



Many cameras make light work: opportunistic photographs of rare species in iNaturalist complement structured surveys of reef fish to better understand species richness

Christopher J. Roberts^{1,2} · Adriana Vergés^{1,2} · Corey T. Callaghan^{2,3} · Alistair G. B. Poore^{1,2}

Received: 25 August 2021 / Revised: 8 February 2022 / Accepted: 21 February 2022 /
Published online: 23 March 2022
© The Author(s) 2022

Abstract

Citizen science is on the rise, with growing numbers of initiatives, participants and increasing interest from the broader scientific community. iNaturalist is an example of a successful citizen science platform that enables users to opportunistically capture and share biodiversity observations. Understanding how data from such opportunistic citizen science platforms compare with and complement data from structured surveys will improve their use in future biodiversity research. We compared the opportunistic fish photographs from iNaturalist to those obtained from structured surveys at eight study reefs in Sydney, Australia over twelve years. iNaturalist recorded 1.2 to 5.5 times more fish species than structured surveys resulting in significantly greater annual species richness at half of the reefs, with the remainder showing no significant difference. iNaturalist likely recorded more species due to having simple methods, which allowed for broad participation with substantially more iNaturalist observation events (e.g., dives) than structured surveys over the same period. These results demonstrate the value of opportunistic citizen science platforms for documenting fish species richness, particularly where access and use of the marine environment is common and communities have the time and resources for expensive recreational activities (i.e., underwater photography). The datasets also recorded different species composition with iNaturalist recording many rare, less abundant, or cryptic species while the structured surveys captured many common and abundant species. These results suggest that integrating data from both opportunistic and structured data sources is likely to have the best outcome for future biodiversity monitoring and conservation activities.

Keywords Reef life survey · Unstructured biodiversity data · Presence-only data · Species occurrence data · Citizen Science · Community Science

Communicated by James Tony Lee.

This article belongs to the Topical Collection: Coastal and marine biodiversity

Extended author information available on the last page of the article

Introduction

Global biodiversity patterns are being fundamentally altered in response to climate change and other human impacts (Blowes et al. 2019). A key component of managing and conserving biodiversity is the ability to monitor species occurrences at both local and global scales in a timely and cost-effective manner (Dickman and Wardle 2012; Sullivan et al. 2017). Species richness, that is the number of species at a given location, is a key measure used in conservation actions such as protecting biodiversity hotspots or identifying habitats of rare and endangered species (Gotelli and Chao 2013; Chao and Chiu 2016). Given that gathering biodiversity data takes a considerable amount of time, effort and resources, citizen science (also termed community science), is increasingly being used to efficiently gather and process large volumes of species occurrence data (Thiel et al. 2014; Follett and Strezov 2015; Theobald et al. 2015; Pocock et al. 2017). In the last decade, new citizen science initiatives have tended towards having simpler methods that encourage mass participation (Pocock et al., 2017) such as gathering observations of living organisms opportunistically (i.e., during normal daily activities) through photographs or recordings. These observations are generally collected in an unstructured format without formal survey methods or guidance from professional scientists.

iNaturalist, one of the most popular citizen science platforms, has over 1.3 million users contributing millions of observations globally each month (Seltzer et al. 2020). The increasing popularity of platforms such as iNaturalist is likely due, at least in part, to participants having freedom to choose both where and when to make observations (i.e., during recreational activities such as bush walking and scuba diving) as well as how (i.e., no restrictive survey protocols). As participation in platforms such as iNaturalist continues to grow and observations rise rapidly (Mesaglio and Callaghan 2021), it becomes increasingly important to explore the potential of opportunistic datasets for biodiversity monitoring.

The reliability of data gathered through citizen science is often regarded with some degree of scepticism among scientists (Riesch and Potter 2014; Burgess et al. 2017), despite numerous studies indicating that citizen science can provide data comparable in quality to data gathered by trained scientists (see review by Aceves-Bueno et al. 2017). Data derived from citizen science projects that use highly structured survey methods such as Reef Life Survey (Edgar and Stuart-Smith 2009) or even semi-structured checklists such as eBird (Sullivan et al. 2014) are increasingly being used in peer-reviewed ecological research (Follett and Strezov 2015). In contrast, the vast amount of valuable biodiversity information contained in databases of opportunistic observations is underutilised due to concerns about data quality and potential biases (Dickinson et al. 2010; Isaac et al. 2014; Rapacciuolo et al. 2021) and uncertainty regarding the use of presence-only data (Giraud et al. 2016; Bradter et al. 2018). Where opportunistic observations have been used, it has predominately been to map species distribution (van Strien et al. 2013; Fourcade 2016; Wang et al. 2018) rather than addressing questions such as quantifying spatial patterns in abundance or species composition.

In the absence of standardised and structured sampling methods, potential biases in opportunistic observation databases include the over-representation of colourful, interesting or rare species (Isaac and Pocock 2015; Prudic et al. 2018; Caley et al. 2020) or the over-sampling of accessible locations such as those closer to roads and/or urban centres (Reddy and Dávalos 2003; Szabo et al. 2007; Tiago et al. 2017). Consequently, the number of obser-

vations may indicate the amount of interest in a species rather than its abundance, and the location of observations may reflect the distribution of observers more than that of the target species (Williams et al. 2002; Snäll et al. 2011; Giraud et al. 2016). Use of data from opportunistic observations will be improved by a greater understanding of how it differs from or complements structured surveys, particularly in terms of potential biases toward, or away from, certain taxa. For example, if opportunistic observers record more rare species, but tend to overlook or undersample common species, then the most effective means of documenting biodiversity is likely to involve a combination of structured and unstructured sampling (Giraud et al. 2016; Soroye et al. 2018; Rapacciuolo et al. 2021).

To date, there have been numerous assessments of the data generated by structured surveys conducted by citizen scientists compared to professionals (Aceves-Bueno et al. 2017). In contrast, studies comparing presence-only data from unstructured opportunistic observations to data generated from structured surveys are limited but examples include comparisons of species richness of birds, ladybeetles and butterflies (Losey et al. 2012; Klemann-Junior et al. 2017; Prudic et al. 2018) and temporal and spatial trends in bird abundance (Snäll et al. 2011; Giraud et al. 2016; Kamp et al. 2016). A recent study of marine intertidal communities demonstrated the value of combining opportunistic observations with structured surveys observations to monitor temporal trends in intertidal species (Rapacciuolo et al. 2021).

Monitoring marine biodiversity is particularly challenging, time-consuming and expensive due to the need for calm ocean conditions and good water clarity, specialised scuba training and equipment, and often a dive vessel. To address this, scientists are increasingly turning to citizen science to gather the data needed for marine life monitoring and biodiversity conservation. In Australia, several marine citizen science projects have been running for many years including Reef Life Survey, Reef Check, Redmap, Eye on the Reef, and the Australasian Fishes project in iNaturalist. These programs have different aims and objectives, use different approaches, vary in sampling effort and generate different data. Individually, these projects have generated much valuable information, however, limited work has been done comparing these data sources. Consequently, the data generated from each program are generally considered and used in isolation (Peterson et al. 2020). The ability to combine data from programs that use different approaches, such as opportunistic observations with structured surveys, could considerably improve the biodiversity data available for marine conservation and ecological research (Ballard et al. 2017; Kelly et al. 2020; Peterson et al. 2020). To facilitate this, it is important to understand how these structured and unstructured approaches differ in terms of the biodiversity data they generate and quantify the differences in sampling effort in given localities through time. Here, we compare the fish species photographed and contributed to the opportunistic observation database iNaturalist to structured data gathered by Reef Life Survey (RLS) (Edgar and Stuart-Smith 2014) at eight dive sites in Sydney, Australia. Specifically, we quantified how opportunistic observations and structured surveys differed in: (1) species richness, (2) species composition, and (3) to what extent sampling effort explained the differences observed.

Methods

Unstructured citizen science data: iNaturalist

iNaturalist ([inaturalist.org](https://www.inaturalist.org)) is an online platform for users to share their nature observations (e.g., photographs) which has been operating since 2008. The platform was designed with the primary intention of engaging people with the natural world, with the potential secondary use of the observations for scientific purposes. It has not been designed to follow any structured scientific sampling methods or techniques and there are few constraints around providing observations. The main constraint is the requirement to provide evidence of an observation, generally a photograph, along with the location and time of the sighting. This does potentially place some limitations on the iNaturalist dataset as the ability to capture an identifiable photograph of some fish species is often challenging or not possible for many encounters.

The iNaturalist platform allows projects to be created that target specific taxa, places, and/or times. The Australasian Fishes project was started by the Australian Museum in late 2016 and targets observations of marine and freshwater fish from Australia, New Zealand and their respective territorial waters ([inaturalist.org/projects/australasian-fishes](https://www.inaturalist.org/projects/australasian-fishes)). Contributions to this project can include any fish photograph within the region including from divers, snorkellers and fishers. It is important to note, however, that the contribution of fishers to the current dataset is likely negligible with only eight fishing-based photographs (i.e., a fish removed from the water) of the approximately 7600 photographs used in this study. Data for this study were downloaded from the Australasian Fishes project on 13 February 2020.

iNaturalist observations are identified to various taxonomic levels based on combination of computer vision suggestions and identifications provided by the iNaturalist community (i.e., citizen scientists). Observations become ‘research grade’ when at least two iNaturalist users provide a consistent species level identification, or if more than two thirds of suggestions are for the same species. The Australasian Fishes project is curated by the Australian Museum and many observations, particularly unusual sightings or difficult identifications, are referred to trained fish taxonomists. The referral process is primarily driven by iNaturalist users who, if necessary, can refer observations to Australian Museum staff or to a taxon specialist (i.e., by mentioning them in an observation by using @UserName), many of whom are active members of the Australasian Fishes project. It is worth highlighting that the data quality assurance is greater for the Australasian Fishes project than it may be for iNaturalist more broadly, due to the association with the Australian Museum and consequently the large number of fish taxonomic experts involved in identifying and checking observations. Data used in this project were restricted to research grade identifications. Research grade iNaturalist observations have previously been found to be between 65% accurate for insects to 91% accurate for birds (Ueda 2019), although fish were not included in this assessment. iNaturalist observations were also excluded if their positional accuracy was reported as >500 m or if the true co-ordinates of an observation were obscured by the contributor for privacy reasons.

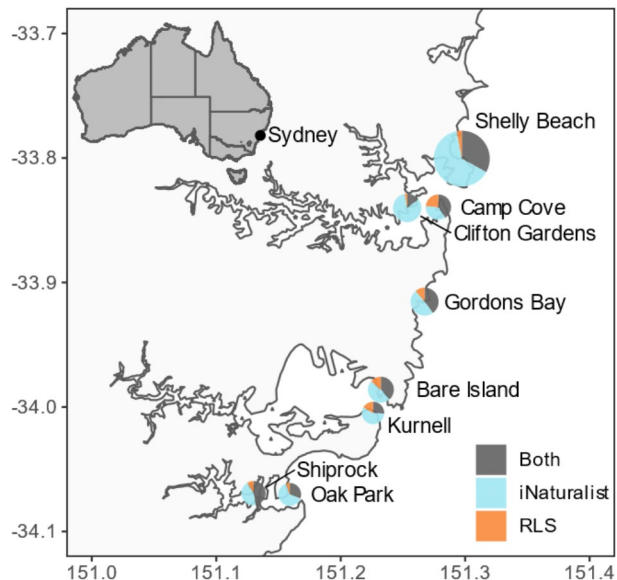
Structured citizen science data: Reef Life Survey

Reef Life Survey (RLS; reeflifesurvey.com) is a citizen science biodiversity monitoring program which started in 2007. The program uses standardised underwater surveys, which are done by a mixture of specialist scientists and experienced recreational scuba divers who undergo a rigorous training and testing program in species identification and underwater surveying techniques (Reef Life Survey Foundation 2019). An assessment of RLS data quality found that volunteers generated fish and invertebrate data indistinguishable from experienced scientists associated with the program (Edgar and Stuart-Smith 2009). RLS database administrators check uploaded data for potential errors such as species outside of their normal region of occurrence.

The use of standardised survey techniques creates a structured data source, although there are generally no constraints on timing of surveys, resulting in a temporally variable dataset. RLS uses two methods to survey fish species along a 50 m transect line. The main method includes all fish species observed 5 m to either side and above the transect line. The counts are done separately on each side of the transect either by two separate divers simultaneously, or on a return swim by the same diver. In addition, a second count is done for cryptic fish, covering an area of 1 m to either side of the transect line. Since only species presence was required for this study, data from the two methods (i.e., the 5 m and 1 m survey) were combined to generate the species list for each survey.

The RLS data were extracted from the data portal (Edgar and Stuart-Smith 2020a, b) on 14 February 2020. The data extracted from RLS were cleaned to exclude individuals not identified to species level as well as non-fish observations (e.g., cephalopods).

Fig. 1 The location of the eight study sites in Sydney, Australia. Pie charts show the proportion of species at each site recorded by iNaturalist only, RLS only or both between 2017 and 2019. Chart size indicates the relative difference in total number of species



Study site selection

Eight popular dive sites in Sydney, Australia, were selected for inclusion in this study: Shelly Beach, Camp Cove, Clifton Gardens, Gordons Bay, Bare Island, Kurnell, Shiprock and Oak Park (Fig. 1). The sites were chosen as they had the greatest number of contributions to iNaturalist in the Sydney region and have been repeatedly sampled by RLS. The selected sites in the Sydney region encompass a wide range of conditions including variable exposure, seabed composition, depth, and marine protected area status. The study was constrained to between 2008 and 2019 (inclusive) as limited RLS surveys or iNaturalist observations were available prior to 2008. Although the Australasian Fishes project only commenced in 2016, observations can be retrospectively added from earlier years. As such, the iNaturalist dataset includes observations from before 2016, but at a much lower rate of contribution than after 2016.

iNaturalist photographs were assigned to sites based on their geographic co-ordinates falling within an approximately 500×500 m bounding box centred on each site. The exact size was varied slightly to encompass the entire “dive site” at each location based on the natural coastline of each site. RLS surveys are repeated at a consistent GPS co-ordinate through time at each site. In some cases, multiple surveys are conducted for different areas within a site and these were included in the analysis.

Contrasting fish communities between datasets

The two datasets were transformed into lists of species recorded at each site during each sampling year (i.e., presence/absence) to allow direct comparison. There is potential for duplication of observations between the two datasets, however, less than 1% of iNaturalist photographs came from the same day and site of an RLS survey. Further, many of these observations were likely not taken by divers involved in the RLS surveys, so on this basis we consider the two datasets to be largely independent of each other.

All data manipulation, statistical analyses and graphing was done in R version 3.6.3 (R Core Team 2020). Species lists generated from both data sources were cross checked against species names in FishBase using the R package ‘rfishbase’ (Boettiger et al. 2012). Species that did not match a record in the FishBase species list were manually inspected and names were changed to be consistent with FishBase for both datasets. Mismatches were generally either due to a change in the accepted name, which had not been adopted by one of the datasets or a spelling discrepancy.

The difference in the average annual species richness between iNaturalist and RLS was tested for the 2017 to 2019 period using a two-factor analysis of variance with dataset and site as fixed factors. This constrained time-period was used as both programs were running, resulting in few sites or years with very low numbers of iNaturalist contributions or no RLS surveys. Plots of annual and cumulative species richness from 2008 to 2019 were also used to compare between the two datasets. Cumulative species richness was calculated using the ‘accumcomp’ function of the R package *BiodiversityR* (Kindt and Coe 2005). The ‘collector’ method was used to add species in order of the sampling year to visualize the actual increase in species richness over the sampling period. As a measure of similarity, the number of species common to both the methods at each site was calculated for all years combined.

The variation in community composition between the two datasets and among sites for the 2017 to 2019 period was visualized with an ordination plot based on a Generalised Linear Latent Variable Model (GLLVM). The GLLVM was fit using two latent variables based on a binomial complementary log log link transformation with random row effects included in the model. The GLLVM model fit was checked using a ‘Residuals vs Linear Predictors’ plot and a ‘Normal Q-Q’ plot. A Multivariate Generalised Linear Model (MGLM) based on 1000 permutations was used to test for statistically significant differences between the datasets (iNaturalist and RLS), sites (eight levels) and for an interaction between dataset and site. Pairwise comparisons for differences between datasets for each site was done by running the MGLM analysis on the data for each site separately. Univariate comparisons, adjusted for multiple comparisons, were done to test which species showed a significant difference between the datasets. The analyses were done using the ‘gllvm’ function in the gllvm package (Niku et al. 2020) and the manyglm function of the mvabund package (Wang et al. 2020).

Variation in sampling effort between datasets

The relative effort was compared between iNaturalist and RLS based on the number of sampling events. An iNaturalist ‘observation event’ was considered as all records submitted by a single observer from one site on the same day, while an RLS observation event was one survey transect. Plots of the number of observation events, and the number of iNaturalist photograph submissions were used to assess trends through time. Analysis of individual submissions was limited to iNaturalist as no meaningful equivalent measure is available for the RLS dataset. The relative sampling efficiency was also compared between iNaturalist and RLS by visually comparing the number of species recorded per observation event. The mean species observed per event at each site was also calculated for the two datasets.

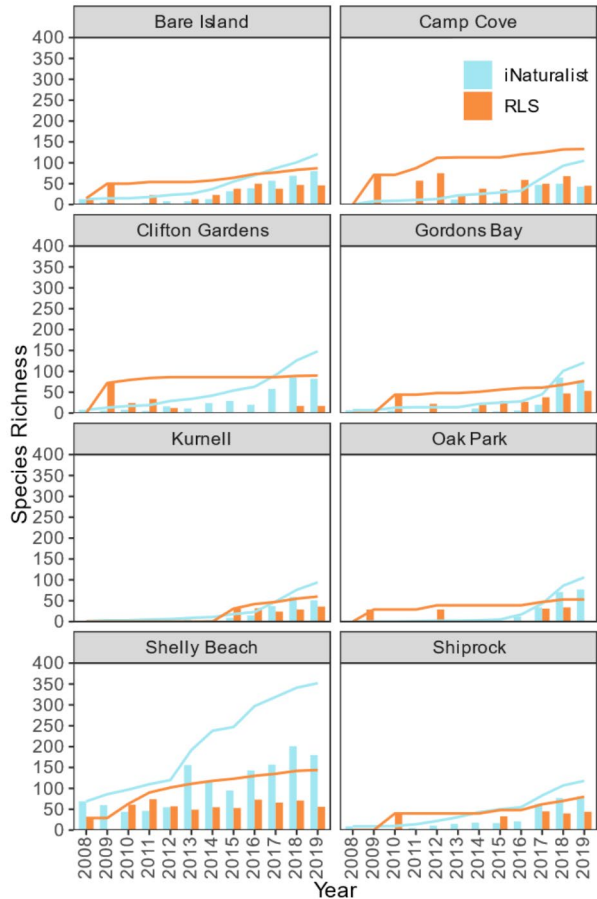
Results

Variation in species richness between datasets

Overall, iNaturalist recorded 363 opportunistic species observations between 2017 and 2019 while structured surveys by RLS recorded 150 species for the eight study sites combined. At a site level, iNaturalist recorded between 1.2 (Camp Cove) and 5.5 times (Clifton Gardens) more species than RLS for the 2017–2019 period (Supplementary Material 1).

Prior to 2017, iNaturalist generally had lower numbers of species recorded per year than RLS at most sites (Fig. 2). The main exception was Shelly Beach where iNaturalist recorded more species than RLS in all surveys except between 2010 and 2012. Since 2017, iNaturalist has recorded more species per year for most sites. The exception was Camp Cove, where RLS recorded more species in all years, and Gordons Bay where RLS had more species in 2017. For the time period 2017 to 2019, when both the iNaturalist Australasian Fishes project and RLS were active, annual species richness was, on average, significantly greater for iNaturalist at Shelly Beach ($F=93.40$, $p<0.0001$), Shiprock ($F=5.84$, $p=0.022$), Clifton Gardens ($F=18.68$, $p=0.0002$), Oak Park ($F=4.616$, $p=0.0399$) and Bare Island ($F=4.22$, $p=0.049$) (Supplementary Material 2). No difference in annual species richness

Fig. 2 Species richness recorded per year (bars) and cumulative species richness (lines) for iNaturalist and RLS



was detected at Kurnell ($F=2.59$, $p=0.12$), Gordons Bay ($F=1.56$, $p=0.22$) and Camp Cove ($F=0.41$, $p=0.53$).

Cumulative species richness increased relatively quickly for RLS at most sites and generally began to flatten after 1–3 years of surveys (Fig. 2). In contrast, species richness for iNaturalist increased gradually until 2016 at most sites, before rapidly increasing between 2017 and 2019. The exception was Shelly Beach, which started with a relatively high number of species observations in 2008 gradually increasing through to 2012 before rapidly growing between 2013 and 2019. This difference in the species accumulation trends between the iNaturalist and RLS programs meant that cumulative species richness was greater for RLS than iNaturalist through to 2017 or 2018 at most sites at which point the cumulative number of species recorded by iNaturalist surpassed that recorded by RLS at most sites. At Shelly Beach, however, iNaturalist consistently recorded a greater cumulative species richness throughout the whole monitoring period. At Camp Cove, the cumulative species richness remained greater for RLS than iNaturalist for the whole 2008–2019 study period.

Total species richness between 2017 and 2019 varied considerably between datasets and sites (Fig. 1, Supplementary Material 1). Shelly Beach reported the greatest species richness

for both datasets, with 261 and 97 species for iNaturalist and RLS, respectively. However, discrepancies occurred at other sites such as Camp Cove, which had the second most species recorded by RLS (79 species) but the second least recorded by iNaturalist (93 species). Conversely, Clifton Gardens had the second highest richness recorded by iNaturalist (117 species) while RLS recorded the lowest species richness (24 species) of all the sites.

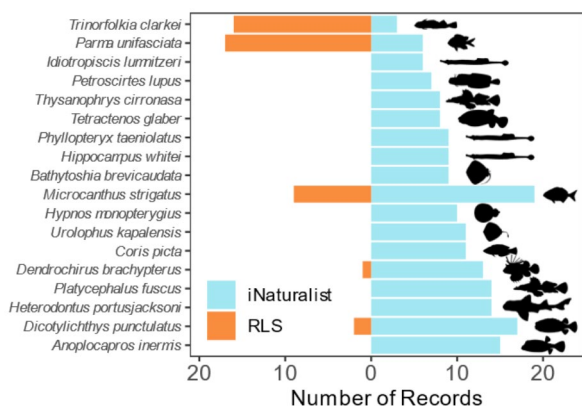
Variation in species composition between datasets

Overall, between 2017 and 2019 there were 142 species, which were recorded by both RLS and iNaturalist across all sites. A further 221 species were recorded exclusively by iNaturalist while RLS recorded only 8 species not submitted to iNaturalist between 2017 and 2019 at any study site. At a site level, the proportion of species shared by the two datasets ranged between 15% at Clifton Gardens (20 of 137 species) and 47% at Shiprock (55 of 117) (Fig. 1, Supplementary Material 1). The proportion of species unique to iNaturalist at each site range between 35% at Camp Cove (43 of 122) to 82% at Clifton Gardens (113 of 137). In contrast, the proportion of species only recorded by RLS ranged from 3% at Shelly Beach and Clifton Gardens (8 of 269 and 4 of 137 respectively) to 24% at Camp Cove (29 of 122).

The species recorded by iNaturalist differed significantly to those recorded by RLS but only at some sites (Supplementary Material 3, significant dataset x site interaction: $Dev=650.6, p \leq 0.001$). Pairwise comparisons showed that datasets were significantly different at Shelly Beach ($Dev=758.0, p=0.04$), Bare Island ($Dev=285.2, p=0.048$) and Kurnell ($Dev=290.2, p=0.03$). There was no evidence for a difference in species composition between datasets at Clifton Gardens ($Dev=286.6, p=0.12$), Gordons Bay ($Dev=0.237.3, p=0.17$) and Oak Park ($Dev=215.3, p=0.13$), Camp Cove ($Dev=308.5, p=0.06$) and Shiprock ($Dev=212.7, p=0.07$).

Overall, 311 species were more frequently recorded by iNaturalist, while only 44 species were recorded more frequently by RLS. Twelve species were recorded the same number of times by both datasets. Univariate analyses contrasting datasets showed 16 species were recorded significantly more often by iNaturalist than RLS while only 2 species were recorded significantly more by RLS (Fig. 3).

Fig. 3 Number of recorded occurrences for species with a significant difference between the iNaturalist and RLS datasets. Most of the significant differences were for species that were only recorded in the iNaturalist dataset. Species sorted by the difference between RLS and iNaturalist. The family of each species is represented by a silhouette to aid visual interpretation of the graph



Comparison of sampling effort between datasets

Almost 7600 unique photographic species records (i.e., unique species observed by a single user from the same day and site) were submitted to iNaturalist for the eight monitoring sites between 2008 and 2019 (Fig. 4). A large proportion of the iNaturalist observations and sampling effort occurred between 2017 and 2019 with nearly 5600 photographic records across all sites from over 1200 observation events (i.e., all photos from a distinct user, site and day) (Fig. 4, Supplementary Material 4, Supplementary Material 5). There were five or fewer iNaturalist sampling events (e.g., dives) occurring in most years until 2016 after which the number of events increased to between 5 and 27 events from 2017 to 2019 (Supplementary Material 4). In contrast, only 71 RLS observation events (i.e., transects) occurred from 2008 to 2019 and there were generally 6 or fewer RLS transects at each site with only a few years with greater numbers of surveys (Supplementary Materials 4 & 5).

iNaturalist was highly skewed towards low numbers of observations per event, with three or fewer species photographed during nearly 65% of events, whereas RLS recorded a minimum of 11 species per event (Fig. 5). Similarly, the average number of species submitted to iNaturalist per observation event ranged from 2 (± 0.2 SE) at Kurnell to 7 (± 0.9 SE) at Shiprock between 2017 and 2019 (Supplementary Material 5). In contrast, the average

Fig. 4 Number of photographic observations (log₁₀ scale) submitted to iNaturalist between 2008 and 2019 at each of the monitoring sites

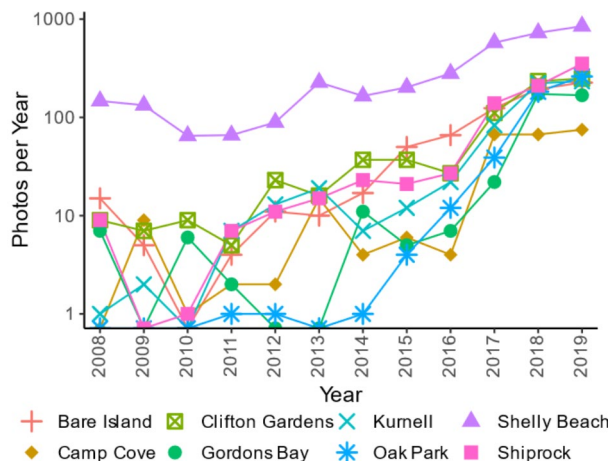
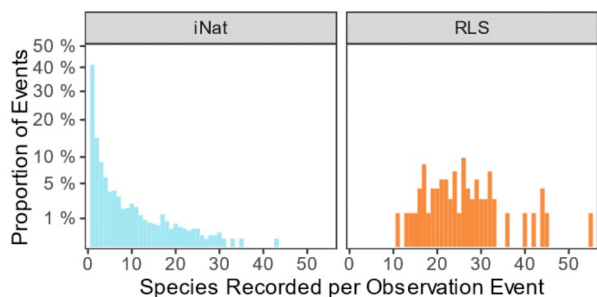


Fig. 5 The number of species recorded per observation event (e.g., iNaturalist dive or RLS survey) as a proportion of the total number of observation events (y-axis square root transformed)



number of species observed per RLS event ranged between 17 (± 0 SE) at Clifton Gardens and 43 (± 1.5 SE) at Shiprock.

Discussion

We found that opportunistic observations by iNaturalist users recorded more species than structured surveys cumulatively at most sites and, on average, more species per year at half the monitoring sites. In addition, iNaturalist recorded a different subset of species, with fewer than half the species observed opportunistically by iNaturalist users being recorded by structured surveys. iNaturalist likely recorded more species in this study, at least in part, due to the substantially greater sampling effort, with the iNaturalist observations being acquired from more than 1200 observation events between 2017 and 2019 (i.e., dives where at least one species was recorded and submitted to iNaturalist) compared to only 71 structured surveys done over the same period. The high number of species recorded by iNaturalist clearly demonstrates the considerable potential of opportunistic observations as an effective tool for documenting species richness. Tiralongo et al. (2021) similarly noted the efficiency of using opportunistic observations to record fish biodiversity, with considerably more species recorded during underwater photography competitions in the Mediterranean than various standardised survey techniques.

Sydney has a large community of predominantly local divers and underwater camera ownership is prevalent, and this likely helped the region accumulate such a substantial number of observations in the relatively short 3-year period since the Australasian Fishes project commenced. The fact that the study region is also dominated by local divers who often revisit the same sites frequently may mean that many contributors have a high degree of familiarity with local species and actively seek out rare or cryptic species. The high number of submissions in Sydney may have resulted in more species being recorded than in less populated areas of Australia or those with less active diving, snorkelling, or fishing communities. This bias towards areas of high population density in opportunistic databases and other citizen science initiatives has been shown previously and discussed extensively (Szabo et al. 2007; Tiago et al. 2017; Callaghan et al. 2019). Despite this, we consider the success of Australasian Fishes in Sydney within a relatively short time period to indicate the potential of iNaturalist in regions with less diving, snorkelling or fishing, given sufficient time and promotional effort to grow the project.

Losey et al. (2012) found that the species richness derived from opportunistic observation of ladybugs was similarly greater than the combined richness of several structured professional taxonomic surveys. However, in that case the difference was attributed to not only the greater number of opportunistic samples but also to a greater geographic spread. A greater spread of sampling effort is likely to have also influenced species richness in this study, but at a localised site scale. That is, the structured surveys were constrained to standardised transects at a consistent depth, with only one 50 m stretch of reef generally sampled at each dive site. In addition, a similar area is sampled on repeat surveys with transects commencing from a consistent GPS coordinate. In contrast, a recreational diver could easily cover several hundred meters of reef on a single dive, and the depths and area covered would vary among different divers and visits. Further, iNaturalist observations come from a range of different types of contributors, including snorkelers and fishers, and these groups

may observe species that are less frequently encountered by scuba divers. Snorkelers, for example, will likely encounter more species that inhabit shallower waters, which may be under-represented in the structured surveys which were done by Scuba diving only. Consequently, most of a site is likely to be covered by the combined efforts of many iNaturalist contributors, which in this dataset included hundreds of visits to some sites. Although fishers have the potential to contribute unique observations of species, which are attracted to bait but may avoid divers or snorkelers, this is unlikely to have occurred in this study as only a very small proportion of observations (8 of 7600 photos) were contributed by fishers, all of which were of species also observed in-situ by divers or snorkelers.

It is important to note that the structured surveys used by Reef Life Survey are not specifically designed to measure species richness, rather, it is a global scale survey with effort primarily directed at sampling many sites with a consistent methodology, instead of sampling individual sites intensively (Edgar and Stuart-Smith 2014). It is also worth highlighting that the structured surveys were considerably more efficient at recording species with approximately five times as many species recorded per dive. This is likely due to the structured surveys recording all species observed within the sampling parameters while iNaturalist users are highly selective about what they photograph and contribute. Importantly, the use of a consistent methodology by RLS and similar structured survey approaches allows for robust comparison of trends through time and across sites, on a global scale. In addition, RLS gathers a suite of information, which is not readily obtainable from iNaturalist photographs such as the relative abundance of species, the size of species, as well as documenting the habitat composition using photo-quadrats. Comparison of iNaturalist, or similar opportunistic observations, to a more intensive structured survey program that is designed to specifically capture biodiversity would be a valuable future research direction. Such a comparison would help better understand how much sampling effort is required to capture similar amounts of biodiversity using structured and unstructured approaches.

The fact that fewer than half of the species recorded at all sites between 2017 and 2019 were present in both datasets demonstrates a considerable difference in the species recorded by the opportunistic observers and structured surveys. In part, this is likely to result from the greater overall species richness recorded by iNaturalist at most sites, which is also reflected by the large proportion of species that were unique to iNaturalist at each site. The large number of species unique to iNaturalist suggests that users are photographing and contributing species that are not readily captured by conventional structured surveys. Several cryptic species such as Weedy Seadragon (*Phyllopteryx taeniolatus*), White's Seahorse (*Hippocampus whitei*), Sydney Pygmy Pipehorse (*Idiotropiscis lumnitzeri*), and Dwarf Lionfish (*Dendrochirus brachypterus*) were recorded frequently by iNaturalist but rarely present in the RLS dataset. In addition, some rare or low abundance species were also recorded more by iNaturalist including Port Jackson Sharks (*Heterodontus portusjacksoni*), Smooth Stingray (*Bathytoshia brevicaudata*), Three Bar Porcupinefish (*Dicotylichthys punctulatus*), and Comb Wrasse (*Coris picta*). In contrast, the two species more frequently detected by RLS, the Girdled Parma (*Parma unifasciata*) and Clark's Threefin (*Trinorfolkia clarkei*), are commonly encountered on Sydney's rocky reefs. A similar result was reported by Tiralongo et al. (2020) who found that underwater photographers were effective at finding rare, small and cryptic fish species while Snäll et al. (2011) found that rare and uncommon bird species were essentially missed by structured surveys but captured by opportunistic citizen records.

Many iNaturalist contributors are likely to spend a substantial part of their dive searching for rare or cryptic species, simply for the challenge and reward of photographing species that are difficult to find. They may also be more likely to contribute photographs of these species to iNaturalist as their perceived value as a biodiversity observation may be greater due to their rarity. In contrast, rare or less abundant species are likely to be missing from the RLS dataset simply due to the reduced sampling effort and consequently a decreased probability of encounter during surveys. Further, although RLS includes a specific method for cryptic species, including looking in caves and overhangs along the transect, a consequence of using standardised transects means that observers are not free to ‘roam’ the dive site searching for certain species. The tendency of opportunistic observers to seek out rare species can be considered as a bias, however as noted by others, the fact that species are recorded that are often missed by structured surveys can equally be viewed as one of the key benefits of such methods (Snäll et al. 2011; Kamp et al. 2016).

In addition to rare species being favoured over common ones, there is potential for bias towards interesting species and away from less remarkable ones (Isaac and Pocock 2015; Prudic et al. 2018; Caley et al. 2020). Indeed, many of the species recorded more frequently by iNaturalist in this study, are also arguably very ‘photogenic’ such as seahorses and other syngnathids or ‘charismatic’ such as sharks and rays. There is also the potential for iNaturalist observations to be skewed towards species, which are more readily photographed, with many of the species more commonly recorded by iNaturalist in this study being benthic or slow-moving species. A recent traits analysis for birds found evidence that large-bodied species and those that occur in large flocks are over-represented in iNaturalist compared to the semi-structured eBird checklists, potentially as they are easier to find and photograph (Callaghan et al. 2021). A similar quantitative assessment of which fish traits affect the likelihood of a species being represented in opportunistic databases such as iNaturalist, although beyond the scope of this study, deserves further exploration as it influences how opportunistic photographs can be utilised for future research and biodiversity monitoring.

Action to conserve biodiversity, such as determining locations for protection, often relies on species occurrence data to identify biodiversity hotspots or areas that contain rare or endangered species. This is particularly important for rare or cryptic species, which can require substantial time and effort to find using conventional structured surveys. The high species richness and rare species recorded by iNaturalist in this study clearly demonstrates the enormous potential of platforms such as iNaturalist as a tool for documenting biodiversity and species conservation. Importantly, a large proportion of the observations were submitted over a relatively short 3-year period, following the launch and active promotion of the Australasian Fishes project, demonstrating the potential to gather large numbers of biodiversity observations through opportunistic observation platforms such as iNaturalist. This is largely the result of the relative ease of gathering and contributing iNaturalist observations, where essentially the only requirement is a photograph, compared to the high level of training and dedication required to gain the knowledge and skills required to do structured surveys. This means that large numbers of people can easily contribute to platforms such as iNaturalist, since the barriers to participation are low, resulting in substantial sampling effort due to greater ‘people power’.

In addition to having a large recreational diving community, the rapid growth of the Australasian Fishes iNaturalist project may be attributable, at least in part, to the various marine citizen science projects that preceded it in Australia (e.g., RLS, Redmap). These

have potentially helped establish a highly engaged diving community, which is willing to contribute to citizen science initiatives. The ability to replicate the success of Australasian Fishes or similar citizen science initiatives may also be limited in lower socioeconomic countries where there is less time and money for expensive activities (Haklay 2013; Walker et al. 2021) like scuba diving and underwater photography. However, iNaturalist is a global platform with high levels of engagement world-wide, and substantial numbers of fish photographs have been contributed for many geographic areas including lower socioeconomic areas such as South-east Asia, Central America and The Caribbean. Importantly, for many of these regions there is often limited monitoring of marine environments by scientists due to a lack of funding, however, they are popular destinations for scuba diving tourists. As such, there is considerable potential to supplement structured survey data in undersampled regions by recruiting tourists (Schaffer and Tham 2020; Callaghan et al. 2021). The relative ease of contributing observations means that platforms such as iNaturalist may be particularly well suited to documenting biodiversity in areas dominated by tourism diving where potential participants are unlikely to have the time or local species knowledge to do more complex surveys (Hermoso et al. 2021). However, given the considerable differences in the experience and motivations between tourist and local divers (Hermoso et al. 2021), it is difficult to know for certain how the results of our study, in a region with a highly active community of local divers would translate to areas dominated by tourist divers. In areas where recreational diving or snorkelling is minimal, including by tourists, it may be possible to gather opportunistic observations by engaging with other users of the marine environment such as commercial or subsistence fishers (Fulton et al. 2019). Expanding the current study to regions dominated by tourist divers, or those used by recreational, commercial or subsistence fishers would be an important future research direction and further exploration of the differences in experience, knowledge and motivation to participate in citizen science would be a valuable addition.

The lack of standardised methods for gathering observations, and the subsequent variability in effort and numbers of observations, is clearly one of the main limitations of opportunistic observation databases. For example, almost two-thirds the iNaturalist observation events (e.g., dives) in this study had three or less fish species yet it is considered likely that in many of these cases more fish were photographed but not submitted. Further, there is likely to also be many additional observation events where users didn't submit any photographs to iNaturalist as they didn't record any species or photographs which they considered worth submitting. If some users are only submitting 'interesting' observations or 'good' photographs, then simply encouraging existing users to share all their observations may improve the representation of more common species. Alternatively, more data could be gathered by capitalising on incidental data (Callaghan et al. 2021) as common species may often be captured in the background of photographs, and this is an area that deserves further exploration.

Ultimately, the greater number of species recorded by iNaturalist than structured surveys does not mean that opportunistic observations are a better way of measuring species richness or monitoring biodiversity. Indeed, relying on opportunistic observations alone for biodiversity conservation decision making could be highly problematic due to the biases of this method. For example, the increase in species recorded with greater observation effort could potentially result in more popular sites being protected, such as those with greater accessibility, instead of more biodiverse ones (Nelson et al. 1990; Reddy and Dávalos 2003). Our

results from Camp Cove illustrate this point, as iNaturalist recorded the second lowest number of species at this site, hypothetically making it a low priority for protection, however it had the second most species based on the structured surveys. The low iNaturalist species count in this case was likely due to Camp Cove being a less popular dive site with both the lowest number of iNaturalist sampling events and the least photographs submitted. As has been suggested and demonstrated by others (Fithian et al. 2015; Giraud et al. 2016; Soroye et al. 2018; Rapacciuolo et al. 2021), integrating opportunistic citizen science observations with structured survey data from more traditional sources (e.g., government monitoring and university research) will help ensure that both common and rarer species are well represented in biodiversity monitoring. It is worth noting however, that combining data sources may not always be the best approach, and where there are sufficient structured surveys it may be more efficient and reliable to use these data alone, especially if there is considerable and unknown bias in the opportunistic observations (Simmonds et al. 2020).

Conclusions

Although the value of a single opportunistic observation may be small, collectively, the vast quantities of opportunistic observations now being shared through platforms such as iNaturalist makes such data sources hard to ignore for biodiversity monitoring. Here we demonstrated the potential of platforms such as iNaturalist to document species, including many not recorded by structured surveys, due largely to the high number of participants who spent considerable time making observations. Although iNaturalist may currently have the greatest potential in regions like Sydney, where many individuals have the time and resources for expensive recreational activities, we expect this success will likely be reflected more broadly as the popularity of iNaturalist continues to grow and spread across the globe. Indeed, the relative simplicity of making opportunistic observations, including during everyday activities, means platforms like iNaturalist are well suited to expand the reach of citizen science into regions and communities where few individuals have the time and resources to dedicate to more complex biodiversity surveys.

The fact that iNaturalist users are unconstrained by survey methods in terms of how (e.g., diving, snorkelling, fishing), where (e.g., different habitats, in caves), and when (e.g., all seasons, nighttime) to look, also greatly enhances their ability to find a much broader suite of species, including rare and cryptic individuals potentially missed by conventional structured surveys. However, opportunistic observers are also less likely to document common and abundant fish than structured surveys as these species may be considered less interesting to photograph or share. The effects of observer bias and selectivity has important implications for the analytical approaches and potential inferences that can be drawn from the data. There is a need for more research, across a range of taxa, into how factors like rarity or colour drive the contribution of opportunistic observations to platforms such as iNaturalist. Ultimately, to account for the different species recorded by opportunistic observations and structured surveys, integrating data from citizen science, research institutions and government initiatives, is likely to have the best outcome for future biodiversity monitoring and conservation activities.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10531-022-02398-6>.

Acknowledgements We thank the Australasian Fishes community for their ongoing contributions to iNaturalist and Mark McGrouther and Amanda Hay for their work managing the project and for providing the unobscured dataset used for this research. We also thank the Reef Life Survey team and their volunteer divers for collecting, and making publicly available, the survey data used in this research. We thank M. Thiel and the anonymous reviewers for comments that improved this manuscript.

Authors' Contributions CRediT Statement: **Christopher J. Roberts:** Conceptualization, Methodology, Formal analysis, Visualization, Writing - Original Draft. **Adriana Vergés:** Conceptualization, Methodology, Writing - Review & Editing Supervision, Funding acquisition. **Corey T. Callaghan:** Conceptualization, Methodology, Writing - Review & Editing, Funding acquisition. **Alistair G. B. Poore:** Conceptualization, Methodology, Writing - Review & Editing, Supervision, Funding acquisition.

Funding This research was supported by an Australian Government Research Training Program (RTP) Scholarship to C.J.R. and by grant SWR/10/2020 provided by Sea World Research & Rescue Foundation Inc (SWRRFI) and the Winifred Violet Scott Charitable Trust to A.V., C.T.C., A.G.B.P. and C.J.R. CTC was supported by a Marie Skłodowska-Curie Individual Fellowship (no. 891052).

Open Access funding enabled and organized by CAUL and its Member Institutions

Availability of data and code. Data used in this study are open access from iNaturalist (<https://www.inaturalist.org/projects/australasian-fishes>) and RLS (<https://reeflifesurvey.com/survey-data/>). The cleaned data used and the code to reproduce our analyses are available on Zenodo (<https://doi.org/10.5281/zenodo.6015499>).

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aceves-Bueno E, Adeleye AS, Feraud M et al (2017) The Accuracy of Citizen Science Data: A Quantitative Review. *Bull Ecol Soc Am* 98:278–290. <https://doi.org/10.1002/bes2.1336>
- Ballard HL, Robinson LD, Young AN et al (2017) Contributions to conservation outcomes by natural history museum-led citizen science: Examining evidence and next steps. *Biol Conserv* 208:87–97. <https://doi.org/10.1016/j.biocon.2016.08.040>
- Blowes SA, Supp SR, Antão LH et al (2019) The geography of biodiversity change in marine and terrestrial assemblages. *Science* 366:339–345. <https://doi.org/10.1126/science.aaw1620>
- Boettiger C, Lang DT, Wainwright PC (2012) rfishbase: exploring, manipulating and visualizing FishBase data from R. *J Fish Biol* 81:2030–2039. <https://doi.org/10.1111/j.1095-8649.2012.03464.x>

- Bradter U, Mair L, Jönsson M et al (2018) Can opportunistically collected Citizen Science data fill a data gap for habitat suitability models of less common species? *Methods Ecol Evol* 9:1667–1678. <https://doi.org/10.1111/2041-210X.13012>
- Burgess HK, DeBey LB, Froehlich HE et al (2017) The science of citizen science: Exploring barriers to use as a primary research tool. *Biol Conserv* 208:113–120. <https://doi.org/10.1016/J.BIOCON.2016.05.014>
- Caley P, Welvaert M, Barry SC (2020) Crowd surveillance: estimating citizen science reporting probabilities for insects of biosecurity concern. *J Pest Sci* (2004) 93:543–550. <https://doi.org/10.1007/s10340-019-01115-7>
- Callaghan CT, Poore AGB, Hofmann M et al (2021) Large-bodied birds are over-represented in unstructured citizen science data. *Sci Rep* 11:1–11. <https://doi.org/10.1038/s41598-021-98584-7>
- Callaghan CT, Poore AGB, Mesaglio T et al (2021) Three Frontiers for the Future of Biodiversity Research Using Citizen Science Data. *Bioscience* 71:55–63. <https://doi.org/10.1093/biosci/biaa131>
- Callaghan CT, Rowley JLL, Cornwell WK et al (2019) Improving big citizen science data: Moving beyond haphazard sampling. *PLOS Biol* 17:e3000357. <https://doi.org/10.1371/journal.pbio.3000357>
- Chao A, Chiu C-H (2016) Species Richness: Estimation and Comparison. In: Wiley StatsRef: Statistics Reference Online. pp 1–26
- Dickinson JL, Zuckerberg B, Bonter DN (2010) Citizen Science as an Ecological Research Tool: Challenges and Benefits. *Annu Rev Ecol Evol Syst* 41:149–172. <https://doi.org/10.1146/annurev-ecolsys-102209-144636>
- Dickman CR, Wardle GM (2012) Monitoring for Improved Biodiversity Conservation in Arid Australia. In: Lindenmayer DB, Gibbons P (eds) *Biodiversity Monitoring in Australia*. CSIRO Publishing, Collingwood, VIC, pp 157–164
- Edgar GJ, Stuart-Smith RD (2009) Ecological effects of marine protected areas on rocky reef communities — a continental-scale analysis. *Mar Ecol Prog Ser* 388:51–62. <https://doi.org/10.3354/meps08149>
- Edgar GJ, Stuart-Smith RD (2014) Systematic global assessment of reef fish communities by the Reef Life Survey program. *Sci Data*. <https://doi.org/10.1038/sdata.2014.7>. 1:140007
- Edgar GJ, Stuart-Smith RD (2020a) Reef Life Survey (RLS): Global reef fish dataset. Institute for Marine and Antarctic Studies (IMAS). <https://reeflifesurvey.com/survey-data/>. Accessed 14 Feb 2020
- Edgar GJ, Stuart-Smith RD (2020b) Reef Life Survey (RLS): Cryptic Fish. Institute for Marine and Antarctic Studies (IMAS). <https://reeflifesurvey.com/survey-data/>. Accessed 14 Feb 2020
- Fithian W, Elith J, Hastie T, Keith DA (2015) Bias correction in species distribution models: Pooling survey and collection data for multiple species. *Methods Ecol Evol* 6:424–438. <https://doi.org/10.1111/2041-210X.12242>
- Follett R, Strezov V (2015) An Analysis of Citizen Science Based Research: Usage and Publication Patterns. *PLoS ONE* 10:e0143687. <https://doi.org/10.1371/journal.pone.0143687>
- Fourcade Y (2016) Comparing species distributions modelled from occurrence data and from expert-based range maps. Implication for predicting range shifts with climate change. *Ecol Inf* 36:8–14. <https://doi.org/10.1016/j.ecoinf.2016.09.002>
- Fulton S, López-Sagástegui C, Weaver AH et al (2019) Untapped Potential of Citizen Science in Mexican Small-Scale Fisheries. *Front Mar Sci* 6:517. <https://doi.org/10.3389/fmars.2019.00517>
- Giraud C, Calenge C, Coron C, Julliard R (2016) Capitalizing on opportunistic data for monitoring relative abundances of species. *Biometrics* 72:649–658. <https://doi.org/10.1111/biom.12431>
- Gotelli NJ, Chao A (2013) Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data. In: Levin S (ed) *Encyclopedia of Biodiversity* (Second Edition), 2nd edn. Academic Press, Waltham, MA, pp 195–211
- Haklay M (2013) Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. In: Sui D, Elwood S, Goodchild M (eds) *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Springer, Dordrecht, pp 105–121
- Hermoso M, Narváez S, Thiel M (2021) Engaging recreational scuba divers in marine citizen science: Differences according to popularity of the diving area. *Aquat Conserv Mar Freshw Ecosyst* 31:441–455. <https://doi.org/10.1002/aqc.3466>

- Isaac NJB, Pocock MJO (2015) Bias and information in biological records. *Biol J Linn Soc* 115:522–531. <https://doi.org/10.1111/bij.12532>
- Isaac NJB, van Strien AJ, August TA et al (2014) Statistics for citizen science: Extracting signals of change from noisy ecological data. *Methods Ecol Evol* 5:1052–1060. <https://doi.org/10.1111/2041-210X.12254>
- Kamp J, Oettel S, Heldbjerg H et al (2016) Unstructured citizen science data fail to detect long-term population declines of common birds in Denmark. *Divers Distrib* 22:1024–1035. <https://doi.org/10.1111/ddi.12463>
- Kelly R, Fleming A, Peel GT et al (2020) Citizen science and marine conservation: a global review. *Philos Trans R Soc B* 375:20190461. <https://doi.org/10.1098/rstb.2019.0461>
- Kindt R, Coe R (2005) *Tree diversity analysis: A manual and software for common statistical methods for ecological and biodiversity studies*. World Agroforestry Centre, Nairobi, Kenya
- Klemann-Junior L, Villegas Vallejos MA, Scherer-Neto P, Vitule JRS (2017) Traditional scientific data vs. uncoordinated citizen science effort: A review of the current status and comparison of data on avifauna in Southern Brazil. *PLoS ONE* 12:e0188819. <https://doi.org/10.1371/journal.pone.0188819>
- Losey J, Allee L, Smyth R (2012) The Lost Ladybug Project: Citizen Spotting Surpasses Scientist's Surveys. *Am Entomol* 58:22–24. <https://doi.org/10.1093/ae/58.1.0022>
- Mesaglio T, Callaghan CT (2021) An overview of the history, current contributions and future outlook of iNaturalist in Australia. *Wildl Res* 48:289–303. <https://doi.org/10.1071/WR20154>
- Nelson BW, Ferreira CAC, da Silva MF, Kawasaki ML (1990) Endemism centres, refugia and botanical collection density in Brazilian Amazonia. *Nature* 345:714–716. <https://doi.org/10.1038/345714a0>
- Niku J, Brooks W, Herliansyah R et al (2020) gllvm: Generalized Linear Latent Variable Models. R package version 1.2.2
- Peterson EE, Santos-Fernández E, Chen C et al (2020) Monitoring through many eyes: Integrating disparate datasets to improve monitoring of the Great Barrier Reef. *Environ Model Softw* 124. <https://doi.org/10.1016/j.envsoft.2019.104557>
- Pocock MJO, Tweddle JC, Savage J et al (2017) The diversity and evolution of ecological and environmental citizen science. *PLoS ONE* 12:e0172579. <https://doi.org/10.1371/journal.pone.0172579>
- Prudic KL, Oliver JC, Brown BV, Long EC (2018) Comparisons of citizen science data-gathering approaches to evaluate urban butterfly diversity. *Insects* 9:186. <https://doi.org/10.3390/insects9040186>
- R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <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. <https://doi.org/10.1111/oik.08215>
- Reddy S, Dávalos LM (2003) Geographical sampling bias and its implications for conservation priorities in Africa. *J Biogeogr* 30:1719–1727. <https://doi.org/10.1046/j.1365-2699.2003.00946.x>
- Reef Life Survey Foundation (2019) Standardised Survey Procedures for Monitoring Rocky & Coral Reef Ecological Communities. https://reeflifesurvey.com/wp-content/uploads/2019/02/NEW-Methods-Manual_150815.pdf
- Riesch H, Potter C (2014) Citizen science as seen by scientists: Methodological, epistemological and ethical dimensions. *Public Underst Sci* 23:107–120. <https://doi.org/10.1177/0963662513497324>
- Seltzer C, Iwane T, Misraraj A, Loarie S (2020) 50 million observations on iNaturalist! <https://www.inaturalist.org/blog/40699-50-million-observations-on-inaturalist/>. Accessed 21 Jan 2021
- Schaffer V, Tham A (2020) Engaging tourists as citizen scientists in marine tourism. *Tour Rev* 75:333–346. <https://doi.org/10.1108/TR-10-2018-0151>
- Simmonds EG, Jarvis SG, Henrys PA et al (2020) Is more data always better? A simulation study of benefits and limitations of integrated distribution models. *Ecography (Cop)* 43:1413–1422. <https://doi.org/10.1111/ecog.05146>
- Snäll T, Kindvall O, Nilsson J, Pärt T (2011) Evaluating citizen-based presence data for bird monitoring. *Biol Conserv* 144:804–810. <https://doi.org/10.1016/j.biocon.2010.11.010>
- Soroye P, Ahmed N, Kerr JT (2018) Opportunistic citizen science data transform understanding of species distributions, phenology, and diversity gradients for global change research. *Glob Chang Biol* 24:5281–5291. <https://doi.org/10.1111/gcb.14358>
- Sullivan BL, Aycrigg JL, Barry JH et al (2014) The eBird enterprise: An integrated approach to development and application of citizen science. *Biol Conserv* 169:31–40. <https://doi.org/10.1016/j.biocon.2013.11.003>
- Sullivan BL, Phillips T, Dayer AA et al (2017) Using open access observational data for conservation action: A case study for birds. *Biol Conserv* 208:5–14. <https://doi.org/10.1016/j.biocon.2016.04.031>
- Szabo JK, Davy PJ, Hooper MJ, Astheimer LB (2007) Predicting spatio-temporal distribution for eastern Australian birds using Birds Australia's Atlas data: survey method, habitat and seasonal effects. *Emu - Austral Ornithol* 107:89–99. <https://doi.org/10.1071/MU06020>

- Theobald EJ, Ettinger AK, Burgess HK et al (2015) Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biol Conserv* 181:236–244. <https://doi.org/10.1016/j.biocon.2014.10.021>
- Thiel M, Penna-Díaz MA, Luna-Jorquera G et al (2014) Citizen Scientists and Marine Research: Volunteer Participants, Their Contributions, and Projection for the Future. In: Hughes RN, Hughes DJ, Smith IP (eds) *Oceanography and Marine Biology: An Annual Review*. Taylor & Francis, pp 257–314
- Tiago P, Ceia-Hasse A, Marques TA et al (2017) Spatial distribution of citizen science casuistic observations for different taxonomic groups. *Sci Rep* 7:12832. <https://doi.org/10.1038/s41598-017-13130-8>
- Tiralongo F, Crocetta F, Riginella E et al (2020) Snapshot of rare, exotic and overlooked fish species in the Italian seas: A citizen science survey. *J Sea Res* 164:101930. <https://doi.org/10.1016/j.seares.2020.101930>
- Tiralongo F, La Mesa G, De Paladini F et al (2021) Underwater photo contests to complement coastal fish inventories: results from two Marine Protected Areas in the Mediterranean. *Mediterr Mar Sci* 22:436–445. <https://doi.org/10.12681/mms.26176>
- Ueda K (2019) Identification Quality On iNaturalist. In: iNatForum. <https://forum.inaturalist.org/t/identification-quality-on-inaturalist/7507>. Accessed 13 Aug 2021
- 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. *J Appl Ecol* 50:1450–1458. <https://doi.org/10.1111/1365-2664.12158>
- Walker DW, Smigaj M, Tani M (2021) The benefits and negative impacts of citizen science applications to water as experienced by participants and communities. *WIREs Water* 8:1–32. <https://doi.org/10.1002/wat2.1488>
- Wang Y, Casajus N, Buddle C et al (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* 13:1–14. <https://doi.org/10.1371/journal.pone.0201094>
- Wang Y, Naumann U, Eddebuettel D et al (2020) mvabund: Statistical Methods for Analysing Multivariate Abundance Data. R package version 4.1.3
- Williams PH, Margules CR, Hilbert DW (2002) Data requirements and data sources for biodiversity priority area selection. *J Biosci* 27:327–338. <https://doi.org/10.1007/BF02704963>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Christopher J. Roberts^{1,2} · Adriana Vergés^{1,2} · Corey T. Callaghan^{2,3} · Alistair G. B. Poore^{1,2}

✉ Christopher J. Roberts
c.roberts@student.unsw.edu.au

¹ Centre for Marine Science and Innovation, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney, NSW, Australia

² Ecology & Evolution Research Centre, School of Biological, Earth and Environmental Sciences, UNSW Sydney, Sydney, NSW, Australia

³ German Centre for Integrative Biodiversity Research (iDiv) – Halle-Jena-Leipzig, Leipzig, Germany