including fish and aquatic invertebrates, were excluded because the citizen science observation process for these species is expected to differ fundamentally from that of terrestrial species. Finally, data points with a recorded coordinate uncertainty of greater than 150 m were excluded.

2.2.2 | Recreational visitor data

Data on recreational trail use were accessed from Strava Metro (https://metro.strava.com). Strava Metro publishes public data from users of Strava, a mobile app used by recreationists to log running, cycling, skiing and other recreational activities. Data were summarized as the number of recorded trips per Open Street Map (https://www.openstreetmap.org) segment, defined as sections of trail or road between intersections. Strava Metro data were available from 2016 through 2020. The study area contained 7153 segments, with a median segment length of 51 m (interquartile range: 91 m).

2.2.3 | Environmental data

We identified 10 environmental variables, broadly related to ease of area access and natural interest, that we expected to relate to citizen science activity in our study area (Table 1).

2.3 | Analysis

2.3.1 | Environmental covariates of citizen science activity and professional data collection: Grid-based analysis

To examine the relationship between the environmental variables and the distribution of citizen science activity, we established a grid of 150 × 150 m^2 cells in the study area, resulting in 4130 cells. We used the number of citizen science observations in each grid cell as a response variable and the 10 environmental variables as model covariates (Table 1). This approach follows other studies that have examined covariates of citizen science activity at a broader spatial scale (e.g. Romo et al., 2006; Tulloch et al., 2013; Tiago et al., 2017). All continuous covariates were centred and scaled.

There were a small number of outlier cells ($n = 7$) with very high citizen science activity (between 10 and 40 standard deviations greater than the mean number of citizen science observations, which is twice the deviation of the next most active cells). Citizen science participation in these highly active cells was most likely driven by processes that fundamentally differ from typical drivers of citizen science participation; for instance, three such cells were located in the vicinity of a birdwatching tower and two were adjacent to a school and a residential neighbourhood at the edge of the study area. The citizen science activity in these outlier cells is likely not representative of the opportunistic process focused on in this study, so they were excluded from the analysis.

We used a multi-model inference approach to explore potential associations between environmental variables and citizen science activity (Tredennick et al., 2021). We fit a negative binomial generalized linear model including the linear effects of the 10 covariates and no interactions. We tested for spatial autocorrelation using Moran's *I* and included a distance-weighted autocovariate in the model, which reduced autocorrelation (Bardos et al., 2015). We used Akaike's information criterion for small sample sizes (AIC_c) to rank all possible models consisting of combinations of our covariates, and we used the evidence weights of each model to calculate a weighted average of each parameter estimate and standard error across all models (Burnham & Anderson, 2002). The ranked models were used to determine the relative importance of each covariate.

To compare the distribution of citizen science activity with comparable data collection processes for the professionally collected data accessed from GBIF, we first used a Pearson rank correlation analysis to compare the distribution of the two activity types and then repeated the modelling analyses using the number of professional biodiversity data observations per grid cell as the response variable.

2.3.2 | Relationship between citizen science and trail network

Because regional trail density has been shown to predict citizen science activity at broad spatial scales (Tiago et al., 2017), we more closely examined the relationship between citizen science activity and trails within our study area. We hypothesized that the locations of citizen science observations would tend to be closer to trails than the locations of professional data collection, as well as closer than a random distribution of points (obtained using the sf::st_sample() function in R), and used a Kruskal–Wallis test to test this hypothesis.

Next, we conducted a small exploratory analysis intended to provide insight into whether citizen science participants tend to make observations from the trail or to leave the trail before making observations. We expected that if participants tend to make observations from the trail, then the distance between the recorded observation coordinates and the nearest trail would be greatest for taxonomic groups that are more often visible and identifiable from a distance (e.g. mammals, birds, plants). If participants tend to leave the trail to make observations, then we would not expect this relationship. We used a Kruskal–Wallis test

TABLE 1 Environmental variables included in analyses, their expected direction of correlation with citizen science activity and their structure as covariates for the grid-based and trail-based negative binomial generalized linear models

Environmental variables	Expected correlation with citizen science	Covariate structure in grid-based analysis	Covariate structure in trail-based analysis
		Access variables	
Proximity to nearest access point[a,b]	Positive	Negative distance (m) from grid centroid to nearest access point.	Negative distance (m) from trail segment centroid to nearest access point.
Proximity to recreational facilities (e.g. public tourist cabins, playgrounds, swimming beaches)[b]	Positive	Binary variable expressing whether grid cell contains a facility.	Negative distance (m) from trail segment centroid to nearest facility.
Elevation[c]	Negative	Maximum elevation (m) of grid cell.	Maximum elevation (m) of trail segment.
Longitude (eastness)[d]	Positive	Longitude (m) of grid centroid.	Longitude (m) of trail segment centroid.
Presence of recreational trails and access roads[b]	Positive	Total length (m) of trail within grid cell.	Function of segment as a main travel route, defined by the percentage of the trail segment characterized by the 'transportation' land cover category.
		Natural interest variables	
Cultivated land cover[e]	Negative	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Developed land cover[e]	Negative	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Forest land cover[e]	Positive	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Wetland land cover[e]	Positive	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Proximity to a freshwater lake or stream[e]	Positive	Binary variable expressing whether each grid cell contains a freshwater body.	Negative distance (m) from trail segment centroid to nearest freshwater body.

[a]Access points were defined by intersections between a road or trail and the boundary of the natural area as well as public parking areas and public transit stops within or adjacent to the area.
[b]Trondheim Municipality (https://kart.trondheim.kommune.no).
[c]Norwegian Digital Elevation Model (https://www.kartverket.no).
[d]Longitude was used to represent distance from the nearest population centre; the study area lies to the west of Trondheim's population centre, so it was expected that eastern longitudes would be accessed more often.
[e]Norwegian Institute for Bioeconomics AR5 1:5000 land cover data (Ahlstrøm et al., 2014).

to examine the relationship between distance to trail and the observed taxonomic group (grouped in the following categories: birds, fungi, invertebrates, mammals, plants and reptiles/amphibians) and a Dunn's post hoc test for pairwise comparisons between taxonomic groups. This analysis was repeated with the professional dataset.

2.3.3 | Environmental covariates of citizen science and other recreational trail use: Trail-based analysis

Trails have generally been found to be positively associated with citizen science, but some studies have indicated that the relationship between trail access and citizen science activity may be more nuanced (Mair & Ruete, 2016). For this reason, we repeated our modelling process using trail segments as a study unit rather than grid cells. This approach allowed us to compare the trail use of citizen science participants with that of other recreational trail users. We hypothesized that the spatial distribution and drivers of citizen science activity along trail corridors, defined as the zone within 150 m on either side of each trail segment, would be positively correlated with that of other trail users.

To model the relationship between citizen science observations and covariates along trail segments in the study area, we fit a new negative binomial generalized linear model. The response variable was the number of citizen science observations per trail segment corridor, standardized by segment length, and the model covariates were derived from the same 10 environmental variables as in the grid-based analysis (Table 1). The model structure, correction for spatial autocorrelation and model averaging followed the same methods as in the grid-based analysis described in Section 2.3.1.

To test the hypothesis that citizen science activity would correlate with other trail activity, we first used a Pearson rank correlation to compare the number of citizen science observations, standardized by segment length, with the total number of Strava activities reported on the segment. We then used the number of Strava activities reported along each trail segment as a response variable to fit a second model with the same structure and covariates.

All analyses were conducted in R version 4.1.2 (R Core Team, 2021), and analysis scripts are available (Mandeville et al., 2022). Key R packages included tidyverse for data management (Wickham et al., 2019), sf for spatial analyses (Pebesma, 2018) and glmulti for multi-model inference (Calcagno & de Mazancourt, 2010).

3 | RESULTS

3.1 | Biodiversity data

The filtered citizen science data consisted of 44,206 observations from seven citizen science platforms. The vast majority (91%) were contributed through the Norwegian Species Observation Service (https://www.biodiversity.no/), which is Norway's main biodiversity citizen science platform. Citizen science data were contributed by 560 participants. As is typical of digital citizen science datasets (Boakes et al., 2016; Rowley et al., 2019; Wood et al., 2011), a small number of highly active participants contributed the majority of the data; the most active 5% of participants contributed 79% of the total data, while the median participant contributed just six observations. The filtered professionally collected data available on GBIF consisted of 2059 observations from 31 data providers.

The citizen science data contained reports of 1524 species and the professional data contained reports of 991 species (Figure 2). Both types of data collection took place year-round with a peak in intensity in the summer months, but annual variation in sampling intensity was more extreme in the professional data, with sampling intensity peaking later in the summer and falling to a lower rate in the winter than in the citizen science data (Figure 2). Observations occurred in all available land cover types (Figure 2).

3.2 | Environmental covariates of citizen science activity and professional data collection: Grid-based analysis

As expected, ease of area access was positively correlated with citizen science activity among grid cells (Figures 3, 4, and S2.1; Table S2.2). The total trail length per grid cell was the most important covariate and had a large positive effect on citizen science activity. Grid cells nearer to an area access point and to the closest population centre were also positively associated with citizen science activity, though the effect of these covariates was smaller. Neither elevation nor the presence of recreational facilities had an important relationship with citizen science activity. Contrary to expectations, the developed and cultivated land cover types had a positive association with citizen science, while the wetland and forest land cover types were unimportant. The presence of freshwater had an important positive relationship to citizen science activity. Parameter estimates were consistent among the highly weighted models; they varied little between the six models that had a substantial level of support ($\Delta AIC_c < 2$), which in total account for 66.3% of the weight of evidence (Table S2.3; Figure S2.4).

Citizen science activity was not correlated with professional data collection among grid cells (Figure 1; Pearson correlation $r = 0.035$, $p = 0.023$). Two access covariates—proximity to access points and to the population centre at the area's eastern edge—were important in the professional data model (Figures 3, 5, and S2.1; Table S2.2). The effect of proximity to the population centre was opposite to its effect on citizen science. As with citizen science, the presence of water and cultivated land had a positive relationship to professional data collection and forest had a small negative effect. Unlike with citizen science, the presence of wetland land cover had a small negative relationship to professional data collection and the developed land cover type did not have an important effect. There was little variation in parameter estimates among the nine models with a substantial level of support ($\Delta AIC_c < 2$), which in total account for 46.7% of the weight of evidence (Table S2.3; Figure S2.4).

3.3 | Relationship between citizen science and trail network

The locations of citizen science observations were a median of 11 m from the nearest trail, which was closer than sites of professional data collection (median: 29 m). Both were closer than a random distribution of sites, which would be expected to have a median distance from the nearest trail of 45 m (Kruskal–Wallis $\chi^2(2) = 1167$, $p < 0.0001$) (Figure 6).

There was high variability in the distance between observation points and the nearest trail within taxonomic groups. Still, taxonomic groups that may be difficult to see from a distance (fungi, reptiles and amphibians) were associated with the smallest mean distance from the trail, while taxonomic groups that tend to be relatively easy to spot from a distance (birds) were associated with the greatest distance (Kruskal–Wallis $\chi^2(5) = 3083$, $p < 0.0001$). Invertebrates were an exception to this trend. These results are partially consistent with the trend expected if observations tended to be made from a trail, though there may be alternative potential explanations for the pattern. Though differences between groups were observed in the professional dataset as well, there was greater variability and less evidence for the hypothesized trend (Kruskal–Wallis $\chi^2(4) = 74$, $p < 0.0001$) (Figure 6).

3.4 | Environmental covariates of citizen science and other recreational trail use: Trail-based analysis

The tested covariates had limited ability to explain variation in citizen science activity among trail segments; four models had a substantial level of support, totalling 24.2% of the weight of evidence (Table S3.3; Figure S3.4). All effect sizes were relatively small compared to the grid-based models (Figures 3, 7, and S3.1; Table S3.2). Notably, most covariates that were important at the grid scale were not important to describe variation between trail segments; proximity to the nearest

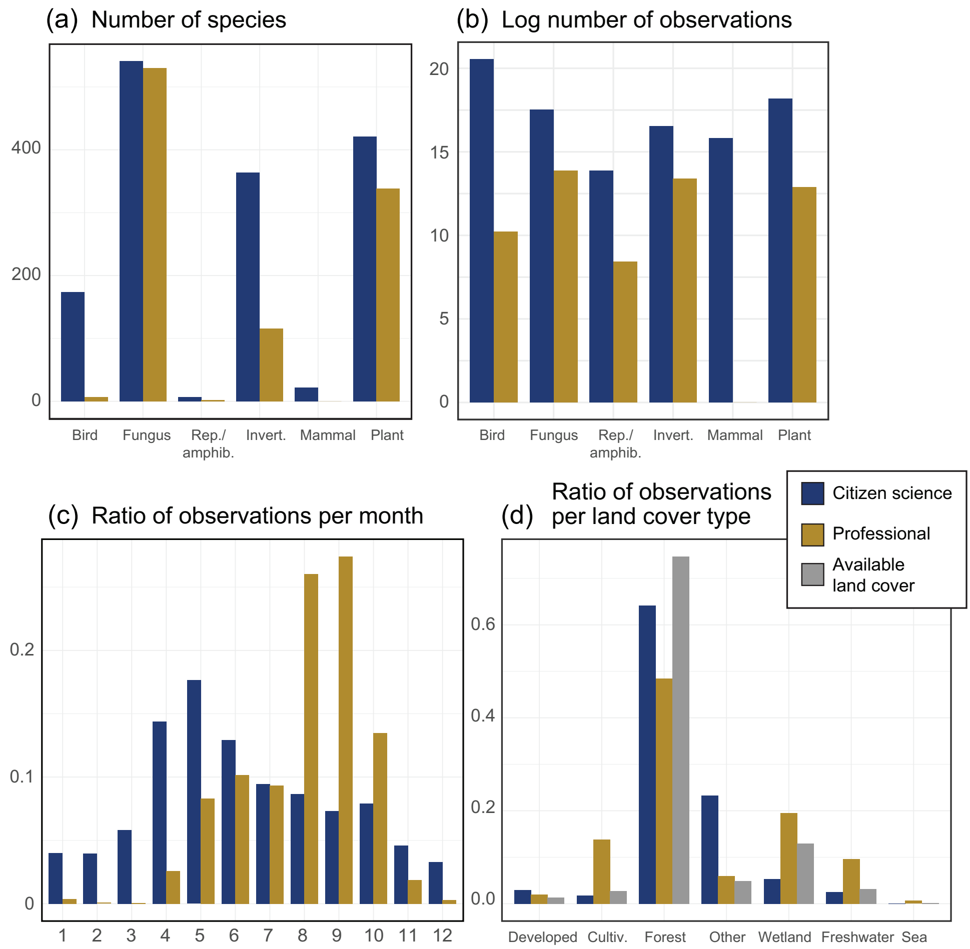


FIGURE 2 (a) Number of observations from each taxonomic group for citizen science and professional data; (b) number of species from each taxonomic group for citizen science and professional data; (c) month of observation for citizen science and professional data; (d) land cover type for citizen science and professional observations, shown relative to the availability of land cover types within the area.

access point, the eastern area edge and freshwater were important at the grid level but had only a small and uncertain relationship to citizen science activity along trail segments. The most important variable was forest cover, which had a small negative relationship to citizen science activity.

The number of citizen science observations per trail segment corridor had no relationship to the number of reported Strava activities (Figure 1; Pearson correlation test, $r = -0.01$, $p = 0.414$). The relationship between the covariates and Strava activity differed substantially from their relationship to citizen science activity. The degree to which a trail segment functioned as a main travel route was the most important covariate, with a large positive relationship to Strava activity (Figures 3, 8, and S3.1; Table S3.2). In contrast, this covariate had only a small positive effect on citizen science activity (Figures 3 and 7). Elevation had a positive association with Strava activity but a small negative association with citizen science activity. Wetland land cover had a positive association, while the relationship with forest and developed land cover was small and uncertain. There was little variation in parameter estimates among the 12 models with a substantial level of support ($\Delta AIC_c < 2$), together accounting for 42.7% of the weight of evidence (Table S3.3; Figure S3.4).

4 | DISCUSSION

We responded to calls for research on citizen science within protected and other natural areas by examining citizen science activity in a small natural area that serves as a regional citizen science hotspot. Our results illustrate that citizen science participation is spatially heterogeneous on a local scale. Ease of area access was the dominant landscape

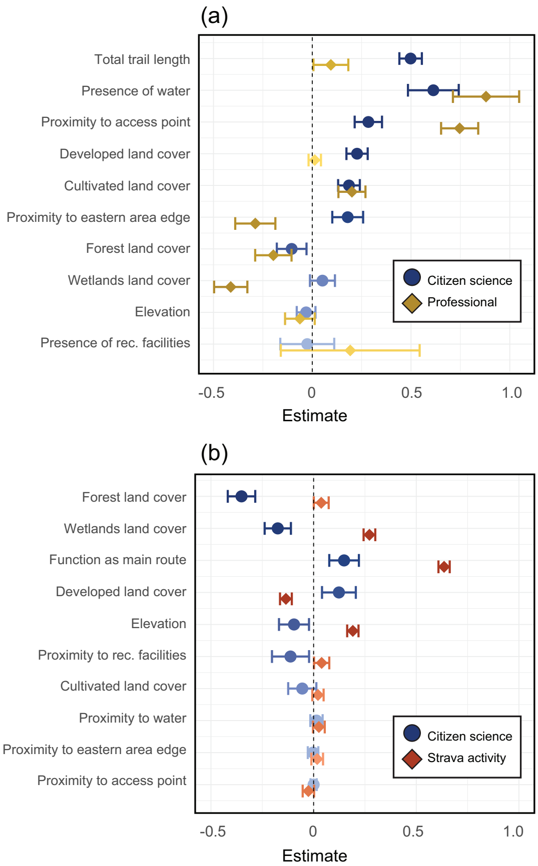


FIGURE 3 Effect of all covariates on the response variable of (a) the grid-based models of citizen science and professional data collection and (b) the trail-based models of citizen science and Strava activity. Decreasing colour intensity indicates decreasing variable importance. All continuous variables have been centred and scaled.

characteristic driving the distribution of citizen science in our study area, and a key component of accessibility is the use of a trail network to access observation sites. However, the distribution of citizen science activity along the trail network differed from that of other trail users. In general, citizen science activity was more evenly dispersed over a wider range of trail characteristics than other trail use; for example, citizen science participants were more likely than other trail users to spend time both in more developed parts of the natural area and also on less well-established paths that do not function as main travel routes.

The importance of area access is a notable result of our study. It is known that accessibility and natural interest are major regional determinants of citizen science activity, but our results are among the first to show that, within a small natural area, accessibility has a stronger relationship to citizen science than particular landscapes perceived as the most natural. To the contrary, citizen science activity was positively associated with cultivated and developed land within the area. This may be partially explained by the increased accessibility afforded by infrastructure in these areas, or by interest in the biodiversity of these land cover types. But it may also stem from an affinity for cultivated and developed land cover, as suggested by recent findings that the integration of biodiversity with cultural and agricultural heritage plays an important role in communities' relationship to natural areas (Cusens et al., 2022). Proximity to water was positively associated with citizen science activity in our study area, as has previously been shown at regional scales and within urban areas (Boakes et al., 2016; Tiago et al., 2017). This could be explained by trends in either participant behaviour (e.g. participants might prefer spending time near water or observing species found near water) or in species availability (e.g. landscapes containing freshwater may be more species rich or afford greater detectability for species that are present).

The strong association between accessibility and citizen science participation offers some possibilities for improving the analysis of citizen science data. First, it may be possible to coarsely model citizen science sampling bias in local-scale analyses by accounting for access opportunities, as has been done previously at broader scales (Cretois et al., 2021; Johnston et al., 2020; Sicacha-Parada et al., 2021). Further, it may be possible to incorporate local-scale sampling bias within citizen science hotspots into regional models. A better understanding of sampling process can support a diverse range of applications that are relevant to local area management, including biodiversity assessments, monitoring of trends, assessment of interventions, invasive species detection, and species distributions analyses (Dobson et al., 2020; Foster et al., 2021; Johnston et al., 2022; Kühl et al., 2020).

At the same time, our results emphasize that mass participation citizen science can be a valuable supplement or, where needed, surrogate for biodiversity data from other data sources. Though the analysis of opportunistic citizen science data is characterized by a range of challenges in addition to spatial and temporal unevenness, including taxonomic bias and accuracy, geographic accuracy and typical lack of non-detection data, they are widely recognized as a critical source of biodiversity data (Callaghan et al., 2021; Cooper et al., 2014; Johnston et al., 2020). The citizen science data on GBIF include a greater number of species from all taxonomic groups than the equivalent professional datasets within our study area, covering a similarly diverse range of land cover types. In some ways, citizen science expands the reach of professional data collection; for instance, citizen science outpaced the professional data available on GBIF in the winter months in our study area. Winter ecology is recognized as understudied yet critical to conservation in the face of climate change (Studd et al., 2021; Sutton et al., 2021), so the contribution to this research area by citizen science is noteworthy.

When comparing citizen science and professionally collected data, it is important to note that the professional data available on GBIF for a natural area are almost certainly not a complete record of professional biodiversity data that have been collected in the area; while the value of openly sharing data is increasingly recognized, barriers still prevent much biodiversity data from being shared (Mandeville et al., 2021). Many small natural areas also support locally managed, place-based citizen science programs that are typically structured to address specific

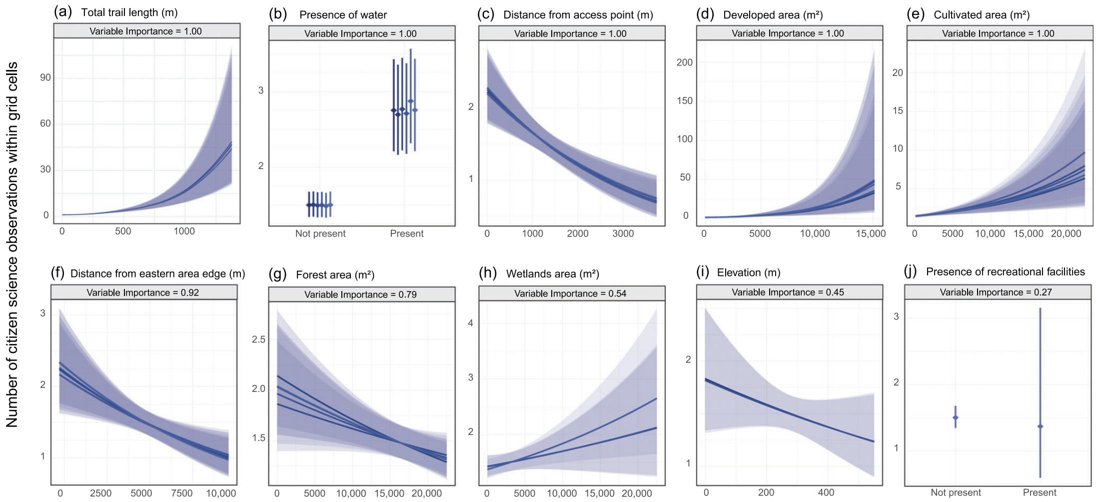


FIGURE 4 Predicted effect of each covariate on the number of citizen science observations per grid cell, modelled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All six models with substantial support ($\Delta AIC_c < 2$) are shown, with decreasing colour intensity indicating decreasing model rank. Relative variable importance, calculated with a weighted average of all models, is indicated above each covariate plot.

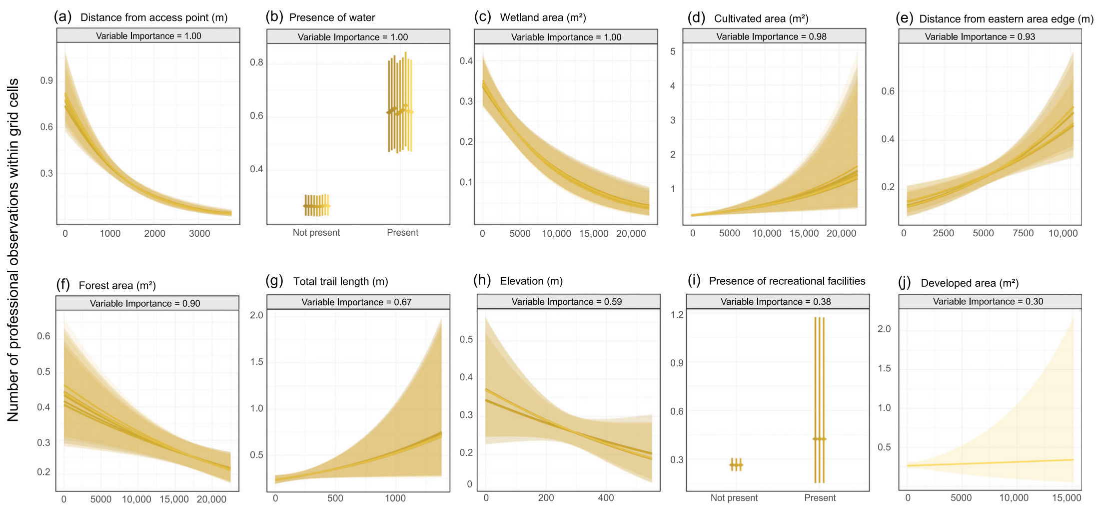


FIGURE 5 Predicted effect of each covariate on the number of professional biodiversity observations per grid cell, modelled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All nine models with substantial support ($\Delta AIC_c < 2$) are shown, with decreasing colour intensity indicating decreasing model rank. Relative variable importance, calculated with a weighted average of all models, is indicated above each covariate plot.

research and monitoring questions (Mandeville & Finstad, 2021; Rosemartin et al., 2021). Such programs are highly valuable but are often resource intensive to coordinate at a local level and therefore may not be feasible to implement in all settings (Alfonso et al., 2022; Rosemartin et al., 2021; Tancoigne, 2019). Further, they often produce data that are not openly shared on GBIF (Mandeville et al., 2021). For this reason, the open biodiversity data collected through opportunistic citizen science platforms are particularly valuable for their relative ease of access, allowing them to fill gaps both when other data do not exist and when other data cannot be accessed. In parallel with increasing the utility of citizen science data for area management, it is critical to continue increasing area managers' access to other existing data sources; among other reasons, this is because opportunistic citizen science data are often most valuable when integrated with other data types (Dobson et al., 2020; Kühl et al., 2020; Mandeville et al., 2021).

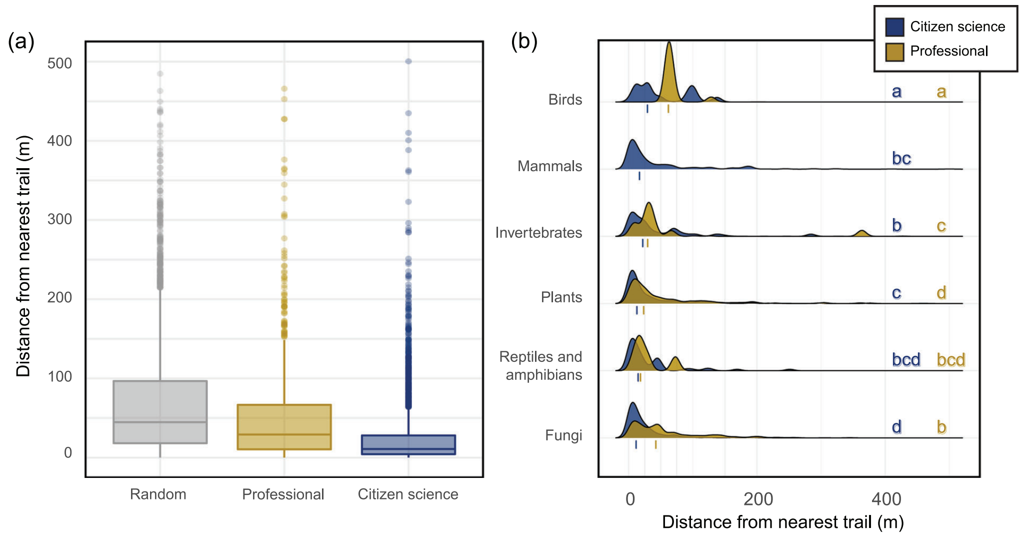


FIGURE 6 (a) Distance between reported observation coordinates and the nearest trail for the locations of citizen science data collection, professional data collection and a random sample of locations in the study area. (b) Distance between reported observation coordinates and the nearest trail for observations within each taxonomic group, for citizen science and professional data. The area under the curve indicates the proportion of the total data with each value along the *x*-axis, and dashes indicate median values. Letters indicate significantly different groups as indicated by a Dunn's post hoc test ($\alpha = 0.05$).

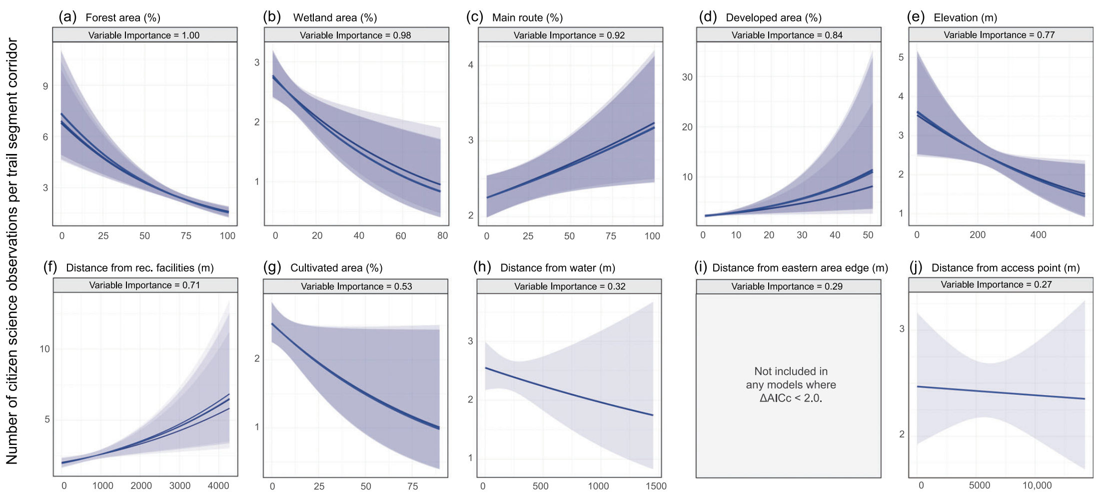


FIGURE 7 Predicted effect of each covariate on the number of citizen science observations per 300-m-wide trail segment corridor, standardized by segment length, modelled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All four models with substantial support ($\Delta AIC_c < 2$) are shown, with decreasing colour intensity indicating decreasing model rank. Relative variable importance, calculated with a weighted average of all models, is indicated above each covariate plot.

In addition to informing more effective analysis of existing citizen science data, knowledge of citizen science activity patterns can be used by area managers to promote and guide future data collection. First, managers could use knowledge about citizen science trends to reach out to current participants to prompt collection of data to meet specific monitoring needs, for example by posting signs in areas regularly frequented by citizen science participants or communicating through customization features offered by citizen science platforms (Callaghan et al., 2021; Gosal et al., 2021; Koen & Newton, 2021). Second, managers could identify areas of low citizen science activity to target for recruiting new participants (Weaver & Lawton, 2017). For instance, recreational facilities were not closely associated with citizen science

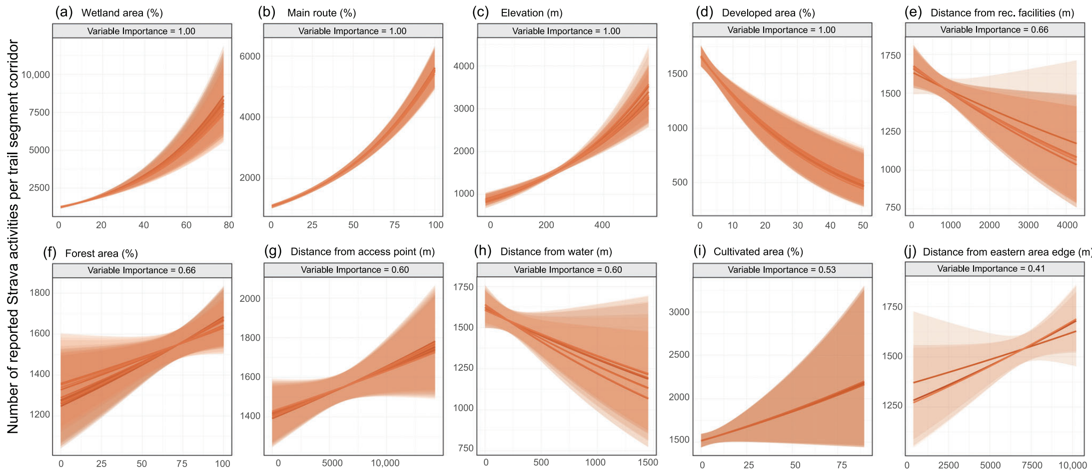


FIGURE 8 Predicted effect of each covariate on the number of reported Strava activities per trail segment corridor, modelled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All 12 models with substantial support ($\Delta AIC_c < 2$) are shown, with decreasing colour intensity indicating decreasing model rank. Relative variable importance, calculated with a weighted average of all models, is indicated above each covariate plot.

participation in our study area; collaboration with relevant recreational organizations and facilities to promote citizen science could more firmly ground local citizen science participation in a sense of place and engage recreational visitors who do not yet participate in citizen science (Allf et al., 2022; Newman et al., 2017). Finally, managers may be able to prioritize professional data collection to complement citizen science by emphasizing areas of low citizen science activity.

Knowledge of spatial trends in citizen science activity can further inform overall management strategies for natural areas and green spaces. The needs and preferences of area visitors are regularly used to make management decisions about natural areas and even to justify ongoing area protection, but because different subsets of visitors prioritize different types of area management, it can be challenging to identify the diverse needs of area visitors (Hornigold et al., 2016; Komossa et al., 2021; Mancini et al., 2019; Muñoz et al., 2020). Our results show that citizen science participants in our study area tend to use the area's trail network differently than other visitors, so their needs may be overlooked if not explicitly considered. Citizen science participants may even serve as a useful proxy to represent a broader group of nature-oriented visitors whose area use might differ in similar ways from the more activity-oriented visitors captured in the Strava Metro data (Cambria et al., 2021; Havinga et al., 2020). The Strava Metro dataset is itself biased toward visitors with a focus on athletic recreation, though a recent study elsewhere in Norway found a high correlation between Strava activities and absolute counts of segment users, suggesting that Strava is relatively representative of the dominant trends in segment use, particularly in areas of high activity (Venter et al., 2021).

Moving forward, there is much left to learn about citizen science participation at a local scale. The knowledge gained from modelling spatial patterns in citizen science participation is especially meaningful when considered alongside studies that directly investigate citizen science participants' motivations, goals and outcomes. Our results demonstrate that trends in citizen science participants' behaviour can manifest in spatial patterns, and also suggest new directions that could be followed up with social science research: for instance, it would be useful to survey citizen science participants about their selection of trail routes or their on- and off-trail activity. Importantly, our goal of understanding the distribution of citizen science activity at a local scale responds to a commonly documented motivation for citizen science participation: participants regularly indicate that they want their data to be used for the conservation and management of places that they value (Bowler et al., 2022; Ganzevoort et al., 2017; Larson et al., 2020; Maund et al., 2020). Through facilitation of improved data analysis and citizen science program implementation, a stronger understanding of citizen science activity can be a step toward increasing the local conservation impact of participants' contributions.

AUTHOR CONTRIBUTIONS

Caitlin P. Mandeville conceived of the idea, analysed the data, and led the writing of the manuscript. Erlend B. Nilsen and Anders G. Finstad supported the conceptual development and writing of the manuscript. All authors contributed to the drafts and gave approval for publication.

ACKNOWLEDGEMENTS

We are grateful to Francesco Frassinelli from the Norwegian Institute for Nature Research for assistance with the Strava Metro data and to Trondheim Municipality, especially Kristian Morin, for assistance with spatial data layers. We also wish to thank the Transforming Citizen Science for Biodiversity research group, of which the authors are a part, for many constructive conversations about biodiversity citizen science. We further appreciate the valuable insights of two anonymous reviewers. Funding for CM was provided by the Digital Transformation Initiative of the Norwegian University of Science and Technology. Finally, we extend deep gratitude to all the citizen science participants and practitioners who have played a role in assembling the biodiversity data used in this paper.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Biodiversity data are obtained from the Global Biodiversity Information Facility and are available here: https://doi.org/10.15468/dl.pd3tce (GBIF, 2021). This report includes aggregated and de-identified data from Strava Metro. R scripts for the analyses are available from Open Science Framework: https://doi.org/10.17605/osf.io/bgf3d (Mandeville et al., 2022).

ORCID

Caitlin P. Mandeville https://orcid.org/0000-0002-1361-607X
Erlend B. Nilsen https://orcid.org/0000-0002-5119-8331
Anders G. Finstad https://orcid.org/0000-0003-4529-6266

PEER REVIEW

The peer review history for this article is available at: https://publons.com/publon/10.1002/2688-8319.12185.

REFERENCES

Adams, V. M., Visconti, P., Graham, V., & Possingham, H. P. (2021). Indicators keep progress honest: A call to track both the quantity and quality of protected areas. *One Earth*, *4*, 901–906.

Ahlstrøm, A., Bjørkelo, K., & Frydenlund, J. (2014). AR5 klassifikasjonssyste: Klassifikasjon av *Arealressurser*. Rapport 6. Norsk Institutt for Skog og Landskap.

Alfonso, L., Gharesifard, M., & Wehn, U. (2022). Analysing the value of environmental citizen generated data: Complementarity and cost per observation. *Journal of Environmental Management*, *303*, 114157.

Allf, B. C., Cooper, C. B., Larson, L. R., Dunn, R. R., Futch, S. E., Sharova, M., & Cavalier, D. (2022). Citizen science as an ecosystem of engagement: Implications for learning and broadening participation. *BioScience*, *72*, 651–663.

Armsworth, P. R., Cantú-Salazar, L., Parnell, M., Davies, Z. G., & Stoneman, R. (2011). Management costs for small protected areas and economies of scale in habitat conservation. *Biological Conservation*, *144*, 423–429.

Baldwin, R. F., & Fouch, N. T. (2018). Understanding the biodiversity contributions of small protected areas presents many challenges. *Land*, *7*, 123.

Ball-Damerow, J. E., Brenskelle, L., Barve, N., Soltis, P. S., Sierwald, P., Bieler, R., LaFrance, R., Ariño, A. H., & Guralnick, R. P. (2019). Research applications of primary biodiversity databases in the digital age. *PLoS ONE*, *14*, e0215794.

Bardos, D. C., Guillera-Arroita, G., & Wintle, B. A. (2015). Valid auto-models for spatially autocorrelated occupancy and abundance data. *Methods in Ecology and Evolution*, *6*, 1137–1149.

Binley, A. D., Proctor, C. A., Pither, R., Davis, S. A., & Bennett, J. R. (2021). The unrealized potential of community science to support research on the resilience of protected areas. *Conservation Science and Practice*, *3*, e376.

Boakes, E. H., Gliozzo, G., Seymour, V., Harvey, M., Smith, C., DB, R., & Haklay, M. (2016). Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. *Scientific Reports*, *6*, 33051.

Bonnet, P., Joly, A., Faton, J.-M., Brown, S., Kimiti, D., Deneu, B., Servajean, M., Affouard, A., Lombardo, J.-C., Mary, L., Vignau, C., & Munoz, F. (2020). How citizen scientists contribute to monitor protected areas thanks to automatic plant identification tools. *Ecological Solutions and Evidence*, *1*, e12023.

Bonney, R. (2021). Expanding the impact of citizen science. *BioScience*, *71*, 448–451.

Bowler, D. E., Bhandari, N., Repke, L., Beuthner, C., Callaghan, C. T., Eichenberg, D., Henle, K., Klenke, R., Richter, A., Jansen, F., Buelheide, H., & Bonn, A. (2022). Decision-making of citizen scientists when recording species observations. *Scientific Reports*, *12*, 11069.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach*. Springer.

Buta, N., Holland, S. M., & Kaplanidou, K. (2014). Local communities and protected areas: The mediating role of place attachment for pro-environmental civic engagement. *Journal of Outdoor Recreation and Tourism*, *5–6*, 1–10.

Calcagno, V., & de Mazancourt, C. (2010). glmulti: An R package for easy automated model selection with (generalized) linear models. *Journal of Statistical Software*, *34*, 1–29.

Callaghan, C., & Gawlik, D. (2015). Efficacy of eBird data as an aid in conservation planning and monitoring. *Journal of Field Ornithology*, *86*, 298–304.

Callaghan, C. T., Rowley, J. J. L., Cornwell, W. K., Poore, A. G. B., & Major, R. E. (2019). Improving big citizen science data: Moving beyond haphazard sampling. *PLoS Biology*, *17*, e3000357.

Callaghan, C. T., Poore, A. G. B., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., Rowley, J. J. L., Vergés, A., Wilshire, J. H., & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, *71*, 55–63.

Cambria, V. E., Campagnaro, T., Trentanovi, G., Testolin, R., Attorre, F., & Sitzia, T. (2021). Citizen science data to measure human use of green areas and forests in European cities. *Forests*, *12*, 779.

Cheung, S. Y., Leung, Y.-F., & Larson, L. R. (2022). Citizen science as a tool for enhancing recreation research in protected areas: Applications and opportunities. *Journal of Environmental Management*, *305*, 114353.

Cooper, C. B., Shirk, J., & Zuckerberg, B. (2014). The invisible prevalence of citizen science in global research: Migratory birds and climate change. *PLoS ONE*, *9*, e106508.

Cretois, B., Simmonds, E. G., Linnell, J. D. C., van Moorter, B., Rolandsen, C. M., Solberg, E. J., Strand, O., Gundersen, V., Roer, O., & Rød, J. K. (2021). Identifying and correcting spatial bias in opportunistic citizen science data for wild ungulates in Norway. *Ecology and Evolution*, *11*, 15191–15204.

Cusens, J., Barraclough, A. M. D., & Måren, I. E. (2022). Participatory mapping reveals biocultural and nature values in the shared landscape of a Nordic UNESCO Biosphere Reserve. *People and Nature*, *4*, 365–381.

Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., & Hurlbert, A. H. (2021). Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *BioScience*, *71*, 1179–1188.

Dobson, A. D. M., Milner-Gulland, E. J., Aebischer, N. J., Beale, C. M., Brozovic, R., Coals, P., Critchlow, R., Dancer, A., Greve, M., Hinsley, A., Ibbett, H., Johnston, A., Kuiper, T., Le Comber, S., Mahood, S. P., Moore, J. F., Nilsen, E. B., Pocock, M. J. O., Quinn, A., … Keane, A. (2020). Making messy data work for conservation. *One Earth*, *2*, 455–465.

Dreiss, L. M., & Malcom, J. W. (2022). Identifying key federal, state, and private lands strategies for achieving 30 × 30 in the United States. *Conservation Letters*, *15*, e12849.

Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., & Fenton, N. J. (2021). Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS ONE*, *16*, e0234587.

Foster, S. D., Vanhatalo, J., Trenkel, V. M., Schulz, T., Lawrence, E., Przeslawski, R., & Hosack, G. R. (2021). Effects of ignoring survey design information for data reuse. *Ecological Applications*, *31*(6), e02360.

Gaget, E., Johnston, A., Pavón-Jordán, D., Lehikoinen, A. S., Sandercock, B. K., Soultan, A., Božič, L., Clausen, P., Devos, K., Domsa, C., Encarnação, V., Faragó, S., Fitzgerald, N., Frost, T., Gaudard, C., Gosztonyi, L., Haas, F., Hornman, M., Langendoen, T., … Brommer, J. E. (2022). Protected area characteristics that help waterbirds respond to climate warming. *Conservation Biology*, *36*, e13877.

Ganzevoort, W., van den Born, R. J. G., Halffman, W., & Turnhout, S. (2017). Sharing biodiversity data: Citizen scientists' concerns and motivations. *Biodiversity and Conservation*, *26*, 2821–2837.

Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, *22*, 1139–1149.

Global Biodiversity Information Facility (GBIF). (2021). *GBIF Occurrence Download*. https://doi.org/10.15468/dl.pd3tce

Gosal, A. S., McMahon, J. A., Bowgen, K. M., Hoppe, C. H., & Ziv, G. (2021). Identifying and mapping groups of protected area visitors by environmental awareness. *Land*, *10*, 560.

Gurney, G. G., Darling, E. S., Ahmadia, G. N., Agostini, V. N., Ban, N. C., Blythe, J., Claudet, J., Epstein, G., Estradivari, H.-C. A., Jonas, H. D., Armitage, D., Campbell, S. J., Cox, C., Friedman, W. R., Gill, D., Lestari, P., Mangubhai, S., McLeod, E., Muthiga, N. A., … Jupiter, S. D. (2021). Biodiversity needs every tool in the box: Use OECMs. *Nature*, *595*, 646–649.

Hagen, D., Vistad, O. I., & Rød-Eriksen, L. (2019). *Sårbarhetsvurdering av utvalgte ferdselslokaliteter I bymarka naturreservat. Kvistingstien, Gråkallen rundt, St. Olavsspranget. NINA Rapport 1680*. Norsk Institutt for Naturforskning.

Häkkilä, M., Johansson, A., Sandren, T., Uusitalo, A., Mönkkönen, M., Puttonen, P., & Savilaakso, S. (2021). Are small protected habitat patches within boreal production forests effective in conserving species richness, abundance and community composition? A systematic review. *Environmental Evidence*, *10*, 2.

Halliwell, P., Whipple, S., & Bowser, G. (2022). Learning to love protected areas: Citizen science projects inspire place attachment for diverse students in United States National Parks. *Journal of Geoscience Education*, *70*, 412–420.

Havinga, I., Bogaart, P. W., Hein, L., & Tuia, D. (2020). Defining and spatially modelling cultural ecosystem services using crowdsourced data. *Ecosystem Services*, *43*, 101091.

Hlásny, T., Mokroš, M., Dobor, L., Merganičová, K., & Lukac, M. (2021). Fine-scale variation in projected climate change presents opportunities for biodiversity conservation in Europe. *Scientific Reports*, *11*, 17242.

Hornigold, K., Lake, I., & Dolman, P. (2016). Recreational use of the countryside: No evidence that high nature value enhances a key ecosystem service. *PLoS ONE*, *11*, e0165043.

IUCN WCPA Task Force on OECMs. (2019). *Recognising and reporting other effective area based conservation measures*. International Union for Conservation of Nature.

Jansujwicz, J. S., Calhoun, A. J. K., Bieluch, K. H., McGreavy, B., Silka, L., & Sponarski, C. (2021). Localism "reimagined": Building a robust localist paradigm for overcoming emerging conservation challenges. *Environmental Management*, *67*, 91–108.

Johnston, A., Moran, N., Musgrove, A., Fink, D., & Baillie, S. R. (2020). Estimating species distributions from spatially biased citizen science data. *Ecological Modelling*, *422*, 108927.

Johnston, A., Matechou, E., & Dennis, E. B. (2022). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, . https://doi.org/10.1111/2041-210X.13834

Kays, R., McShea, W. J., & Wikelski, M. (2020). Born-digital biodiversity data: Millions and billions. *Diversity and Distributions*, *26*, 644–648.

Kendal, D., Zeeman, B. J., Ikin, K., Lunt, I. D., McDonnell, M. J., Farrar, A., Pearce, L. M., & Morgan, J. W. (2017). The importance of small urban reserves for plant conservation. *Biological Conservation*, *213*, 146–153.

Knape, J., Coulson, S. J., van der Wal, R., & Arlt, D. (2022). Temporal trends in opportunistic citizen science reports across multiple taxa. *Ambio*, *51*, 183–198.

Koen, E. L., & Newton, E. J. (2021). Outreach increases detections of an invasive species in a crowdsourced monitoring program. *Biological Invasions*, *23*, 2611–2620.

Komossa, F., Wartmann, F. M., & Verburg, P. H. (2021). Expanding the toolbox: Assessing methods for local outdoor recreation planning. *Landscape and Urban Planning*, *212*, 104105.

Kühl, H. S., Bowler, D. E., Bösch, L., Bruelheide, H., Dauber, J., Eichenberg, D., Eisenhauer, N., Fernández, N., Guerra, C. A., Henle, K., Herbinger, I., Isaac, N. J. K., Jansen, F., König-Ries, B., Kühn, I., Nilsen, E. B., Pe'er, G., Richter, A., Schulte, R., … Bonn, A. (2020). Effective biodiversity monitoring needs a culture of integration. *One Earth*, *3*, 462–474.

Larson, L. R., Cooper, C. B., Futch, S., Singh, D., Shipley, N. J., Dale, K., LeBaron, G. S., & Takekawa, J. Y. (2020). The diverse motivations of citizen scientists: Does conservation emphasis grow as volunteer participation progresses? *Biological Conservation*, *242*, 108428.

Leung, Y.-F., Spenceley, A., Hvenegaard, G., Buckley, R., & Groves, C. (2018). *Tourism and visitor management in protected areas: Guidelines for sustainability* (1st ed.). International Union for Conservation of Nature.

Mair, L., & Ruete, A. (2016). Explaining spatial variation in the recording effort of citizen science data across multiple taxa. *PLoS ONE*, *11*, e0147796.

Mancini, F., Coghill, G. M., & Lusseau, D. (2019). Quantifying wildlife watchers' preferences to investigate the overlap between recreational and conservation value of natural areas. *Journal of Applied Ecology*, *56*, 387–397.

Mandeville, C. P., & Finstad, A. G. (2021). Community science supports research on protected area resilience. *Conservation Science and Practice*, *3*, e442.

Mandeville, C. P., Koch, W., Nilsen, E. B., & Finstad, A. G. (2021). Open data practices among users of primary biodiversity data. *BioScience*, *71*, 1128–1147.

Mandeville, C. P., Nilsen, E. B., & Finstad, A. G. (2022). Spatial distribution of biodiversity citizen science in a natural area. *Open Science Framework*, . https://doi.org/10.17605/osf.io/bgf3d

Maund, P. R., Irvine, K. N., Lawson, B., Steadman, J., Risely, K., Cunningham, A. A., & Davies, Z. G. (2020). What motivates the masses: Understanding why people contribute to conservation citizen science projects. *Biological Conservation*, *246*, 108587.

Maxwell, S. L., Cazalis, V., Dudley, N., Hoffmann, M., Rodrigues, A. S. L., Stolton, S., Visconti, P., Woodley, S., Kingston, N., Lewis, E., Maron, M., Strassburg, B. B. N., Wenger, A., Jonas, H. D., Venter, O., & Watson, J. E. M. (2020). Area-based conservation in the twenty-first century. *Nature*, *586*, 217–227.

Millar, E. E., Hazell, E. C., & Melles, S. J. (2019). The 'cottage effect' in citizen science? Spatial bias in aquatic monitoring programs. *International Journal of Geographical Information Science*, *33*, 1612–1632.

Moen, A. (1998). *Nasjonalatlas for Norge: Vegetasjon*. Statens Kartverk.

Mueller, M. A., Drake, D., & Allen, M. L. (2019). Using citizen science to inform urban canid management. *Landscape and Urban Planning*, *189*, 362–371.

Muñoz, L., Hausner, V. H., Runge, C., Brown, G., & Diagle, R. (2020). Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in Southern Norway. *People and Nature*, *2*, 437–449.

Newman, G., Chandler, M., Clyde, M., McGreavy, B., Haklay, M., Ballard, H., Gray, S., Scarpino, R., Hauptfeld, R., Mellor, D., & Gallo, J. (2017). Leveraging the power of place in citizen science for effective conservation decision making. *Biological Conservation*, *208*, 55–64.

Oldekop, J. A., Holmes, G., Harris, W. E., & Evans, K. L. (2016). A global assessment of the social and conservation outcomes of protected areas. *Conservation Biology*, *30*, 133–141.

Pebesma, E. (2018). Simple features for R: Standardized support for spatial vector data. *The R Journal*, *10*, 439–446.

Petersen, T. K., Speed, J. D. M., Grøtan, V., & Austrheim, G. (2021). Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. *Ecological Solutions and Evidence*, *2*, e12048.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rapacciuolo, G., Young, A., & Johnson, R. (2021). Deriving indicators of biodiversity change from unstructured community-contributed data. *Oikos*, *130*, 1225–1239.

Rodríguez-Rodríguez, D., Sánchez-Espinosa, A., & Abdul Malak, D. (2021). Potential contribution of OECMs to international area-based conservation targets in a biodiversity rich country Spain. *Journal for Nature Conservation*, *62*, 126019.

Romo, H., García-Barros, E., & Lobo, J. M. (2006). Identifying recorder-induced geographic bias in an Iberian butterfly database. *Ecography*, *29*, 873–885.

Rosemartin, A., Watkins, T., & Miller-Rushing, A. J. (2021). Monitoring phenology in US national parks through citizen science: Some preliminary lessons and prospects for protected areas. *Parks Stewardship Forum*, *37*, 552–560.

Rowley, J. J. L., Callaghan, C. T., Cutajar, T., Portway, C., Potter, K., Mahoney, S., Trembath, D. F., Flemons, P., & Woods, A. (2019). FrogID: Citizen scientists provide validated biodiversity data on frogs of Australia. *Herpetological Conservation and Biology*, *14*, 155–170.

Salmon, R. A., Rammell, S., Emeny, M. T., & Hartley, S. (2021). Citizens, scientists, and enablers: A tripartite model for citizen science projects. *Diversity*, *13*, 309.

Schmeller, D. S., Böhm, M., Arvanitidis, C., Barber-Meyer, S., Brummitt, N., Chandler, M., Chatzinikolaou, E., Costello, M., Ding, H., García-Moreno, J., Gill, M., Haase, P., Jones, M., Juillard, R., Magnusson, W. E., Martin, C. S., McGeoch, M., Mihoub, J., Pettorelli, N., … Belnap, J. (2017). Building capacity in biodiversity monitoring at the global scale. *Biodiversity and Conservation*, *26*, 2765–2790.

Sicacha-Parada, J., Steinsland, I., Cretois, B., & Borgelt, J. (2021). Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. *Spatial Statistics*, *42*, 100446.

Studd, E. K., Bates, A. E., Bramburger, A. J., Fernandes, T., Hayden, B., Henry, H. A. L., Humphries, M. M., Martin, R., McMeans, B. C., Moise, E. R. D., O'Sullivan, A. M., Sharma, S., Sinclair, B. J., Sutton, A. O., Templer, P. H., & Cooke, S. J. (2021). Nine maxims for the ecology of cold-climate winters. *BioScience*, *71*, 820–830.

Sutton, A. O., Studd, E. K., Fernandes, T., Bates, A. E., Bramburger, A. J., Cooke, S. J., Hayden, B., Henry, H. A. L., Humphries, M. M., Martin, R., McMeans, B., Moise, E., O'Sullivan, A. M., Sharma, S., & Templer, P. H. (2021). Frozen out: Unanswered questions about winter biology. *Environmental Reviews*, *29*, 1–12.

Tancoigne, E. (2019). Invisible brokers: "citizen science" on Twitter. *Journal of Science Communication*, *18*, A05.

Tiago, P., Ceia-Hasse, A., Marques, T. A., Capinha, C., & Pereira, H. M. (2017). Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Scientific Reports*, *7*, 12832.

Tredennick, A. T., Hooker, G., Ellner, S. P., & Adler, P. B. (2021). A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*, e03336.

Tulloch, A. I. T., Mustin, K., Possingham, H. P., Szabo, J. K., & Wilson, K. A. (2013). To boldly go where no volunteer has gone before: Predicting volunteer activity to prioritize surveys at the landscape scale. *Diversity and Distributions*, *19*, 465–480.

Venter, Z. S., Barton, D. N., Gundersen, V., Figari, H., & Nowell, M. S. (2021). Back to nature: Norwegians sustain increased recreational use of urban green space months after the COVID-19 outbreak. *Landscape and Urban Planning*, *214*, 104175.

Vimal, R., Navarro, L. M., Jones, Y., Wolf, F., Le Moguédec, G., & Réjou-Méchain, M. (2021). The global distribution of protected areas management strategies and their complementarity for biodiversity conservation. *Biological Conservation*, *256*, 109014.

Volenec, Z. M., & Dobson, A. P. (2020). Conservation value of small reserves. *Conservation Biology*, *34*, 66–79.

Weaver, D. B., & Lawton, L. J. (2017). A new visitation paradigm for protected areas. *Tourism Management*, *60*, 140–146.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., Fraçois, F., Gralemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*, 1686.

Wintle, B. A., Kujala, H., Whitehead, A., Cameron, A., Veloz, S., Kukkala, A., Moilanen, A., Gordon, A., Lentini, P. E., Cadenhead, N. C. R., & Bekessy, S. A. (2019). Global synthesis of conservation studies reveals the importance of small habitat patches for biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 909–914.

Wood, C., Sullivan, B., Iliff, M., Fink, D., & Kelling, S. (2011). eBird: Engaging birders in science and conservation. *PLoS Biology*, *9*(12), e1001220.

Zulian, V., Miller, D. A. W., & Ferraz, G. (2021). Integrating citizen-science and planned-survey data improves species distribution estimates. *Diversity and Distributions*, *27*, 2498–2509.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

S1.1 Data contributors to the biodiversity data accessed from GBIF, after filtering for inclusion in this study. *n* indicates the number of included data points contributed by the indicated data source. Data sources where $n = 0$ were present in the study area on GBIF but all data from these sources were excluded through the filtering described in the Methods section.

S2.2 Importance and model-averaged estimates and standard error for each covariate in the (a) citizen science and (b) professional model of biodiversity observations among grid cells in the study area.

S2.3 All negative binomial generalized linear models of (a) citizen science and (b) professional biodiversity observations within grid cells with a substantial level of support (ΔAICc < 2).

S3.2 Importance and model-averaged estimates and standard error for each covariate in the (a) citizen science and (b) Strava model among trail segment corridors in the study area.

S3.3 All negative binomial generalized linear models of (a) citizen science observations, standardized by trail length, and (b) reported Strava activities within trail segment corridors with a substantial level of support (ΔAICc < 2).

How to cite this article: Mandeville, C. P., Nilsen, E. B., & Finstad, A. G. (2022). Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use. *Ecological Solutions and Evidence*, *3*, e12185.
https://doi.org/10.1002/2688-8319.12185

SUPPORTING INFORMATION: Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use

Supporting information contents

TABLE SI 3.2. Importance and model-averaged estimates and standard error for each covariate in the (a) citizen science and (b) Strava model among trail segment corridors in the study area.

TABLE SI 3.3. All negative binomial generalized linear models of (a) citizen science observations, standardized by trail length, and (b) reported Strava activities within trail segment corridors with a substantial level of support ($\Delta AIC_c < 2$).

FIGURE SI 3.4. AIC_c weights of the 2000 highest rated negative binomial generalized linear models for the number of (a) citizen science observations, standardized by trail length, and (b) recorded Strava activities per trail segment, out of a set consisting of all possible combinations of the ten covariates with no interactions. Models below the red line have substantial support ($\Delta AIC_c < 2$).

TABLE SI 1.1. Data contributors to the biodiversity data accessed from GBIF, after filtering for inclusion in this study. *n* indicates the number of included data points contributed by the indicated data source. Data sources where *n* = 0 were present in the study area on GBIF but all data from these sources were excluded through the filtering described in the Methods section.

Data source	**Type of source**	**n (included)**
Norwegian Species Observation Service	Citizen science - opportunistic	40376
eBird Observation Dataset	Citizen science - opportunistic	3450
iNaturalist Research-grade Observations	Citizen science - opportunistic	299
Pl@ntNet	Citizen science - opportunistic	50
Skandobs	Citizen science - opportunistic	11
Naturgucker	Citizen science - opportunistic	11
Observation.org	Citizen science - opportunistic	6
Vascular plant herbarium TRH, NTNU University Museum	Professional	492
Lichen herbarium TRH, NTNU University Museum	Professional	291
Terrestrial and limnic invertebrates systematic collection, NTNU University Museum	Professional	246
Mycology herbarium TRH, NTNU University Museum	Professional	218
Fungi field notes, Oslo (O)	Professional	170
NINA insect database	Professional	120
International Barcode of Life project (iBOL)	Professional	100
BioFokus	Professional	99
Geographically tagged INSDC sequences	Professional	67
Bryophyte herbarium TRH, NTNU University Museum	Professional	53
Lichen field notes, Oslo (O)	Professional	29
Royal Botanic Garden Edinburgh Living Plant Collections (E)	Professional	28
Bird collection NTNU University Museum	Professional	27
Lichen herbarium, Oslo (O) UiO	Professional	25
Mycology herbarium, Oslo (O) UiO	Professional	23
NHMO DNA Bank Vascular plants collection	Professional	19
Vascular Plant Herbarium, Oslo (O) UiO	Professional	15
NHMO DNA Bank Fungi and Lichens collection	Professional	7
Danish Mycological Society, fungal records database	Professional	6
Artsprosjekt: hypogeous_macrofungi	Professional	4
Bryophyte Herbarium, Oslo (O) UiO	Professional	4
Herpetile collection NTNU University Museum	Professional	4
Entomological collections, UiB	Professional	3
Lichen herbarium, UiB	Professional	2

Algae herbarium TRH, NTNU University Museum	Professional	1
Mycology collection, Norwegian Forest and Landscape Institute	Professional	1
Reptilia notes, NTNU University Museum	Professional	1
Seabirds in Norway - Estimated population sizes	Professional	1
The cryptogamy collection (PC) at the Herbarium of the Muséum national d'Histoire Naturelle (MNHN - Paris)	Professional	1
Tropicos Specimen Data	Professional	1
Vascular plant herbarium (KMN) UiA	Professional	1
Norwegian Biodiversity Information Centre - Other datasets	Citizen science - structured	0
Algae collection, Oslo (O) UiO	Professional	0
Algae, Norwegian College of Fishery Science	Professional	0
Birds ringed with Norwegian rings 1914-1960	Professional	0
Birds ringed with Norwegian rings 1961-1990	Professional	0
Bryophyte herbarium, UiT Tromsø Museum	Professional	0
Collembola collection of Arne Fjellberg, Norway	Professional	0
Entomology collection, UiT Tromsø Museum	Professional	0
Entomology Division, Yale Peabody Museum	Professional	0
Entomology, Natural History Museum, University of Oslo	Professional	0
Fish collection NTNU University Museum	Professional	0
Herbarium GB, University of Gothenburg	Professional	0
Huitfeldt Kaas: Freswhater fish distribution in Norway 1918	Professional	0
Ims fish tag database	Professional	0
Lichen herbarium, UiT Tromsø Museum	Professional	0
Limnic freshwater benthic invertebrates biogeographical mapping/inventory NTNU University Museum	Professional	0
Limnic freshwater pelagic invertebrates biogeographical mapping/inventory NTNU University Museum	Professional	0
Lund Botanical Museum (LD	Professional	0
Mammal collection NTNU University Museum	Professional	0
Marine invertebrate collection NTNU University Museum	Professional	0
Mycology herbarium, UiT Tromsø Museum	Professional	0
fNational fish tag database	Professional	0

NHMO DNA Bank Fish and Herptile collection	Professional	0
NINA Vanndata fisk	Professional	0
NINA Vanndata øvrige arter	Professional	0
NMNH Extant Specimen Records	Professional	0
Notes from the Mycology Herbarium, Oslo (O)	Professional	0
NSW AVH data	Professional	0
Provincial Museum of Alberta, Edmonton, AB, Canada. Birds (Aves)	Professional	0
SEAPOP - Last observation per locality in breeding season	Professional	0
Thrips (Thysanoptera) in Norway	Professional	0
Vascular plant field notes, NTNU University Museum	Professional	0
Vascular plant herbarium, UiT Tromsø Museum	Professional	0
Vascular Plants, Field notes, Oslo (O)	Professional	0
Vascular Plants, Museum of Archaeology, University of Stavanger	Professional	0

FIGURE SI 1.2. Map indicates the year in which citizen science biodiversity data used in the study were collected. The 2016-2020 time period, which contains 41% of the citizen science data, corresponds with the dates of available Strava Metro data.

We chose to use citizen science data from 2000-2021 because we do not expect the distribution of citizen science or recreational trail use to have changed substantially from 2000 onwards; the trail network, access points, and other relevant environmental variables have remained largely unchanged during that time and there is no reason to believe user behavior would have changed in a systematic way. Given this expectation, we preferred to use the wider range of citizen science data because it allows for a larger sample size. We briefly tested this expectation by comparing the distribution of the citizen science data collected in the 2016-2020 time period with the rest of the citizen science data. Data from the two time periods were positively correlated (Pearson correlation $r = 0.45$, $p = < 0.0001$).

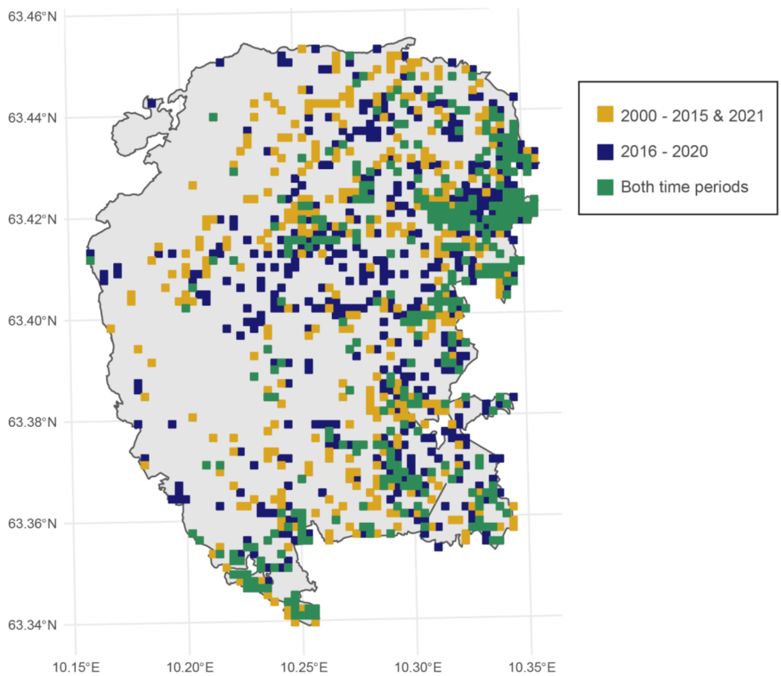


FIGURE SI 2.1. Model-averaged relative variable importance of each covariate for the models of (a) citizen science and (b) professional biodiversity data observations among grid cells in the study area.

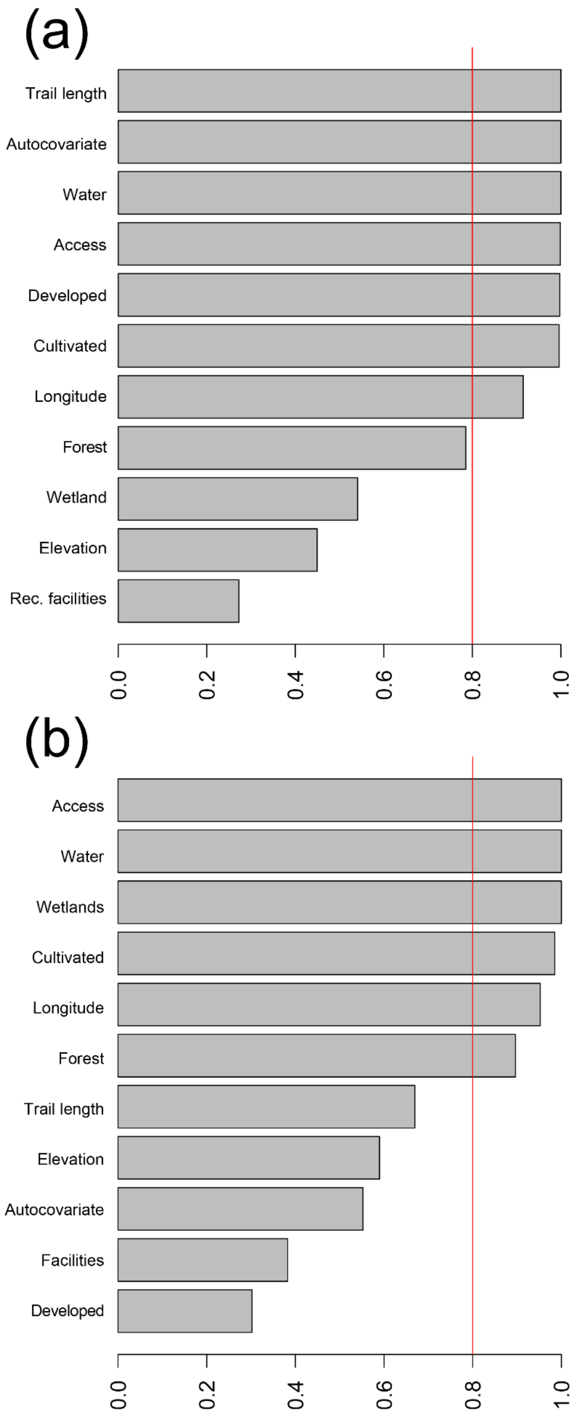


TABLE SI 2.2. Importance and model-averaged estimates and standard error for each covariate in the (a) citizen science and (b) professional model of biodiversity observations among grid cells in the study area.

	(a) Citizen science			(b) Professional		
	Importance	Estimate	Std. error	Importance	Estimate	Std. error
Intercept	1.00	0.21	0.06	1.00	-1.34	0.01
Total trail length	1.00	0.50	0.06	0.67	0.09	0.09
Presence of water	1.00	0.61	0.13	1.00	0.88	0.17
Proximity to access	1.00	0.28	0.07	1.00	0.75	0.09
Developed land cover	1.00	0.22	0.05	0.30	0.01	0.03
Cultivated land cover	1.00	0.19	0.05	0.98	0.20	0.07
Proximity to eastern edge	0.92	0.18	0.08	0.93	-0.29	0.10
Forest land cover	0.79	-0.10	0.08	0.90	-0.20	0.09
Wetlands land cover	0.54	0.05	0.06	1.00	-0.41	0.08
Elevation	0.45	-0.03	0.05	0.59	-0.06	0.08
Presence of facilities	0.27	-0.03	0.14	0.38	0.19	0.35

TABLE SI 2.3. All negative binomial generalized linear models of (a) citizen science and (b) professional biodiversity observations within grid cells with a substantial level of support (ΔAIC_c < 2).

(a) Citizen science				
Model	AIC_c	ΔAIC_c	k	Evidence weight
water + access + trails + longitude + developed + cultivated + forest + ac	10764.62	0.00	9	0.15
water + access + trails + longitude + elevation + developed + cultivated + forest + ac	10764.93	0.14	10	0.14
water + access + trails + longitude + developed + cultivated + forest + wetlands + ac	10765.07	0.45	10	0.13
water + access + trails + longitude + elevation + developed + cultivated + forest + wetlands + ac	10765.44	0.68	11	0.10
water + access + trails + longitude + developed + cultivated + wetlands + ac	10765.95	1.33	9	0.08
facilities + water + access + trails + longitude + developed + cultivated + forest + ac	10766.58	1.96	10	0.06
(b) Professional				
water + access + trails + longitude + elevation + cultivated + forest + wetlands	4914.83	0.00	9	0.09
water + access + trails + longitude + elevation + cultivated + forest + wetlands + ac	4915.29	0.46	10	0.07
water + access + trails + longitude + cultivated + forest + wetlands + ac	4915.44	0.61	9	0.07
facilities + water + access + trails + longitude + elevation + cultivated + forest + wetlands	4915.91	1.08	10	0.05
facilities + water + access + trails + longitude + elevation + cultivated + forest + wetlands + ac	4916.33	1.50	11	0.04
facilities + water + access + trails + longitude + cultivated + forest + wetlands + ac	4916.51	1.68	10	0.04
water + access + longitude + elevation + cultivated + forest + wetlands	4916.67	1.84	8	0.04
water + access + trails + longitude + cultivated + forest + wetlands	4916.79	1.96	8	0.03
water + access + trails + longitude + elevation + developed + cultivated + forest + wetlands	4916.80	1.97	10	0.03

FIGURE SI 2.4. AIC_c weights of the 2000 highest rated negative binomial generalized linear models for the number of (a) citizen science and (b) professional observations per grid cell, out of a set consisting of all possible combinations of the ten covariates with no interactions. Models below the red line have substantial support ($\Delta AIC_c < 2$).

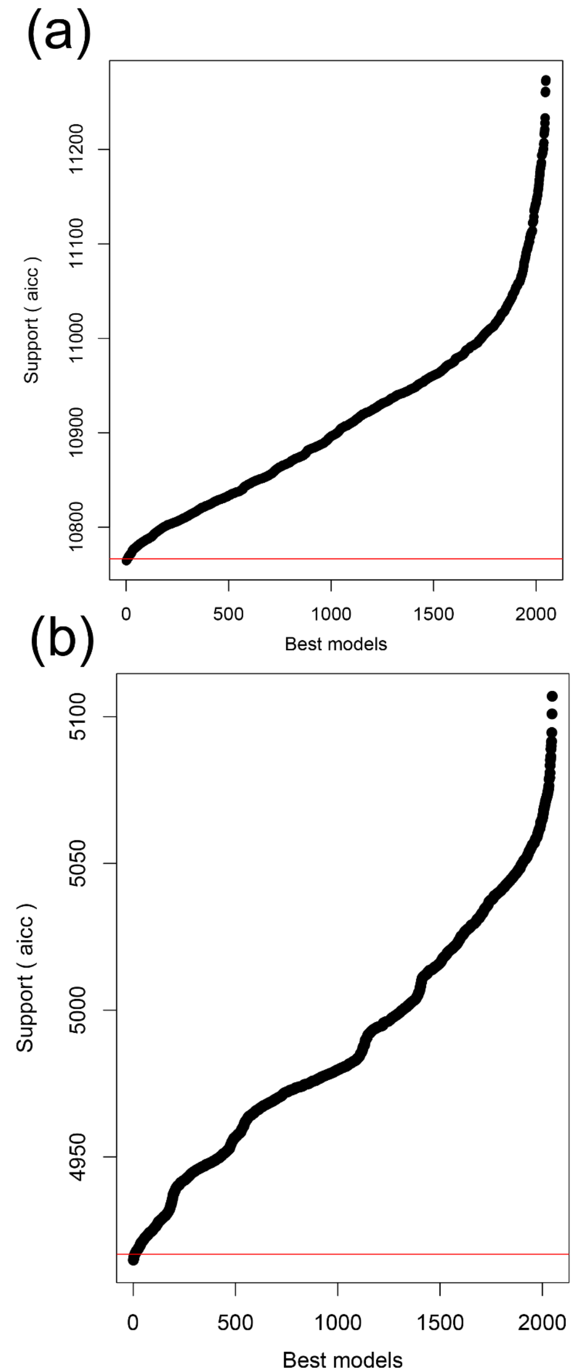


FIGURE SI 3.1. Model-averaged relative variable importance of each covariate for the models of (a) citizen science observations, standardized by trail length, and (b) reported Strava activities among trail segment corridors in the study area.

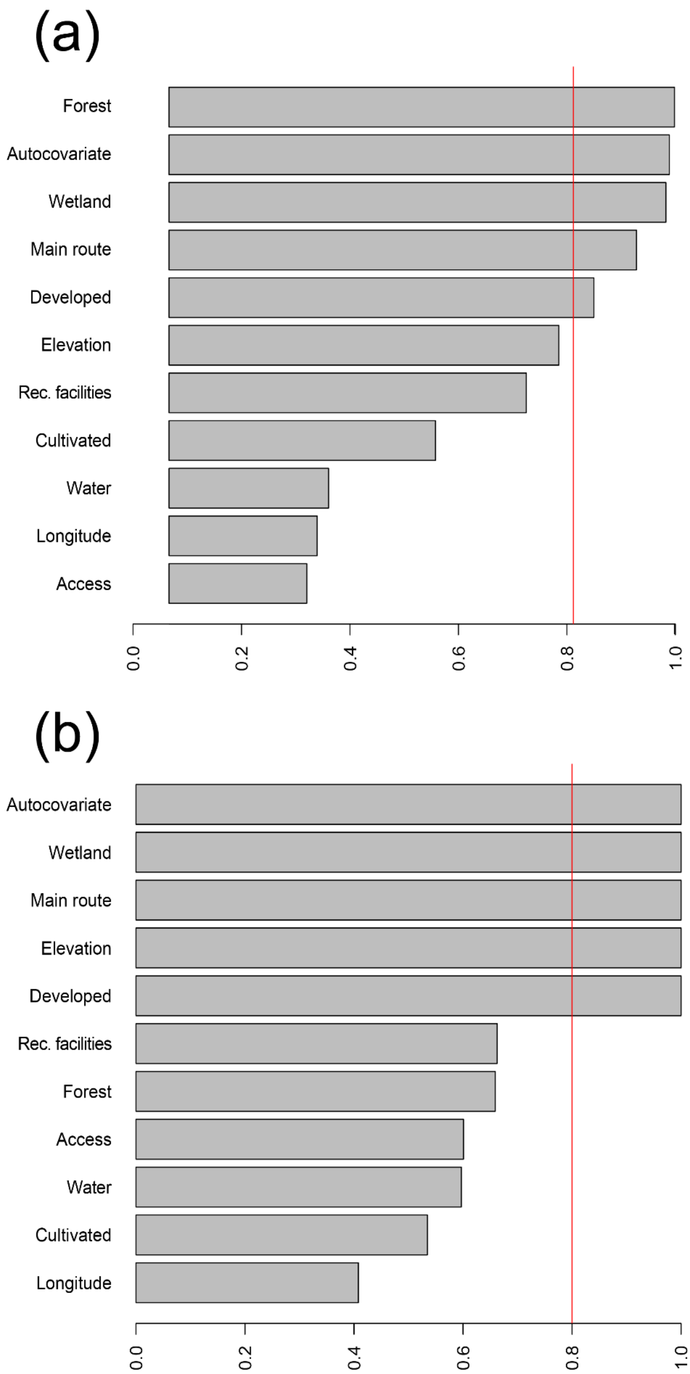


TABLE SI 3.2. Importance and model-averaged estimates and standard error for each covariate in the (a) citizen science and (b) Strava model among trail segment corridors in the study area.

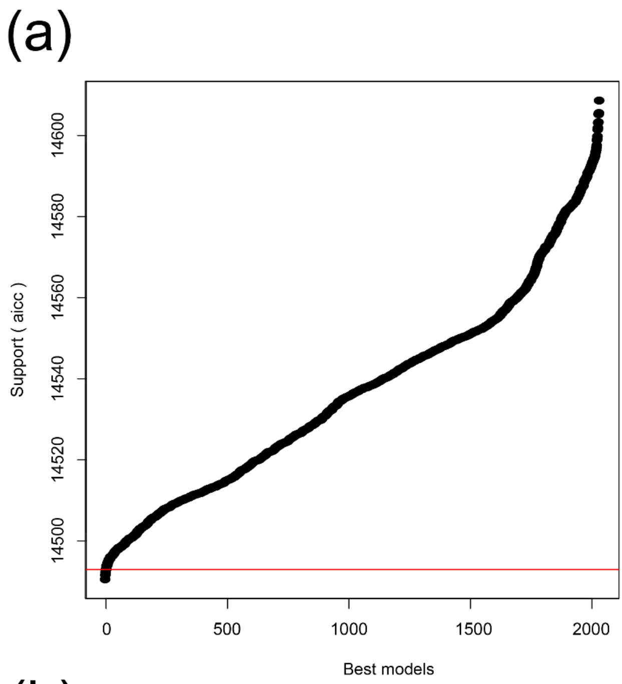
	(a) Citizen science			(b) Strava		
	Importance	Estimate	Std. error	Importance	Estimate	Std. error
Intercept	1.00	0.71	0.07	1.00	6.60	0.04
Forest land cover	1.00	-0.35	0.07	0.66	0.04	0.04
Wetland land cover	0.98	-0.17	0.06	1.00	0.27	0.03
Function as main route	0.92	0.15	0.07	1.00	0.64	0.03
Developed land cover	0.84	0.12	0.08	1.00	-0.13	0.03
Elevation	0.77	-0.10	0.07	1.00	0.19	0.03
Proximity to facilities	0.71	-0.11	0.09	0.66	0.04	0.04
Cultivated land cover	0.53	-0.06	0.07	0.53	0.02	0.03
Proximity to water	0.32	0.01	0.03	0.60	0.03	0.03
Proximity to eastern edge	0.29	0.00	0.03	0.41	0.02	0.03
Proximity to access	0.27	0.00	0.02	0.60	-0.03	0.03

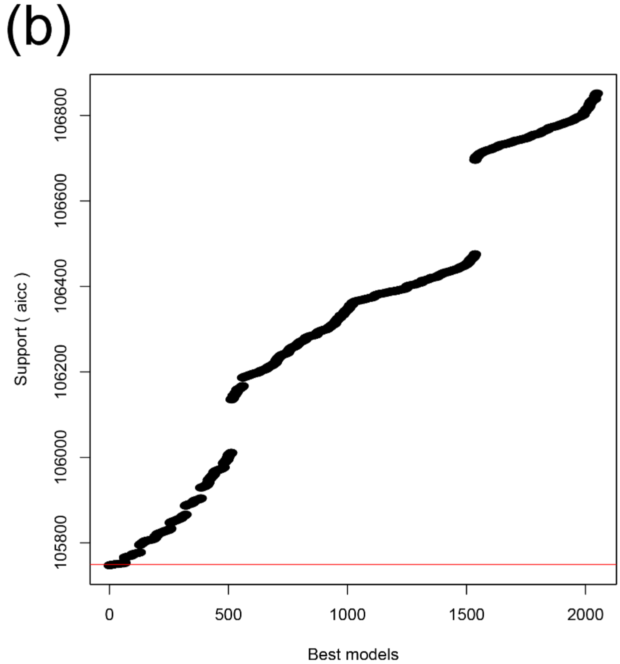
TABLE SI 3.3. All negative binomial generalized linear models of (a) citizen science observations, standardized by trail length, and (b) reported Strava activities within trail segment corridors with a substantial level of support ($\Delta AIC_c < 2$).

(a) Citizen science				
Model	AIC_c	ΔAIC_c	k	Evidence weight
mainPath + facilities + elevation + wetlands + forest + developed. + cultivated + ac	14490.37	0.00	9	0.10
mainPath + facilities + elevation + wetlands + forest + developed + ac	14491.27	0.90	8	0.06
mainPath + facilities + elevation + wetlands + forest + developed + cultivated + water + ac	14491.84	1.47	10	0.05
access + mainPath + facilities + elevation + wetlands + forest + developed + cultivated + ac	14492.35	1.98	10	0.04
(b) Strava				
access + mainPath + facilities + elevation + wetlands + forest + developed + cultivated + water + ac	105746.8	0.7	11	0.06
access + mainPath + facilities + elevation + wetlands + forest + developed + cultivated + ac	105747.5	1.0	10	0.05
access + mainPath + longitude + elevation + wetlands + forest + developed + cultivated + water + ac	105747.8	1.1	11	0.04
mainPath + facilities + elevation + wetlands + forest + developed + cultivated + water + ac	105747.9	1.1	10	0.04
access + mainPath + longitude + facilities + elevation + wetlands + forest + developed + cultivated + water + ac	105747.9	1.1	12	0.04
access + mainPath + facilities + elevation + wetlands + forest + developed + water + ac	105748.0	1.2	10	0.04
mainPath + facilities + elevation + wetlands + forest + developed + cultivated + ac	105748.1	1.3	9	0.03
access + mainPath + facilities + elevation + wetlands + forest + developed + ac	105748.5	1.7	9	0.03
access + mainPath + elevation + wetlands + forest + developed + cultivated + water + ac	105748.6	1.8	10	0.03

access + mainPath + longitude + elevation + wetlands + forest + developed + water + ac	105748.7	1.9	10	0.03
mainPath + facilities + elevation + wetlands + forest + developed + water + ac	105748.8	2.0	9	0.02
access + mainPath + facilities + elevation + wetlands + developed + ac	105748.8	2.0	8	0.02

FIGURE SI 3.4. AIC_c weights of the 2000 highest rated negative binomial generalized linear models for the number of (a) citizen science observations, standardized by trail length, and (b) recorded Strava activities per trail segment, out of a set consisting of all possible combinations of the ten covariates with no interactions. Models below the red line have substantial support ($\Delta AIC_c < 2$).





Article 3

Interspecific competition impacts the occupancy and range limits of two ptarmigan species along the elevation gradient in Norway

Caitlin P. Mandeville[1,2]*, Anders G. Finstad[1,2], Erlend B. Nilsen[3,4]

[1] Department of Natural History, Norwegian University of Science and Technology
[2] Centre for Biodiversity Dynamics, Norwegian University of Science and Technology
[3] Department of Terrestrial Biodiversity, Norwegian Institute for Nature Research
[4] Faculty of Biosciences and Aquaculture, Nord University, Steinkjer, Norway
* Corresponding author

Abstract

Many alpine species are expected to respond to climate change through upslope shifts of their range limits, but competition may restrict or alter this response. Under traditional range-limit theory, it is expected that lower-elevation species are better competitors than closely related higher-elevation species. However, recent work finds that this prediction is often unmet. We investigated evidence for the impact of competition during breeding season on the elevational range limits of a pair of closely related alpine bird species, willow ptarmigan (*Lagopus lagopus*) and rock ptarmigan (*L. muta*), in mainland Norway. The species share overlapping ranges loosely divided by the treeline ecotone, with willow ptarmigan generally occupying lower sites and rock ptarmigan occupying higher sites. We used multi-species occupancy models to test four competing hypotheses for how competition may affect the range limit between willow ptarmigan and rock ptarmigan: (1) asymmetric competition that restricts the lower range limit of rock ptarmigan; (2) asymmetric competition that restricts the upper range limit of willow ptarmigan; (3) condition-dependent competition that restricts both species' range limits; (4) range limits unaffected by competition. We found evidence for a negative pairwise interaction between the two species. Changes in interaction strength along the elevation gradient suggested evidence for condition-specific competition. However, a strong positive correlation between rock ptarmigan and higher-elevation habitat resulted in a highly asymmetric outcome, where the upper range limit of willow ptarmigan was restricted but rock ptarmigan occupancy was fairly independent of willow ptarmigan. This outcome is opposite to the prediction of traditional range-limit theory and may suggest a greater climate threat to willow ptarmigan than has been previously projected. Thus, our results demonstrate the importance of considering biotic interactions at both the higher and lower ends of species' range limits along elevation gradients.

This paper is awaiting publication and is not included in NTNU Open

Article 4

Open Data Practices among Users of Primary Biodiversity Data

CAITLIN P. MANDEVILLE, WOUTER KOCH, ERLEND B. NILSEN, AND ANDERS G. FINSTAD

Presence-only biodiversity data are increasingly relied on in biodiversity, ecology, and conservation research, driven by growing digital infrastructures that support open data sharing and reuse. Recent reviews of open biodiversity data have clearly documented the value of data sharing, but the extent to which the biodiversity research community has adopted open data practices remains unclear. We address this question by reviewing applications of presence-only primary biodiversity data, drawn from a variety of sources beyond open databases, in the indexed literature. We characterize how frequently researchers access open data relative to data from other sources, how often they share newly generated or collated data, and trends in metadata documentation and data citation. Our results indicate that biodiversity research commonly relies on presence-only data that are not openly available and neglects to make such data available. Improved data sharing and documentation will increase the value, reusability, and reproducibility of biodiversity research.

Keywords: applied ecology, biodiversity, informatics, monitoring and mapping, publication practices

Biodiversity data are increasingly made openly available, facilitated by extensive digital infrastructures that support data standardization and publication (Farley et al. 2018, Anderson et al. 2020, Kays et al. 2020). There is growing recognition that this open sharing of biodiversity data is critical for advancing biodiversity research (Farley et al. 2018). Some of the primary benefits of open biodiversity data include enhanced reproducibility of research (Alston and Rick 2021); making data available for reuse in new research applications (Chawinga and Zinn 2019); enabling researchers to receive credit, in the form of citations, for their efforts producing and sharing data sets (Costello et al. 2013, Brown 2021); and minimizing the duplication of research effort, enabling researchers to prioritize new data collection that fills research gaps (Troudet et al. 2017). As data sharing continues to become normalized, best practices have developed for the sharing of biodiversity data (Kühl et al. 2020). The FAIR data principles, for instance, outline four key attributes of effectively shared data: findable, accessible, interoperable, and reusable (Wilkinson et al. 2016). Specific practices have been developed to implement biodiversity data sharing in accordance with FAIR data principles. For example, global data aggregators such as the Global Biodiversity Information Facility (GBIF) provide a central location for aggregated data sets, ensuring that they will be findable and accessible (Robertson et al. 2014), and standardization schemes such as Darwin Core provide a mechanism for researchers to improve interoperability (Wieczorek et al. 2012). Such innovations support extensive data reuse; for example, the GBIF currently enables integrated data searches of nearly 1.7 billion species records from diverse sources around the world and has facilitated data reuse in thousands of publications (Heberling et al. 2021).

Although any type of data can be openly shared, the biodiversity data type most readily associated with open data sharing is presence-only occurrence data (König et al. 2019, Anderson et al. 2020, Wüest et al. 2020, Gadelha et al. 2021). Presence-only data consist of the taxonomic identification and location of an organism, often with the time of observation but without further information about species abundance, sampling design, or sites at which the species was not observed. The quantity of presence-only data aggregated in open biodiversity data repositories is immense and continuing to grow rapidly (Peterson et al. 2018, Ball-Damerow et al. 2019). This growth has been driven in large part by two simultaneous trends: the increasing popularity of citizen science platforms through which the public submit opportunistic observations to centralized databases (Theobald et al. 2015, Amano et al. 2016, Sullivan et al. 2017) and the digitization and aggregation of historical records and museum specimens (Speed et al. 2018, Nelson and Ellis 2019, Hedrick et al. 2020, Miller et al. 2020). The growing volume of openly shared presence-only data is also driven by characteristics of the data type itself: It is relatively simple and is easily standardized within existing best practices for data sharing (Anderson et al. 2020). Presence-only occurrence data now

https://doi.org/10.1093/biosci/biab072 Advance Access publication 18 August 2021

offer greater spatial, temporal, and taxonomic coverage on a global scale than other biodiversity data types and are often less costly and time intensive to collect (Tulloch et al. 2013, Bayraktarov et al. 2019).

As presence-only biodiversity data have grown in volume and accessibility, they have become increasingly common in biodiversity research (Peterson et al. 2018, Heberling et al. 2021). The open availability of massive modern and historical biodiversity data sets has contributed to a wide range of research areas, including ecology, biogeography, global change, and conservation (James et al. 2018, Ball-Damerow et al. 2019, Heberling et al. 2021). But the analysis of presence-only data is not without challenges; both historical and modern presence-only data are associated with limitations and biases that are distinct from other data types, both because of the lack of absence data and also because of the opportunistic collection process frequently associated with presence-only data (James et al. 2018, Støa et al. 2018, Gelfand and Shirota 2019, Grimmett et al. 2020, Sicacha-Parada et al. 2020, Johnston et al. 2021, Petersen et al. 2021). Further biases, errors, and limitations can be introduced in the processes of data preparation, publishing, and long-term maintenance (Tessarolo et al. 2017, Mesibov 2018), including the issues of data leakage (Peterson et al. 2018) and data obsolescence (Escribano et al. 2016). In response to these challenges, the growing application of presence-only data has been paralleled by an explosion of innovation in approaches to assess and improve both data accessibility and quality (Ball-Damerow et al. 2019) and also analysis methods that account for the specific limitations associated with this data type (Araújo et al. 2019, Kelling et al. 2019). As the development of analysis approaches for presence-only data continues, there is broad consensus that the documentation of metadata that details the study protocol, including information about sampling design or effort, allows for greater inference and also greater data reuse and reproducibility of analyses (Huettmann 2009, Kelling et al. 2019, Dobson et al. 2020, Foster et al. 2021). Open biodiversity data repositories commonly encourage the publishing of metadata (Poisot et al. 2019), but in practice the quality and amount of documented metadata varies widely (Peterson et al. 2018, Bishop et al. 2019, Anderson et al. 2020).

Although presence-only biodiversity data are reported and analyzed extensively in the traditional peer-reviewed literature, they are not restricted to it. In particular, authors who publish or access openly accessible biodiversity data may be more likely to seek out alternative outlets for research publication, such as preprint servers and journals with novel publishing models, because of their emphasis on free sharing of scientific information. Furthermore, biodiversity data are likely reported and analyzed often in gray literature and conference proceedings. Still, because a great deal of biodiversity data are reported and analyzed in the traditional peer-reviewed literature, it is important to understand the role that this literature plays in either facilitating or hindering the open sharing of biodiversity data. In this review we consider the extent of and barriers to the adoption of open data sharing practices within the traditional peer-reviewed literature, represented by the set of journals indexed by the Web of Science Core Collection.

Many aspects of the sharing and reuse of openly accessible biodiversity data in the peer-reviewed literature have been characterized, including common research applications of open data, taxonomic and spatial trends in open data, persistence of data stored in open databases, and current citation practices for open data (Troudet et al. 2017, Escribano et al. 2018, Ball-Damerow et al. 2019, Heberling et al. 2021, Luo et al. 2021). These studies make it clear that openly shared presence-only biodiversity data are foundational to a large body of biodiversity research. Still, many data go unshared. Earlier in the open data movement, it was widely recognized that open data formed just a small portion of the total biodiversity data known to exist (Ariño 2010, Amano et al. 2016, Peterson et al. 2018). But the current volume of presence-only data that are not openly shared, despite being presented and analyzed in the literature, is unknown. The concept of data sources and sinks can be helpful to conceptualize this issue; publication approaches that generate or perpetuate openly shared data can act as sources for continued data reuse, whereas publication approaches that entail a single use of data with no means for open access or reuse can be thought of as data sinks.

In the present article, we examine a broad cross section of the traditional peer-reviewed literature to assess the degree to which it serves as a source or sink for open presence-only biodiversity data. Our goal is to provide insight into the current adoption of open data practices among users of presence-only biodiversity data in journals indexed by the Web of Science Core Collection. To our knowledge, this is the first review of open data practices to be broadly defined by the presence-only data type, rather than by a particular type of data source, such as open databases. We focus on the following questions: How commonly does research published in articles indexed by the Web of Science Core Collection rely on presence-only data from open sources, and how commonly does it rely on data that are newly generated or compiled from other sources? To what extent do articles indexed by the Web of Science Core Collection serve as a data source for open presence-only biodiversity data; that is, are newly generated or compiled data made openly available, and are open data analyzed, documented, and cited in a way that supports continued reuse?

We identify both successes and challenges in the open sharing of presence-only biodiversity data, finding that the sharing of presence-only biodiversity data is overall increasing but that there is ample room for improvement in adherence to many data sharing best practices. We compare these findings with those of other recent reviews of the biodiversity literature, discussing trends that may be distinct to the presence-only data type, as well as new patterns that may be emerging within open data sharing practices. Because presence-only data are the biodiversity data type most

Box 1. The search string used to query the Web of Science Core Collection to obtain literature.

(((TS = ("presence-only" OR "presence only" OR "opportunistic observation*" OR "opportunistic species observation*" OR "opportunistic occurrence*" OR "opportunistic distribution*" OR "opportunistic species occurrence*" OR "opportunistic species distribution*" OR "pseudo-absence*" OR "pseudoabsence*" OR "inferred absence*" OR "presence-background" OR "presence background" OR "citizen science" OR "community science" OR "participatory science" or "ad hoc data" OR "ad hoc collection" OR "ad hoc method*" OR "incidental data" OR "incidental sighting*" OR "incidentally collected" OR "geographic one-class data" OR "incidental detection*" OR "opportunistic detection*" OR "primary biodiversity data*" OR "occurrence record*" OR "atlas data" OR "unstructured occurrence data" OR "unstructured species observation" OR "unstructured biodiversity data"))

AND (TS = ("distribution" OR "species" OR "biodiversity" OR "habitat*" OR "niche*")))

AND LANGUAGE: (English) *AND* DOCUMENT TYPES: (Article)

Indexes = SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan = All years

commonly associated with open data sharing, they can serve as an early indicator to illustrate the developing state of data sharing more broadly in the related fields of biodiversity, ecology, and conservation. Therefore, our characterization of current practices in presence-only data sharing can illuminate successes, challenges, and barriers to the adoption of data sharing practices that may be of growing relevance to the greater biodiversity research community.

Review of the presence-only biodiversity data literature

We searched the Web of Science Core Collection to target all scholarly articles that report on the application of presence-only biodiversity occurrence data. Our search targeted articles whose titles, abstracts, or keywords contained any of 31 terms commonly used in the literature to indicate presence-only data as well as any of 5 terms used to indicate biodiversity (box 1). We screened the abstracts of all returned articles and retained those that demonstrated the analysis or reporting of presence-only occurrence data. After screening, a total of 2151 articles were included in the review (see the extended methods description in supplemental file S1). Data management and bibliometric summary statistics were conducted in part with the bibliometrix package in R (Aria and Cuccurullo 2017).

To identify broad trends in applications of presence-only data, we classified all included articles into three topic clusters using latent dirichlet allocation (LDA) topic modeling. LDA topic modeling uses word associations within a corpus to identify topic clusters and assigns documents to the topic clusters on the basis of word frequency within each document (Westgate 2019). We classified each document on the basis of the words in the abstract and title. LDA topic modeling requires the desired number of clusters to be defined, so to select a number of topic clusters we conducted LDA analysis six times, each time producing a different number of clusters ranging from three to eight. We used two criteria to select the number of clusters in our final topic model: First, we assessed the clusters for lack of redundancy in an ordination of all articles by their highest rated topic classification, and, second, we assessed the redundancy and interpretability of the sets of most highly weighted words in each set of clusters (see supplemental file S2; Asmussen and Møller 2019, Westgate 2019). The modeling iteration that produced three topic clusters was least redundant and most interpretable. The topic clusters were assigned descriptive names on the basis of the words most characteristic of each cluster: *methodological* articles were characterized by terms related to the application and assessment of analysis methods; *applied* articles were characterized by terms related to topics in biodiversity science, conservation, and related fields; and *records* articles were characterized by terms related to the collection and reporting of occurrence data (figure 1). Topic modeling was conducted with the revtools package in R (Westgate 2019).

A subset of 300 articles randomly selected from the included articles was read in full and coded according to a standardized data sheet (see supplemental files S3 and S4). The 300-article subset was representative of the full data set in terms of publication year and topic area (figure 2). For each article read in full, we recorded information on 10 fields: taxa, study system, study and author region, sample size, study scale, sampling design, analysis approach, data source, and data publication (see supplemental file S3). For all data fields except for study region and author region, the classifications were not mutually exclusive; each article was tagged with all applicable responses. Such classification is a common approach in descriptive literature reviews (e.g., Ball-Damerow et al. 2019, Hao et al. 2019). All data management and analyses were conducted with R version 4.0.2 (R Core Team 2020), and data and R scripts are available online (Mandeville 2021).

Broad trends in the presence-only biodiversity literature

The literature relying on presence-only biodiversity occurrence data has grown steadily since the mid-2000s, maintaining an average annual growth rate that exceeds that of the biodiversity literature as a whole (Stork and Astrin 2014). This literature has seen a shift in recent years from

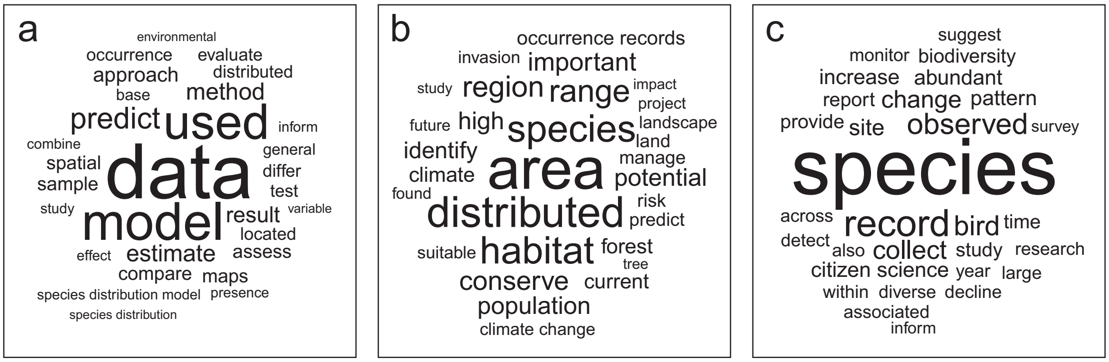


Figure 1. The articles were classified into three topic areas using latent dirichlet allocation (LDA) topic modeling, which uses word frequencies to cluster articles by topic. The 30 most heavily weighted words in (a) the methodological topic (n = 641), (b) the applied topic (n = 753), and (c) the records topic (n = 757) are shown in the present figure. Word size indicates relative weight within each topic.

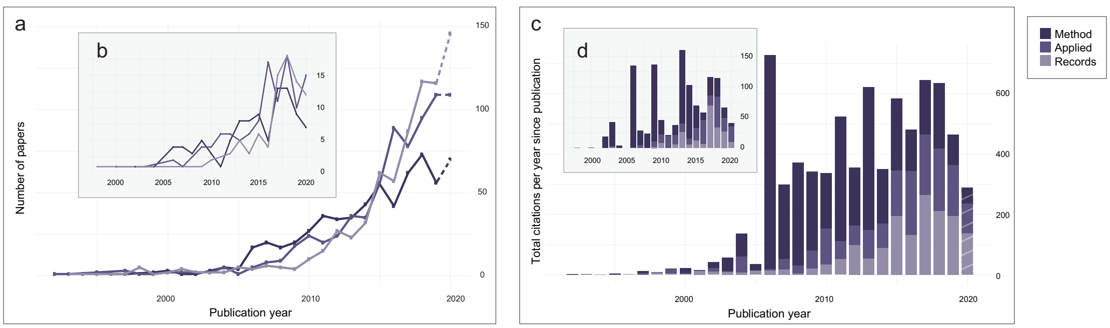


Figure 2. The number of articles published per year in each topic area within (a) the full set of 2151 articles and (b) the 300-article subset; the total citations per year since publication in each topic area within (c) the full set of 2151 articles and (d) the 300-article subset. 2020 is indicated with dashed lines because the results for 2020 may be less complete than those for other years; although the set of articles was obtained with a search on 4 January 2021, some articles with a 2020 publication date may not yet have been indexed by journals or the Web of Science.

a focus on methodological research to data sharing and applied analyses, as is evidenced by both the number of articles published and the citations obtained by articles in each topic area (figure 2). The *methodological* topic area was most common from the mid-2000s through 2015. From 2015 to 2020, the frequency of articles within the *methodological* topic area remained relatively constant, whereas the frequency of *applied* and *records* articles increased rapidly. *Methodological* articles are overall the most highly cited, but the relative citation rate has declined since 2015 (figure 2). The shifting distribution of topic areas suggests that there are two distinct eras in the presence-only data literature: an era focused on methodological developments, which lasted from approximately 2005–2015 and an era with a greater focus on applications that began in 2015 and continues today. A similar trend has been reported among articles that rely on GBIF-mediated data (Heberling et al. 2021).

The increase in articles focused on simple reports of occurrence is likely due to an increase in infrastructure and incentivization for data papers in recent years (Chavan and Penev 2011, Ball-Damerow et al. 2019, Li et al. 2020), and the parallel increase in applied research may indicate that presence-only approaches are being used more frequently to address issues of relevance to conservation and management (Guisan et al. 2013, Tulloch et al. 2018, Bayraktarov et al. 2019). The decline of methodological articles in terms of relative frequency and citation rate might suggest that applied researchers are using more established analysis methods more often than they are adopting newer approaches.

As a whole, the literature relying on presence-only biodiversity data is relatively decentralized and young. Its influence, as was measured by citations, is still growing; just a small number of the reviewed articles were highly cited, with a median of six citations per article. Unsurprisingly, *methodological* articles made up the majority of the 89 articles cited more than 100 times (figure 2; see supplemental file S5). The average author contributed to just 1.3 of the reviewed articles, which aligns with trends reported in the biodiversity literature (Stork and Astrin 2014) but is substantially lower than authorship rates in the biological sciences overall (Fanelli and Larivière 2016). Articles were published in a wide range of outlets, with 482 distinct journals represented in our review. The relative lack of common references is a further indicator of the varied scope of the presence-only biodiversity literature (see supplemental file S5). This is likely due to specialization among biodiversity researchers within many distinct research areas, defined for example by taxon of interest, geographic region, or scientific subdiscipline. Nevertheless, it may indicate a challenge to the efficient sharing of information regarding best practices for biodiversity data sharing.

Using complementary reviews to build a more complete picture of the biodiversity literature

All efforts to systematically review literature contain tradeoffs and biases introduced by the strategy used to search the literature, including search terms, search platform, and screening protocol. Therefore, efforts to characterize a body of literature are most informative when complementary reviews are considered alongside one another to form a more complete picture of the literature as a whole. We expect that this is particularly true for rapidly expanding research areas, including the presence-only biodiversity data literature; reviews of presence-only biodiversity data are complicated by the broad and rapidly developing variety of ways that this data type is accessed, analyzed, and referred to in the literature. To this end, we conducted a small test of the similarity of our search results to those of two recently published complementary reviews: Ball-Damerow and colleagues (2019) and the 2019 GBIF Science Review (GBIF Secretariat 2019). Each of these reviews used a search strategy and platform that complements our own, targeting a distinct subset of the literature on applications of presence-only biodiversity data (figure 3).

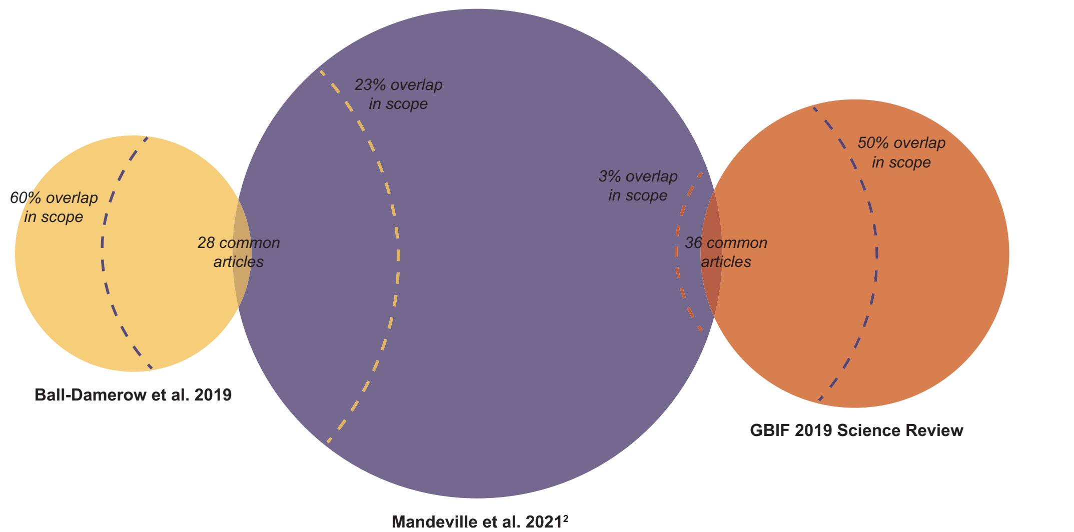
For this test, we identified the articles from our review that met the inclusion criteria defined for each of the other two reviews, screened the abstracts of 50 articles randomly selected from each of the other reviews according to our own inclusion criteria, and identified the percentage of articles that were common to our review and each of the complementary reviews. There was relatively little overlap between the articles in our review and the other two reviews (figure 3). The lack of overlap illustrates the importance of considering complementary reviews alongside one another. Although other recent reviews, including the two considered in the present article, have focused largely on applications of presence-only biodiversity data known to be accessed from open sources, our review fills a key knowledge gap by characterizing a broad set of the traditional literature with an as yet unknown reliance on open databases.

Comparison of basic study characteristics with trends in biodiversity research. Our review joins several recent studies in identifying trends in basic characteristics of the biodiversity literature, including taxonomic focus, study domain, and study region (Tydecks et al. 2018, Ball-Damerow et al. 2019, Heberling et al. 2021). We found that the articles in our review align some general trends in the biodiversity literature, including an emphasis on terrestrial settings (figures 4 and 5; Tydecks et al. 2018, Ball-Damerow et al. 2019, Heberling et al. 2021). Still, there are some distinct trends associated with the articles in our review: vertebrates—and, to a lesser extent, invertebrates—are better represented among our reviewed articles than in other reviews of the biodiversity literature, whereas plants and the freshwater domain are underrepresented (figure 4; Tydecks et al. 2018, Ball-Damerow et al. 2019, Heberling et al. 2021). The overrepresentation of vertebrates in our review is primarily due to their prevalence in reviewed articles that did not use data from open databases, suggesting that the range of vertebrate data available from open databases may not be as aligned with research needs as data from other taxonomic groups. On the other hand, the relative underrepresentation of freshwater and marine studies in our review was consistent between articles that did and did not rely on open data. This suggests that the presence-only data type as a whole may be less common in freshwater and marine domains, likely because many freshwater and marine species are not as easily detected via opportunistic observation.

The global distribution of studies in our review aligns closely with trends in the biodiversity literature (Tydecks et al. 2018, Heberling et al. 2021). The largest number of articles were authored by researchers based in Europe, followed by North America (figure 4). Alignment between study region and author region was uneven; articles that addressed Europe and North America were written by first authors based at institutions in the same region in respectively 98% and 95% of cases, whereas articles that addressed study regions in other parts of the world were less likely to have been written by first authors based in the focus region (figure 6). The uneven global distribution of biodiversity research reflects the greater coverage of biodiversity data in North America, Europe, and Australia relative to much of the rest of the world (Serra-Diaz et al. 2017, Pelayo-Villamil et al. 2018, Wüest et al. 2020) and is also partially explained by the less frequent publication of ecological research conducted in the Global South in journals that are indexed by major databases (Nuñez et al. 2019). It is critical that the field of biodiversity advances to better represent and support researchers based in underrepresented global regions in the international academic

	Ball-Damerow et al. 2019	Mandeville et al. 2021	GBIF 2019 Science Review
Number of reviewed articles	501 articles	2151 articles (300 screened in greater detail)	854 articles
Search platform	• Google Scholar • Selected a predetermined number of returned articles	• Web of Science Core Collection	• GBIF literature tracking programme[1]
Article inclusion criteria	• Primary biodiversity data accessed from openly accessible online database • Published between 2010 and April 2017	• Presence-only biodiversity occurrence data • Published before January 2021	• Mention or citation of GBIF or GBIF data • Published in 2018

[1] Draws from Google Scholar, Scopus, Wiley Online Library, SpringerLink, NCBI Pubmed, and bioRxiv



[2] Circle size refers to the 2151 articles used in a portion of analyses; 300 of these were screened in greater detail for further analyses.

Figure 3. The Venn diagram indicates the overlap between articles included in this review and two complementary reviews. The circle size corresponds to review sample size; it should be noted that only a portion of the analyses reported in Mandeville (2021) were conducted on the full article set, whereas the remaining analyses were conducted on a subset of 300 samples chosen randomly from the full set. The overlap between the circles indicates the overlap in articles included in each review, and the dotted lines indicate the estimated overlap in targeted articles according to the reviews' described inclusion criteria. The inset table indicates the inclusion criteria and search strategy of each review.

literature (Ramirez et al. 2018, Nuñez et al. 2019, Pettorelli et al. 2021). It has been shown that international collaborations are often inequitable, with European and North American researchers gaining more benefits in terms of publications and reputation than collaborators in the Global South (Boshoff 2009, Habel et al. 2014, Di Marco et al. 2017, Tydecks et al. 2018, Heberling et al. 2021). This trend should prompt caution in the growing open data

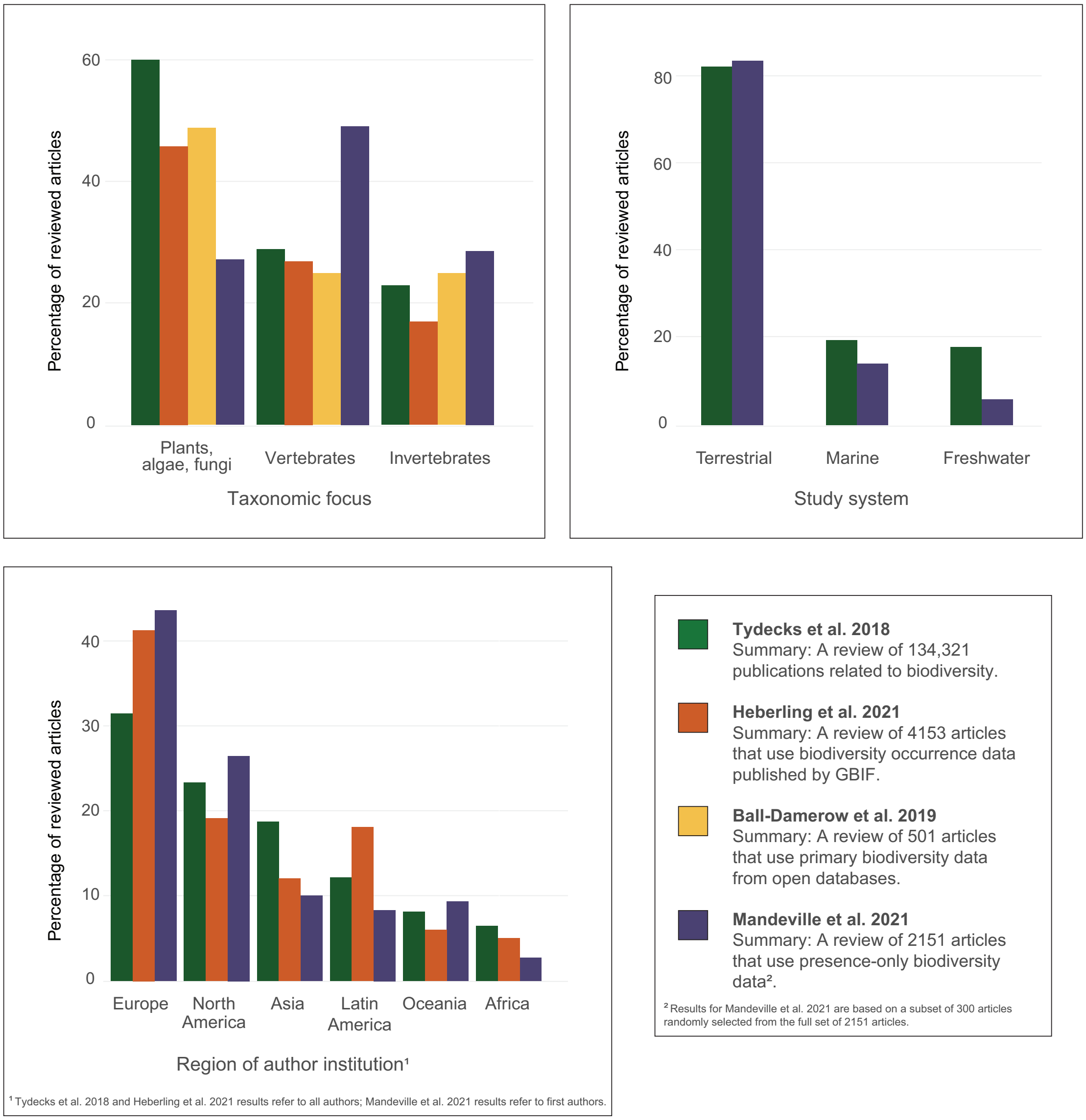


Figure 4. A comparison of trends in taxonomic focus, study system, and geographic region of the biodiversity literature identified by this review and three complementary reviews covering different aspects of the biodiversity literature. See each cited paper for specific methods and results, because the methods of defining and measuring each trend may differ slightly between articles.

movement; it will be essential to ensure that open sharing of data is supportive rather than exploitative of Global South researchers (Serwadda et al. 2018, Eichhorn et al. 2020, Pettorelli et al. 2021, Trisos et al. 2021). One example of an approach to this issue from within the biodiversity data community is the ongoing effort to repatriate biodiversity data that have been collected within a historically exploited region but stored and managed elsewhere, in order to transfer primary data custody and decision-making power back to the communities from which the data were collected (Dias et al. 2017, Eichhorn et al. 2020, Heberling et al. 2021).

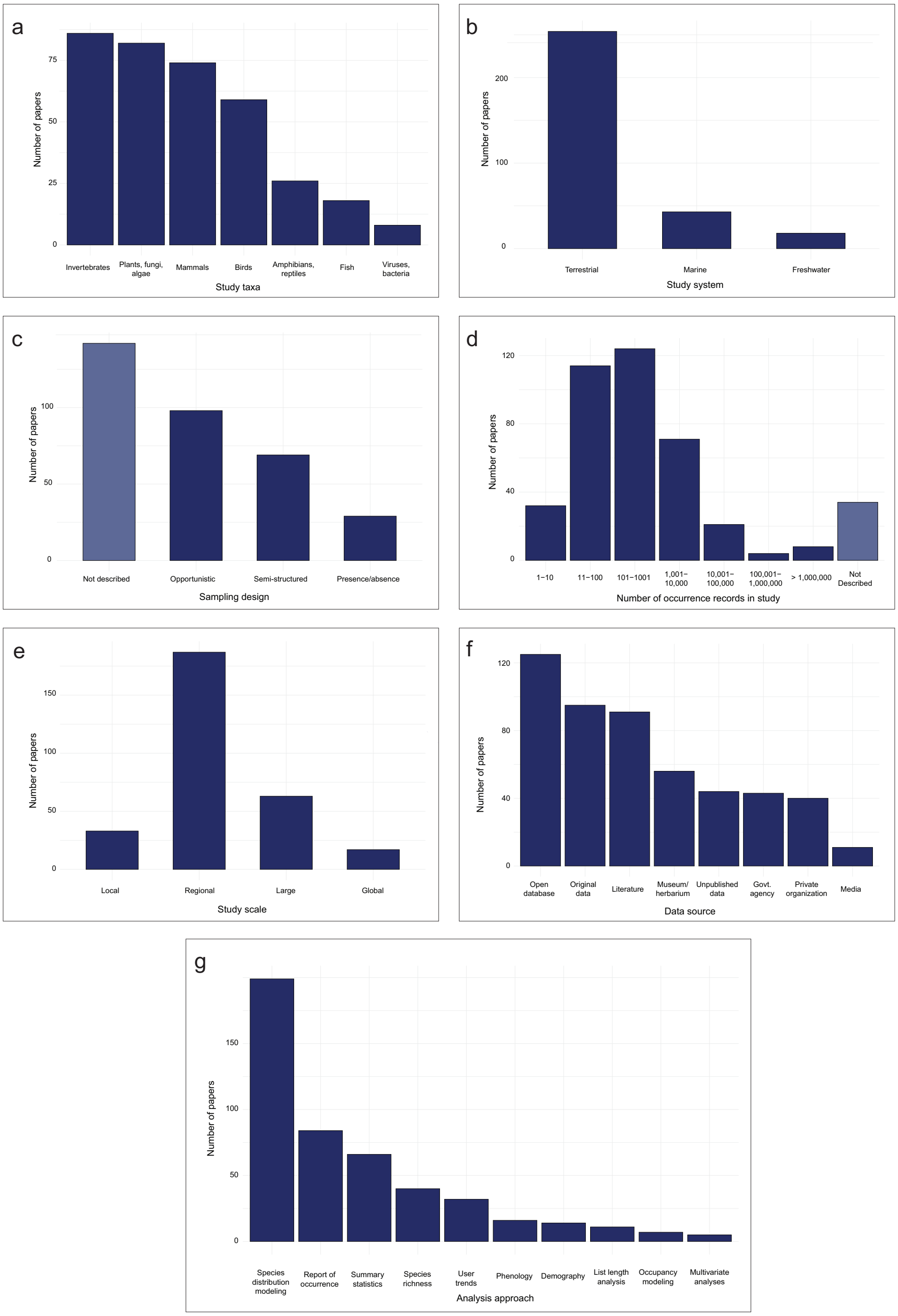


Figure 5. The frequency of characteristics among the subset of 300 randomly selected articles: (a) study taxa, (b) study system, (c) sampling design, (d) sample size, (e) study scale, (f) direct data source, and (g) analysis approach. Characteristics are not mutually exclusive; multiple responses per characteristic can apply to an article.

Figure 6. The study regions of the subset of 300 articles are indicated on the y-axis and the region of the first author of each article, defined by institutional affiliation, is indicated on the x-axis. The number in each cell indicates the number of articles written about the region on the y-axis by a first author based in the corresponding region on the x-axis.

Presence-only data: A lens into current trends in the access, analysis, and publishing of openly accessible biodiversity data. As the biodiversity research literature continues to grow, the open sharing of biodiversity data is increasingly recognized as necessary and is quickly becoming normalized (Peterson et al. 2018, Ball-Damerow et al. 2019, Heberling et al. 2021). Presence-only biodiversity data are relatively representative of broad taxonomic and geographic trends associated with the field of biodiversity as a whole, but they differ in the ease with which they can be shared in accordance with currently recognized best practices (König et al. 2019, Anderson et al. 2020, Wüest et al. 2020, Gadelha et al. 2021). Therefore, as practices continue to be developed to facilitate the sharing of a wide range of data types (Anderson et al. 2020), presence-only data can serve as an early indicator to illustrate the progress, challenges, and limitations to the adoption of biodiversity data sharing practices. The work of recent reviews focused on presence-only data from open databases (e.g., Ball-Damerow et al. 2019 and the GBIF Science Review series) makes it clear that open data infrastructure actively supports a large body of research. But to understand the extent to which biodiversity research in the traditional peer-reviewed literature serves to facilitate or slow the progress toward open data, it is necessary to consider presence-only data from a wider range of sources.

In the sections that follow, we focus on three aspects of the presence-only biodiversity data literature indexed in the Web of Science Core Collection, with an emphasis on open data practices. We first consider the sources of presence-only data in this body of literature. Next, we consider how presence-only data are analyzed and whether these analyses are supported by well-documented metadata. Finally, we characterize the data publication practices associated with the presence-only biodiversity data in this set of literature. Our objective is to delineate the current state of data sharing practices and to identify areas for growth, many of which will apply to both presence-only data and also more generally to a range of biodiversity data types.

Sources of presence-only biodiversity data

Openly accessible databases—that is, searchable online repositories in which biodiversity data from many original sources are aggregated—make billions of biodiversity data points freely available for anyone to access and use (Peterson et al. 2018, Ball-Damerow et al. 2019). Researchers may choose to access data from openly accessible databases for many reasons: to avoid duplicating research effort that has been undertaken in the past, to access data on a larger temporal and spatial scale than could be collected through original field work, to synthesize data from disparate sources, or to replicate or build on a previous study. So it is unsurprising that openly accessible databases were the most common direct data source in our review, accessed by 42% of the reviewed articles. However, only 19% of the reviewed articles used data exclusively from open databases; the vast majority accessed some or all of their data from sources other than open databases. Other common data sources include original fieldwork, the literature, and museums and herbaria (figure 5). Ball-Damerow and colleagues (2019) identified these same three sources of occurrence data as the most commonly integrated with occurrence data accessed from open databases.

In many cases, it is likely that researchers choose to collect new data or compile data from a variety of original sources because the data they need are not available in an openly accessible database (Troudet et al. 2017, Ball-Damerow et al. 2019). For instance, articles in our review were substantially more likely to address vertebrate species than in reviews in which all articles rely at least partially on open data (figure 4). In particular, a large percentage of the articles in our review addressed mammals (figure 5). Although mammals are considered overrepresented in open databases on a per-species basis, they make up a relatively small portion of the total volume of data available from open databases, likely because of many mammal species' lower detection probability, wider-ranging distributions, and relatively lower dedicated citizen science interest than some other taxa (Troudet et al. 2017, Parsons et al. 2018). This may explain why articles that addressed mammal species were relatively unlikely to obtain data from an open database and more likely to obtain data from government agencies, private organizations, and through original data collection. Overall, the relatively small percentage of articles based on open presence-only data corroborates a growing sentiment from the literature: Although the volume of openly accessible biodiversity data continues to grow, there are substantial taxonomic and spatial gaps for which there is minimal open data (Pino-Del-Carpio et al. 2014, Chambers et al. 2017, Troudet et al. 2017, Ondei et al. 2018, Wetzel et al. 2018, Ball-Damerow et al. 2019, Hochkirch et al. 2020). Our results corroborate the many studies that have identified gaps in biodiversity data, making it clear that the majority of researchers who conduct presence-only analyses do not find the data they need in open databases. This highlights the need for the biodiversity research community to continue ongoing efforts to identify and fill critical taxonomic and spatial knowledge gaps in open databases.

Data gaps can be filled through both novel data collection and mobilization of existing data that are not yet openly accessible. Many large pools of data exist outside the open data infrastructure—for example, in government agencies and private organizations (Stephenson et al. 2017, Wetzel et al. 2018, Cretois et al. 2020). Identifying these sources of data, supporting policies and infrastructure that facilitate their access and reuse, and incentivizing data sharing at an institutional level is needed to facilitate more open access to these data (Voříšek et al. 2018). This is critical for establishing the long-term records that are essential for studying trends across space and time and informing conservation interventions in the face of global change (Wetzel et al. 2018). Opening existing data for reuse is also necessary to avoid duplication of data collection effort and research waste, freeing research resources to target true data gaps (Grainger et al. 2020). Consider, for example, that 13% of the articles in our review accessed data from 10 or more nonopen sources, some accessing well over one thousand distinct sources. The collation of data from multiple sources represents an extensive research effort that will likely need to be repeated by future researchers if the data are not made more openly accessible. Reducing inefficiencies by supporting the access and reuse of data will allow researchers to prioritize generation of data that will fill gaps in the available knowledge. To achieve this, efforts to build relationships between data aggregators and the research community will continue to be essential.

In other cases, openly accessible data may be available to replace or supplement data from other sources but authors may neglect to use it, either because they are not aware of it or because they do not trust its quality (Faith et al. 2013). Even when data are aggregated in an open database, some researchers may choose to access the data from their original sources rather than from the open database (Singer et al. 2020). In some cases, researchers may be aware of open data but believe they lack the skills to access and use it effectively (Poisot et al. 2019). Indeed, a broad survey of researchers found that the perceived value and efficiency of reusing open data were major factors in whether researchers chose to access open data (Curty et al. 2017). Finally, it is also important to note that inequities in technological infrastructure, competence, and training mean that access to digital platforms is also inequitable (Johnson et al. 2021). Finding solutions to the barriers that keep researchers from accessing open biodiversity data should be a goal of the biodiversity research community.

Practices for accessing and citing open data vary widely. Among open databases, data sources varied widely. We identified 117 open databases that were used to access presence-only occurrence data (see supplemental file S6). We classified nine of these as large open databases, defined as relatively well known, established databases that contain data covering a very large geographic range, a wide range of taxa, or both. The most commonly accessed was the GBIF, which was accessed by 37 articles, followed by eBird (9 articles). The remaining 108 open databases, classified as small databases, had a narrower geographic or disciplinary scope and were each accessed by an average of 1.2 articles. Of the articles that accessed open data from at least one source, 55% accessed a large database and 65% accessed a small database. Two thirds directly accessed just one database, whereas the remaining third accessed between two and 10 distinct open databases. Of course, because many open data sources serve to aggregate many smaller databases, data users that accessed just one database may still have obtained data from a wide range of original sources. These results are similar to the findings of Ball-Damerow and colleagues (2019), who also found that a small number of open data sources were cited by many articles, whereas a large number of open data sources were cited very few times.

The frequent reliance on small open databases is probably due in large part to the prevalence of small databases within specific research areas (Costello and Wieczorek 2014, Ball-Damerow et al. 2019, Singer et al. 2020) and may also be partially explained by a lack of familiarity with or trust in large databases (Faith et al. 2013). We recognize many values of small databases, including responsiveness to specific disciplinary requirements (Franz and Sterner 2018) and the cultivation of strong relationships between data curators and communities of data users (Blair et al. 2020, Monfils et al. 2020). However, small open databases may lack the standardization and interoperability that are built into larger data aggregators (Poisot et al. 2019), they may lack consistent leadership to maintain growing content and keep up with developing best practices (Costello et al. 2013), and they are more likely to become technologically obsolete, rendering the data inaccessible (Vines et al. 2014, Tessarolo et al. 2017, Ball-Damerow et al. 2019, Blair et al. 2020).

We attempted to access all of the databases referred to in our reviewed articles and found that we could not locate or access 9% of the small databases from which articles in our review had obtained data. In a few other cases, the database website could be accessed, but it was not clear that the data were still accessible; for example, data could be visualized but the link to download data was broken, or it was requested that visitors contact the database managers to request access. Although still concerning, it is perhaps a cause for cautious optimism that the proportion of inaccessible databases in our review is considerably lower than the 26% of databases found to be inaccessible by Ball-Damerow and colleagues (2019), who reviewed articles published through April 2017. An additional 15% of the small databases had been consolidated into a different database but were still accessible. All nine large databases remained accessible. Because of the important role played by small databases, we do not intend to suggest that authors avoid them; rather, we caution the biodiversity data community to be cognizant that these small databases are strongly relied on and to be proactive about supporting them over time (Costello and Wieczorek 2014). The true reliance on small databases is likely to be even higher than identified in our study because small regional databases may be cited more frequently by articles published in regional journals and gray literature, which may not be indexed by the Web of Science and so may have been underrepresented in our search (Calver et al. 2017).

The proliferation of open data aggregators, along with the rapidly evolving best practices for their use, has resulted in an uneven landscape of how such data are cited in the literature (Escribano et al. 2018, Ball-Damerow et al. 2019, Luo et al. 2021). Citation of a digital object identifier (DOI) that is uniquely connected to the full data set analyzed in an article has emerged as the best practice in this area (Brown 2021, Heberling et al. 2021); this practice enables the data set to be clearly replicated and all original sources to be credited (Escribano et al. 2018, Luo et al. 2021). But not all researchers are yet aware of this best practice, because it is relatively new. Furthermore, not all open databases have a clear mechanism for producing a citable DOI (Altman and Crosas 2014, Penev et al. 2017). We found a great deal of variation in how open databases were cited among the articles in our review. The vast majority of articles simply listed the names of the databases from which they obtained data, sometimes accompanied by a brief description of the type of original sources from which the data were aggregated. Only 4% of the data sets accessed from an open database were cited with a DOI, and another 3% were not cited but, instead, were described in the text of the article with a direct link to the full data set or other thorough directions that would enable a reader to replicate the data retrieval process. Interestingly, the proportion of articles in our review that included a database citation with a URL or DOI was much lower than the 34% observed by Ball-Damerow and colleagues (2019). This may reflect a difference in search strategy; the search terms used by Ball-Damerow and colleagues (2019) ensured that all reviewed articles at least mentioned the type of database accessed, whereas our search terms required only that articles mentioned the type of data. The differing results obtained by these two searches suggest that the use of appropriate citation practices may be correlated with authors' use of specific terminology to refer to open databases, perhaps signaling their perception of their work as related to the open data movement.

A small number of authors in our review found alternative ways to recognize original providers of data even when there was no mechanism to do so through the open database—for example, by listing all original data sources in the supplemental material. Giving credit to the original providers of open data is critical for incentivizing data sharing to researchers, institutions, and funders (Escribano et al. 2018, Ball-Damerow et al. 2019, Groom et al. 2020) and for recognizing and supporting the diverse landscape of organizations and institutions that engage in biodiversity monitoring (Kühl et al. 2020). This may be especially true when data were collected through public involvement in citizen science. Thirty-four percent of the articles in our review identified citizen science as the original source of some or all of their data, although the true percentage of articles that derived data from citizen science is likely higher because citizen science data are frequently reused without their source being clearly described (Cooper et al. 2014). Citizen science plays an important role in biodiversity data collection but long-term funding and support for many citizen science programs may be dependent on the demonstrated impact, so appropriate citation is critical (Chandler et al. 2017, Pearce-Higgins et al. 2018, MacPhail and Colla 2020, Mandeville and Finstad 2021).

Analysis and reporting of presence-only biodiversity data and associated metadata

The growth of interest in presence-only data in the mid-2000s was paralleled by innovation in species distribution modeling approaches tailored to this data type (Vaz et al. 2015, Araújo et al. 2019, Ball-Damerow et al. 2019), so it is unsurprising that species distribution modeling was the dominant analysis approach in our review (figure 5). These methods have become increasingly sophisticated and widely popular (Hao et al. 2019, Norberg et al. 2019, Zurell et al. 2020). A large review of articles that use GBIF data found a similar prevalence of species distribution modeling and identified a recent transition in focus from methodological developments to widespread application similar to that seen in our overall set of reviewed articles (Heberling et al. 2021). Although the initial development of species distribution modeling approaches for presence-only data was at least partially a response to the increased availability of the data type, we suggest that their subsequent wide adoption has created a positive feedback effect whereby researchers, driven by the growing ease of analyzing presence-only data, have increasingly begun to seek out presence-only data from a wider range of sources.

Despite its prevalence, however, species distribution modeling is far from the only analysis method applicable to presence-only data. Our results illustrate a wide range of analysis approaches, including both inferential statistics and a variety of descriptive statistics. Presence-only data are also occasionally used indirectly—for example, to validate the results of another analysis or to inform a sampling design. Methodological innovation in inferential approaches is ongoing, and since 2012, a number of articles have applied a variety of less common inferential approaches, including phenology analyses, demography analyses, list length

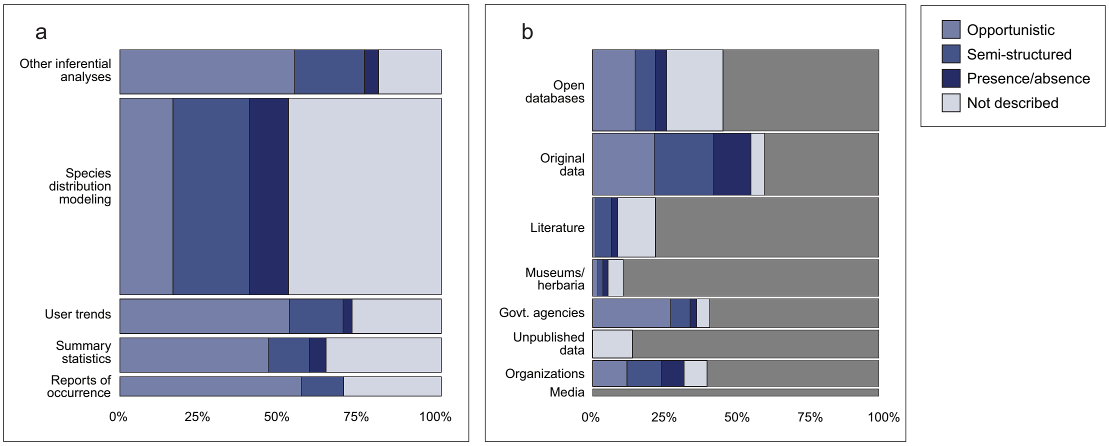


Figure 7. The percentage of the 300-article subset that is associated with each type of data structure, as a function of (a) analysis approach and (b) direct data source accessed by study authors. In panel (a), the* y*-axis categories represent all articles for which the indicated analysis approach was the most complex approach applied (with the exception of "user trends," in which case all articles using this approach are represented). The bar widths indicate the number of articles in the 300-article subset within each category. In panel (b), the* y*-axis categories represent all articles that use data from the indicated data source. The bar widths indicate the overall proportion of the 300-article subset that used each data type. The gray portions of the bars represent articles that integrated data from the indicated source with data from other sources; because of the confounding effect of data integration on metadata reporting, metadata reporting trends are not reported for these articles. The portions of the bars shaded according to the legend represent articles for which the indicated source was the only source accessed by the article.

analysis, occupancy modeling, and multivariate statistics (figure 5). In particular, the integration of presence-only data with other types of biodiversity data is of growing interest in the literature (Pacifici et al. 2017, Fletcher et al. 2019, Miller et al. 2019, Isaac et al. 2020, Simmonds et al. 2020, Zipkin et al. 2021). In our review, articles that integrated presence-only data with other types of biodiversity data were nearly three times as likely to employ an uncommon inferential analysis approach as the articles that used only presence-only data, indicating that data integration can open a wider range of analysis options for presence-only data.

Clearly documented metadata, particularly an explicit description of the data structure and original sampling design, also enable a wider range of analytical approaches, including data integration (Isaac et al. 2014, Araújo et al. 2019, Dobson et al. 2020). This trend is reflected in our results, with articles that employed more complex analysis approaches being correspondingly more likely to describe the underlying data structure (figure 7). Articles that employ species distribution modeling are the major exception to this trend; despite the relative statistical complexity of species distribution modeling, articles that modeled species distributions were the least likely to document data structure (figure 7). This likely reflects the growing accessibility of species distribution modeling approaches, which have become increasingly straightforward to implement through user-friendly platforms that can be implemented as a black box by researchers without a clear understanding of the method (Joppa et al. 2013, Merow et al. 2013, Kass et al. 2018). Although the growing accessibility of species distribution modeling offers great potential for research and conservation (Rapacciuolo 2019, Sofaer et al. 2019), we caution that it is still essential to share metadata whenever possible to aid in interpretation and evaluation of results (Soranno et al. 2020, Zurell et al. 2020, Muscatello et al. 2021, Sillero and Barbosa 2021, Foster et al. 2021). Relatedly, it is important to check for and correct data quality errors in data and metadata, particularly when data are obtained from open databases or collated from several sources (Ball-Damerow et al. 2019). In addition to supporting data interpretation and analysis, the reporting of high quality metadata facilitates a wide range of potential future data uses.

Reporting of metadata is inconsistent. Despite the value of clear metadata, around half of the articles that we reviewed did not explicitly describe the structure or sampling design of all of their data, corroborating previously reported trends (figure 5; Kervin et al. 2013, Roche et al. 2015). Of course, researchers can only report metadata if they have access to this information, and researchers reusing data may simply not have information on the original data structure. For instance, 118 articles obtained data from museums, herbaria, and the literature and 77% of these did not report the structure of their data; in the vast majority of these cases, metadata on the original sampling design were likely unavailable. Users of open data also have inconsistent access to metadata, and around half of the articles that obtained data exclusively from open sources did not describe data structure (figure 7). Although many openly accessible databases enable and encourage metadata standardization and sharing, most prominently through the Darwin Core standard (Wieczorek et al. 2012), many data available through open databases have been digitized from historical records, for which such metadata may be unavailable or may have been lost over time (Specht et al. 2018). Articles that rely on data collected by government agencies and private organizations describe data structure more frequently (figure 7). In the instances in which the structure of data from these sources is not described, it may be due to the loss of information that occurs when complete information was not passed from the data owners to the data users. Standardizing the methods used by governmental and private institutions to share data with researchers may reduce instances of data loss associated with more informal sharing of data (Kühl et al. 2020). Unsurprisingly, articles exclusively based on original field work were most consistent in documenting data structure (figure 7). The combination of data from multiple sources is an additional barrier to describing presence-only data because of practical challenges associated with describing a large number of separate sampling schemes. For each additional source accessed by an article in our review, the likelihood of data structure being described decreased by 12%. Although authors may have little recourse when working with data sets for which metadata are unavailable or with large data sets for which it may be impractical to describe a large number of separate sampling schemes, improving data citation practices may provide a partial solution by making it possible to trace data to its original source to gather any available metadata.

Of articles that described the structure of their data, most described one or more data source as opportunistic (i.e., collected with no predefined sampling design), followed by semistructured (*sensu* Dobson et al. 2020), and finally a smaller percentage used presence or absence data and discarded the absence records before analysis. Of the articles that converted presence or absence data to presence-only format before analysis, one third did this for the purpose of comparing different modeling approaches. The remaining two thirds discarded the absence data and conducted analyses exclusively in a presence-only framework. Previous authors have cautioned that it is not advisable to analyze presence or absence data in a presence-only framework (Yackulic et al. 2013), so it is concerning that some articles in our review took this approach. In some cases researchers may be motivated to convert presence or absence data to presence-only to facilitate merging presence or absence and presence-only data sets, but many recent studies suggest approaches for integrating various data types without reducing data structure (Pacifici et al. 2017, Fletcher et al. 2019, Miller et al. 2019, Isaac et al. 2020, Zipkin et al. 2021).

The articles in our review were more consistent in reporting the scope of their presence-only data set, in terms of both sample size and study scale. The sample size varied considerably between articles, but the majority of studies were small to mid-size (figure 5). The studies' geographic scale followed a similar trend, with the majority addressing a regional scale (figure 5). The small number of articles that did not explicitly state a sample size tended to involve several separate analyses of a large number of species and stated a total sample size and total number of species rather than the sample size for each analysis. The tendency toward mid-size studies has remained relatively consistent over time, with the exception of studies with a sample size of over one hundred thousand occurrence records. These very large studies were absent from our reviewed articles until 2014. This recent increase in large studies likely reflects growing infrastructure for and interest in big data macroecology (Hampton et al. 2013, Wüest et al. 2020). Such large studies are more likely to rely on open data than studies with a smaller scope.

How often are presence-only data made available for reuse?

Our results suggest that the majority of data used in presence-only analyses are not made available after the analyses are published, although there is a recent trend toward increased data sharing. To characterize trends in data sharing, we excluded the 19% of articles that were based entirely on data accessed from open sources. Of the remaining articles that used data from at least one source other than an open database, just 21% made all data used in the study openly available on publication of the article. Of these, 18% published their data in an openly accessible online database, whereas the rest used a different form of publication, such as supplementary material or an online repository (figure 8). The most common means of sharing data was to directly include it in the article, either the main text or the supplemental material. Data formats varied from those that facilitate reuse relatively easily (e.g., CSV files, spatial data files) to those that pose challenges for reuse (e.g., PDF files). Online repositories, including Dryad, Figshare, and GitHub, were also used by a small number of articles to share data. Only nine articles indicated that their data sets had been shared in an openly accessible database, although it is possible that the authors of some articles in our review published their data to an open database but neglected to mention this in the article. Of course, the data analyzed in the 19% of reviewed articles that obtained data exclusively from open databases remained openly available as long as the databases from which the authors accessed their data were still accessible.

To maximize their research value, data must be published in a way that is both searchable and persistent (Wilkinson et al. 2016, Bishop et al. 2019). Therefore, publication of data in aggregated databases is preferable to publication in supplemental material. In particular, larger

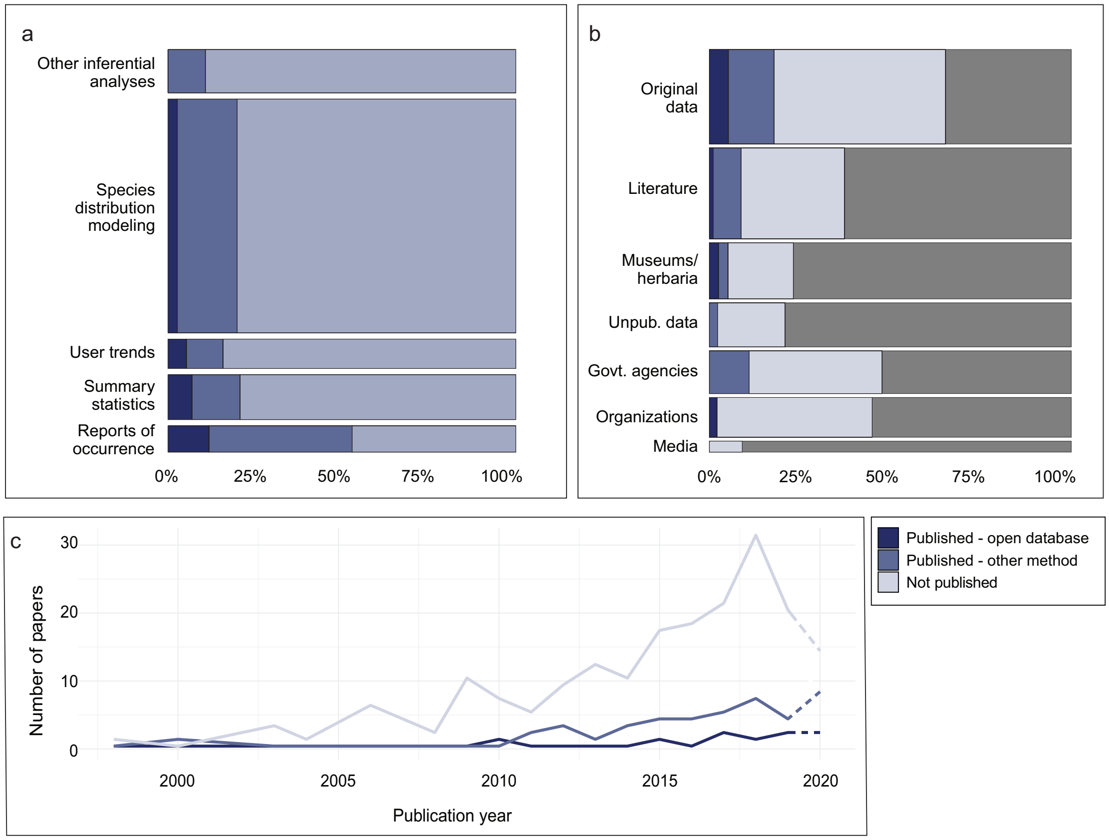


Figure 8. The percentage of the 300-article subset that is associated with the three levels of data availability as a function of (a) analysis approach and (b) direct data source accessed by study authors. For all panels of this figure, articles based entirely on data accessed from open databases have been excluded, leaving a subset of 242 articles that access data from at least one source other than an open database. In panel (a), the y-*axis categories represent all articles for which the indicated analysis approach was the most complex approach applied (with the exception of "user trends," in which case all articles using this approach are represented). The bar widths indicate the total number of articles within each category. In panel (b), the* y-*axis categories represent all articles in which the indicated direct data source was accessed. The bar widths indicate the overall proportion of the 242-article subset that used each data type. The portions of the bars shaded according to the legend represent articles for which the indicated source was the only source accessed by the article or which integrated the indicated source with open data. The gray portions of the bars represent articles that integrated data from the indicated source with data from other sources; because of the confounding effect of data integration on data sharing, data sharing trends are not reported for these articles. Panel (c) indicates trends in data availability over time. 2020 is indicated with dashed lines because the results for 2020 may be less complete than those for other years; although the set of articles was obtained with a search on 4 January 2021, some articles with a 2020 publication date may not yet have been indexed by journals or the Web of Science.*

databases are more likely to have greater longevity, stability, and infrastructure to maintain current best practices for data management in this rapidly developing field (Costello and Wieczorek 2014, Poisot et al. 2019). Much like small open databases, it has been demonstrated that data in supplementary material often become inaccessible over time (Vines et al. 2014, Stodden et al. 2018). We attempted to access all data shared by our reviewed articles and found that it was largely, but not entirely, still accessible: 7% of the data sets shared in journal supplementary materials were no longer available, and 22% of the data sets shared in an open database were no longer available. The inaccessible data from open databases were exclusively shared in small databases.

Although the overall accessibility of openly available presence-only data has increased dramatically in recent years, our results make it clear that the traditional peer-reviewed literature still largely serves as a sink for presence-only biodiversity data rather than facilitating its sharing and reuse. Making presence-only data more accessible should be a clear priority. Because strong infrastructure and clear best practices already exist for sharing presence-only occurrence data (Costello and Wieczorek 2014, Peterson et al. 2018, Hackett et al. 2019, Anderson et al. 2020) this should be achievable. However, several barriers can stand in the way of data sharing, including researchers' lack of incentive and ability, data ownership, and data set complexity. The strategies for overcoming these barriers will differ on the basis of the original source, ownership, and structure of the data.

Data sharing considerations for different types of presence-only data. The most straightforward type of presence-only data to target for increased data sharing are likely those collected by the study authors. Our results do indicate that original data are the most frequently shared, but the sharing rate is still low, at just 27% (figure 8). The publishing rate of original data collected with citizen science was somewhat higher than average, although still fewer than half of the articles based on original citizen science published their data. This is problematic, because studies have shown that citizen science participants generally expect and want their data to be made available for research, conservation, and policymaking (Chandler et al. 2017, Ganzevoort et al. 2017, Groom et al. 2017, Fox et al. 2019, Larson et al. 2020). Further integration of citizen science with open biodiversity data aggregators should therefore be a priority.

We anticipated lower rates of data publication from articles that compiled data from third party data owners, including the literature and museums and herbaria, and our results indicated rates of data publication that were just slightly lower than that of original data (figure 8). We suggest two major reasons why authors may not share data they have collated from other data owners. First, they may lack (or perceive that they lack) the permission to do so. And second, they may perceive that data sharing is unnecessary, assuming that readers wishing to reproduce their data set could retrace the data acquisition methods described in the paper to reassemble the data set from its original sources. Although this may sometimes be true, collating data from multiple sources takes a great deal of time and effort, so it is not a trivial process for a reader to reassemble a data set following a process described in the literature. And even if original data sources are well documented and still accessible, it cannot be assumed that a reader will be able to replicate the steps taken to collect data; literature is often behind paywalls, and access to institutional databases may be limited. Therefore, researchers working with data compiled from museums, herbaria, and journal articles should strive to provide as thorough a description as possible of their exact process of compiling their data set or, better yet, publish their complete data set whenever possible (Cousijn et al. 2018). Widespread progress on this issue will depend in part on the support of institutions: Institutions that host data should institute mechanisms to generate citations when data are accessed, making data easier to cite (Mooney and Newton 2012, Fenner et al. 2019, Powers and Hampton 2019), and journals that publish research should outline clear policies that support and facilitate data sharing and citation (Hrynaszkiewicz et al. 2020).

Finally, there are circumstances in which researchers may be unable to share data because of its proprietary or sensitive nature. We expect that this issue is most relevant to data obtained from private organizations or government agencies; in the present review, articles that accessed data primarily from one of these sources were characterized by low rates of data publication (figure 8). This is a complex issue, but we would encourage owners of sensitive data to use existing decision tools and prioritization schemes to consider whether there is a suitable way to make these data available for reuse, even in a more limited format (Clements et al. 2018, Tulloch et al. 2018, Chapman 2020). Because 37% of reviewed articles derive at least a portion of their data from sources that are assumed to generally be nonopen (e.g., data provided by government agencies, private organizations, or personal communications), and 41% derive some or all of their data from sources that are potentially accessible but cannot be assumed to be available to all readers (e.g., museums, literature, media), it is clear that a large portion of the presence-only biodiversity literature relies on data that are not accessible, hampering the replicability of these studies and the reusability of the data on which they are based.

A separate but related issue concerns data ethics and ownership. Issues of data ownership and governance are inherently related to social governance, and it is essential that the ethics of data sharing be held in the forefront at all stages of data management (Carroll et al. 2021, Rubert-Nason et al. 2021, Trisos et al. 2021). Data relevant to local communities must be made accessible to community members and must not be used in ways that are counter to community priorities (Johnson et al. 2021). This is particularly essential when it comes to Indigenous data; the CARE Principles for Indigenous Data Governance are a critical framework for ensuring Indigenous peoples' rights to the control of Indigenous data (GIDA 2019, Carroll et al. 2021). In addition, when data are collected by community members, as with citizen science, it is important to understand and respect volunteers' motivations for and concerns about the use of data they have contributed (Ganzevoort et al. 2017, Lynn et al. 2019, Tengö et al. 2021). The continued normalization of open data sharing must center scholarship and practice that respects ethical data governance, stewardship, and access.

The future of presence-only biodiversity data sharing. Data sharing practices in the presence-only biodiversity literature have until recently remained relatively constant over time, but the proportion of reviewed articles that publish their data has increased somewhat since 2016 (figure 8). This is cause for optimism and continued efforts to normalize open sharing of biodiversity data. Recent studies document overwhelmingly positive attitudes to data sharing (Tenopir et al. 2020, Soeharjono and Roche 2021), so if practical barriers can be overcome, there is a high likelihood that data sharing will continue to increase. Increased sharing of biodiversity data may even produce a ripple effect across disciplines; biodiversity research has historically exhibited a higher rate of open data sharing than closely related scientific disciplines such as ecology and conservation science (Michener 2015, Osawa 2019, Shin et al. 2020), but given the broad and growing application of presence-only biodiversity data across many related scientific disciplines (Ball-Damerow et al. 2019, Heberling et al. 2021), continued improvements in open sharing of presence-only biodiversity data may serve to spread awareness of open data practices across disciplines.

Past studies have indicated that the majority of biodiversity researchers support data sharing but may be held back by lack of sufficient incentive, lack of familiarity with data aggregators, lack of information on data set structure or ownership, and lack of trust in public databases (Huang et al. 2012, Tenopir et al. 2020). We compared articles that did and did not publish their data to examine the relative impact of some potential barriers to data sharing. First, we anticipated that two measures of data set complexity might negatively correlate with data sharing: first, the number of data sources accessed to compile a data set and, second, whether the original sampling design was reported. We expected that authors might be held back from sharing data by the complexity of crediting multiple original sources or by their own lack of complete information on data structure. However, we did not find either of these relationships in our results. This finding suggests that data set complexity may not be the primary factor prohibiting researchers from publishing their data sets. It is a concern but is more likely secondary to other barriers. Because lack of familiarity with open databases has also been cited as a barrier to data sharing, we expected that authors' familiarity with open data, as has been demonstrated by the integration of data from open databases with presence-only data from other sources, would correlate with greater rates of data publication. This was not the case: Of the articles that integrated data from open databases and other sources, 76% did not publish the data that were not already open.

These findings suggest that other concerns, including lack of researcher incentive and concern about receiving appropriate credit for shared data, may be more serious barriers to data sharing (Escribano et al. 2018, Tenopir et al. 2020). Some developments have begun to address the issue of researcher incentive: Data sharing is increasingly incentivized through journal policies, funding agency requirements, and the promotion of data citations (Mills et al. 2015, Colavizza et al. 2020, Walters 2020). Continuing to normalize these incentives may help overcome existing barriers to data sharing, especially in situations in which data users are the original data owners (Chavan and Penev 2011, Mooney and Newton 2012, Kattge et al. 2014, Escribano et al. 2018). Furthermore, researchers are increasingly taking ownership over the process of data sharing, establishing grassroots collaborations that organize specific research communities to engage with open data infrastructure and practices (Aubin et al. 2020). This integration of open data practices into local networks of biodiversity researchers has great potential to incentivize open data sharing by establishing it as a key component of network building and collaboration within specific research areas. As open data sharing becomes increasingly normalized, it will be essential that practitioners of open science maintain a supportive, rather than critical, approach to encouraging researchers who are taking their first steps into open data sharing. Researchers do not all have equal access to the resources, training, technical capacity, and institutional support to fully engage in open data practices, and small steps toward open data sharing must be welcomed while the field as a whole shifts to become more equitably supportive of open data practices (Bahlai et al. 2019, Chawinga and Zinn 2019, Powers and Hampton 2019, Soeharjono and Roche 2021).

Conclusions

Open access to high quality biodiversity occurrence data is key to many emerging themes in biodiversity research and conservation, including development and implementation of international biodiversity assessments and targets (Hochkirch et al. 2020), research synthesis for conservation decision-making (Nakagawa et al. 2020), and near-term ecological forecasting of species abundance in space and time (Callaghan et al. 2021), so continued efforts to increase the open sharing of biodiversity data will be critical. This will require increased incentivization, institutional support, ongoing shifts in cultural norms, and a growing emphasis on an ethical, equitable framework for data sharing. Recent trends toward increased sharing of presence-only biodiversity data are a cause for optimism, but there is still a great deal of work to be done in normalizing the use of best practices in data access, documentation, citation, and sharing. Still, we see evidence in the trends reported in the present article for an often-reported survey result: Researchers generally feel positively toward reusing and sharing data, despite persistent uncertainty about best practices and concern about credit and incentives (Ross-Hellauer et al. 2017, Tenopir et al. 2020, Soeharjono and Roche 2021). Such evidence includes the recent increase in the proportion of articles that produce open data, the efforts taken by some authors to credit original data providers even when no clear mechanism had yet been developed to do so, and the above-average sharing rate for citizen science data.

For researchers looking to begin or continue their journey into reuse and sharing of open biodiversity data, there are many excellent resources that offer an entry point into accessing and sharing open data; we particularly point such researchers to Hampton and colleagues (2015), Wilkinson and colleagues (2016), Boland and colleagues (2017), Alston and Rick (2021), and to guides such as the FAIR Principles (GO FAIR 2021), the CARE Principles of Indigenous Data Governance (GIDA 2019), and the Quick Guide to Publishing Data Through GBIF.org (GBIF 2021). To those beginning to engage with open data, we echo the wisdom of Bahlai and colleagues (2019), Alston and Rick (2021), and others in encouraging researchers to begin with any first steps, however small, that are feasible given their circumstances. Increased open data sharing will rely on both the progressive adoption of data sharing practices by individual researchers and ultimately on broad cultural shifts within biodiversity and related fields (Chawinga and Zinn 2019). This shift to a culture of ethical open data sharing will be essential to meet challenges associated with the growing biodiversity crisis and to support a growing need for biodiversity assessment, monitoring, and conservation.

Acknowledgments

This work is part of the Norwegian University of Science and Technology's Transforming Citizen Science for Biodiversity project, and we thank the other project members for their valuable feedback on earlier stages of this work. We also greatly appreciate the feedback of three anonymous reviewers and an anonymous editor. Funding for CPM was provided by the Transforming Citizen Science for Biodiversity program within the Digital Transformation Initiative of the Norwegian University of Science and Technology. WK was funded by the Norwegian Research Council (grant no. 272947) and the Norwegian Biodiversity Information Centre. EBN was funded by the Norwegian Institute for Nature Research.

Supplemental material

Supplemental data are available at *BIOSCI* online.

References cited

Alston JM, Rick JA. 2021. A beginner's guide to Conducting reproducible research. Bulletin of the Ecological Society of America 102: e01801.

Altman M, Crosas M. 2014. The evolution of data citation: From principles to implementation. IASSIST Quarterly 37: 62.

Amano T, Lamming JDL, Sutherland WJ. 2016. Spatial Gaps in global biodiversity information and the role of citizen science. BioScience 66: 393–400.

Anderson RP, Araújo MB, Guisan A, Lobo JM, Martínez-Meyer E, Peterson AT, Soberón JM. 2020. Optimizing biodiversity informatics to improve information flow, data quality, and utility for science and society. Frontiers of Biogeography 12.

Araújo MB et al. 2019. Standards for distribution models in biodiversity assessments. Science Advances 5: eaat4858.

Aria M, Cuccurullo C. 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics 11: 959–975.

Ariño AH. 2010. Approaches to estimating the universe of natural history collections data. Biodiversity Informatics 7: 81–92.

Asmussen CB, Møller C. 2019. Smart literature review: A practical topic modelling approach to exploratory literature review. Journal of Big Data 6: 93.

Aubin I, Cardou F, Boisvert-Marsh L, Garnier E, Strukelj M, Munson AD. 2020. Managing data locally to answer questions globally: The role of collaborative science in ecology. Journal of Vegetation Science 31: 509–517.

Bahlai CA, Bartlett LJ, Burgio KR, Fournier AMV, Keiser CN, Poisot T, Whitney KS. 2019. Open science isn't always open to all scientists. American Scientist 107: 78–82.

Ball-Damerow JE, Brenskelle L, Barve N, Soltis PS, Sierwald P, Bieler R, LaFrance R, Ariño AH, Guralnick RP. 2019. Research applications of primary biodiversity databases in the digital age. PLOS ONE 14: e0215794.

Bayraktarov E, Ehmke G, O'Connor J, Burns EL, Nguyen HA, McRae L, Possingham HP, Lindenmayer DB. 2019. Do big unstructured biodiversity data mean more knowledge? Frontiers in Ecology and Evolution 6: 239.

Bishop BW, Hank C, Webster J, Howard R. 2019. Scientists' data discovery and reuse behavior: (Meta)data fitness for use and the FAIR data principles. Proceedings of the Association for Information Science and Technology 56: 21–31.

Blair J, Gwiazdowski R, Borrelli A, Hotchkiss M, Park C, Perrett G, Hanner R. 2020. Towards a catalogue of biodiversity databases: An ontological case study. Biodiversity Data Journal 8: e32765.

Boland MR, Karczewski KJ, Tatonetti NP. 2017. Ten simple rules to enable multi-site collaborations through data sharing. PLOS Computational Biology 13: e1005278.

Boshoff N. 2009. Neo-colonialism and research collaboration in Central Africa. Scientometrics 81: 413–434.

Brown RF. 2021. The importance of data citation. BioScience 71: 211–211.

Callaghan CT, Poore AGB, Mesaglio T, Moles AT, Nakagawa S, Roberts C, Rowley JJL, Vergés A, Wilshire JH, Cornwell WK. 2021. Three frontiers for the future of biodiversity research using citizen science data. BioScience 71: 55–63.

Calver MC, Goldman B, Hutchings PA, Kingsford RT. 2017. Why discrepancies in searching the conservation biology literature matter. Biological Conservation 213: 19–26.

Carroll SR, Herczog E, Hudson M, Russell K, Stall S. 2021. Operationalizing the CARE and FAIR Principles for Indigenous data futures. Scientific Data 8: 108.

Chambers LE, Barnard P, Poloczanska ES, Hobday AJ, Keatley MR, Allsopp N, Underhill LG. 2017. Southern Hemisphere biodiversity and global change: Data gaps and strategies. Austral Ecology 42: 20–30.

Chandler M et al. 2017. Contribution of citizen science towards international biodiversity monitoring. Biological Conservation 213: 280–294.

Chapman AD. 2020. Current Best Practices for Generalizing Sensitive Species Occurrence Data. Global Biodiversity Information Facility.

Chavan V, Penev L. 2011. The data paper: A mechanism to incentivize data publishing in biodiversity science. BMC Bioinformatics 12: S2.

Chawinga WD, Zinn S. 2019. Global perspectives of research data sharing: A systematic literature review. Library Information Science Research 41: 109–122.

Clements HS, Selinske MJ, Archibald CL, Cooke B, Fitzsimons JA, Groce JE, Torabi N, Hardy MJ. 2018. Fairness and transparency are required for the inclusion of privately protected areas in publicly accessible conservation databases. Land 7: 96.

Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B. 2020. The citation advantage of linking publications to research data. PLOS ONE 15: e0230416.

Cooper CB, Shirk J, Zuckerberg B. 2014. The invisible prevalence of citizen science in global research: Migratory birds and climate change. PLOS ONE 9.

Costello MJ, Wieczorek J. 2014. Best practice for biodiversity data management and publication. Biological Conservation 173: 68–73.

Costello MJ, Michener WK, Gahegan M, Zhang Z-Q, Bourne PE. 2013. Biodiversity data should be published, cited, and peer reviewed. Trends in Ecology Evolution 28: 454–461.

Cousijn H et al. 2018. A data citation roadmap for scientific publishers. Scientific Data 5: 180259.

Cretois B, Linnell JDC, Grainger M, Nilsen EB, Rød JK. 2020. Hunters as citizen scientists: Contributions to biodiversity monitoring in Europe. Global Ecology and Conservation 23: e01077.

Curty RG, Crowston K, Specht A, Grant BW, Dalton ED. 2017. Attitudes and norms affecting scientists' data reuse. PLOS ONE 12: e0189288.

Di Marco M et al. 2017. Changing trends and persisting biases in three decades of conservation science. Global Ecology and Conservation 10: 32–42.

Dias D, Baringo Fonseca C, Correa L, Soto N, Portela A, Juarez K, Tumolo Neto RJ, Ferro M, Gonçalves J, Junior J. 2017. Repatriation data: More than two million species occurrence records added to the Brazilian Biodiversity Information Facility Repository (SiBBr). Biodiversity Data Journal 2017: e12012.

Dobson ADM et al. 2020. Making messy data work for conservation. One Earth 2: 455–465.

Eichhorn MP, Baker K, Griffiths M. 2020. Steps towards decolonising biogeography. Frontiers of Biogeography 12: e44795e.

Escribano N, Ariño AH, Galicia D. 2016. Biodiversity data obsolescence and land uses changes. PeerJ 4: e2743.

Escribano N, Galicia D, Ariño AH. 2018. The tragedy of the biodiversity data commons: A data impediment creeping nigher? Database 2018: bay033.

Faith D, Collen B, Ariño A, Koleff PKP, Guinotte J, Kerr J, Chavan V. 2013. Bridging the biodiversity data gaps: Recommendations to meet users' data needs. Biodiversity Informatics 8.

Fanelli D, Larivière V. 2016. Researchers' individual publication rate has not increased in a century. PLOS ONE 11: e0149504.

Farley SS, Dawson A, Goring SJ, Williams JW. 2018. Situating ecology as a big-data science: Current advances, challenges, and solutions. BioScience 68: 563–576.

Fenner M et al. 2019. A data citation roadmap for scholarly data repositories. Scientific Data 6: 28.

Fletcher RJ, Hefley TJ, Robertson EP, Zuckerberg B, McCleery RA, Dorazio RM. 2019. A practical guide for combining data to model species distributions. Ecology 100: e02710.

Foster SD, Vanhatalo J, Trenkel VM, Schulz T, Lawrence E, Przeslawski R, Hosack GR. 2021. Effects of ignoring survey design information for data reuse. Ecological Applications: e2360.

Fox R, Bourn NAD, Dennis EB, Heafield RT, Maclean IMD, Wilson RJ. 2019. Opinions of citizen scientists on open access to UK butterfly and moth occurrence data. Biodiversity and Conservation 28: 3321–3341.

Franz NM, Sterner BW. 2018. To increase trust, change the social design behind aggregated biodiversity data. Database 2018: bax100.

Gadelha LMR et al. 2021. A survey of biodiversity informatics: Concepts, practices, and challenges. WIREs Data Mining and Knowledge Discovery 11: e1394.

Ganzevoort W, Van den Born RJG, Halffman W, Turnhout S. 2017. Sharing biodiversity data: Citizen scientists' concerns and motivations. Biodiversity and Conservation 26: 2821–2837.

[GBIF] Global Biodiversity Information Facility. 2019. GBIF Science Review 2019. GBIF. https://doi.org/10.15468/QXXG-7K93.

[GBIF] Global Biodiversity Information Facility. 2021. Quick Guide to Publishing Data Through GBIF.org. GBIF. www.gbif.org/publishing-data.

Gelfand AE, Shirota S. 2019. Preferential sampling for presence/absence data and for fusion of presence/absence data with presence-only data. Ecological Monographs 89: E01372.

[GIDA] Global Indigenous Data Alliance. 2019. CARE Principles for Indigenous Data Governance. GIDA. GIDA-global.org.

GO FAIR. 2021. FAIR Principles. GO FAIR. www.go-fair.org/fair-principles.

Grainger MJ, Bolam FC, Stewart GB, Nilsen EB. 2020. Evidence synthesis for tackling research waste. Nature Ecology and Evolution 4: 495–497.

Grimmett L, Whitsed R, Horta A. 2020. Presence-only species distribution models are sensitive to sample prevalence: Evaluating models using spatial prediction stability and accuracy metrics. Ecological Modelling 431: 109194.

Groom Q, Weatherdon L, Geijzendorffer IR. 2017. Is citizen science an open science in the case of biodiversity observations? Journal of Applied Ecology 54: 612–617.

Groom Q, Güntsch A, Huybrechts P, Kearney N, Leachman S, Nicolson N, Page RDM, Shorthouse DP, Thessen AE, Haston E. 2020. People are essential to linking biodiversity data. Database 2020: baaa072.

Guisan A et al. 2013. Predicting species distributions for conservation decisions. Ecology Letters 16: 1424–1435.

Habel JC et al. 2014. Towards more equal footing in north–south biodiversity research: European and sub-Saharan viewpoints. Biodiversity and Conservation 23: 3143–3148.

Hackett RA, Belitz MW, Gilbert EE, Monfils AK. 2019. A data management workflow of biodiversity data from the field to data users. Applications in Plant Sciences 7: e11310.

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.

Hampton SE, et al. 2015. The Tao of open science for ecology. Ecosphere 6: 120.

Hao T, Elith J, Guillera-Arroita G, Lahoz-Monfort JJ. 2019. A review of evidence about use and performance of species distribution modelling ensembles like BIOMOD. Diversity and Distributions 25: 839–852.

Heberling JM, Miller JT, Noesgaard D, Weingart SB, Schigel D. 2021. Data integration enables global biodiversity synthesis. Proceedings of the National Academy of Sciences 118: e2018093118.

Hedrick BP et al. 2020. Digitization and the future of natural history collections. BioScience 70: 243–251.

Hochkirch A et al. 2020. A strategy for the next decade to address data deficiency in neglected biodiversity. Conservation Biology 35: 502–509. doi:10.1111/cobi.13589

Hrynaszkiewicz I, Simons N, Hussain A, Grant R, Goudie S. 2020. Developing a research data policy framework for all journals and publishers. Data Science Journal 19: 5.

Huang J, Frimpong EA. 2015. Using historical atlas data to develop high-resolution distribution models of freshwater fishes. PLOS ONE 10: e0129995.

Huang X, Hawkins BA, Lei F, Miller GL, Favret C, Zhang R, Qiao G. 2012. Willing or unwilling to share primary biodiversity data: Results and implications of an international survey. Conservation Letters 5: 399–406.

Huettmann F. 2009. The global need for, and appreciation of, high-quality metadata in biodiversity database work. Pages 25–28 in Spehn EM, Korner C, eds. Data Mining for Global Trends in Mountain Biodiversity. Taylor and Francis.

Isaac NJB et al. 2020. Data integration for large-scale models of species distributions. Trends in Ecology and Evolution 35: 56–67.

Isaac NJB, Strien AJ van, August TA, Zeeuw MP de, Roy DB. 2014. Statistics for citizen science: Extracting signals of change from noisy ecological data. Methods in Ecology and Evolution 5: 1052–1060.

James SA, Soltis PS, Belbin L, Chapman AD, Nelson G, Paul DL, Collins M. 2018. Herbarium data: Global biodiversity and societal botanical needs for novel research. Applications in Plant Sciences 6: e1024.

Johnson N, Druckenmiller ML, Danielsen F, Pulsifer PL. 2021. The use of digital platforms for community-based monitoring. BioScience 71: 452–466.

Johnston A, Hochachka WM, Strimas-Mackey ME, Gutierrez VR, Robinson OJ, Miller ET, Auer T, Kelling ST, Fink D. 2021. Analytical guidelines to increase the value of community science data: An example using eBird data to estimate species distributions. Diversity and Distributions 27: 1265–1277.

Joppa LN, McInerny G, Harper R, Salido L, Takeda K, O'Hara K, Gavaghan D, Emmott S. 2013. Troubling trends in scientific software use. Science 340: 814–815.

Kattge J, Díaz S, Wirth C. 2014. Of carrots and sticks. Nature Geoscience 7: 778–779.

Kass JM, Vilela B, Aiello-Lammens ME, Muscarella R, Merow C, Anderson RP. 2018. Wallace: A flexible platform for reproducible modeling of species niches and distributions built for community expansion. Methods in Ecology and Evolution 9: 1151–1156.

Kays R, McShea WJ, Wikelski M. 2020. Born-digital biodiversity data: Millions and billions. Diversity and Distributions 26: 644–648.

Kelling S et al. 2019. Using semistructured surveys to improve citizen science data for monitoring biodiversity. BioScience 69: 170–179.

Kervin K, Michener W, Cook R. 2013. Common errors in ecological data sharing. Journal of eScience Librarianship 2: 1024.

König C, Weigelt P, Schrader J, Taylor A, Kattge J, Kreft H. 2019. Biodiversity data integration: The significance of data resolution and domain. PLOS Biology 17: e3000183.

Kühl HS et al. 2020. Effective biodiversity monitoring needs a culture of integration. One Earth 3: 462–474.

Larson LR, Cooper CB, Futch S, Singh D, Shipley NJ, Dale K, LeBaron GS, Takekawa JY. 2020. The diverse motivations of citizen scientists: Does conservation emphasis grow as volunteer participation progresses? Biological Conservation 242: 108428.

Li K, Greenberg J, Dunic J. 2020. Data objects and documenting scientific processes: An analysis of data events in biodiversity data papers. Journal of the Association for Information Science and Technology 71: 172–182.

Luo M, Xu Z, Hirsch T, Aung TS, Xu W, Ji L, Qin H, Ma K. 2021. The use of Global Biodiversity Information Facility (GBIF)-mediated data in publications written in Chinese. Global Ecology and Conservation 25: e01406.

Lynn SJ, Kaplan N, Newman S, Scarpino R, Newman G. 2019. Designing a platform for ethical citizen science: A case study of CitSci.org. Citizen Science: Theory and Practice 4: 14.

MacPhail VJ, Colla SR. 2020. Power of the people: A review of citizen science programs for conservation. Biological Conservation 249: 108739.

Mandeville CP. 2021. Open data practices among users of primary biodiversity data. Open Science Framework. https://osf.io/jueqc. doi:10.17605/OSF.IO/JUEQC

Mandeville CP, Finstad AG. 2021. Community science supports research on protected area resilience. Conservation Science and Practice 2021: e442.

Merow C, Smith MJ, Silander JA. 2013. A practical guide to MaxEnt for modeling species' distributions: What it does, and why inputs and settings matter. Ecography 36: 1058–1069.

Mesibov R. 2018. An audit of some processing effects in aggregated occurrence records. ZooKeys 751: 129–146.

Michener WK. 2015. Ten simple rules for creating a good data management plan. PLOS Computational Biology 11: e1004525.

Miller DAW, Pacifici K, Sanderlin JS, Reich BJ. 2019. The recent past and promising future for data integration methods to estimate species' distributions. Methods in Ecology and Evolution 10: 22–37.

Miller SE et al. 2020. Building natural history collections for the twenty-first century and beyond. BioScience 70: 674–687.

Mills JA et al. 2015. Archiving primary data: Solutions for long-term studies. Trends in Ecology and Evolution 30: 581–589.

Monfils AK et al. 2020. Regional collections are an essential component of biodiversity research infrastructure. BioScience 70: 1045–1047.

Mooney H, Newton MP. 2012. The anatomy of a data citation: Discovery, reuse, and credit. Scholarly Communication 1: eP1035–eP1035.

Muscatello A, Elith J, Kujala H. 2021. How decisions about fitting species distribution models affect conservation outcomes. Conservation Biology: 10.1111/cobi.13669.

Nakagawa S et al. 2020. A new ecosystem for evidence synthesis. Nature Ecology and Evolution 4: 498–501.

Nelson G, Ellis S. 2019. The history and impact of digitization and digital data mobilization on biodiversity research. Philosophical Transactions of the Royal Society B 374: 20170391.

Norberg A et al. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecological Monographs 89: e01370.

Nuñez MA, Barlow J, Cadotte M, Lucas K, Newton E, Pettorelli N, Stephens PA. 2019. Assessing the uneven global distribution of readership, submissions and publications in applied ecology: Obvious problems without obvious solutions. Journal of Applied Ecology 56: 4–9.

Ondei S, Brook BW, Buettel JC. 2018. Nature's told stories: An overview on the availability and type of on-line data on long-term biodiversity monitoring. Biodiversity and Conservation 27: 2971–2987.

Osawa T. 2019. Perspectives on biodiversity informatics for ecology. Ecological Research 34: 446–456.

Pacifici K, Reich BJ, Miller DAW, Gardner B, Stauffer G, Singh S, McKerrow A, Collazo JA. 2017. Integrating multiple data sources in species distribution modeling: A framework for data fusion. Ecology 98: 840–850.

Parsons AW, Goforth C, Costello R, Kays R. 2018. The value of citizen science for ecological monitoring of mammals. PeerJ 6: e4536.

Pearce-Higgins JW, et al. 2018. Overcoming the challenges of public data archiving for citizen science biodiversity recording and monitoring schemes. Journal of Applied Ecology 55: 2544–2551.

Pelayo-Villamil P, Guisande C, Manjarrés-Hernández A, Jiménez LF, Granado-Lorencio C, García-Roselló E, González-Dacosta J, Heine J, González-Vilas L, Lobo JM. 2018. Completeness of national freshwater fish species inventories around the world. Biodiversity and Conservation 27: 3807–3817.

Penev L et al. 2017. Strategies and guidelines for scholarly publishing of biodiversity data. Research Ideas and Outcomes 3: e12431.

Petersen TK, Speed JDM, Grøtan V, Austrheim G. 2021. Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. Ecological Solutions and Evidence 2: e12048.

Peterson AT, Asase A, Canhos D, Souza S de, Wieczorek J. 2018. Data leakage and loss in biodiversity informatics. Biodiversity Data Journal 6: e26826.

Pettorelli N, Barlow J, Nuñez MA, Rader R, Stephens PA, Pinfield T, Newton E. 2021. How international journals can support ecology from the Global South. Journal of Applied Ecology 58: 4–8.

Pino-Del-Carpio A, Ariño AH, Villarroya A, Puig J, Miranda R. 2014. The biodiversity data knowledge gap: Assessing information loss in the management of Biosphere Reserves. Biological Conservation 173: 74–79.

Poisot T, Bruneau A, Gonzalez A, Gravel D, Peres-Neto P. 2019. Ecological data should not be so hard to find and reuse. Trends in Ecology and Evolution 34: 494–496.

Powers SM, Hampton SE. 2019. Open science, reproducibility, and transparency in ecology. Ecological Applications 29: e01822.

R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing. www.R-project.org.

Ramirez KS et al. 2018. The future of ecology is collaborative, inclusive and deconstructs biases. Nature Ecology and Evolution 2: 200–200.

Rapacciuolo G. 2019. Strengthening the contribution of macroecological models to conservation practice. Global Ecology and Biogeography 28: 54–60.

Robertson T, Döring M, Guralnick R, Bloom D, Wieczorek J, Braak K, Otegui J, Russell L, Desmet P. 2014. The GBIF Integrated Publishing Toolkit: Facilitating the efficient publishing of biodiversity data on the internet. PLOS ONE 9: e102623.

Roche DG, Kruuk LEB, Lanfear R, Binning SA. 2015. Public data archiving in ecology and evolution: How well are we doing? PLOS Biology 13: e1002295.

Ross-Hellauer T, Deppe A, Schmidt B. 2017. Survey on open peer review: Attitudes and experience among editors, authors and reviewers. PLOS ONE 12: e0189311.

Rubert-Nason K, Casper AA, Jurjonas M, Mandeville C, Potter R, Schwarz K. 2021. Ecologist engagement in translational science is imperative for building resilience to global change threats. Rethinking Ecology 6: 65–92.

Serra-Diaz JM, Enquist BJ, Maitner B, Merow C, Svenning J-C. 2017. Big data of tree species distributions: How big and how good? Forest Ecosystems 4: 30.

Serwadda D, Ndebele P, Grabowski MK, Bajunirwe F, Wanyenze RK. 2018. Open data sharing and the Global South: Who benefits? Science 359: 642–643.

Shin N, Shibata H, Osawa T, Yamakita T, Nakamura M, Kenta T. 2020. Toward more data publication of long-term ecological observations. Ecological Research 35: 700–707.

Sicacha-Parada J, Steinsland I, Cretois B, Borgelt J. 2020. Accounting for spatial varying sampling effort due to accessibility in Citizen Science data: A case study of moose in Norway. Spatial Statistics 100446.
Sillero N, Barbosa AM. 2021. Common mistakes in ecological niche models. International Journal of Geographical Information Science 35: 213–226.
Simmonds EG, Jarvis SG, Henrys PA, Isaac NJB, O'Hara RB. 2020. Is more data always better? A simulation study of benefits and limitations of integrated distribution models. Ecography 43: 1413–1422.
Singer RA, Ellis S, Page LM. 2020. Awareness and use of biodiversity collections by fish biologists. Journal of Fish Biology 96: 297–306.
Soeharjono S, Roche DG. 2021. Reported individual costs and benefits of sharing open data among Canadian Academic Faculty in ecology and evolution. BioScience. https://doi.org/10.1093/biosci/biab024.
Sofaer HR et al. 2019. Development and delivery of species distribution models to inform decision-making. BioScience 69: 544–557.
Soranno PA et al. 2020. Ecological prediction at macroscales using big data: Does sampling design matter? Ecological Applications 30: e02123.
Specht A, Bolton MP, Kingsford B, Specht RL, Belbin L. 2018. A story of data won, data lost and data re-found: The realities of ecological data preservation. Biodiversity Data Journal 6: e28073.
Speed JDM, Bendiksby M, Finstad AG, Hassel K, Kolstad AL, Prestø T. 2018. Contrasting spatial, temporal and environmental patterns in observation and specimen based species occurrence data. PLOS ONE 13: e0196417.
Stephenson PJ et al. 2017. Unblocking the flow of biodiversity data for decision-making in Africa. Biological Conservation 213: 335–340.
Støa B, Halvorsen R, Mazzoni S, Gusarov VI. 2018. Sampling bias in presence-only data used for species distribution modelling: Theory and methods for detecting sample bias and its effects on models. Sommerfeltia 38: 1–53.
Stodden V, Seiler J, Ma Z. 2018. An empirical analysis of journal policy effectiveness for computational reproducibility. Proceedings of the National Academy of Sciences 115: 2584–2589.
Stork H, Astrin JJ. 2014. Trends in biodiversity research—a bibliometric assessment. Open Journal of Ecology 4: 354–370.
Sullivan BL et al. 2017. Using open access observational data for conservation action: A case study for birds. Biological Conservation 208: 5–14.
Tengö M, Austin BJ, Danielsen F, Fernández-Llamazares Á. 2021. Creating synergies between citizen science and indigenous and local knowledge. BioScience 71: 503–518.
Tenopir C, Rice NM, Allard S, Baird L, Borycz J, Christian L, Grant B, Olendorf R, Sandusky RJ. 2020. Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. PLOS ONE 15: e0229003.
Tessarolo G, Ladle R, Rangel T, Hortal J. 2017. Temporal degradation of data limits biodiversity research. Ecology and Evolution 7: 6863–6870.
Theobald EJ et al. 2015. Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. Biological Conservation 181: 236–244.
Trisos CH, Auerbach J, Katti M. 2021. Decoloniality and anti-oppressive practices for a more ethical ecology. Nature Ecology and Evolution 2021: s41559-021-01460-w.
Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. 2017. Taxonomic bias in biodiversity data and societal preferences. Scientific Reports 7: 9132.
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.
Tulloch AIT et al. 2018. A decision tree for assessing the risks and benefits of publishing biodiversity data. Nature Ecology and Evolution 2: 1209–1217.
Tydecks L, Jeschke JM, Wolf M, Singer G, Tockner K. 2018. Spatial and topical imbalances in biodiversity research. PLOS ONE 13: e0199327.
Vaz UL, Cunha HF, Nabout JC, Vaz UL, Cunha HF, Nabout JC. 2015. Trends and biases in global scientific literature about ecological niche models. Brazilian Journal of Biology 75: 17–24.
Vines TH, Albert AYK, Andrew RL, Débarre F, Bock DG, Franklin MT, Gilbert KJ, Moore J-S, Renaut S, Rennison DJ. 2014. The availability of research data declines rapidly with article age. Current Biology 24: 94–97.
Voříšek P, Gregory RD, Keller V, Herrando S, Lindström Å, Nagy S, Burfield IJ, Noble D, Ramírez I, Foppen RPB. 2018. Wetzel et al. fail to identify the real gaps in European bird monitoring. Biological Conservation 225: 245–246.
Walters WH. 2020. Data journals: Incentivizing data access and documentation within the scholarly communication system. Insights 33: 18.
Westgate MJ. 2019. revtools: An R package to support article screening for evidence synthesis. Research Synthesis Methods 10: 606–614.
Wetzel FT, et al. 2018. Unlocking biodiversity data: Prioritization and filling the gaps in biodiversity observation data in Europe. Biological Conservation 221: 78–85.
Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D. 2012. Darwin Core: An evolving community-developed biodiversity data standard. PLOS ONE 7: e29715.
Wilkinson MD et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data 3: 160018.
Wüest RO et al. 2020. Macroecology in the age of Big Data: Where to go from here? Journal of Biogeography 47: 1–12.
Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Grant EHC, Veran S. 2013. Presence-only modelling using MAXENT: When can we trust the inferences? Methods in Ecology and Evolution 4: 236–243.
Zipkin EF, Zylstra ER, Wright AD, Saunders SP, Finley AO, Dietze MC, Itter MS, Tingley MW. 2021. Addressing data integration challenges to link ecological processes across scales. Frontiers in Ecology and the Environment 19: 30–38.
Zurell D et al. 2020. A standard protocol for reporting species distribution models. Ecography 43: 1261–1277.

Caitlin P. Mandeville (caitlin.mandeville@ntnu.no) is a PhD candidate in the Centre for Biodiversity Dynamics and Department of Natural History at the Norwegian University of Science and Technology, in Trondheim, Norway. Wouter Koch is a senior advisor of biodiversity informatics at the Norwegian Biodiversity Information Centre, as well as a PhD candidate in the Centre for Biodiversity Dynamics and Department of Natural History at the Norwegian University of Science and Technology, in Trondheim, Norway. Erlend B. Nilsen is a senior research scientist at the Norwegian Institute for Nature Research in Trondheim, Norway, as well as a professor of ecology at the Faculty of Biosciences and Aquaculture at Nord University, in Steinkjer, Norway. Anders G. Finstad is a professor in the Centre for Biodiversity Dynamics and Department of Natural History at the Norwegian University of Science and Technology, in Trondheim, Norway.

SUPPORTING INFORMATION:

Open data practices among users of primary biodiversity data

Supporting information contents

S1. Extended methods for literature search and screening

We searched the Web of Science Core Collection to target all scholarly articles that report on the application of presence-only biodiversity occurrence data, targeting articles whose titles, abstracts, or keywords contained any of 31 terms commonly used in the literature to indicate presence-only data as well as any of five terms used to indicate biodiversity:

(((TS=("presence-only" OR *"presence only"* OR *"opportunistic observation*"* OR *"opportunistic species observation*"* OR *"opportunistic occurrence*"* OR *"opportunistic distribution*"* OR *"opportunistic species occurrence*"* OR *"opportunistic species distribution*"* OR *"pseudo-absence*"* OR *"pseudoabsence*"* OR *"inferred absence*"* OR *"presence-background"* OR *"presence background"* OR *"citizen science"* OR *"community science"* OR *"participatory science" or "ad hoc data"* OR *"ad hoc collection"* OR *"ad hoc method*"* OR *"incidental data"* OR *"incidental sighting*"* OR *"incidentally collected"* OR *"geographic one-class data"* OR *"incidental detection*"* OR *"opportunistic detection*"* OR *"primary biodiversity data*"* OR *"occurrence record*"* OR *"atlas data"* OR *"unstructured occurrence data"* OR *"unstructured species observation"* OR *"unstructured biodiversity data"))*

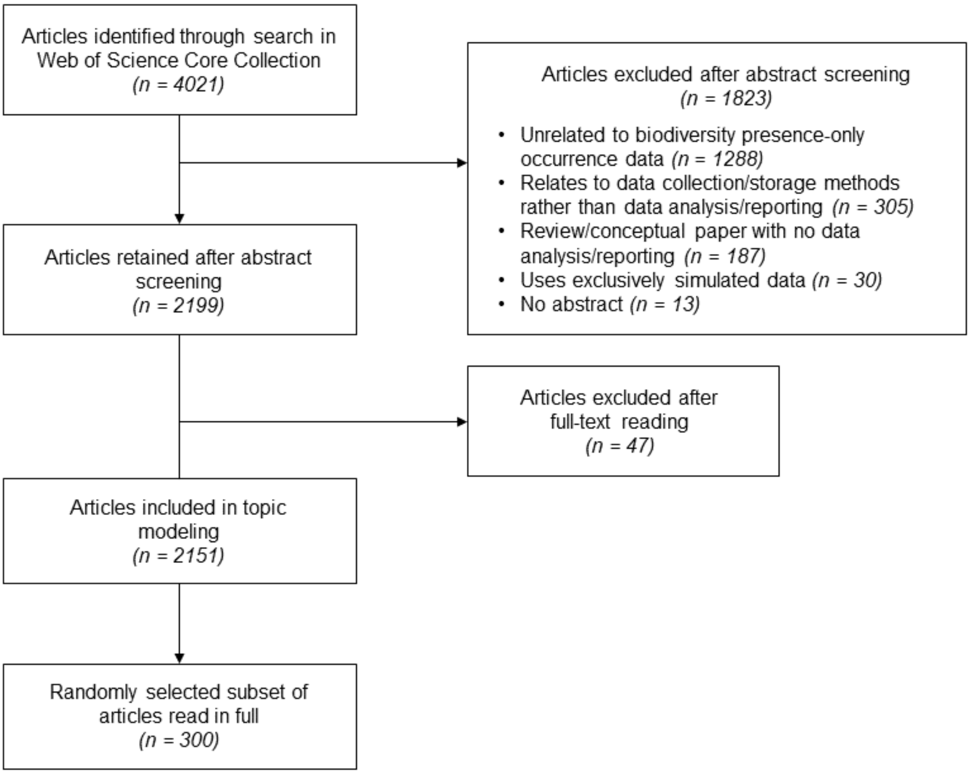
AND (TS=("distribution" OR *"species"* OR *"biodiversity"* OR *"habitat*"* OR *"niche*")))*

AND LANGUAGE: (English) *AND* DOCUMENT TYPES: (Article)

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

The search, conducted on January 4, 2021, returned 4021 peer-reviewed English-language articles.

We screened the abstracts of all returned articles and retained those that demonstrated the analysis or reporting of presence-only occurrence data. in the following categories were excluded: 1) articles unrelated to use of presence-only biodiversity occurrence data; 2) review or conceptual articles that did not perform data analysis or reporting; 3) articles that focused on the storage or management, rather than analysis or reporting, of occurrence data; and 4) articles that used exclusively simulated data. The article screening process is reported in the following diagram, modified from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) scheme (Moher et al. 2009). After screening, a total of 2151 articles were included in the review. Data management and bibliometric summary statistics were conducted in part with the bibliometrix package in R (Aria and Cuccurullo 2017).

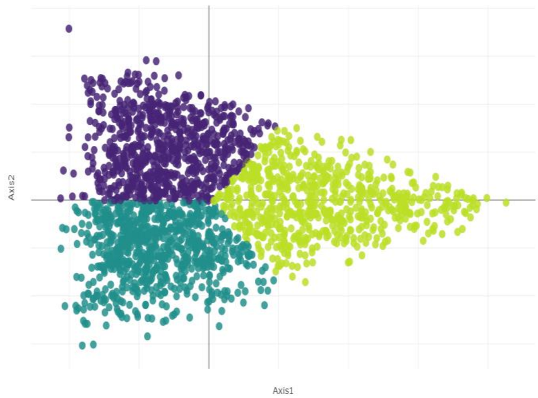


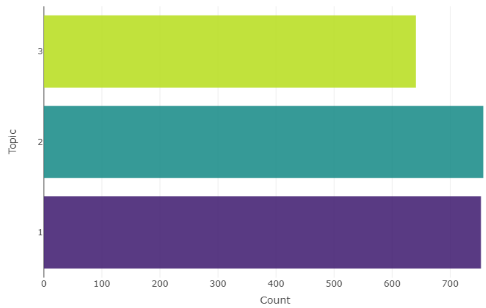
S2. Sets of topic clusters produced by LDA topic modeling

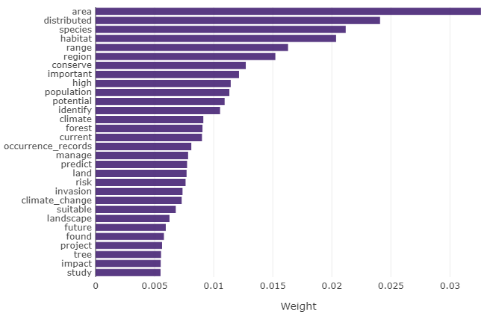
We ran Latent Dirichlet Allocation (LDA) topic modeling six times to produce sets of clusters ranging from three through eight clusters per set. We assessed each set of results for redundancy and interpretability and selected the set of three clusters as the most interpretable and least redundant. All six sets of modeling results are shown here. All topic modeling was conducted and LDA figures were produced using the revtools package in R (Westgate 2019).

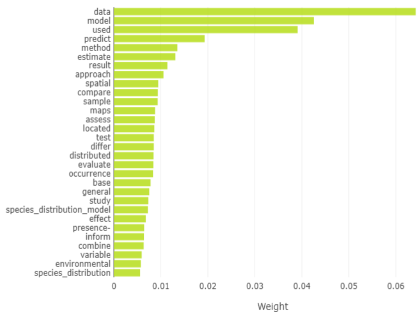
For each set of clusters, the biplot indicates the arrangement of articles relative to each other in terms of topic similarity. Each point represents an article and proximity indicates topical similarity. Colors indicate clusters. The topic bar charts represent the number of articles classified into each topic cluster. The word bar charts indicate the words most strongly associated with each cluster; these word associations were used to assign interpretations to each cluster.

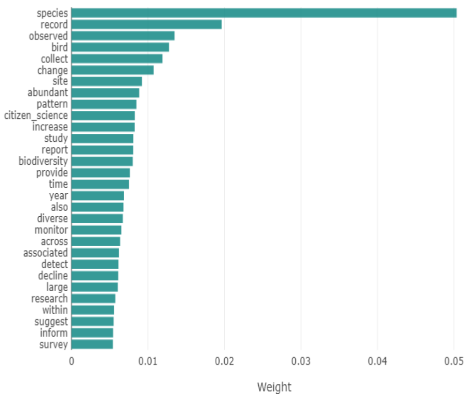
Three clusters



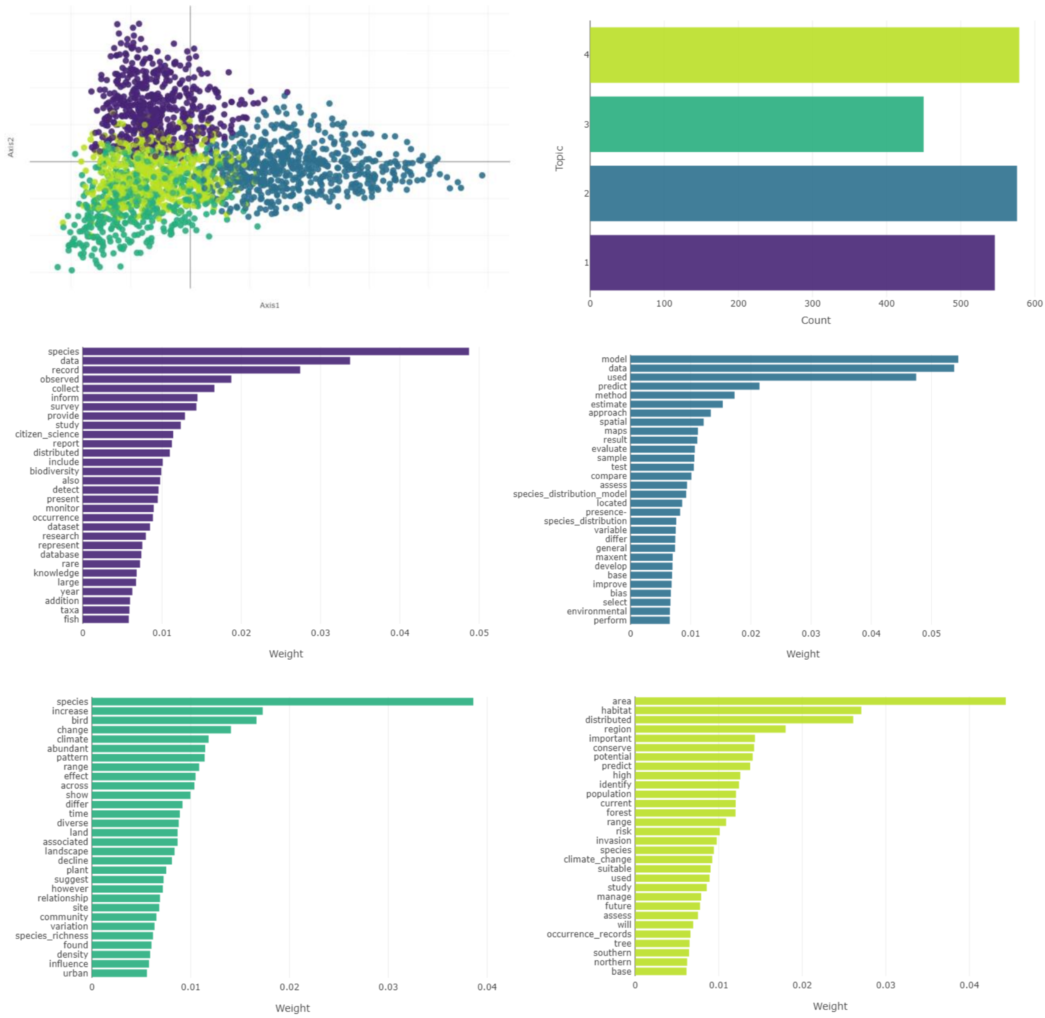




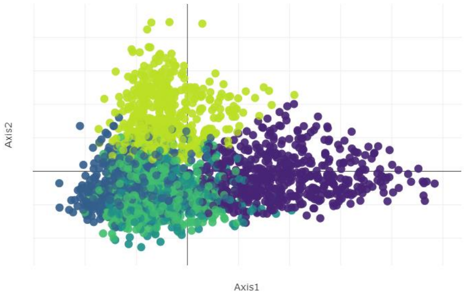


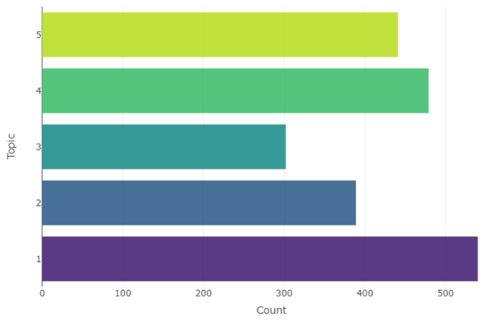


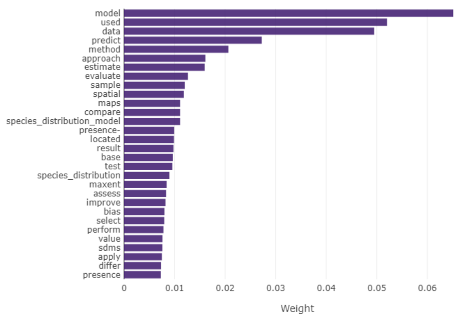
Four clusters

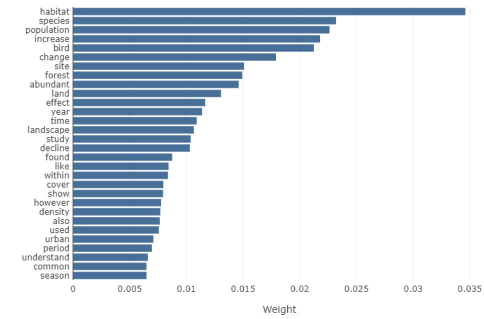


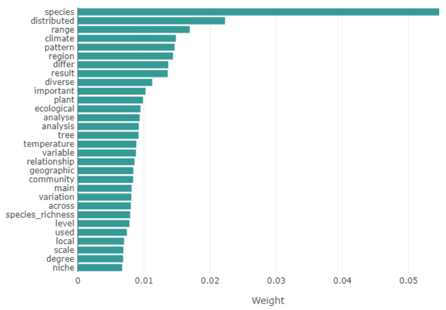
Five clusters

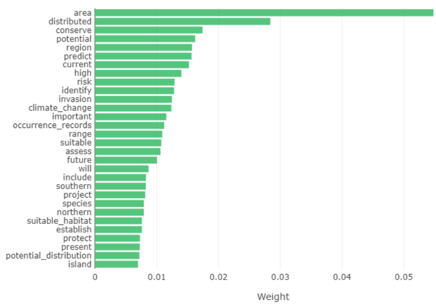


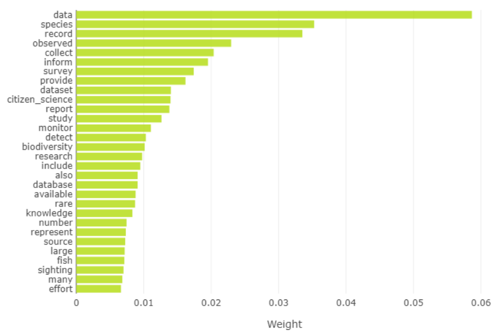




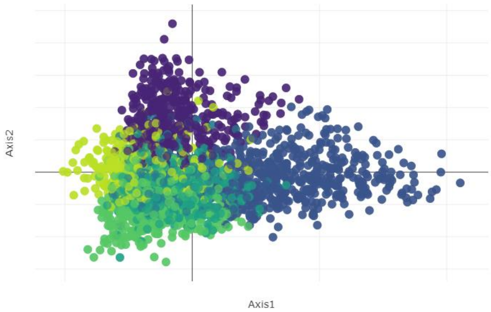


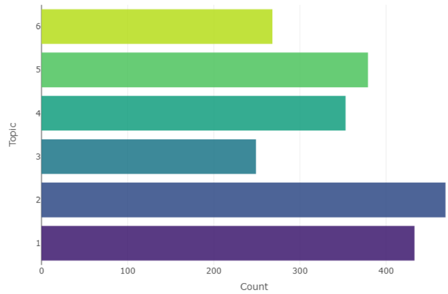


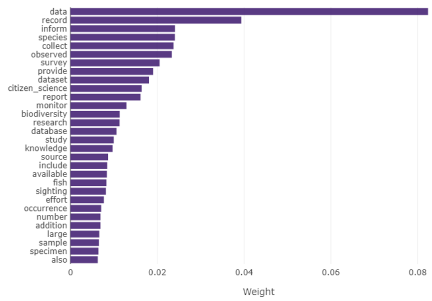


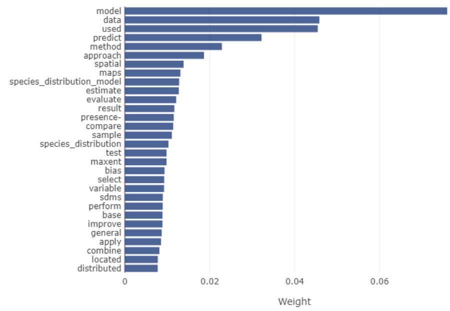


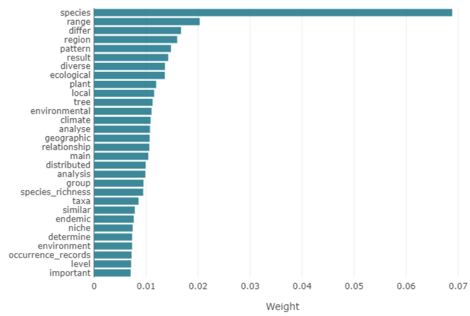
Six clusters

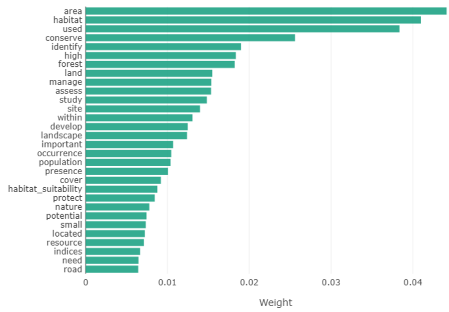


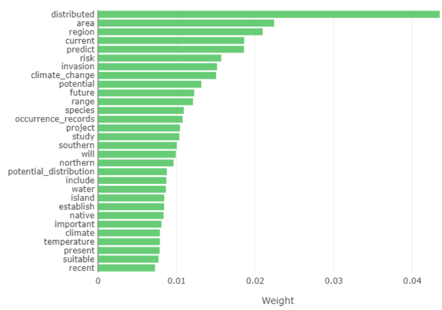


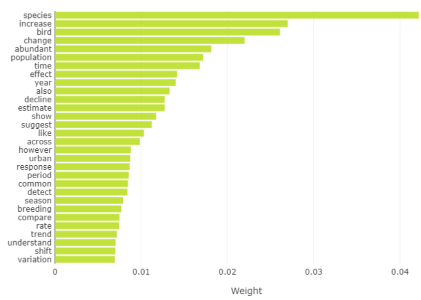




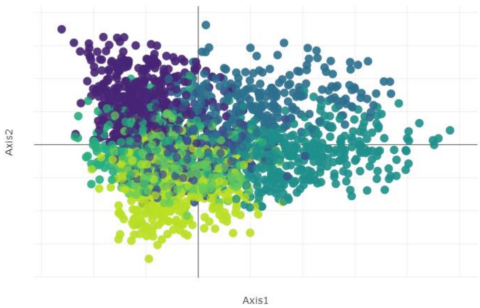


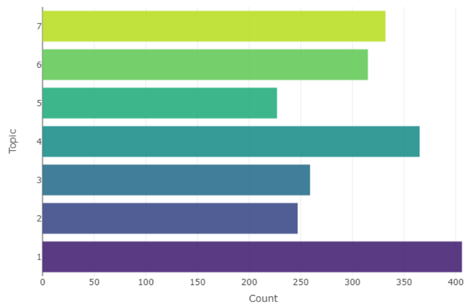


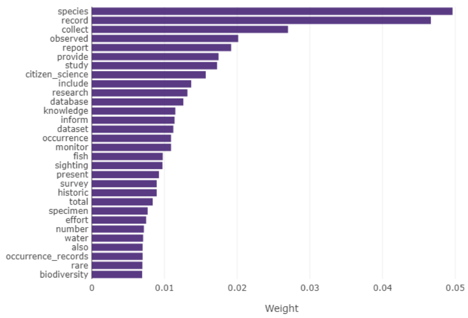


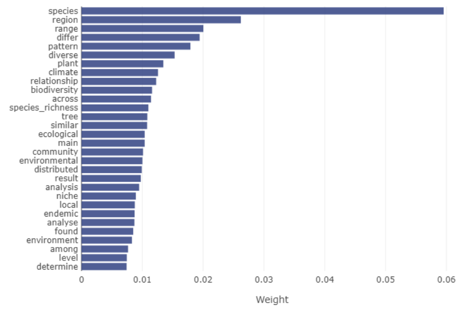


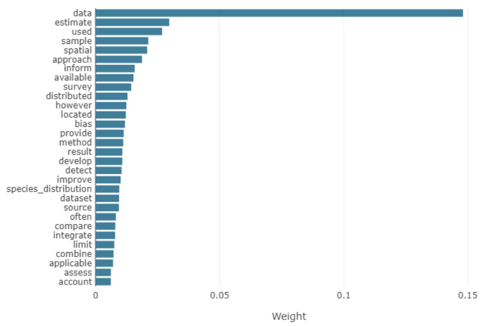
Seven clusters

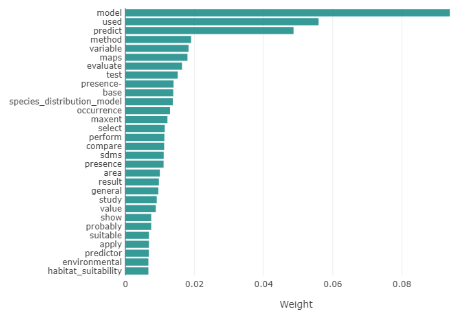


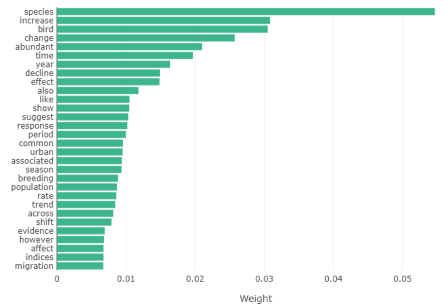


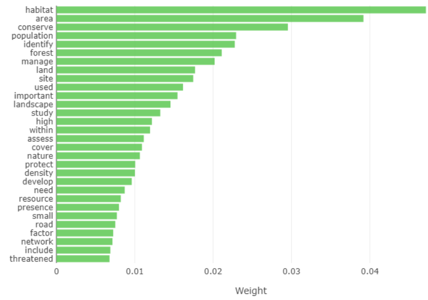


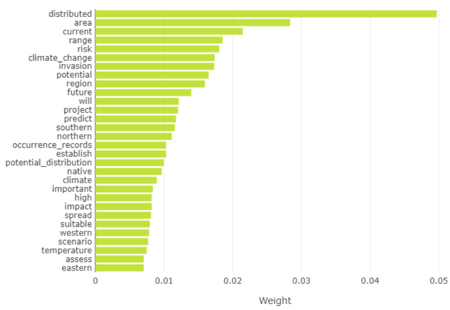




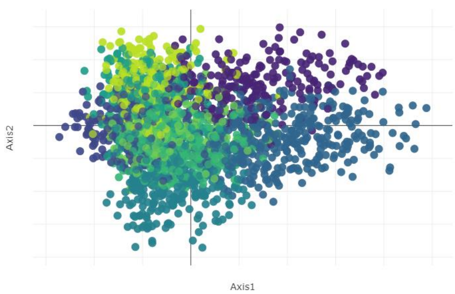


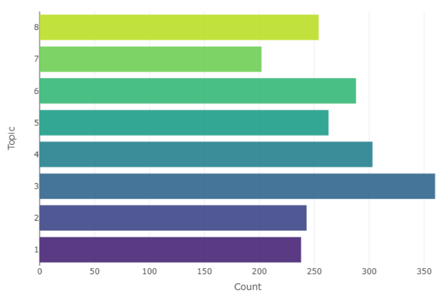


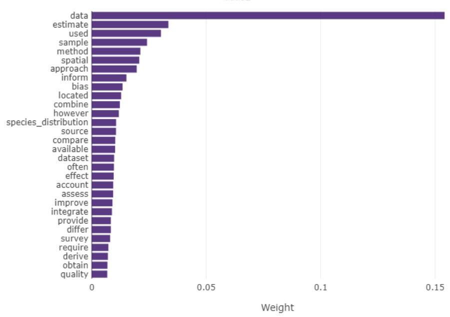


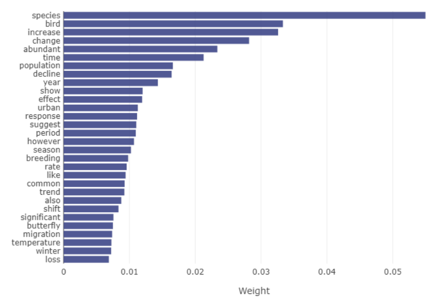


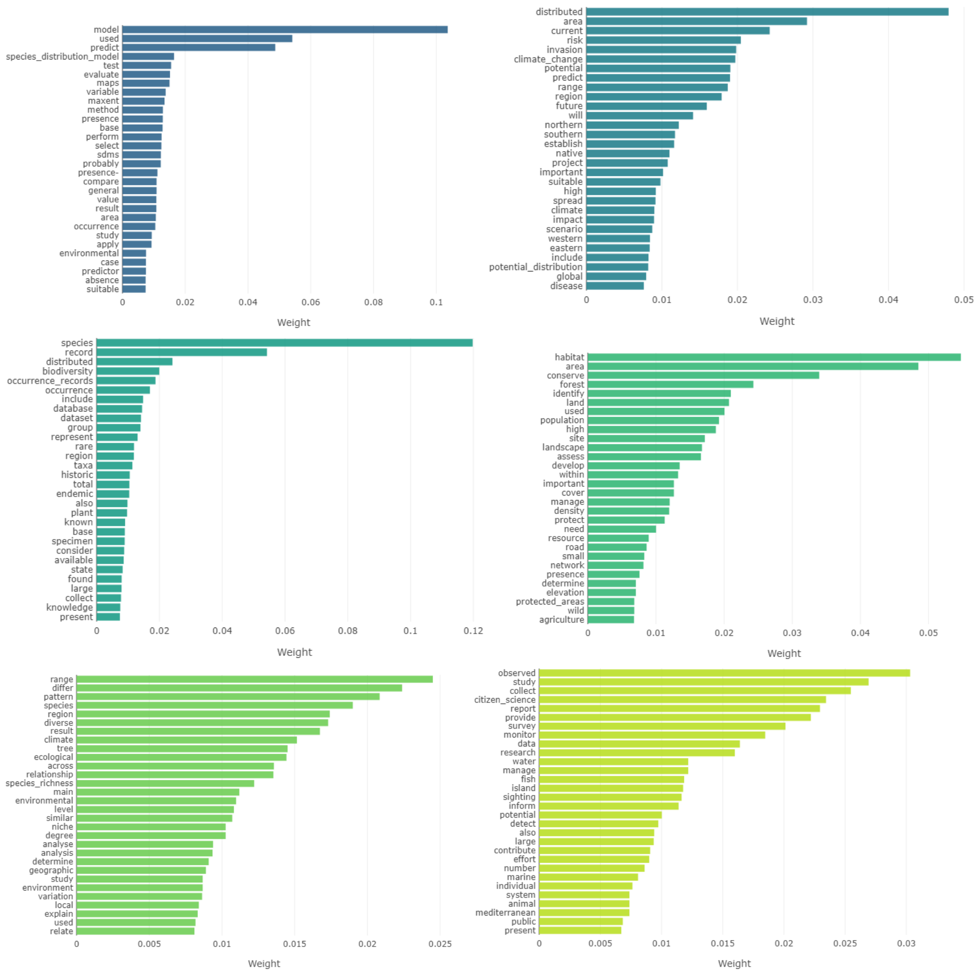
Eight clusters










model
used
predict
species_distribution_model
test
evaluate
maps
variable
maxent
method
presence
base
perform
select
sdms
probably
presence-
compare
general
value
result
area
occurrence
study
apply
environmental
case
predictor
absence
suitable
0
0.02
0.04
0.06
0.08
0.1
Weight
distributed
area
current
risk
invasion
climate_change
potential
predict
range
region
future
will
northern
southern
establish
native
project
important
suitable
high
spread
climate
impact
scenario
western
eastern
include
potential_distribution
global
disease
0
0.01
0.02
0.03
0.04
0.05
Weight
species
record
distributed
biodiversity
occurrence_records
occurrence
include
database
dataset
group
represent
rare
region
taxa
historic
total
endemic
also
plant
known
base
specimen
consider
available
state
found
large
collect
knowledge
present
0
0.02
0.04
0.06
0.08
0.1
0.12
Weight
habitat
area
conserve
forest
identify
land
used
population
high
site
landscape
assess
develop
within
important
cover
manage
density
protect
need
resource
road
small
network
presence
determine
elevation
protected_areas
wild
agriculture
0
0.01
0.02
0.03
0.04
0.05
Weight
range
differ
pattern
species
region
diverse
result
climate
tree
ecological
across
relationship
species_richness
main
environmental
level
similar
niche
degree
analyse
analysis
determine
geographic
study
environment
variation
local
explain
used
relate
0
0.005
0.01
0.015
0.02
0.025
Weight
observed
study
collect
citizen_science
report
provide
survey
monitor
data
research
water
manage
fish
island
sighting
inform
potential
detect
also
large
contribute
effort
number
marine
individual
system
animal
mediterranean
public
present
0
0.005
0.01
0.015
0.02
0.025
0.03
Weight

S3. Data sheet categories used to categorize the set of 300 papers read in full

All variable categories, except for study region, were not mutually exclusive; that is, an article could be coded with as many variable responses as applicable.

Variable category	Variable (True/False Response)
Topic categories mentioned in abstract [True/false]	Invasive species
	Land use change
	Climate change
	Overexploitation
	Pollution
	Other conservation issues
	Other basic ecology topics
	New methods development
	Comparing multiple presence-only approaches
	Comparing presence-only with more structured approaches
	Testing methodological choices within one presence-only approach
	Testing new technology for analyzing/reporting presence-only data
Taxa [True/false]	Bird
	Mammal
	Amphibian/reptile
	Fish
	Invertebrate
	Virus/bacteria/similar
	Plant/similar
Study system [True/false]	Terrestrial
	Marine
	Freshwater
Study region [True/false][1]	Africa
	Asia
	Europe
	Latin America
	North America
	Oceania
	Oceans
	Polar regions
	Multiple/global
Author region [True/false][1]	Africa
	Asia
	Europe
	Latin America

	North America Oceania
Study scale [True/false]	Local (upper size limit defined as municipality) Regional (upper size limit defined as large state/province and/or small nation) Large (defined as large national to continental scale) Global (defined as multiple continents)
Sample size [True/false]	1-10 11-100 101-1,000 1,001-10,000 10,001-100,000 100,001-1,000,000 > 1,000,000 Not described
Sampling design [True/false]	Explicitly described as opportunistic Semi-structured sampling design Structured presence/absence data Not described
Direct data source [Number of each type of source, unless otherwise noted]	Original data Large openly accessible database Small openly accessible database Literature Social media Unpublished data/personal communication Private organization/nonprofit Government agency Museum/herbarium/collections [For open databases] Name of database [open-ended response field] [For open databases] Is open database still available? [True/false]
Original data source [True/false]	Citizen science
Data availability [True/false, unless otherwise noted]	All data shared after publication – in an open database All data shared after publication – other method Location/format of shared data [open-ended response field] Is shared data still accessible?
Analysis approach [True/false]	Report of occurrence Spatial summary statistics Analysis of user trends Species distribution/ecological niche modeling Occupancy modeling List length analysis Species richness/diversity measures

	Phenology Population dynamics/demographic modeling Multivariate analyses
Other analysis information [True/false]	Comparison with more structured analysis types Integration with more structured data types Presence-only data used to evaluate a different type of analysis Presence-only data used to design a different type of analysis Biases associated with presence-only data discussed

[1] Study and author region categories were derived from the GBIF Regions (GBIF Secretariat 2019).

S4. Data collected from the set of 300 articles

Data are available here:

DOI: **10.17605/OSF.IO/JUEQC**

S5. Ten most cited articles and most commonly cited references among included articles

Table S5a. The ten most cited articles from within the articles included in our review.

Article	Times cited
Phillips et al. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling.	7546
Phillips and Dudík 2008. Modeling of species distributions with Maxent. Ecography.	3063
Elith et al. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions.	2658
Pearson et al. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography.	1540
Hernandez et al. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography.	1258
Phillips et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications.	1251
Merow et al. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. Ecography.	1129
Anderson et al. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models Ecological Modelling.	712
Engler et al. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology.	576
Pearson et al. 2006. Model-based uncertainty in species range prediction. Journal of Biogeography.	556

Table S5b. The most common references cited by articles included in our review. These references are not necessarily within the set of articles included in our review.

Article	Times referenced
Phillips et al. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling.	695
Elith et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography.	509
Hijmans et al. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology.	429
Phillips and Dudík 2008. Modeling of species distributions with Maxent. Ecography.	347
Fielding and Bell 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environmental Conservation.	322
Elith et al. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions.	306
Guisan and Zimmermann 2000. Predictive habitat distribution models in ecology. Ecological Modelling.	291
Phillips et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications.	289
Elith and Leathwick 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. Annual Review of Ecology, Evolution, and Systematics.	263
Guisan and Thuiller 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters.	261

S6. Openly accessible databases used by articles in the set of 300

Asterisk indicates databases considered ‘large' for the purpose of this review.

Database	Times used
*Global Biodiversity Information Facility (GBIF)	37
*eBird	9
*Atlas of Living Australia	8
*iNaturalist	8
*Tropicos	8
*OBIS	4
*speciesLink	4
Butterflies for the New Millenium	3
*FishBase	3
Victorian Biodiversity Atlas	3
Birdlife Australia	2
Biodiversity Information Serving Our Nation (US) (BISON)	2
BugGuide.net	2
Butterfly Conservation	2
Chinese Virtual Herbarium	2
Dutch National Database Flora and Fauna	2
EDDMaps	2
*iDigBio	2
Joint Nature Conservancy Council Seabird Censuses	2
ManisNet.org	2
National Specimen Information Infrastructure (China)	2
SEINet Portal Network	2
Swedish Lifewatch	2
Taiwan Roadkill Observation Network	2
UK Biological Records Centre	2
VertNet	2
WikiAves	2
AK Libellen NRW	1
AquaNIS	1
ArtDatabanken (Swedish Species Observation System)	1
Artsdatabanken (Norwegian Biodiversity Information Centre)	1
Atlas of New South Wales Wildlife	1
Aves de Chile	1
Base de Datos sobre Scarabaeidae (BANDASCA)	1
Basking Shark Watch (UK Marine Conservation Society)	1
Biodiversity Databank of Catalonia	1
BioObs	1

Bird Conservation Society of Thailand (BCST)	1
BirdLife Finland Tiira database	1
Birdlife International	1
BOLD (Barcode of Life)	1
British Dragonfly Society Recording Scheme	1
British Trust for Ornithology	1
Butterflies and Moths of North America	1
CalOdes	1
CardObs	1
Centre for Agriculture and Biosciences International (CABI)	1
Centre of Environmental Data and Recording (CEDaR) (North Ireland)	1
Centre Suisse de la Faune	1
cloudbirders.com	1
COL (National Colombian Herbarium of the Instituto de Ciencias Naturales)	1
Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO)	1
Database for Ecosystems and Ecosystem Service Zoning in China	1
Datenbank Artenschutzkartierung	1
Données d'Observations pour la Reconnaissance et l'Identification de la faune et la flore Subaquatiques (DORIS)	1
Dutch Butterfly Monitoring Scheme	1
Dutch Dragonfly Monitoring Scheme	1
eButterfly	1
EPPO Global Database (European and Mediterranean Plant Protection Organization)	1
EUFORGEN (European Forest Genetic Resources Programme)	1
European Environment Agency (http://eunis.eea.europa.eu)	1
falterfunde.de (science4you)	1
Faune-Aquitaine	1
Flora of Cyprus	1
Flora-On	1
Flotrop	1
Global Ant Biodiversity Informatics (GABI)	1
Global Mammal Parasite Database	1
HOLOS Ecoinformatics Engine	1
http://magicicada.org/	1
http://mammiferimarini.unipv.it/ Strandings Database	1
https://www.geocetus.it/ Stranding Information System	1
https://www.ornitho.at/	1
https://www.ornitho.ch/	1
https://www.ornitho.it/	1
Influenza Research Database (FluDB)	1
INPN Espèces	1
Insect Database (Finnish Museum of Natural History)	1
insecte.org	1
iSeahorse	1
JABOT (Rio de Janeiro Botanical Garden)	1

Jaguar GIS (http://www.savethejaguar.org)	1
JellyWatch (http://www.jellywatch.org)	1
JSTOR	1
LANDFIRE reference data base 2010, v1.2.0	1
Malaysian Nature Society Bird i-Witness database	1
Massachusetts Audubon Butterfly Atlas	1
MosquitoMap	1
Natagora	1
National Biodiversity Data Centre (Ireland)	1
National Indigenous Vegetation Survey Database (New Zealand)	1
National Institute of Invasive Species Science (NIISS) database (US)	1
NeoTropTree	1
New Zealand Herpetofauna Database	1
North American Breeding Bird Survey	1
OBIS-SEAMAP	1
Observadores del Mar	1
Odonata Central	1
PERSEUS (Policy-oriented marine Environmental Research in the Southern European Seas)	1
REBIOMA (Réseau de la Biodiversité de Madagascar)	1
Red de Observadores de Libélulas de Andalucía	1
Redmap	1
Reef Life Survey	1
Seaquest Southwest, Cornwall Wildlife Trust	1
SIG-Ivoire	1
Société française d'Odonatologie	1
SOMBASE (Southern Ocean Mollusc Database)	1
SWEMP (Southwest Exotic Mapping Program)	1
The Database on Taxonomy of Drosophilidae	1
Tokyo Butterfly Monitoring	1
UK National Biodiversity Network	1
University of British Columbia E-fauna	1
USGS Nonindigenous Aquatic Species database (US)	1
VectorMap	1
West African Vegetation database of the Senckenberg Research Institute	1
www.naturbeobachtung.at	1
xeno-canto	1

S7. Bibliography of the subset of 300 articles that were randomly selected from the full set of 2151 articles to be read in full and coded for analysis

Abood F, Murphy RJ. 2006. World distribution of Minthea rugicollis (coleoptera: lyctidae). Journal of Tropical Forest Science 18: 250-254.

Acosta LE. 2014. Bioclimatic profile and potential distribution of the Mesopotamian harvestman Discocyrtus testudineus (Holmberg, 1876) (Opiliones, Gonyleptidae). Zootaxa 3821: 301.

Adhikari D, Tiwary R, Singh PP, Upadhaya K, Singh B, Haridasan KE, Bhatt BB, Chettri A, Barik SK. 2019. Ecological niche modeling as a cumulative environmental impact assessment tool for biodiversity assessment and conservation planning: A case study of critically endangered plant Lagerstroemia minuticarpa in the Indian Eastern Himalaya. Journal of Environmental Management 243: 299–307.

Alahmed AM, Naeem M, Kheir SM, Sallam MF. Ecological Distribution Modeling of Two Malaria Mosquito Vectors Using Geographical Information System in Al-Baha Province, Kingdom of Saudi Arabia. 10.

Albuquerque F, Benito B, Rodriguez MÁM, Gray C. 2018. Potential changes in the distribution of Carnegiea gigantea under future scenarios. PeerJ 6: e5623.

Alimi TO, Fuller DO, Qualls WA, Herrera SV, Arevalo-Herrera M, Quinones ML, Lacerda MVG, Beier JC. 2015. Predicting potential ranges of primary malaria vectors and malaria in northern South America based on projected changes in climate, land cover and human population. Parasites & Vectors 8: 431.

Alkhamis MA, Arruda AG, Vilalta C, Morrison RB, Perez AM. 2018. Surveillance of porcine reproductive and respiratory syndrome virus in the United States using risk mapping and species distribution modeling. Preventive Veterinary Medicine 150: 135–142.

Allouche O, Steinitz O, Rotem D, Rosenfeld A, Kadmon R. 2008. Incorporating distance constraints into species distribution models. Journal of Applied Ecology 45: 599–609.

Anderson RP, Lew D, Peterson AT. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecological Modelling 162: 211–232.

Armstrong AJ, Armstrong AO, Bennett MB, McGregor F, Abrantes KG, Barnett A, Richardson AJ, Townsend KA, Dudgeon CL. 2020. The geographic distribution of reef and oceanic manta rays (Mobula alfredi and Mobula birostris) in Australian coastal waters. Journal of Fish Biology 96: 835–840.

Ashrafzadeh MR, Naghipour AA, Haidarian M, Kusza S, Pilliod DS. 2019. Effects of climate change on habitat and connectivity for populations of a vulnerable, endemic salamander in Iran. Global Ecology and Conservation 19: e00637.

Aubry KB, Lewis JC. 2003. Extirpation and reintroduction of fishers (Martes pennanti) in Oregon: implications for their conservation in the Pacific states. Biological Conservation 114: 79–90.

Ay J-S, Guillemot J, Martin-StPaul N, Doyen L, Leadley P. 2017. The economics of land use reveals a selection bias in tree species distribution models: Selection bias in species distribution models. Global Ecology and Biogeography 26: 65–77.

Aylesworth L, Phoonsawat R, Suvanachai P, Vincent ACJ. 2017. Generating spatial data for marine conservation and management. Biodiversity and Conservation 26: 383–399.

Ballesteros-Mejia L, Kitching IJ, Jetz W, Beck J. 2017. Putting insects on the map: near-global variation in sphingid moth richness along spatial and environmental gradients. Ecography 40: 698–708.

Baltensperger AP, Morton JM, Huettmann F. 2017. Expansion of American marten (Martes americana) distribution in response to climate and landscape change on the Kenai Peninsula, Alaska. Journal of Mammalogy 98: 703–714.

Bariche M. 2018. First confirmed record of the white-spotted puffer Arothron hispidus (Linnaeus, 1758) in the Mediterranean Sea. BioInvasions Records 7: 433–436.

Beasley-Hall PG, Lee TRC, Rose HA, Lo N. 2018. Multiple abiotic factors correlate with parallel evolution in Australian soil burrowing cockroaches. Journal of Biogeography 45: 1515–1528.

Bedriñana-Romano L, Hucke-Gaete R, Viddi FA, Morales J, Williams R, Ashe E, Garcés-Vargas J, Torres-Florez JP, Ruiz J. 2018. Integrating multiple data sources for assessing blue whale abundance and distribution in Chilean Northern Patagonia. Diversity and Distributions 24: 991–1004.

Belkhiria J, Hijmans RJ, Boyce W, Crossley BM, Martínez-López B. 2018. Identification of high risk areas for avian influenza outbreaks in California using disease distribution models. PLOS ONE 13: e0190824.

Beneš J, Konvička M, Vrabec V, Zámečník J. 2003. Do the sibling species of small whites, Leptidea sinapis and L. reali (Lepidoptera, Pieridae) differ in habitat preferences? Biologia 58: 943-951.

Benito X, Trobajo R, Ibáñez C. 2014. Modelling Habitat Distribution of Mediterranean Coastal Wetlands: The Ebro Delta as Case Study. Wetlands 34: 775–785.

Berzaghi F, Engel JE, Plumptre AJ, Mugabe H, Kujirakwinja D, Ayebare S, Bates JM. 2018. Comparative niche modeling of two bush-shrikes (Laniarius) and the conservation of mid-elevation Afromontane forests of the Albertine Rift. The Condor 120: 803–814.

Bezuijen MR. 2000. The occurrence of the flat-headed cat Prionailurus planiceps in south-east Sumatra. Oryx 34: 222-226.

Birkmanis CA, Partridge JC, Simmons LW, Heupel MR, Sequeira AMM. 2020. Shark conservation hindered by lack of habitat protection. Global Ecology and Conservation 21: e00862.

Boakes EH, Fuller RA, McGowan PJK, Mace GM. 2016. Uncertainty in identifying local extinctions: the distribution of missing data and its effects on biodiversity measures. Biology Letters 12: 20150824.

Bocksberger G, Schnitzler J, Chatelain C, Daget P, Janssen T, Schmidt M, Thiombiano A, Zizka G. 2016. Climate and the distribution of grasses in West Africa. Journal of Vegetation Science 27: 306–317.

Bolliger J, Kienast F, Soliva R, Rutherford G. 2007. Spatial sensitivity of species habitat patterns to scenarios of land use change (Switzerland). Landscape Ecology 22: 773–789.

Bonnet-Lebrun A -S., Karamanlidis AA, de Gabriel Hernando M, Renner I, Gimenez O. 2020. Identifying priority conservation areas for a recovering brown bear population in Greece using citizen science data. Animal Conservation 23: 83–93.

Bosch S, Tyberghein L, Deneudt K, Hernandez F, De Clerck O. 2018. In search of relevant predictors for marine species distribution modelling using the MarineSPEED benchmark dataset. Diversity and Distributions 24: 144–157.

Botella C, Joly A, Bonnet P, Monestiez P, Munoz F. 2018. Species distribution modeling based on the automated identification of citizen observations. Applications in Plant Sciences 6: e1029.

Bottin M, Peyre G, Vargas C, Raz L, Richardson JE, Sanchez A. 2020. Phytosociological data and herbarium collections show congruent large-scale patterns but differ in their local descriptions of community composition. Journal of Vegetation Science 31: 208–219.

Botts EA, Erasmus BFN, Alexander GJ. 2015. Observed range dynamics of South African amphibians under conditions of global change: Changes in Amphibian Distribution. Austral Ecology 40: 309–317.

Bradsworth N, White JG, Isaac B, Cooke R. 2017. Species distribution models derived from citizen science data predict the fine scale movements of owls in an urbanizing landscape. Biological Conservation 213: 27–35.

Bradter U, Mair L, Jönsson M, Knape J, Singer A, Snäll T. 2018. Can opportunistically collected Citizen Science data fill a data gap for habitat suitability models of less common species? Methods in Ecology and Evolution 9: 1667–1678.

Brambilla M, Falco R, Negri I. 2012. A spatially explicit assessment of within-season changes in environmental suitability for farmland birds along an altitudinal gradient: Within-season changes in habitat suitability in birds. Animal Conservation 15: 638–647.

Brambilla M, Ficetola GF. 2012. Species distribution models as a tool to estimate reproductive parameters: a case study with a passerine bird species: Distribution models and reproductive parameters. Journal of Animal Ecology 81: 781–787.

Brambilla M, Scridel D, Bazzi G, Ilahiane L, Iemma A, Pedrini P, Bassi E, Bionda R, Marchesi L, Genero F, Teufelbauer N, Probst R, Vrezec A, Kmecl P, Mihelič T, Bogliani G, Schmid H, Assandri G, Pontarini R, Braunisch V, Arlettaz R, Chamberlain D. 2020. Species interactions and climate change: How the disruption of species co-occurrence will impact on an avian forest guild. Global Change Biology 26: 1212–1224.

Braz AG, de Viveiros Grelle CE, de Souza Lima Figueiredo M, Weber M de M. 2020. Interspecific competition constrains local abundance in highly suitable areas. Ecography 43: 1560–1570.

Breed GA, Stichter S, Crone EE. 2013. Climate-driven changes in northeastern US butterfly communities. Nature Climate Change 3: 142–145.

Briscoe D, Hiatt S, Lewison R, Hines E. 2014. Modeling habitat and bycatch risk for dugongs in Sabah, Malaysia. Endangered Species Research 24: 237–247.

Brown DM, Sieswerda PL, Parsons ECM. 2019. Potential encounters between humpback whales (Megaptera novaeangliae) and vessels in the New York Bight apex, USA. Marine Policy 106: 103527.

Brown JL, Yoder AD. 2015. Shifting ranges and conservation challenges for lemurs in the face of climate change. Ecology and Evolution 5: 1131–1142.

Brown S, Clarke M, Clarke R. 2009. Fire is a key element in the landscape-scale habitat requirements and global population status of a threatened bird: The Mallee Emu-wren (Stipiturus mallee). Biological Conservation 142: 432–445.

Byers JE, McDowell WG, Dodd SR, Haynie RS, Pintor LM, Wilde SB. 2013. Climate and pH Predict the Potential Range of the Invasive Apple Snail (Pomacea insularum) in the Southeastern United States. PLoS ONE 8: e56812.

Bystriakova N, Peregrym M, Dragicevic S. 2015. Effect of environment on distributions of rock ferns in the Mediterranean climate: The case of the genus Asplenium in Montenegro. Flora - Morphology, Distribution, Functional Ecology of Plants 215: 84–91.

Camacho C. 2016. Birding trip reports as a data source for monitoring rare species. Animal Conservation 19: 430–435.

Cao B, Bai C, Xue Y, Yang J, Gao P, Liang H, Zhang L, Che L, Wang J, Xu J, Duan C, Mao M, Li G. 2020. Wetlands rise and fall: Six endangered wetland species showed different patterns of habitat shift under future climate change. Science of The Total Environment 731: 138518.

Carota C, Nava CR, Ghiglione C, Schiaparelli S. 2017. A Bayesian semiparametric GLMM for historical and newly collected presence-only data: An application to species richness of Ross Sea Mollusca: A Bayesian semiparametric GLMM for presence-only data. Environmetrics 28: e2462.

Carrascal LM, Aragón P, Palomino D, Lobo JM. 2015. Predicting regional densities from bird occurrence data: validation and effects of species traits in a Macaronesian Island. Diversity and Distributions 21: 1284–1294.

Castuera-Oliveira L, Oliveira-Filho AT de, Eisenlohr PV. 2020. Emerging hotspots of tree richness in Brazil. Acta Botanica Brasilica 34: 117–134.

Cerasoli F, Thuiller W, Guéguen M, Renaud J, D'Alessandro P, Biondi M. 2020. The role of climate and biotic factors in shaping current distributions and potential future shifts of European Neocrepidodera (Coleoptera, Chrysomelidae). Insect Conservation and Diversity 13: 47–62.

César de Sá N, Marchante H, Marchante E, Cabral JA, Honrado JP, Vicente JR. 2019. Can citizen science data guide the surveillance of invasive plants? A model-based test with Acacia trees in Portugal. Biological Invasions 21: 2127–2141.

Changeux T, Blazy C, Ruitton S. 2020. The use of citizen science for marine biodiversity surveys: from species identification to ecologically relevant observations. Hydrobiologia 847: 27–43.

Chapman D, Pescott OL, Roy HE, Tanner R. 2019. Improving species distribution models for invasive non-native species with biologically informed pseudo-absence selection. Journal of Biogeography 46: 1029–1040.

Chen H, Chen L, Albright TP. 2007. Predicting the potential distribution of invasive exotic species using GIS and information-theoretic approaches: A case of ragweed (Ambrosia artemisiifolia L.) distribution in China. Chinese Science Bulletin 52: 1223–1230.

Cheyne SM, Mohamed A, Hearn AJ, Ross J, Samejima H, Heydon M, Augeri DM, van Berkel T, Boonratana R, Fredriksson G, Hon J, Marshall J, Macdonald DW, Belant JL, Kramer-Schadt S, Wilting A. 2016. Predicted distribution of the otter civet Cynogale bennettii (Mammalia: Carnivora: Viverridae) on Borneo. The Raffles Bulletin of Zoology 33: 126-131.

Christodoulou CS, Griffiths GH, Vogiatzakis IN. 2018. Using threatened plant species to identify conservation gaps and opportunities on the island of Cyprus. Biodiversity and Conservation 27: 2837–2858.

Chyn K, Lin T-E, Chen Y-K, Chen C-Y, Fitzgerald LA. 2019. The magnitude of roadkill in Taiwan: Patterns and consequences revealed by citizen science. Biological Conservation 237: 317–326.

Číhal L, Kaláb O, Plášek V. 2017. Modeling the distribution of rare and interesting moss species of the family Orthotrichaceae (Bryophyta) in Tajikistan and Kyrgyzstan. Acta Societatis Botanicorum Poloniae 86: 3543.

Clements GR, Rayan DM, Aziz SA, Kawanishi K, Traeholt C, Magintan D, Yazi MFA, Tingley R. 2012. Predicting the distribution of the Asian tapir in Peninsular Malaysia using maximum entropy modeling. Integrative Zoology 7: 400–406.

Coleman T, Mentch L, Fink D, La Sorte FA, Winkler DW, Hooker G, Hochachka WM. 2020. Statistical inference on tree swallow migrations with random forests. Journal of the Royal Statistical Society: Series C (Applied Statistics) 69: 973–989.

Colli-Silva M, Pirani JR. 2020. Estimating bioregions and undercollected areas in South America by revisiting Byttnerioideae, Helicteroideae and Sterculioideae (Malvaceae) occurrence data. Flora 271: 151688.

Collins SD, Abbott JC, McIntyre NE. 2017. Quantifying the degree of bias from using county-scale data in species distribution modeling: Can increasing sample size or using county-averaged environmental data reduce distributional overprediction? Ecology and Evolution 7: 6012–6022.

Condet M, Dulau-Drouot V. 2016. Habitat selection of two island-associated dolphin species from the south-west Indian Ocean. Continental Shelf Research 125: 18–27.

Cooley JR, Arguedas N, Bonaros E, Bunker G, Chiswell SM, DeGiovine A, Edwards M, Hassanieh D, Haji D, Knox J, Kritsky G, Mills C, Mozgai D, Troutman R, Zyla J, Hasegawa H, Sota T, Yoshimura J, Simon C. 2018. The periodical cicada four-year acceleration hypothesis revisited and the polyphyletic nature of Brood V, including an updated crowd-source enhanced map (Hemiptera: Cicadidae: Magicicada). PeerJ 6: e5282.

Coro G, Pagano P, Ellenbroek A. 2013. Combining simulated expert knowledge with Neural Networks to produce Ecological Niche Models for Latimeria chalumnae. Ecological Modelling 268: 55–63.

Coron C, Calenge C, Giraud C, Julliard R. 2018. Bayesian estimation of species relative abundances and habitat preferences using opportunistic data. Environmental and Ecological Statistics 25: 71–93.

Costa H, Medeiros V, Azevedo EB, Silva L. 2013. Evaluating ecological-niche factor analysis as a modelling tool for environmental weed management in island systems. Weed Research 53: 221–230.

Costa H, Ponte NB, Azevedo EB, Gil A. 2015. Fuzzy set theory for predicting the potential distribution and cost-effective monitoring of invasive species. Ecological Modelling 316: 122–132.

Couturier T, Besnard A, Bertolero A, Bosc V, Astruc G, Cheylan M. 2014. Factors determining the abundance and occurrence of Hermann's tortoise Testudo hermanni in France and Spain: Fire regime and landscape changes as the main drivers. Biological Conservation 170: 177–187.

Coxen CL, Frey JK, Carleton SA, Collins DP. 2017. Species distribution models for a migratory bird based on citizen science and satellite tracking data. Global Ecology and Conservation 11: 298–311.

Crawford PHC, Hoagland BW. 2010. Using species distribution models to guide conservation at the state level: the endangered American burying beetle (Nicrophorus americanus) in

Oklahoma. Journal of Insect Conservation 14: 511–521.
Cruz-Cárdenas G, López-Mata L, Villaseñor JL, Ortiz E. 2014. Potential species distribution modeling and the use of principal component analysis as predictor variables. Revista Mexicana de Biodiversidad 85: 189–199.
Dallas T, Huang S, Nunn C, Park AW, Drake JM. 2017. Estimating parasite host range. Proceedings of the Royal Society B: Biological Sciences 284: 20171250.
De Wysiecki AM, Sánchez-Carnero N, Irigoyen AJ, Milessi AC, Colonello JH, Bovcon ND, Cortés F, Barbini SA, Cedrola PV, Coller NM, Jaureguizar AJ. 2020. Using temporally explicit habitat suitability models to infer the migratory pattern of a large mobile shark. Canadian Journal of Fisheries and Aquatic Sciences 77: 1529–1539.
Deb JC, Rahman HMT, Roy A. 2016. Freshwater Swamp Forest Trees of Bangladesh Face Extinction Risk from Climate Change. Wetlands 36: 323–334.
Deidun A, Castriota L, Falautano M, Maggio T. 2017a. Yet another angelfish species for the Mediterranean – the first record of Holacanthus africanus Cadenat, 1951 from Maltese waters, central Mediterranean. BioInvasions Records 6: 373–376.
Deidun A, Sciberras J, Sciberras A, Gauci A, Balistreri P, Salvatore A, Piraino S. 2017b. The first record of the white-spotted Australian jellyfish Phyllorhiza punctata von Lendenfeld, 1884 from Maltese waters (western Mediterranean) and from the Ionian coast of Italy. BioInvasions Records 6: 119–124.
Dennis EB, Morgan BJT, Freeman SN, Ridout MS, Brereton TM, Fox R, Powney GD, Roy DB. 2017. Efficient occupancy model-fitting for extensive citizen-science data. PLOS ONE 12: e0174433.
Dias D, Baringo Fonseca C, Correa L, Soto N, Portela A, Juarez K, Tumolo Neto RJ, Ferro M, Gonçalves J, Junior J. 2017. Repatriation Data: More than two million species occurrence records added to the Brazilian Biodiversity Information Facility Repository (SiBBr). Biodiversity Data Journal 5: e12012.
Dimson M, Lynch SC, Gillespie TW. 2019. Using biased sampling data to model the distribution of invasive shot-hole borers in California. Biological Invasions 21: 2693–2712.
Ding F, Ma T, Hao M, Wang Q, Chen S, Wang D, Huang L, Zhang X, Jiang D. 2020. Mapping Worldwide Environmental Suitability for Artemisia annua L. Sustainability 12: 1309.
Dorazio RM. 2012. Predicting the Geographic Distribution of a Species from Presence-Only Data Subject to Detection Errors. Biometrics 68: 1303–1312.
Drake JM, Randin C, Guisan A. 2006. Modelling ecological niches with support vector machines. Journal of Applied Ecology 43: 424–432.
Duan R-Y, Kong X-Q, Huang M-Y, Fan W-Y, Wang Z-G. 2014. The Predictive Performance and Stability of Six Species Distribution Models. PLoS ONE 9: e112764.
Dwyer RG, Carpenter-Bundhoo L, Franklin CE, Campbell HA. 2016. Using citizen-collected wildlife sightings to predict traffic strike hot spots for threatened species: a case study on the southern cassowary. Journal of Applied Ecology 53: 973–982.
Dyer RJ, Gillings S, Pywell RF, Fox R, Roy DB, Oliver TH. 2017. Developing a biodiversity-based indicator for large-scale environmental assessment: a case study of proposed shale gas extraction sites in Britain. Journal of Applied Ecology 54: 872–882.
El-Gabbas A, Dormann CF. 2018. Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. Ecography 41: 1161–1172.

Etherington TR, Ward AI, Smith GC, Pietravalle S, Wilson GJ. 2009. Using the Mahalanobis distance statistic with unplanned presence-only survey data for biogeographical models of species distribution and abundance: a case study of badger setts. Journal of Biogeography 36: 845–853.

Falk BG, Snow RW, Reed RN. 2016. Prospects and Limitations of Citizen Science in Invasive Species Management: A Case Study with Burmese Pythons in Everglades National Park. Southeastern Naturalist 15: 89–102.

Ferreira S, Tierno de Figueroa JM, Martins F, Verissimo J, Quaglietta L, Grosso-Silva JM, Lopes P, Sousa P, Paupério J, Fonseca N, Beja P. 2020. The InBIO Barcoding Initiative Database: contribution to the knowledge on DNA barcodes of Iberian Plecoptera. Biodiversity Data Journal 8: e55137.

Flaherty M, Lawton C. 2019. The regional demise of a non-native invasive species: the decline of grey squirrels in Ireland. Biological Invasions 21: 2401–2416.

Forero-Medina G, Cárdenas-Arévalo G, Castaño-Mora OV. Habitat modeling of Dahl's toad-headed turtle (Mesoclemmys dahli) in Colombia. Herpetological Conservation and Biology 7: 313-322.

Frey JK, Calkins MT. 2014. Snow cover and riparian habitat determine the distribution of the short-tailed weasel (Mustela erminea) at its southern range limits in arid western North America. 78: 45-56.

Froese JG, Smith CS, Durr PA, McAlpine CA, van Klinken RD. 2017. Modelling seasonal habitat suitability for wide-ranging species: Invasive wild pigs in northern Australia. PLOS ONE 12: e0177018.

Gamliel I, Buba Y, Guy-Haim T, Garval T, Willette D, Rilov G, Belmaker J. 2020. Incorporating physiology into species distribution models moderates the projected impact of warming on selected Mediterranean marine species. Ecography 43: 1090–1106.

Gebrewahid Y, Abrehe S, Meresa E, Eyasu G, Abay K, Gebreab G, Kidanemariam K, Adissu G, Abreha G, Darcha G. 2020. Current and future predicting potential areas of Oxytenanthera abyssinica (A. Richard) using MaxEnt model under climate change in Northern Ethiopia. Ecological Processes 9: 6.

Gerstner BE, Kass JM, Kays R, Helgen KM, Anderson RP. 2018. Revised distributional estimates for the recently discovered olinguito (Bassaricyon neblina), with comments on natural and taxonomic history. Journal of Mammalogy 99: 321–332.

Ghisbain G, Williams PH, Michez D, Branstetter MG, Rasmont P. 2020. Contribution to the knowledge of the bumblebee fauna of Afghanistan (Hymenoptera, Apidae, Bombus Latreille). ZooKeys 973: 69–87.

Gibson L, Barrett B, Burbidge A. 2007. Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot: Uncertain absences in habitat modelling of a rare bird. Diversity and Distributions 13: 704–713.

Girardello M, Chapman A, Dennis R, Kaila L, Borges PAV, Santangeli A. 2019. Gaps in butterfly inventory data: A global analysis. Biological Conservation 236: 289–295.

Goberville E, Hautekèete N-C, Kirby RR, Piquot Y, Luczak C, Beaugrand G. 2016. Climate change and the ash dieback crisis. Scientific Reports 6: 35303.

Golding N, Purse BV. 2016. Fast and flexible Bayesian species distribution modelling using Gaussian processes. Methods in Ecology and Evolution 7: 598–608.

Goldstein EA, Lawton C, Sheehy E, Butler F. 2014. Locating species range frontiers: a cost and efficiency comparison of citizen science and hair-tube survey methods for use in tracking an invasive squirrel. Wildlife Research 41: 64.

Gómez-Martínez MA, Klem D, Rojas-Soto O, González-García F, MacGregor-Fors I. 2019. Window strikes: bird collisions in a Neotropical green city. Urban Ecosystems 22: 699–708.

Gooliaff T, Weir RD, Hodges KE. 2018. Estimating bobcat and Canada lynx distributions in British Columbia: Estimating Bobcat and Lynx Distributions. The Journal of Wildlife Management 82: 810–820.

Gottwald J, Appelhans T, Adorf F, Hillen J, Nauss T. 2017. High-Resolution MaxEnt Modelling of Habitat Suitability for Maternity Colonies of the Barbastelle Bat Barbastella barbastellus (Schreber, 1774) in Rhineland-Palatinate, Germany. Acta Chiropterologica 19: 389–398.

Griffin SC, Taper ML, Hoffman R, Mills LS. 2010. Ranking Mahalanobis Distance Models for Predictions of Occupancy From Presence-Only Data. Journal of Wildlife Management 74: 1112–1121.

Guevara L, León-Paniagua L. 2019. How to survive a glaciation: the challenge of estimating biologically realistic potential distributions under freezing conditions. Ecography 42: 1237–1245.

Guisan A, Zimmermann NE, Elith J, Graham CH, Phillips S, Peterson AT. 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? Ecological Monographs 77: 615–630.

Hanberry BB, He HS, Palik BJ. 2012. Pseudoabsence Generation Strategies for Species Distribution Models. PLoS ONE 7: e44486.

Hann C, Stelle L, Szabo A, Torres L. 2018. Obstacles and Opportunities of Using a Mobile App for Marine Mammal Research. ISPRS International Journal of Geo-Information 7: 169.

Haque MdM, Nipperess DA, Gallagher RV, Beaumont LJ. 2017. How well documented is Australia's flora? Understanding spatial bias in vouchered plant specimens. Austral Ecology 42: 690–699.

Harvey DS, Platenberg RJ. Predicting habitat use from opportunistic observations: a case study of the Virgin Islands tree boa (Epicrates granti). Herpetological Journal 19: 111-118.

Hefley TJ, Hooten MB. 2015. On the existence of maximum likelihood estimates for presence-only data. Methods in Ecology and Evolution 6: 648–655.

Hengl T, Sierdsema H, Radović A, Dilo A. 2009. Spatial prediction of species' distributions from occurrence-only records: combining point pattern analysis, ENFA and regression-kriging. Ecological Modelling 220: 3499–3511.

Heringer G, Almeida TE, Dittrich VA de O, Salino A. 2020. Assessing the effectiveness of protected areas for the conservation of ferns and lycophytes in the Brazilian state of Minas Gerais. Journal for Nature Conservation 53: 125775.

Hernandez PA, Graham CH, Master LL, Albert DL. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography 29: 773–785.

Herrera Campo BV, Hyman G, Bellotti A. 2011. Threats to cassava production: known and potential geographic distribution of four key biotic constraints. Food Security 3: 329–345.

Hertzog LR, Besnard A, Jay-Robert P. 2014. Field validation shows bias-corrected pseudo-absence selection is the best method for predictive species-distribution modelling. Diversity and Distributions 20: 1403–1413.

Higa M, Yamaura Y, Koizumi I, Yabuhara Y, Senzaki M, Ono S. 2015. Mapping large-scale bird distributions using occupancy models and citizen data with spatially biased sampling effort. Diversity and Distributions 21: 46–54.

Hill NJ, Tobin AJ, Reside AE, Pepperell JG, Bridge TCL. 2016. Dynamic habitat suitability modelling reveals rapid poleward distribution shift in a mobile apex predator. Global Change Biology 22: 1086–1096.

Hughes AC, Satasook C, Bates PJJ, Soisook P, Sritongchuay T, Jones G, Bumrungsri S. 2010. Echolocation Call Analysis and Presence-Only Modelling as Conservation Monitoring Tools for Rhinolophoid Bats in Thailand. Acta Chiropterologica 12: 311–327.

Isaac B, White J, Ierodiaconou D, Cooke R. 2013. Response of a cryptic apex predator to a complete urban to forest gradient. Wildlife Research 40: 427.

Iturbide M, Bedia J, Gutiérrez JM. 2018. Background sampling and transferability of species distribution model ensembles under climate change. Global and Planetary Change 166: 19–29.

Jackson MM, Gergel SE, Martin K. 2015. Effects of Climate Change on Habitat Availability and Configuration for an Endemic Coastal Alpine Bird. PLOS ONE 10: e0142110.

Jarnevich CS, Talbert M, Morisette J, Aldridge C, Brown CS, Kumar S, Manier D, Talbert C, Holcombe T. 2017. Minimizing effects of methodological decisions on interpretation and prediction in species distribution studies: An example with background selection. Ecological Modelling 363: 48–56.

Jaźwa M, Jędrzejczak E, Klichowska E, Pliszko A. Predicting the potential distribution area of Solidago ×niederederi (Asteraceae). Turkish Journal of Botany 42: 51-56.

Jezkova T, Wiens JJ. 2018. Testing the role of climate in speciation: New methods and applications to squamate reptiles (lizards and snakes). Molecular Ecology 27: 2754–2769.

Jiménez-Valverde A, Decae AE, Arnedo MA. 2011. Environmental suitability of new reported localities of the funnelweb spider Macrothele calpeiana: an assessment using potential distribution modelling with presence-only techniques: Environmental suitability of new localities of Macrothele calpeiana. Journal of Biogeography 38: 1213–1223.

Jiménez-Valverde A, Ortuño VM, Lobo JM. 2007. Exploring the distribution of Sterocorax Ortuño, 1990 (Coleoptera, Carabidae) species in the Iberian Peninsula: Distribution of Sterocorax species in the Iberian Peninsula. Journal of Biogeography 34: 1426–1438.

Jinga P, Ashley MV. 2019. Climate change threatens some miombo tree species of sub-Saharan Africa. Flora 257: 151421.

Jinga P, Palagi J. 2020. Dry and wet miombo woodlands of south-central Africa respond differently to climate change. Environmental Monitoring and Assessment 192: 372.

Jones A, Bruce E, Davies KP, Blewitt M, Sheehan S. 2019. Assessing potential environmental influences on killer whale (Orcinus orca) distribution patterns in the Bremer Canyon, south-west Australia. Australian Geographer 50: 381–405.

Joyce M, Barnes MD, Possingham HP, Van Rensburg BJ. 2018. Understanding avian assemblage change within anthropogenic environments using citizen science data. Landscape and Urban Planning 179: 81–89.

Justine J-L, Winsor L, Gey D, Gros P, Thévenot J. 2018. Giant worms chez moi! Hammerhead

flatworms (Platyhelminthes, Geoplanidae, Bipalium spp., Diversibipalium spp.) in metropolitan France and overseas French territories. PeerJ 6: e4672.
Kaminski DJ, Poole KE, Clark KB, Harms TM. 2020. Predicting landscape-scale summer resource selection for the northern long-eared bat (Myotis septentrionalis) in Iowa. Journal of Mammalogy 101: 172–186.
Kanagaraj R, Wiegand T, Kramer-Schadt S, Anwar M, Goyal SP. 2011. Assessing habitat suitability for tiger in the fragmented Terai Arc Landscape of India and Nepal. Ecography 34: 970–981.
Keil P, Wilson AM, Jetz W. 2014. Uncertainty, priors, autocorrelation and disparate data in downscaling of species distributions. Diversity and Distributions 20: 797–812.
Kienberger K, Prieto L. 2018. The jellyfish Rhizostoma luteum (Quoy & Gaimard, 1827): not such a rare species after all. Marine Biodiversity 48: 1455–1462.
Kojola I, Heikkinen S, Holmala K. 2018. Balancing costs and confidence: volunteer-provided point observations, GPS telemetry and the genetic monitoring of Finland's wolves. Mammal Research 63: 415–423.
Kolanowska M. 2013. Glacial refugia and migration routes of the Neotropical genus Trizeuxis (Orchidaceae). Acta Societatis Botanicorum Poloniae 82: 225-230.
Kramer-Schadt S, Niedballa J, Pilgrim JD, Schröder B, Lindenborn J, Reinfelder V, Stillfried M, Heckmann I, Scharf AK, Augeri DM, Cheyne SM, Hearn AJ, Ross J, Macdonald DW, Mathai J, Eaton J, Marshall AJ, Semiadi G, Rustam R, Bernard H, Alfred R, Samejima H, Duckworth JW, Breitenmoser-Wuersten C, Belant JL, Hofer H, Wilting A. 2013. The importance of correcting for sampling bias in MaxEnt species distribution models. Diversity and Distributions 19: 1366–1379.
Kress WJ, Garcia-Robledo C, Soares JVB, Jacobs D, Wilson K, Lopez IC, Belhumeur PN. 2018. Citizen Science and Climate Change: Mapping the Range Expansions of Native and Exotic Plants with the Mobile App Leafsnap. BioScience 68: 348–358.
Laaksonen M, Sajanti E, Sormunen JJ, Penttinen R, Hänninen J, Ruohomäki K, Sääksjärvi I, Vesterinen EJ, Vuorinen I, Hytönen J, Klemola T. 2017. Crowdsourcing-based nationwide tick collection reveals the distribution of Ixodes ricinus and I. persulcatus and associated pathogens in Finland. Emerging Microbes & Infections 6: 1–7.
Leibovici D, Williams J, Rosser J, Hodges C, Chapman C, Higgins C, Jackson M. 2017. Earth Observation for Citizen Science Validation, or Citizen Science for Earth Observation Validation? The Role of Quality Assurance of Volunteered Observations. Data 2: 35.
Leidenberger S, Boström S, Wayland M. 2020. Host records and geographical distribution of Corynosoma magdaleni, C. semerme and C. strumosum (Acanthocephala: Polymorphidae). Biodiversity Data Journal 8: e50500.
Lenanton RCJ, Dowling CE, Smith KA, Fairclough DV, Jackson G. 2017. Potential influence of a marine heatwave on range extensions of tropical fishes in the eastern Indian Ocean—Invaluable contributions from amateur observers. Regional Studies in Marine Science 13: 19–31.
Leung B, Hudgins EJ, Potapova A, Ruiz-Jaen MC. 2019. A new baseline for countrywide α-diversity and species distributions: illustration using >6,000 plant species in Panama. Ecological Applications 29: e01866.
Li E, Parker SS, Pauly GB, Randall JM, Brown BV, Cohen BS. 2019. An Urban Biodiversity Assessment Framework That Combines an Urban Habitat Classification Scheme and

Citizen Science Data. Frontiers in Ecology and Evolution 7: 277.
Linhoss AC, Camacho R, Ashby S. 2016. Oyster Habitat Suitability in the Northern Gulf of Mexico. Journal of Shellfish Research 35: 841–849.
Longcore T, Noujdina N, Dixon PJ. 2018. Landscape Modeling of the Potential Natural Vegetation of Santa Catalina Island, California. Western North American Naturalist 78: 617.
Ma B, Sun J. 2018. Predicting the distribution of Stipa purpurea across the Tibetan Plateau via the MaxEnt model. BMC Ecology 18: 10.
MacLean MG, Congalton RG. 2015. A comparison of landscape fragmentation analysis programs for identifying possible invasive plant species locations in forest edge. Landscape Ecology 30: 1241–1256.
Mair L, Harrison PJ, Räty M, Bärring L, Strandberg G, Snäll T. 2017. Forest management could counteract distribution retractions forced by climate change. Ecological Applications 27: 1485–1497.
Maistrello L, Dioli P, Bariselli M, Mazzoli GL, Giacalone-Forini I. 2016. Citizen science and early detection of invasive species: phenology of first occurrences of Halyomorpha halys in Southern Europe. Biological Invasions 18: 3109–3116.
Mäkinen J, Vanhatalo J. 2018. Hierarchical Bayesian model reveals the distributional shifts of Arctic marine mammals. Diversity and Distributions 24: 1381–1394.
Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Albán J, Chilquillo E, Rønsted N, Antonelli A. 2015. Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases?: Species diversity and distribution in the era of Big Data. Global Ecology and Biogeography 24: 973–984.
Malek R, Tattoni C, Ciolli M, Corradini S, Andreis D, Ibrahim A, Mazzoni V, Eriksson A, Anfora G. 2018. Coupling Traditional Monitoring and Citizen Science to Disentangle the Invasion of Halyomorpha halys. ISPRS International Journal of Geo-Information 7: 171.
Mammola S, Cardoso P, Angyal D, Balázs G, Blick T, Brustel H, Carter J, Ćurčić S, Danflous S, Dányi L, Déjean S, Deltshev C, Elverici M, Fernández J, Gasparo F, Komnenov M, Komposch C, Kováč L, Kunt K, Mock A, Moldovan O, Naumova M, Pavlek M, Prieto C, Ribera C, Rozwałka R, Růžička V, Vargovitsh R, Zaenker S, Isaia M. 2019. Continental data on cave-dwelling spider communities across Europe (Arachnida: Araneae). Biodiversity Data Journal 7: e38492.
Manenti R, Mori E, Di Canio V, Mercurio S, Picone M, Caffi M, Brambilla M, Ficetola GF, Rubolini D. 2020. The good, the bad and the ugly of COVID-19 lockdown effects on wildlife conservation: Insights from the first European locked down country. Biological Conservation 249: 108728.
Mang T, Essl F, Moser D, Karrer G, Kleinbauer I, Dullinger S. 2017. Accounting for imperfect observation and estimating true species distributions in modelling biological invasions. Ecography 40: 1187–1197.
Marcer A, Pino J, Pons X, Brotons L. 2012. Modelling invasive alien species distributions from digital biodiversity atlases. Model upscaling as a means of reconciling data at different scales. Diversity and Distributions 18: 1177–1189.
Marchante H, Morais MC, Gamela A, Marchante E. 2017. Using a WebMapping Platform to Engage Volunteers to Collect Data on Invasive Plants Distribution. Transactions in GIS 21: 238–252.

Marra S, de Lucia GA, Camedda A, Espinosa F, Coppa S. 2016. New records of the distribution and conservation status of the endangered limpet Patella ferruginea in Sardinia (Italy, W Mediterranean): Status and distribution of P. ferruginea in central-western Sardinia. Aquatic Conservation: Marine and Freshwater Ecosystems 26: 607–612.

Maslo B, Lockwood JL, Leu K. 2015. Land ownership patterns associated with declining forest birds: targeting the right policy and management for the right birds. Environmental Conservation 42: 216–226.

Mason SC, Hill JK, Thomas CD, Powney GD, Fox R, Brereton T, Oliver TH. 2018. Population variability in species can be deduced from opportunistic citizen science records: a case study using British butterflies. Insect Conservation and Diversity 11: 131–142.

da Mata RA, Tidon R, de Oliveira G, Vilela B, Diniz-Filho JAF, Rangel TF, Terribile LC. 2017. Stacked species distribution and macroecological models provide incongruent predictions of species richness for Drosophilidae in the Brazilian savanna. Insect Conservation and Diversity 10: 415–424.

McLeod SR, Pople AR. 2010. Modelling the distribution and relative abundance of feral camels in the Northern Territory using count data. The Rangeland Journal 32: 21.

van der Meer E. 2018. Carnivore conservation under land use change: the status of Zimbabwe's cheetah population after land reform. Biodiversity and Conservation 27: 647–663.

Metcalfe-Smith JL, Staton SK, Mackie GL, Lane NM. 1998. Selection of candidate species of freshwater mussels (Bivalvia: Unionidae) to be considered for national status designation by COSEWIC. Canadian Field Naturalist 112: 425-440.

Mononen L, Auvinen A-P, Packalen P, Virkkala R, Valbuena R, Bohlin I, Valkama J, Vihervaara P. 2018. Usability of citizen science observations together with airborne laser scanning data in determining the habitat preferences of forest birds. Forest Ecology and Management 430: 498–508.

Monsarrat S, Novellie P, Rushworth I, Kerley G. Shifted distribution baselines: neglecting long-term biodiversity records risks overlooking potentially suitable habitat for conservation management. 374: 20190215.

Morán-Ordóñez A, Lahoz-Monfort JJ, Elith J, Wintle BA. 2017. Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions?: Temporal transferability of species distribution model predictions. Global Ecology and Biogeography 26: 371–384.

Moro S, Jona-Lasinio G, Block B, Micheli F, De Leo G, Serena F, Bottaro M, Scacco U, Ferretti F. 2020. Abundance and distribution of the white shark in the Mediterranean Sea. Fish and Fisheries 21: 338–349.

Morzaria-Luna HN, Cruz-Piñón G, Brusca RC, López-Ortiz AM, Moreno-Báez M, Reyes-Bonilla H, Turk-Boyer P. 2018. Biodiversity hotspots are not congruent with conservation areas in the Gulf of California. Biodiversity and Conservation 27: 3819–3842.

Mossman HL, Panter CJ, Dolman PM. 2015. Modelling biodiversity distribution in agricultural landscapes to support ecological network planning. Landscape and Urban Planning 141: 59–67.

Mostafavi H, Pletterbauer F, Coad BW, Mahini AS, Schinegger R, Unfer G, Trautwein C, Schmutz S. 2014. Predicting presence and absence of trout (Salmo trutta) in Iran. Limnologica 46: 1–8.

Moua Y, Roux E, Girod R, Dusfour I, de Thoisy B, Seyler F, Briolant S. 2016. Distribution of the Habitat Suitability of the Main Malaria Vector in French Guiana Using Maximum Entropy Modeling. Journal of Medical Entomology 54: 606-621.
Moulin N. 2020. When Citizen Science highlights alien invasive species in France: the case of Indochina mantis, Hierodula patellifera (Insecta, Mantodea, Mantidae). Biodiversity Data Journal 8: e46989.
Mueller MA, Drake D, Allen ML. 2019. Using citizen science to inform urban canid management. Landscape and Urban Planning 189: 362–371.
Muscarella R, Galante PJ, Soley-Guardia M, Boria RA, Kass JM, Uriarte M, Anderson RP. 2014. ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecology and Evolution 5: 1198–1205.
Mweya CN, Kimera SI, Stanley G, Misinzo G, Mboera LEG. 2016. Climate Change Influences Potential Distribution of Infected Aedes aegypti Co-Occurrence with Dengue Epidemics Risk Areas in Tanzania. PLOS ONE 11: e0162649.
Myers AT, Gibbs JP. 2013. Landscape-level Factors Influencing Bog Turtle Persistence and Distribution in Southeastern New York State. Journal of Fish and Wildlife Management 4: 255–266.
Nimbs MJ, Willan RC, Smith SDA. 2017. Is Port Stephens, eastern Australia, a global hotspot for biodiversity of Aplysiidae (Gastropoda: Heterobranchia)? Molluscan Research 37: 45–65.
Norris D. 2014. Model Thresholds are More Important than Presence Location Type: Understanding the Distribution of Lowland tapir (Tapirus Terrestris) in a Continuous Atlantic Forest of Southeast Brazil. Tropical Conservation Science 7: 529–547.
Nottingham AC, Thompson JA, Wood F, Edwards PJ, Strager MP. 2019. Mapping pedomemory of spodic morphology using a species distribution model. Geoderma 352: 330–341.
Obidziński A, Pabjanek P, Mędrzycki P. 2013. Determinants of badger Meles meles sett location in Białowieża Primeval Forest, northeastern Poland. Wildlife Biology 19: 48–68.
Obsomer V, Defourny P, Coosemans M. 2012. Predicted Distribution of Major Malaria Vectors Belonging to the Anopheles dirus Complex in Asia: Ecological Niche and Environmental Influences. PLoS ONE 7: e50475.
Ofori BY, Stow AJ, Baumgartner JB, Beaumont LJ. 2017. Combining dispersal, landscape connectivity and habitat suitability to assess climate-induced changes in the distribution of Cunningham's skink, Egernia cunninghami. PLOS ONE 12: e0184193.
Olson J, Wood J, Osborne R, Barrett-Lennard L, Larson S. 2018. Sightings of southern resident killer whales in the Salish Sea 1976-2014: the importance of a long-term opportunistic dataset. Endangered Species Research 37: 105–118.
Olsson O, Rogers DJ. 2009. Predicting the distribution of a suitable habitat for the white stork in Southern Sweden: identifying priority areas for reintroduction and habitat restoration. Animal Conservation 12: 62–70.
Osawa T, Watanabe K, Ikeda H, Yamamoto S. 2014. New approach for evaluating habitat stability using scarce records for both historical and contemporary specimens: a case study using Carabidae specimen records: Evaluating habitat with limited records. Entomological Science 17: 425–431.
Ostrowski M-F, Prosperi J-M, David J. 2016. Potential Implications of Climate Change on

Aegilops Species Distribution: Sympatry of These Crop Wild Relatives with the Major European Crop Triticum aestivum and Conservation Issues. PLOS ONE 11: e0153974.
Ottaviani D, Lasinio GJ, Boitani L. 2004. Two statistical methods to validate habitat suitability models using presence-only data. Ecological Modelling 179: 417–443.
Oviedo L, Fernández M, Herra-Miranda D, Pacheco-Polanco JD, Hernández-Camacho CJ, Aurioles-Gamboa D. 2018. Habitat partitioning mediates the coexistence of sympatric dolphins in a tropical fjord-like embayment. Journal of Mammalogy 99: 554–564.
Pace DS, Giacomini G, Campana I, Paraboschi M, Pellegrino G, Silvestri M, Alessi J, Angeletti D, Cafaro V, Pavan G, Ardizzone G, Arcangeli A. 2019. An integrated approach for cetacean knowledge and conservation in the central Mediterranean Sea using research and social media data sources. Aquatic Conservation: Marine and Freshwater Ecosystems 29: 1302–1323.
Pantaleoni R. 2019. Going overseas: from island to continent colonization in the Mediterranean snakefly Fibla maclachlani (Albarda, 1891). BioInvasions Records 8: 442–451.
Pauly K, Jupp BP, De Clerck O. 2011. Modelling the distribution and ecology of Trichosolen blooms on coral reefs worldwide. Marine Biology 158: 2239–2246.
Pearson RG, Thuiller W, Araújo MB, Martinez-Meyer E, Brotons L, McClean C, Miles L, Segurado P, Dawson TP, Lees DC. 2006. Model-based uncertainty in species range prediction. Journal of Biogeography 33: 1704–1711.
Penado A, Rebelo H, Goulson D. 2016. Spatial distribution modelling reveals climatically suitable areas for bumblebees in undersampled parts of the Iberian Peninsula. Insect Conservation and Diversity 9: 391–401.
Pereira AJ, Francisco A, Porto M. 2016. Flora-On: Occurrence data of the vascular flora of mainland Portugal. PhytoKeys 69: 105–119.
Périquet S, Roxburgh L, le Roux A, Collinson WJ. 2018. Testing the Value of Citizen Science for Roadkill Studies: A Case Study from South Africa. Frontiers in Ecology and Evolution 6: 15.
Pertierra LR, Bartlett JC, Duffy GA, Vega GC, Hughes KA, Hayward SAL, Convey P, Olalla-Tarraga MA, Aragón P. 2020. Combining correlative and mechanistic niche models with human activity data to elucidate the invasive potential of a sub-Antarctic insect. Journal of Biogeography 47: 658–673.
Pettorelli N, Hilborn A, Broekhuis F, Durant SM. 2009. Exploring habitat use by cheetahs using ecological niche factor analysis. Journal of Zoology 277: 141–148.
Phillips SJ, Dudík M, Elith J, Graham CH, Lehmann A, Leathwick J, Ferrier S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecological Applications 19: 181–197.
Phipps WL, Diekmann M, MacTavish LM, Mendelsohn JM, Naidoo V, Wolter K, Yarnell RW. 2017. Due South: A first assessment of the potential impacts of climate change on Cape vulture occurrence. Biological Conservation 210: 16–25.
Platenberg RJ, Harvey DS. Endangered species and land use conflicts: a case study of the Virgin Islands boa (Epicrates granti). Herpetological Conservation and Biology 5: 548-554.
Polgar CA, Primack RB, Williams EH, Stichter S, Hitchcock C. 2013. Climate effects on the flight period of Lycaenid butterflies in Massachusetts. Biological Conservation 160: 25–31.

Porter WT, Motyka PJ, Wachara J, Barrand ZA, Hmood Z, McLaughlin M, Pemberton K, Nieto NC. 2019. Citizen science informs human-tick exposure in the Northeastern United States. International Journal of Health Geographics 18: 9.

Puan CL, Yong DL, Azhar B, Phua MH, Lim KC. 2015. Ecological correlations of nocturnal bird assemblages in Malaysian Borneo. Forktail 31: 82-86.

Qin Z, Zhang JE, Jiang YP, Wang RL, Wu RS. 2020. Predicting the potential distribution of Pseudomonas syringae pv. actinidiae in China using ensemble models. Plant Pathology 69: 120–131.

Quevedo M, Bañuelos MJ, Obeso JR. 2006. The decline of Cantabrian capercaillie: How much does habitat configuration matter? Biological Conservation 127: 190–200.

Rapacciuolo G, Ball-Damerow JE, Zeilinger AR, Resh VH. 2017. Detecting long-term occupancy changes in Californian odonates from natural history and citizen science records. Biodiversity and Conservation 26: 2933–2949.

Ready J, Kaschner K, South AB, Eastwood PD, Rees T, Rius J, Agbayani E, Kullander S, Froese R. 2010. Predicting the distributions of marine organisms at the global scale. Ecological Modelling 221: 467–478.

Rebelo H, Jones G. 2010. Ground validation of presence-only modelling with rare species: a case study on barbastelles Barbastella barbastellus (Chiroptera: Vespertilionidae). Journal of Applied Ecology 47: 410–420.

Rebelo H, Tarroso P, Jones G. 2010. Predicted impact of climate change on European bats in relation to their biogeographic patterns. Global Change Biology 16: 561–576.

Redhead JW, Fox R, Brereton T, Oliver TH. 2016. Assessing species' habitat associations from occurrence records, standardised monitoring data and expert opinion: A test with British butterflies. Ecological Indicators 62: 271–278.

Reese GC, Skagen SK. 2017. Modeling nonbreeding distributions of shorebirds and waterfowl in response to climate change. Ecology and Evolution 7: 1497–1513.

Ribas LG dos S, de Cássia-Silva C, Petsch DK, Silveira MJ, Lima-Ribeiro MS. 2018. The potential invasiveness of an aquatic macrophyte reflects founder effects from native niche. Biological Invasions 20: 3347–3355.

Ringvold (HR) H, Hatlevik (AH) A, Hevrøy (JH) J, Hughes (MH) M, Aukan (NA) N. 2020. Encounters with the rare genus Helicosalpa (Chordata, Thaliacea, Salpida), using citizen science data. Marine Biology Research 16: 369–379.

Robinson JL, Fordyce JA, Parker CR. 2016. Conservation of aquatic insect species across a protected area network: null model reveals shortfalls of biogeographical knowledge. Journal of Insect Conservation 20: 565–581.

Rockhill AP, Sollman R, Powell RA, DePerno CS. 2016. A Comparison of Survey Techniques for Medium- to Large-Sized Mammals in Forested Wetlands. Southeastern Naturalist 15: 175–187.

Rodríguez-Soto C, Monroy-Vilchis O, Maiorano L, Boitani L, Faller JC, Briones MÁ, Núñez R, Rosas-Rosas O, Ceballos G, Falcucci A. 2011. Predicting potential distribution of the jaguar (Panthera onca) in Mexico: identification of priority areas for conservation: Jaguar potential distribution in Mexico. Diversity and Distributions 17: 350–361.

Routson KJ, Volk GM, Richards CM, Smith SE, Nabhan GP, Wyllie de Echeverria V. 2012. Genetic Variation and Distribution of Pacific Crabapple. Journal of the American Society for Horticultural Science 137: 325–332.

Rumi A, Vogler RE, Beltramino AA. 2017. The South-American distribution and southernmost record of Biomphalaria peregrina —a potential intermediate host of schistosomiasis. PeerJ 5: e3401.
Saito M, Momose H, Mihira T, Uematsu S. 2012. Predicting the risk of wild boar damage to rice paddies using presence-only data in Chiba Prefecture, Japan. International Journal of Pest Management 58: 65–71.
Sallam MF, Al Ahmed AM, Abdel-Dayem MS, Abdullah MAR. 2013. Ecological Niche Modeling and Land Cover Risk Areas for Rift Valley Fever Vector, Culex tritaeniorhynchus Giles in Jazan, Saudi Arabia. PLoS ONE 8: e65786.
Santori C, Spencer R-J, Van Dyke JU, Thompson MB. 2018. Road mortality of the eastern long-necked turtle (Chelodina longicollis) along the Murray River, Australia: an assessment using citizen science. Australian Journal of Zoology 66: 41.
Sardà-Palomera F, Vieites DR. 2011. Modelling Species' Climatic Distributions Under Habitat Constraints: A Case Study with Coturnix coturnix. Annales Zoologici Fennici 48: 147–160.
Scarnati L, Attorre F, Farcomeni A, Francesconi F, Sanctis M. 2009. Modelling the spatial distribution of tree species with fragmented populations from abundance data. Community Ecology 10: 215–224.
Scholte RGC, Carvalho OS, Malone JB, Utzinger J, Vounatsou P. 2012. Spatial distribution of Biomphalaria spp., the intermediate host snails of Schistosoma mansoni, in Brazil. Geospatial health 6: 95.
Schubert SC, Manica LT, Guaraldo ADC. 2019. Revealing the potential of a huge citizen-science platform to study bird migration. Emu - Austral Ornithology 119: 364–373.
Senay SD, Worner SP, Ikeda T. 2013. Novel Three-Step Pseudo-Absence Selection Technique for Improved Species Distribution Modelling. PLoS ONE 8: e71218.
Shabani F, Kumar L. 2015. Should species distribution models use only native or exotic records of existence or both? Ecological Informatics 29: 57–65.
Sharifi M, Karami P, Akmali V, Afroosheh M, Vaissi S. Modeling Geographic Distribution for the Endangered Yellow Spotted Mountain Newt, Neurergus microspilotus (Amphibia: Salamandridae) in Iran and Iraq. Herpetological Conservation and Biology 12: 488–497.
Sharma S, Arunachalam K, Bhavsar D, Kala R. 2018. Modeling habitat suitability of Perilla frutescens with MaxEnt in Uttarakhand—A conservation approach. Journal of Applied Research on Medicinal and Aromatic Plants 10: 99–105.
Shatz AJ, Rogan J, Sangermano F, Ogneva-Himmelberger Y, Chen H. 2013. Characterizing the potential distribution of the invasive Asian longhorned beetle (Anoplophora glabripennis) in Worcester County, Massachusetts. Applied Geography 45: 259–268.
Shirakihara M. 2012. Long-Distance Movements of Indo-Pacific Bottlenose Dolphins (Tursiops aduncus) and Habitat Preference of Two Species of Bottlenose Dolphins in Eastern Kyushu, Japan. Aquatic Mammals 38: 145–152.
Shumba T, Montgomery RA, Rasmussen GSA, Macdonald DW. 2018. African Wild Dog Habitat Use Modelling Using Telemetry Data and Citizen Scientist Sightings: Are the Results Comparable? African Journal of Wildlife Research 48: 013002.
Sindato C, Stevens KB, Karimuribo ED, Mboera LEG, Paweska JT, Pfeiffer DU. 2016. Spatial Heterogeneity of Habitat Suitability for Rift Valley Fever Occurrence in Tanzania: An Ecological Niche Modelling Approach. PLOS Neglected Tropical Diseases 10: e0005002.

Skejo J, Deranja M, Adžić K. 2020. Pygmy Hunchback of New Caledonia: Notredamia dora gen. n. et sp. n. – A New Cladonotin (Caelifera: Tetrigidae) Genus and Species from Oceania. Entomological News 129: 170.

Soley-Guardia M, Gutiérrez EE, Thomas DM, Ochoa-G J, Aguilera M, Anderson RP. 2016. Are we overestimating the niche? Removing marginal localities helps ecological niche models detect environmental barriers. Ecology and Evolution 6: 1267–1279.

Soroye P, Ahmed N, Kerr JT. 2018. Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research. Global Change Biology 24: 5281–5291.

Sousa-Baena MS, Garcia LC, Townsend Peterson A. 2014. Knowledge behind conservation status decisions: Data basis for "Data Deficient" Brazilian plant species. Biological Conservation 173: 80–89.

Speed JDM, Bendiksby M, Finstad AG, Hassel K, Kolstad AL, Prestø T. 2018. Contrasting spatial, temporal and environmental patterns in observation and specimen based species occurrence data. PLOS ONE 13: e0196417.

Stas M, Aerts R, Hendrickx M, Dendoncker N, Dujardin S, Linard C, Nawrot T, Van Nieuwenhuyse A, Aerts J-M, Van Orshoven J, Somers B. 2020. An evaluation of species distribution models to estimate tree diversity at genus level in a heterogeneous urban-rural landscape. Landscape and Urban Planning 198: 103770.

Stafford R, Hart AG, Goodenough AE. 2013. A visual method to identify significant latitudinal changes in species' distributions. Ecological Informatics 15: 74–84.

Stefanescu C, Páramo F, Åkesson S, Alarcón M, Ávila A, Brereton T, Carnicer J, Cassar LF, Fox R, Heliölä J, Hill JK, Hirneisen N, Kjellén N, Kühn E, Kuussaari M, Leskinen M, Liechti F, Musche M, Regan EC, Reynolds DR, Roy DB, Ryrholm N, Schmaljohann H, Settele J, Thomas CD, van Swaay C, Chapman JW. 2013. Multi-generational long-distance migration of insects: studying the painted lady butterfly in the Western Palaearctic. Ecography 36: 474–486.

Stone AI, Lima EM, Aguiar GFS, Camargo CC, Flores TA, Kelt DA, Marques-Aguiar SA, Queiroz JAL, Ramos RM, Silva Júnior JS. 2009. Non-volant mammalian diversity in fragments in extreme eastern Amazonia. Biodiversity and Conservation 18: 1685–1694.

Strebel N, Kéry M, Schaub M, Schmid H. 2014. Studying phenology by flexible modelling of seasonal detectability peaks. Methods in Ecology and Evolution 5: 483–490.

Supp SR, La Sorte FA, Cormier TA, Lim MCW, Powers DR, Wethington SM, Goetz S, Graham CH. 2015. Citizen-science data provides new insight into annual and seasonal variation in migration patterns. Ecosphere 6: 15.

van Strien AJ, Boomsluiter M, Noordeloos ME, Verweij RJT, Kuyper TW. 2018. Woodland ectomycorrhizal fungi benefit from large-scale reduction in nitrogen deposition in the Netherlands. Journal of Applied Ecology 55: 290–298.

van Strien AJ, van Swaay CAM, Termaat T. 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. Journal of Applied Ecology 50: 1450–1458.

Suazo CG, Yates O, Azócar J, Díaz P, González-But JC, Cabezas LA. 2017. Emerging platforms to monitor the occurrence and threats to critically endangered seabirds: The waved albatross in Chile and the Southeast Pacific. Revista de biología marina y oceanografía 52: 245–254.

Sun Y, Li L, Li L, Zou J, Liu J. 2015. Distributional dynamics and interspecific gene flow in Picea likiangensis and P. wilsonii triggered by climate change on the Qinghai-Tibet Plateau. Journal of Biogeography 42: 475–484.

Syfert MM, Smith MJ, Coomes DA. 2013. The Effects of Sampling Bias and Model Complexity on the Predictive Performance of MaxEnt Species Distribution Models. PLoS ONE 8: e55158.

Szabo JK, Vesk PA, Baxter PWJ, Possingham HP. 2010. Regional avian species declines estimated from volunteer-collected long-term data using List Length Analysis. Ecological Applications 20: 2157–2169.

Szabo ND, Colla SR, Wagner DL, Gall LF, Kerr JT. 2012. Do pathogen spillover, pesticide use, or habitat loss explain recent North American bumblebee declines?: Causes of bumblebee declines. Conservation Letters 5: 232–239.

Tanner AM, Tanner EP, Papeş M, Fuhlendorf SD, Elmore RD, Davis CA. 2020. Using aerial surveys and citizen science to create species distribution models for an imperiled grouse. Biodiversity and Conservation 29: 967–986.

Tantipisanuh N, Gale GA. 2018. Identification of biodiversity hotspot in national level – Importance of unpublished data. Global Ecology and Conservation 13: e00377.

Termaat T, van Strien AJ, van Grunsven RHA, De Knijf G, Bjelke U, Burbach K, Conze K, Goffart P, Hepper D, Kalkman VJ, Motte G, Prins MD, Prunier F, Sparrow D, van den Top GG, Vanappelghem C, Winterholler M, WallisDeVries MF. 2019. Distribution trends of European dragonflies under climate change. Diversity and Distributions 25: 936–950.

Thompson PM, Brookes KL, Cordes LS. 2015. Integrating passive acoustic and visual data to model spatial patterns of occurrence in coastal dolphins. ICES Journal of Marine Science 72: 651–660.

Tiffin HS, Peper ST, Wilson-Fallon AN, Haydett KM, Cao G, Presley SM. 2019. The Influence of New Surveillance Data on Predictive Species Distribution Modeling of Aedes aegypti and Aedes albopictus in the United States. Insects 10: 400.

Tingley R, Thompson MB, Hartley S, Chapple DG. 2016. Patterns of niche filling and expansion across the invaded ranges of an Australian lizard. Ecography 39: 270–280.

Tovaranonte J, Blach-Overgaard A, Pongsattayapipat R, Svenning J-C, Barfod AS. 2015. Distribution and diversity of palms in a tropical biodiversity hotspot (Thailand) assessed by species distribution modeling. Nordic Journal of Botany 33: 214–224.

Townsend Peterson A, Ortega Huerta MA. 2008. Modelado de nichos ecológicos y predicción de distribuciones geográficas: comparación de seis métodos. Revista Mexicana de Biodiversidad 79: 205-216.

Troudet J, Grandcolas P, Blin A, Vignes-Lebbe R, Legendre F. 2017. Taxonomic bias in biodiversity data and societal preferences. Scientific Reports 7: 9132.

Václavík T, Meentemeyer RK. 2009. Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? Ecological Modelling 220: 3248–3258.

Vale MM, Tourinho L, Lorini ML, Rajão H, Figueiredo MSL. 2018. Endemic birds of the Atlantic Forest: traits, conservation status, and patterns of biodiversity. Journal of Field Ornithology 89: 193–206.

Vásquez-Ordóñez AA, Hazzi NA, Escobar-Prieto D, Paz-Jojoa D, Parsa S. 2015. A geographic

distribution database of the Neotropical cassava whitefly complex (Hemiptera, Aleyrodidae) and their associated parasitoids and hyperparasitoids (Hymenoptera). ZooKeys 545: 75–87.
Vergara J, Acosta LE. 2015. More on the Mesopotamian-Yungas disjunction in subtropical and temperate Argentina: Bioclimatic distribution models of the harvestman Discocyrtus dilatatus (Opiliones: Gonyleptidae). Zoologia (Curitiba) 32: 445–456.
Verlaque M, Breton G. 2019. Biological invasion: Long term monitoring of the macroalgal flora of a major European harbor complex. Marine Pollution Bulletin 143: 228–241.
Vihervaara P, Mononen L, Auvinen A-P, Virkkala R, Lü Y, Pippuri I, Packalen P, Valbuena R, Valkama J. 2015. How to integrate remotely sensed data and biodiversity for ecosystem assessments at landscape scale. Landscape Ecology 30: 501–516.
Villordon A, Njuguna W, Gichuki S, Ndolo P, Kulembeka H, Jeremiah SC, LaBonte D, Yada B, Tukamuhabwa P, Mwanga ROM. 2006. Using GIS-Based Tools and Distribution Modeling to Determine Sweetpotato Germplasm Exploration and Documentation Priorities in Sub-Saharan Africa. HortScience 41: 1377–1381.
Vray S, Rollin O, Rasmont P, Dufrêne M, Michez D, Dendoncker N. 2019. A century of local changes in bumblebee communities and landscape composition in Belgium. Journal of Insect Conservation 23: 489–501.
Wan X, Jiang G, Yan C, He F, Wen R, Gu J, Li X, Ma J, Stenseth NChr, Zhang Z. 2019. Historical records reveal the distinctive associations of human disturbance and extreme climate change with local extinction of mammals. Proceedings of the National Academy of Sciences 116: 19001–19008.
Wang Y, Casajus N, Buddle C, Berteaux D, Larrivée M. 2018. Predicting the distribution of poorly-documented species, Northern black widow (Latrodectus variolus) and Black purse-web spider (Sphodros niger), using museum specimens and citizen science data. PLOS ONE 13: e0201094.
Wang Y-H, Yang K-C, Bridgman CL, Lin L-K. 2008. Habitat suitability modelling to correlate gene flow with landscape connectivity. Landscape Ecology s10980-008-9262–3.
Warton DI, Renner IW, Ramp D. 2013. Model-Based Control of Observer Bias for the Analysis of Presence-Only Data in Ecology. PLoS ONE 8: e79168.
Washitani I, Nagai M, Yasukawa M, Kitsuregawa M. 2020. Testing a butterfly commonness hypothesis with data assembled by a citizen science program "Tokyo Butterfly Monitoring". Ecological Research 35: 1087–1094.
Watts SM, McCarthy TM, Namgail T. 2019. Modelling potential habitat for snow leopards (Panthera uncia) in Ladakh, India. PLOS ONE 14: e0211509.
Webb MH, Terauds A, Tulloch A, Bell P, Stojanovic D, Heinsohn R. 2017. The importance of incorporating functional habitats into conservation planning for highly mobile species in dynamic systems: Habitat Function in Dynamic Systems. Conservation Biology 31: 1018–1028.
Webster RP, de Tonnancour P, Sweeney JD, Webster VL, Kostanowicz CA, Hughes C, Anderson RS, Klymko J, Chantal C, Vigneault R. 2020. New Coleoptera records from eastern Canada, with additions to the fauna of Manitoba, British Columbia, and Yukon Territory. ZooKeys 946: 53–112.
Wei B, Wang R, Hou K, Wang X, Wu W. 2018. Predicting the current and future cultivation regions of Carthamus tinctorius L. using MaxEnt model under climate change in China.

Global Ecology and Conservation 16: e00477.
Wepfer PH, Guénard B, Economo EP. 2016. Influences of climate and historical land connectivity on ant beta diversity in East Asia. Journal of Biogeography 43: 2311–2321.
Westwood R, Westwood AR, Hooshmandi M, Pearson K, LaFrance K, Murray C. 2020. A field-validated species distribution model to support management of the critically endangered Poweshiek skipperling (Oarisma poweshiek) butterfly in Canada. Conservation Science and Practice 2.
Whiting SD, Macrae I, Thorn R, Murray W, Whiting AU. 2014. Sea turtles of the Cocos (Keeling) Islands, Indian Ocean. The Raffles Bulletin of Zoology 30: 168-183.
Williams HF, Bartholomew DC, Amakobe B, Githiru M. 2018. Environmental factors affecting the distribution of African elephants in the Kasigau wildlife corridor, SE Kenya. African Journal of Ecology 56: 244–253.
Wilting A, Hearn AJ, Eaton J, Belant JL, Kramer-Schadt S. 2016. Predicted distribution of the Bornean ferret badger Melogale everetti (Mammalia: Carnivora: Mustelidae) on Borneo. The Raffles Bulletin of Zoology 33: 55-60.
Witt M, Hardy T, Johnson L, McClellan C, Pikesley S, Ranger S, Richardson P, Solandt J, Speedie C, Williams R, Godley B. 2012. Basking sharks in the northeast Atlantic: spatio-temporal trends from sightings in UK waters. Marine Ecology Progress Series 459: 121–134.
Wonham MJ, Hart MW. 2018. El Niño Range Extensions of Pacific Sand Crab (Emerita analoga) in the Northeastern Pacific. Northwest Science 92: 53–60.
Xing S, Au TF, Dufour PC, Cheng W, Landry Yuan F, Jia F, Vu LV, Wang M, Bonebrake TC. 2019. Conservation of data deficient species under multiple threats: Lessons from an iconic tropical butterfly (Teinopalpus aureus). Biological Conservation 234: 154–164.
Yu F, Wang T, Groen TA, Skidmore AK, Yang X, Ma K, Wu Z. 2019. Climate and land use changes will degrade the distribution of Rhododendrons in China. Science of The Total Environment 659: 515–528.
Yue S, Bonebrake TC, Gibson L. 2019. Informing snake roadkill mitigation strategies in Taiwan using citizen science: Snake Roadkill Mitigation Using Citizen Science. The Journal of Wildlife Management 83: 80–88.
Zaniewski AE, Lehmann A, Overton JM. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. Ecological Modelling 157: 261–280.
Zapponi L, Cini A, Bardiani M, Hardersen S, Maura M, Maurizi E, Redolfi De Zan L, Audisio P, Bologna MA, Carpaneto GM, Roversi PF, Sabbatini Peverieri G, Mason F, Campanaro A. 2017. Citizen science data as an efficient tool for mapping protected saproxylic beetles. Biological Conservation 208: 139–145.
Zarzo-Arias A, Penteriani V, Delgado M del M, Peón Torre P, García-González R, Mateo-Sánchez MC, Vázquez García P, Dalerum F. 2019. Identifying potential areas of expansion for the endangered brown bear (Ursus arctos) population in the Cantabrian Mountains (NW Spain). PLOS ONE 14: e0209972.
Zimmermann H, Von Wehrden H, Damascos MA, Bran D, Welk E, Renison D, Hensen I. 2011. Habitat invasion risk assessment based on Landsat 5 data, exemplified by the shrub Rosa rubiginosa in southern Argentina: HABITAT INVASION RISK ASSESSMENT. Austral Ecology 36: 870–880.

S8. R scripts

All scripts used in data management and analysis for our review are available here:
DOI: **10.17605/OSF.IO/JUEQC**

S9. References for supplementary materials

Aria M, Cuccurullo C. 2017. bibliometrix : An R-tool for comprehensive science mapping analysis. Journal of Informetrics 11: 959–975.

GBIF Secretariat. 2019. GBIF Science Review 2019. Available at https://doi.org/10.15468/QXXG-7K93.

Moher D, Liberati A, Tetzlaff J, Altman DG. 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339: b2535.

Westgate MJ. 2019. revtools: An R package to support article screening for evidence synthesis. Research Synthesis Methods 10: 606–614.

ISBN 978-82-326-6039-1 (printed ver.)
ISBN 978-82-326-6685-0 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology