


The influence of sampling design on spatial data quality in a geographic citizen science project

Greg Brown^{1,2}  | Jonathan Rhodes² | Daniel Lunney^{3,4} |
 Ross Goldingay⁵ | Kelly Fielding⁶ | Nicole Garofano² |
 Scott Hetherington⁷ | Marama Hopkins⁷ | Jo Green⁸ |
 Skye McNamara⁹ | Angie Brace¹⁰ | Lorraine Vass¹¹ |
 Linda Swankie¹¹ | Clive McAlpine²

¹Department of Natural Resources Management and Environmental Sciences, California Polytechnic State University, San Luis Obispo, California

²School of Earth and Environmental Sciences, University of Queensland, Brisbane, Australia

³School of Life and Environmental Sciences, University of Sydney, Sydney, Australia

⁴Office of Environment and Heritage, Sydney, New South Wales, Australia

⁵School of Environment, Southern Cross University, East Lismore, Australia

⁶School of Communication and Arts, University of Queensland, Brisbane, Australia

⁷Tweed Shire Council, Murwillumbah, Australia

⁸Byron Shire Council, Mullumbimby, Australia

⁹Ballina Shire Council, Ballina, Australia

¹⁰Lismore City Council, Goonellabah, Australia

¹¹Friends of the Koala, Lismore, Australia

Correspondence

Greg Brown, Department of Natural Resources Management and Environmental Sciences, California Polytechnic State University, San Luis Obispo, CA 93407, USA.
 Email: ggbrown@calpoly.edu

Funding information

Australian Research Council, ARC Linkage Project, Grant/Award Number: LP160100186

Abstract

Geographic citizen science has much potential to assist in wildlife research and conservation, but the quality of observation data is a key concern. We examined the effects of sampling design on the quality of spatial data collected for a koala citizen science project in Australia. Data were collected from three samples—*volunteers* ($n = 454$), an Internet *panel* ($n = 103$), and *landowners* ($n = 35$)—to assess spatial data quality, a dimension of citizen science projects rarely considered. The locational accuracy of koala observations among the samples was similar when benchmarked against authoritative data (i.e., an expert-derived koala distribution model), but there were differences in the quantity of data generated. Fewer koala location data were generated per participant by the Internet *panel* sample than the *volunteer* or *landowner* samples. Spatial preferences for land uses affecting koala conservation were also mapped, with *landowners* more likely to map locations for residential and tourism development and *volunteers* less likely. These spatial preferences have the potential to influence the social acceptability of future koala conservation proposals. With careful sampling design, both citizen observations and land use

preferences can be included within the same project to augment scientific assessments and identify conservation opportunities and constraints.

1 | INTRODUCTION

The number of participatory mapping methods—including participation GIS (PPGIS), participatory GIS (PGIS), and volunteered geographic information (VGI)—has increased significantly over the last two decades (Brown & Kyttä, 2018). However, the rapid pace of mapping application development and innovation has exceeded the capacity to adequately evaluate the validity and impact of the methods used. The diversity of spatial content generated from participatory and crowdsourced methods requires evaluations that transcend traditional geospatial assessment based on authoritative data (Antoniou & Skopeliti, 2015). Sampling design (i.e., who is selected from the population and how) and response bias are recognized as key issues for data quality in participatory mapping (Brown, 2017; Brown & Kyttä, 2014; Bubalo, van Zanten, & Verburg, 2019), which contribute to uncertainties in inferring relationships between participant mapping data and the geographic mapping context (Robertson & Feick, 2018). Further, there is some evidence that participants map responses that reflect their general, non-spatial values and preferences (Brown, 2013).

Participatory mapping data can be evaluated from a traditional GIS *validity-as-accuracy* perspective (Spielman, 2014), such as spatial resolution, position and attribute accuracy, logical consistency, and completeness. The best example is OpenStreetMap (OSM), a VGI system that has been effectively benchmarked against authoritative geospatial data (Girres & Touya, 2010; Haklay, 2010). However, many participatory mapping applications generate subjective spatial attributes such as place values, experiences, and preferences that lack authoritative criteria to assess spatial accuracy. In this situation, the credibility of the mapping participants (i.e., *validity-as-credibility*) is of central importance to the quality of the spatial data. Thus, sampling design and participant characteristics emerge as key variables to evaluate the quality of spatial data generated. For example, in a study of public land management, Brown, Kelly, and Whitall (2014) found that volunteer participants mapped more utilitarian values and consumptive use preferences compared to participants from a random household sample. In another study, Brown, Strickland-Munro, Kobryn, and Moore (2016) found that participants from different stakeholder groups mapped different management preferences. Thus, “who” participates is essential to understand and interpret mapping outcomes.

Without authoritative data to evaluate subjective mapping data, Brown, Weber, Zanon, and de Bie (2012) proposed mapping effort as one potential variable to assess subjective spatial data quality. In the first, multi-study analysis of participatory mapping data quality, Brown et al. (2012) found that participants from an online Internet panel, a group of pre-screened respondents, who are willing to participate in surveys, demonstrated less mapping effort, resulting in lower-quality spatial data compared to a random *household* sample, an *in situ* sample (individuals selected on site in the study area), or a *volunteer* sample. In an expanded study of sampling design and response bias, Brown (2017) found that Internet panel samples had higher response rates but lower spatial data quality than *household* samples, which had greater mapping effort, on average, compared to *volunteer* samples.

In this study we assess the effects of sampling design and response bias in the participatory mapping context of geographic citizen science (i.e., citizen science that uses VGI), where the collection of location information is central to the activity (*sensu* Haklay, 2013). Specifically, we evaluate sampling effects (i.e., whether samples differ in their responses) and response bias in a geographic citizen science project (Brown et al., 2018) that sought to identify the location and frequency of sightings in Australia of one of the nation's most iconic mammals—the koala (*Phascolarctos cinereus*)—and to identify land use preferences with potential impact on koala conservation. As

McAlpine et al. (2015) point out, the koala presents the problem of managing a wide-ranging species that now primarily occurs in human-modified landscapes. Koalas are one of a small number of species that are ideal for citizen science studies because they are so readily identifiable and memorable, but they are also cryptic, being difficult to observe in their forest habitats. Over the last three decades, a series of successful community-based surveys of koalas, now called citizen science surveys, have identified locations of koalas across a wide landscape in New South Wales, as well as focused locations, such as a shire or group of shires. Further, repeat surveys have identified changes in the distribution of koalas, as well as rates of decline or increase (Lunney, Crowther, Shannon, & Bryant, 2009; Lunney et al., 2016; Lunney, O'Neill, Matthews, & Coburn, 2000; Predavec et al., 2016, 2018; Reed, Lunney, & Walker, 1990). It is these surveys that have been the underpinning of government policies to protect koalas, or their habitat, and a major contribution to the koala being formally recognized as a threatened species. In this study, we assess the quality of koala observation data across three samples (*volunteers*, an Internet *panel*, and *landowners*) against an authoritative koala distribution likelihood model for the study region. We further examine the data collected for potential sources of bias related to demographic characteristics, mapping responses and behavior, and participant attitudes and beliefs toward koalas and their conservation.

Within the context of geographic citizen science, this study provides an opportunity to evaluate both objective (koala observations) and subjective spatial attributes (land use preferences such as where koala conservation should occur, or where various development activities should occur) from multiple samples. We posit that the mapping of koala observations would be less subject to sampling and response bias than the mapping of preferences that are related to participant values, beliefs, and motivations for participation.

The purpose of this study is to identify sampling effects that arise from different samples which can influence the quality of spatial data and related text-based survey data. This study builds on previous research by examining the following research questions:

1. How are participant characteristics and mapping responses similar or different across samples and which attributes appear important to assess data quality in a project for research and conservation?
2. How does mapping effort differ by samples and does this effort affect the quality of the data (quantity and accuracy)?
3. What are the potential sources of response bias from sampling design in geographic citizen science?

2 | METHODS

2.1 | Study design and data collection

In 2018, we implemented a geographic citizen science survey to identify the locations of koalas in a study area on the far north coast of the state of New South Wales (NSW), Australia (Figure 1). In the study area, the koala is listed as vulnerable under both state and commonwealth legislation. A major NSW state-wide citizen science survey in 2006 identified the coastal shires on the far north coast as being a koala stronghold (Lunney et al., 2009), while Adams-Hosking et al. (2016) confirmed that the study area lies within a bioregion that contains an important koala population, but one with a predicted steep decline. The study area contains four local governments with the following population estimates: Byron Shire (31,556); Ballina Shire (41,790); City of Lismore (43,135); and Tweed Shire (91,371) (Australian Bureau of Statistics, 2016).

A Google® maps application was developed where participants were asked to drag and drop markers representing koala observations and land use preferences in the study area. The first "tab" panel had markers allowing participants to record places where they had observed koalas in the following categories: *weekly*, *monthly*, *yearly*, *once*, and *dead/injured*. The second and third tab panels allowed participants to record their land use preferences. The second tab panel contained markers that participants could use to identify places

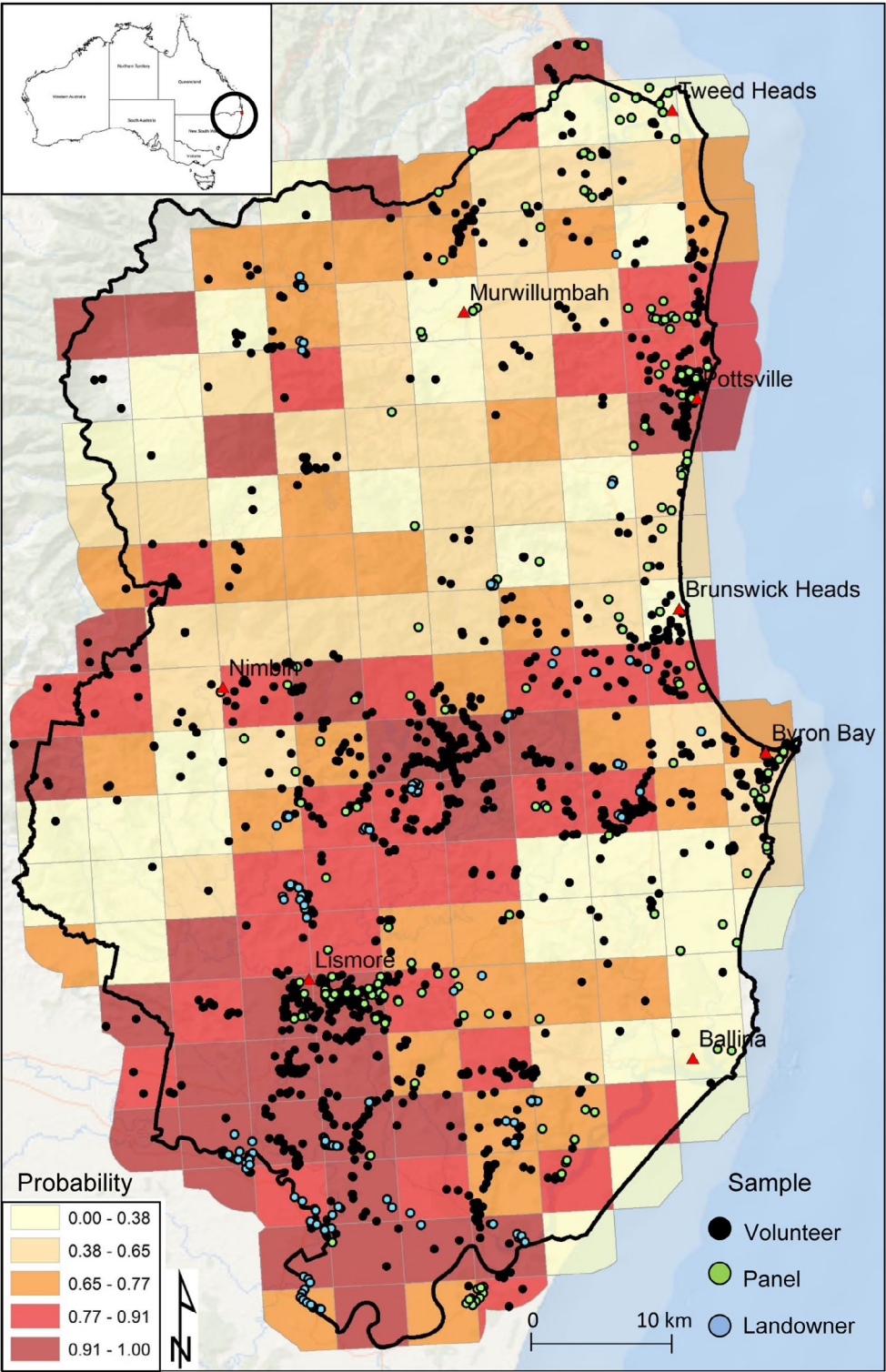


FIGURE 1 Koala observations by sample plotted on koala likelihood map where koala presence is calculated as a probability ranging from 0 to 1 (Predavec et al., 2015)



FIGURE 2 Website interface showing three tab panels of markers to be dragged and dropped onto a digital map of the study area: (a) koala observations; (b) acceptable land use and regulatory preferences; (c) unacceptable land use and regulatory preferences; and (d) a visual example of markers placed on a Google Map

where they supported specific land uses (e.g., for residential and commercial development), while the third panel contained markers that participants could use to indicate places where they opposed specific land uses (Figure 2). After completing the mapping part of the study, participants went on to complete a survey that comprised questions about participant demographic variables, including home location, age, gender, and formal education. There were also survey questions that assessed participants' knowledge of koalas and places in the study area, attitudes towards koalas, threats to koalas, and level of support for various koala conservation efforts (Figure 3).

2.1.1 | Volunteer sample

Participants were crowdsourced through: (a) local government newsletters and websites; (b) a koala-specific wildlife care and rehabilitation organization, "Friends of the Koala"; (c) Facebook® promotions; (d) news stories in local media; and (e) friend and relative referrals. The volunteer data collection ran from December 2017 to March 2018.

1. How would you rate your level of knowledge about Koalas?

- No knowledge at all
- A little knowledge
- Moderate knowledge
- High knowledge
- Very high level of knowledge

2. How would you rate your knowledge of places in the New South Wales North Coast area?

- Excellent
- Good
- Average
- Below average
- Poor/little knowledge

3. For each statement below, please indicate whether you agree or disagree with the statement. (Please check one response per item) Response scale [strongly disagree, disagree, somewhat disagree, neither, somewhat agree, agree, strongly agree]

- a. I would be pleased to be living amongst Koalas
- b. I consider the bushlands in our community as an important area for Koala conservation
- c. Koalas are a bother to have near my home
- d. I consider the future conservation of the Koala as important
- e. Current legislation is effective in conserving local Koala populations
- f. It would be better for residents if Koalas in suburban areas were moved to bushland
- g. I consider Koalas as an important part of my community
- h. Being environmentally-friendly is an important part of who I am
- i. I see myself as an environmentally-friendly person
- j. Most people who are important to me take some action to protect Koalas
- k. Most people in my community take some action to protect koalas.
- l. I am very attached to my community (the place where you live)

4. In your opinion, how threatening is each of the following to the long-term survival of Koalas in your community? (Response scale 1=No threat at all to 5=An extreme threat).

- a. Roads and traffic
- b. Housing development
- c. Wild dogs
- d. Roaming domestic dogs
- e. Tree clearing
- f. Bushfire
- g. Logging native forests
- h. Disease

5. How likely are you to participate in each of the following activities? (Response scale=highly unlikely, unlikely, likely, highly likely, unsure)

- a. Join a community conservation group
- b. Restore native bush on your property
- c. Drive slowly in Koala areas
- d. Keep your dog(s) indoors at night (even if you don't currently have a dog)
- e. Report Koala sightings
- f. Council-led initiatives for conservation on your property
- g. Caring for and/or rescuing wildlife
- h. Advocating (e.g. signing petition, writing letters) to government for Koala protection

6. How likely are you to support the following Koala management strategies in your region? (Response scale=highly unlikely, unlikely, likely, highly likely, unsure)

- a. Restrict dogs from roaming outside at night
- b. Prevent development in Koala habitat areas
- c. Using public money (from rates or taxes) to buy land for Koala reserves
- d. Increasing habitat for Koalas (e.g. planting more Koala food trees)
- e. Infrastructure to reduce speed in Koala areas (e.g., rumble strips, speed bumps)
- f. Decreased speed limits within Koala zones
- g. Apply existing regulations to protect koala habitat
- h. Do nothing
- i. Funding more research into Koala diseases
- j. Using public money (from rates or taxes) to fund Koala rescue services

FIGURE 3 Post-mapping survey questions

2.1.2 | Panel sample

We contracted a commercial vendor, Pureprofile®, to administer the survey to an Internet *panel* of residents in the study area. We required at least 100 completed surveys that included both the mapping and text-based survey questions. Panel data were collected from October 2017 through March 2018.

2.1.3 | Landowner sample

With agricultural interests under-represented in the *volunteer* and *panel* samples, we sampled rural landowners who were geographically distributed across the study area, who owned 20 acres (8 ha) or more, and who used their land for horticulture and/or cattle grazing. Participants were identified using a snowball approach by contacting koala field researchers, local government representatives, and industry associations. The mapping survey was completed using a “facilitated” mode, where a member of the research team interviewed the participant face-to-face and recorded the responses online. The *landowner* data were collected from November through mid-December 2018, resulting in 35 completed surveys.

2.2 | Analyses

We assessed the representativeness of the samples by comparing demographic characteristics (i.e., age, gender, education, residence location) to population data from the 2016 Australian census for NSW. We analyzed mapping behavior by tabulating the number and type of digital markers placed in the study area by each sample. Mapping effort was calculated as the elapsed time between the first and last marker placement, adjusted for obvious time breaks in mapping activity. The propensity for participants in a sample to map more or less of a specific type of marker was analyzed by coding whether a participant mapped one or more markers (yes) or not (no). Chi-square analysis (test for independence) with standardized residuals was calculated for these binary frequencies to determine if the frequency of individuals mapping a specific marker type deviated significantly from the expected distribution across the three samples. A residual is the difference in the observed frequency and the expected frequency (i.e., the value that would be expected under the null hypothesis of no association). A standardized residual is calculated by dividing the residual value by the standard error of the residual. Standardized residuals are a normalized score, like a z score without units, and if greater than 2.0, indicate that a significantly greater number of individuals mapped a specific marker type than would be expected, while standardized residuals less than -2.0 indicate that fewer individuals mapped a specific marker type than expected.

Questions measuring attitudes towards koalas, perceived threats, and support for various koala conservation measures were operationalized on Likert scales and analyzed by sample using analysis of variance (ANOVA) with Tukey *honestly significant difference* (HSD) post-hoc comparisons. A Likert-scale question measures a respondent's opinion or attitude on a given subject, such as a respondent's attitude toward living amongst koalas, typically recorded using a five or seven-point agreement scale. The responses are quantified and analyzed using statistics. ANOVA tests were used to find statistically significant differences in the means of three sample groups (i.e., *volunteer*, *panel*, *landowner*). With a statistically significant result, the Tukey HSD post-hoc test was run to identify pairwise differences between sample groups. The Tukey HSD test compares all possible pairs of means to find those that are significantly different from each other.

The quality of koala observation data by samples was assessed by: (a) examining the number and distribution of koala observations for each sample; and (b) benchmarking sample participant observations against a koala likelihood model (Predavec et al., 2015) that indicates probable koala occurrence within 5-km grid cells located in the study area. The 5-km grid was selected for the likelihood model based on acceptable levels of spatial autocorrelation in the reporting of koala observations in the study region. The likelihood model generates a confidence level (high, medium, low) associated with the probability of koalas being present and was derived from historical observation records of koalas and other mammals based on a 20-year data window (1997–2017). To assess the consistency between participant observations and the model, cross-tabulations between the likelihood map and participant observations were tabulated for the 173 grid cells that were coded based on the presence or absence of one or more koala observations in each cell. The phi coefficient (ϕ), a measure of the degree of association between two binary variables that ranges between +1.0 and -1.0, was calculated to provide an overall measure of the association between the presence and absence of cells. Fitz-Gibbon and Morris (1987) suggest the following

TABLE 1 Sampling profiles based on survey responses

Mapping behavior	Volunteer	Panel	Landowner	Findings
No. of participants	454	103	35	Volunteer and landowner samples had greater mapping effort than the panel sample in both number of markers and time spent mapping
No. completing post-mapping text survey	397	103	35	
No. of total markers	6,362	921	492	
No. of koala observations mapped	1,749	182	130	
Range of all markers	1–366	1–64	2–29	
Range of observation-only markers	0–173	0–14	0–13	
Mean (median) all markers mapped	15 (7)	9 (4)	14 (14)	
Mean (median) observation-only markers mapped	4 (2)	2 (1)	4 (3)	
Range of mapping time (h:min)	0:00–1:26	0:00–0:32	0:00–0:46	
Mean time per marker (min:s)	00:38	00:19	01:14	Landowners have more knowledge of study area than volunteer or panel samples
Knowledge of study area (%)				
Excellent	14	12	21	
Good	45	44	58	
Average	33	36	21	
Below average	7	7	0	
Poor	2	1	0	Volunteer sample has most knowledge of koalas; panel sample least knowledge
Knowledge of koalas (%)				
Very high knowledge	3	0	3	
High knowledge	33	4	9	
Moderate knowledge	48	35	61	
A little knowledge	13	58	24	
No knowledge	3	3	3	Landowners lived in region significantly longer than volunteer or panel samples
Years lived in study area (mean, median)	20, 18	18, 18	45, 43	
Derive some income from farming (%)	10	2	88	

(Continues)

TABLE 1 (Continued)

Mapping behavior	Volunteer	Panel	Landowner	Findings
Participant distribution by local government area (%)				
Ballina (percentage of study area population = 20%)	8	10	19	Proportion of participants deviate from actual population distribution in local government areas
Byron (percentage of study area population = 15%)	24	11	31	
Lismore (percentage of study area population = 21%)	35	19	38	
Tweed (percentage of study area population = 44%)	33	60	13	
Demographics				
Gender (ABS for NSW 2016: male 49.3%)				Volunteer and panel biased toward females; landowner biased toward males
Female (%)	70	75	34	
Male (%)	30	25	66	
Age in years (mean/median) (ABS for NSW 2016: mean 48, median 47) ^a	52/53	43/41	61/61	Volunteer and landowner samples biased toward older participants; panel biased toward younger participants
Education (%) (ABS for NSW 2016: 23.4% Bachelor's/postgraduate)				All samples biased toward higher formal education; panel and landowner samples most representative of population
Less than Bachelor's	35	64	62	
Bachelor's/postgraduate	65	36	38	

Note: Selected census demographics from the 2016 ABS Census are provided for comparison. Not all percentages total 100% due to rounding.

^aEstimates from 2016 grouped census data for individuals aged 20 or older.

interpretation of the phi coefficient: $\varphi < 0.2$ (little or no association); $0.2 \leq \varphi < 0.4$ (weak association); $0.4 \leq \varphi < 0.6$ (moderate association); $\varphi \geq 0.6$ (strong association). We also cross-tabulated the distribution of participant observations by cell confidence levels.

3 | RESULTS

3.1 | Participant characteristics and mapping effort

The number of participants in each sample that mapped one or more locations was 454 *volunteer*, 103 *panel*, and 35 *landowner* (Table 1). Conventional survey response rates could not be calculated because it is not known how many people received an invitation to take part in the study. However, completion rates were calculated from those who accessed the website and completed the survey through the mapping activity. The completion rate was 56% for the *volunteer* sample, 71% for the *panel* sample, and 97% for the *landowner* sample.

Analysis of demographic variables indicates significant sample bias compared to the general population (i.e., census data). The *volunteer* and *landowner* participants were older, while the *panel* participants were younger than the general population. The *volunteer* and *panel* participants were disproportionately female, while the *landowner* participants were disproportionately male. All samples had a higher level of formal education than the general population, but the *volunteer* sample was especially biased toward a higher level of formal education.

The geographic distribution of the participants in all samples deviated from the general population distribution by local government area. For example, participants in the Tweed local government area were over-represented in the *panel* sample but under-represented in the *volunteer* sample. *Landowner* participants lived in the study area the longest (mean 45 years) compared to *volunteers* (20 years) and the *panel* (18 years). The *landowners* sample expressed the highest self-rated knowledge of the study area, while *volunteers* expressed the highest self-rated knowledge of koalas. *Panel* participants appeared the least knowledgeable about koalas.

The greatest mapping effort was observed with *volunteers*, who averaged 15 markers per person (Table 1), taking about 40 s per marker. The least mapping effort was from the *panel*, where participants mapped 9 markers each at about 20 s per marker. *Landowner* mapping effort was similar to that of *volunteers*, averaging 14 markers and a clock time of over 1 min per marker. The longer average clock time for *landowners* can be attributed to the “facilitated” mode of data collection, where the interviewer recorded the map locations for the *landowner*.

3.2 | Sample attitudes and beliefs

All samples expressed positive attitudes toward koalas and their conservation, but *volunteers* expressed significantly more positive attitudes than the *panel* (Table 2). The positive attitudes of *landowners* toward koalas were similar to those of *volunteers* and greater than those of the *panel*. *Volunteers* and *landowners* considered themselves to be strongly environmentally oriented, significantly more than *panel* participants. Overall, the stronger environmental orientation and positive attitudes of *volunteers* manifested in a significantly higher likelihood of supporting koala conservation strategies (e.g., increasing habitat, preventing development, and using public money to buy reserves) and participating in specific activities related to koala protection (e.g., joining a community conservation group or advocating for government protection of koalas). Of the three samples, the *panel* was most likely to “do nothing.” As a caveat, these reported differences between samples in scores were statistically significant, but the differences are in fact small and do not reflect substantially different responses between the samples.

With respect to the long-term survival of the koala, *volunteers* and the *panel* perceived most threats (e.g., roads/traffic, housing development, tree clearing) to be significantly greater than did the *landowners*. Of all threats, *landowners* perceived the logging of native forests to be the least threat and wild dogs to be the greatest threat. The *panel* perceived bushfire to be a significantly greater threat than did either *volunteers* or *landowners*.

TABLE 2 Comparison of sampling groups' (volunteer, panel, landowner) responses to survey questions

	Volunteer	Panel	Landowner	Significant findings ^d
<i>Attitudes toward koalas^a</i>	Higher score = more favorable attitude			
I would be pleased to be living amongst koalas	6.7	6.3	6.3	Volunteer > Panel; Landowner
Koalas are a bother to have near my home	1.4	1.9	2.0	Volunteer < Panel; Landowner
I consider the future conservation of the koala as important	6.8	6.4	6.5	Volunteer > Panel
I consider koalas as an important part of my community	6.5	6.1	6.4	Volunteer > Panel
<i>Environmental orientation^a</i>	Higher score = stronger environmental orientation			
Being environmentally friendly is an important part of who I am	6.6	5.8	6.6	Volunteer, Landowner > Panel
I see myself as an environmentally friendly person	6.5	5.8	6.5	Volunteer, Landowner > Panel
Most people who are important to me take some action to protect koalas	5.6	4.8	5.6	Volunteer, Landowner > Panel
<i>Likely to support koala management strategies^b</i>	Higher score = more likely to support			
Restrict dogs from roaming outside at night	4.7	4.4	4.8	Volunteer, Landowner > Panel
Prevent development in koala habitat areas	4.6	4.2	4.0	Volunteer > Panel, Landowner
Use public money (from rates or taxes) to buy land for koala reserves	4.5	4.0	3.7	Volunteer > Panel, Landowner
Increase habitat for koalas (e.g., planting more koala food trees)	4.8	4.5	4.5	Volunteer > Panel, Landowner
Infrastructure to reduce speed in koala areas (e.g., rumble strips, speed bumps)	4.6	4.4	4.1	Volunteer > Panel, Landowner
Decrease speed limits within koala zones	4.5	4.3	4.0	Volunteer > Landowner
Apply existing regulations to protect koala habitat	4.7	4.4	4.1	Volunteer > Panel, Landowner
Do nothing	1.5	2.0	1.6	Panel > Volunteer, Landowner
Fund more research into koala diseases	4.4	3.9	4.4	Volunteer, Landowner > Panel
Use public money (from rates or taxes) to fund koala rescue services	4.5	4.0	4.3	Volunteer > Panel
<i>Likely to participate in activity^b</i>	Higher score = more likely to participate			

(Continues)

TABLE 2 (Continued)

	Volunteer	Panel	Landowner	Significant findings ^d
Join a community conservation group	3.7	3.0	3.4	Volunteer > Panel
Restore native bush on your property	4.3	3.6	4.2	Volunteer > Panel
Drive slowly in koala areas	4.7	4.6	4.7	No difference
Keep your dog(s) indoors at night (even if you don't currently have a dog)	4.7	4.6	4.7	No difference
Report koala sightings	4.5	4.0	3.5	Volunteer > Panel > Landowner
Council-led initiatives for conservation on your property	4.3	3.8	3.8	Volunteer > Panel, Landowner
Caring for and/or rescuing wildlife	4.0	3.8	3.9	No difference
Advocating (e.g., signing petition, writing letters) to government for koala protection	4.3	4.1	3.6	Volunteer > Landowner
<i>Perceived threat to the long-term survival of koalas^c</i>	Higher score = greater threat			
Roads and traffic	4.5	4.5	3.8	Volunteer, Panel > Landowner
Housing development	4.5	4.5	3.8	Volunteer, Panel > Landowner
Wild dogs	4.5	4.2	4.7	Volunteer, Landowner > Panel
Roaming domestic dogs	4.3	3.8	4.0	Volunteer > Panel
Tree clearing	4.7	4.6	3.7	Volunteer, Panel > Landowner
Bushfire	4.0	4.3	3.7	Panel > Landowner
Logging native forests	4.5	4.4	2.9	Volunteer, Panel > Landowner
Disease	4.6	4.3	4.3	Volunteer > Panel

^aSeven-point scale: 1 = Strongly disagree; 2 = Disagree; 3 = Somewhat disagree; 4 = Neither disagree nor agree; 5 = Somewhat agree; 6 = Agree; 7 = Strongly agree.

^bFive-point scale: 1 = Highly unlikely; 2 = Unlikely; 3 = Unsure; 4 = Likely; 5 = Highly likely.

^cFive-point scale: 1 = No threat at all; 5 = Extreme threat.

^dANOVA, mean difference with Tukey HSD post-hoc test ($p < 0.05$).

3.3 | Quality of observation data

There were large differences in sample sizes, resulting in significantly fewer koala observations in the *panel* (179) and *landowner* (129) samples than in the *volunteer* sample (1,695), with the *volunteer* observations covering a larger geographic area within the study region. The nearest-neighbor ratio is a measure of observation point dispersion. Ratios less than 1 indicate observation clustering, while ratios greater than 1 indicate a tendency toward observation dispersion. The nearest-neighbor ratios were 0.37, 0.59, and 0.25 for the *volunteer*, *panel*, and *landowner* samples, respectively, indicating that koala observations were clustered in the study area. The *landowner* observations were the most clustered and the *panel* observations were the least clustered (Figure 1). These results highlight the

TABLE 3 Presence/absence analysis for *volunteer* sample, *panel* sample, and *landowner* sample with the likelihood model. Presence is one or more koala observations in the grid cell and absence is the lack of any observations

Volunteer sample ^a					
			Absent	Present	
Likelihood model	Absent	Count	6	5	11
		Percent	54.5%	45.5%	100.0%
	Present	Count	25	137	162
		Percent	15.4%	84.6%	100.0%
Total	Count		31	142	173
	Percent		17.9%	82.1%	100.0%
Model confidence levels for inconsistent cells		High	Medium	Low	Total
Count		6	3	21	30
Percent		20	10	70	100%
Panel sample ^b					
			Absent	Present	
Likelihood model	Absent	Count	11	0	11
		Percent	100.0%	0.0%	100.0%
	Present	Count	103	59	162
		Percent	63.6%	36.4%	100.0%
Total	Count		114	59	173
	Percent		65.9%	34.1%	100.0%
Model confidence levels for inconsistent cells		High	Medium	Low	Total
Count		38	17	47	103
Percent		37	18	46	100%
Landowner sample ^c					
			Absent	Present	
Likelihood model	Absent	Count	11	0	11
		Percent	100.0%	0.0%	100.0%
	Present	Count	128	34	162
		Percent	79.0%	21.0%	100.0%
Total	Count		139	34	173
	Percent		80.3%	19.7%	100.0%
Model confidence levels for inconsistent cells		High	Medium	Low	Total
Count		53	23	52	128
Percent		41	18	41	100%

^a $\varphi = 0.25$, $p < 0.01$. ^b $\varphi = 0.19$, $p < 0.05$. ^c $\varphi = 0.13$, $p < 0.10$.

importance of sampling for broad geographic coverage of participants in the study region because koala observations are related (more proximate) to participant home location.

The accuracy of koala observation data was benchmarked against a koala likelihood model (Predavec et al., 2015, with updated data from 2017) using presence/absence analysis and by examining the distribution of koala observations by cell confidence levels. The statistical association between *volunteer* observations and the likelihood model was weak (Table 3), but significant ($\varphi = 0.25$, $p < 0.01$). The presence/absence analysis for *volunteers* indicated 96% cell consistency (137/142) with the model where participants had observed koalas. Where participants had not observed koalas, consistency was 19% (6/31). Overall, volunteers did not record koala observations in 15% (25/162) of the model cells indicating historical presence of koalas. For the *volunteers*, 70% of the inconsistent cells were classified as low confidence in the likelihood model. In practical terms, this finding reflects positively on the putative accuracy of participant observations. If the inconsistency in observations were associated with high rather than low model confidence locations, this would suggest larger error in participant observations.

The statistical association between *panel* observations and the likelihood model was weak (Table 3), but significant ($\varphi = 0.19$, $p < 0.05$). There was 100% consistency (59/59) where participants had observed koalas. *Panel* consistency with the model was only 10% (11/114) where participants had not observed koalas. Overall, *panel* participants did not record koala observations in 64% (103/162) of the model cells indicating historical presence of koalas. Of the inconsistent cells, 46% were classified as low confidence.

The statistical association between *landowner* observations and the likelihood model (Table 3) was the weakest of the samples ($\varphi = 0.13$, $p < 0.10$). There was 100% consistency (34/34) where *landowner* participants had observed koalas. The consistency was only 8% (11/139) where participants had not observed koalas. Overall, *landowners* did not identify 79% (128/162) of the model cells indicating historical presence of koalas. Where cells were inconsistent, 41% were classified as low confidence.

The lower spatial concurrence of *panel* and *landowner* observations with the model can largely be attributed to the smaller quantity of mapped koala observation data arising from the smaller size of the *panel* and *landowner* samples. Small sample sizes limit the geographic coverage of koala observations and thus reduce overall agreement with the model. Where the *panel* and *landowners* did record koala observations (61% and 62%, respectively), they were located in cells with a high confidence level compared to 49% of cells for *volunteers*. These findings allow the suggestion to be made that the quantity (i.e., too few participant observations) is a greater threat to data quality than the locational accuracy of the observations.

3.4 | Propensity to map by marker type

We assessed the propensity of participants to map more (or fewer) of specific types of markers by examining the frequency of individuals in each sample that mapped one or more markers using a chi-square test of independence. With respect to koala observations, *volunteers* were more likely to map *weekly* observations of koalas and less likely to map *yearly* or *once* observations compared to the *panel* (Table 4). With respect to preferences, *volunteers* were more likely to map preferences *wanting more koalas*, with fewer preferences favoring *tourism development* (Table 4). *Landowners* were more likely to map preferences favoring both *residential* and *tourism development*. *Landowners* were also more likely to map preferences opposing *prescribed fire*. These propensities to map specific types of markers were consistent with the positive attitudes of *volunteers* toward koalas and the lower perceived threat to koalas from development among *landowners*.

4 | DISCUSSION

The purpose of this study was to analyze the influence of sampling design on the quality of spatial data in the context of a geographic citizen science project for koala conservation. To do this we compared three samples: *volunteers*, *panel members*, and *landowners*. Consistent with previous mapping research (Brown, 2017), we found that *volunteer* participants demonstrated greater mapping effort than an Internet *panel*, resulting in significantly

TABLE 4 Propensity of individuals to map specific observation and preference categories by sample with significant chi-square associations indicated in **bold** ($p < 0.05$)

Sampling group	Observations	Preferences (affirmative)	Preferences (negative)
Volunteer	Weekly (+)	Want more koalas (+)	Want fewer koalas (0)
	Monthly (0)	Residential development-yes (0)	Residential development-no (0)
	Yearly (-)	Tourism development-yes (-)	Tourism development-no (0)
	Once (-)	Commercial development-yes (0)	Commercial development-no (0)
	Dead/injured (0)	New parks/reserves-yes (0)	New parks/reserves-no (0)
	Increasing trend (0)	Dogs allowed-yes (0)	Dogs allowed-no (0)
	Decreasing trend (0)	Driving rules-yes (0)	Driving rules-no (0)
		Prescribed fire-yes (0)	Prescribed fire-no (0)
Panel	Weekly (-)	Want more koalas (-)	Want fewer koalas (0)
	Monthly (0)	Residential development-yes (0)	Residential development-no (0)
	Yearly (+)	Tourism development-yes (0)	Tourism development-no (0)
	Once (+)	Commercial development-yes (0)	Commercial development-no (0)
	Dead/injured (0)	New parks/reserves-yes (0)	New parks/reserves-no (0)
	Increasing trend (-)	Dogs allowed-yes (0)	Dogs allowed-no (0)
	Decreasing trend (0)	Driving rules-yes (0)	Driving rules-no (0)
		Prescribed fire-yes (0)	Prescribed fire-no (0)
Landowner	Weekly (0)	Want more koalas (0)	Want fewer koalas (0)
	Monthly (0)	Residential development-yes (+)	Residential development-no (0)
	Yearly (0)	Tourism development-yes (+)	Tourism development-no (0)
	Once (0)	Commercial development-yes (0)	Commercial development-no (0)
	Dead/injured (-)	New parks/reserves-yes (0)	New parks/reserves-no (0)
	Increasing trend (0)	Dogs allowed-yes (0)	Dogs allowed-no (0)
	Decreasing trend (0)	Driving rules-yes (0)	Driving rules-no (0)
		Prescribed fire-yes (0)	Prescribed fire-no (+)

(+) Indicates greater frequency; (-) indicates lesser frequency; (0) no difference in frequency compared to other samples.

more spatial data, both observation and preference data. Also consistent with previous research (Zolkafli, Brown, & Liu, 2017), we found that a “facilitated” mode of participatory mapping with *landowners* generated more spatial data per participant than the self-administered mode with the *panel* sample. Overall, the results demonstrate the importance of sampling design because different groups within the target population contribute different but equally valid data, thereby increasing the inferential power of geographic citizen science. Specifically, the three samples contributed different numbers of koala observations, but showed 96% (*volunteer*), 100% (*panel*), and 100% (*landowner*) consistency with the model where participants had observed koalas.

4.1 | Sampling implications for spatial data quality

Spatial data quality for citizen science projects involving wildlife observations is a function of both locational accuracy and data completeness. The latter is affected by the quantity of spatial data generated, as has been noted previously for koalas in citizen science surveys (Lunney et al., 2009, 2016). Previous analysis of the volunteer data found that the koala observations were sufficiently accurate to augment field-based koala research, with the added benefit of engaging a broader public in conservation efforts (Brown et al., 2018; Predavec et al., 2016,

2018). For geographic citizen science, the larger concern is the quantity of koala observations generated by the samples. The quantity of observation data from the *panel* was significantly less per participant than for the *volunteer* or *landowner* samples. For an Internet *panel* to generate a similar number of koala observations as the *volunteer* sample would require more than 900 *panel* participants, while the *landowner* sample would require at least 400 participants. Sampling designs based exclusively on an Internet *panel* or interviews with *landowners* would not have been feasible (insufficient number of *panel* members living in the study area) or cost-effective (*landowner* interviews across a large geographic area would be particularly time-consuming and expensive). Thus, *volunteer* sampling was a pragmatic choice for this geographic citizen science project focused on koala observations across a large geographic area. Indeed, one of the outstanding strengths of citizen science is its capacity to cover a wide area quickly, and in that context focusing on recruiting *volunteer* participants is the most effective.

The sample size needed to achieve acceptable observational data quality is not a statistical determination, but rather is guided by the sample size needed to obtain adequate geographic coverage of the study area. This sample size will vary based on the size of the study area and the geographic distribution of people living in the study area. As a general principle, the larger the study area and the more heterogeneous the spatial distribution of the human population, the larger the sample size needed to achieve geographic coverage, with greater recruitment effort needed for sparsely populated areas. While we considered the sample size of 450+ volunteers sufficient to meet a minimum threshold of observational data quality for the study area, there was still relatively poor geographic coverage of the hinterland in the study area, which prompted our decision to supplement the sample with *landowners* from these areas. As a sampling strategy in future geographic citizen science, we would suggest intentional participant recruitment in rural areas, which are unlikely to achieve enough observations based purely on volunteerism.

4.2 | Sample representativeness

When a geographic citizen science project also contains subjective preference variables, as in this study, it is possible to assess the extent to which the samples are biased, that is, unrepresentative of the population from which they are drawn. For example, a koala citizen science project in South Australia found that project volunteers were not representative of the entire South Australian population (Hollow, Roetman, Walter, & Daniels, 2015). Similarly, in this study, *volunteers* were not representative of the study area population and were biased toward older, highly educated females, with positive attitudes toward koalas and their conservation. They were also most knowledgeable about koalas. In contrast, the *panel* was younger, had less formal education, and was significantly less knowledgeable about koalas. Their attitudes toward koalas and conservation, while positive, were less positive than those of the *volunteer* sample. The *landowner* sample represented an important perspective that was largely absent in the *volunteer* and *panel* samples—rural farmers who considered that they would be impacted by koala conservation measures involving habitat. This sample was disproportionately male, older, and had lived in the region the longest. Their attitudes about living near to koalas were less positive and they were less likely to engage in koala conservation efforts that might adversely influence their livelihoods.

4.3 | Importance of sample attitudes

The differences in attitudes among the samples are important if the data are to be used to guide conservation efforts. Clearing of koala habitat for urban or regional development is a significant threat to the viability of the koala population in the study area. *Landowners* were more likely to map spatial preferences in support of residential and tourism development than the *volunteer* or *panel* samples. These types of development run counter to koala conservation. A sampling design that is exclusively rural *landowners* may be biased toward more residential and tourism development, while an exclusive *volunteer* sample may be biased toward conservation action that limits residential and tourism development, particularly if a significant proportion of the *volunteers* are members of a

conservation organization such as “Friends of the Koala” and potentially hostile to the interests of landowners. The *panel* sample may most closely approximate general public sentiment toward koala conservation activities, but this supposition cannot be confirmed or rejected without representative sampling of the general population.

4.4 | Combining sample responses to increase inferential power

An important question for geographic citizen science is whether spatial data from multiple sample groups can be combined to increase the inferential power of the spatial data. For koala observation data, the responses of samples can be aggregated because there was no difference in locational accuracy. For preference data, the question of sample aggregation is less clear. For most spatial preferences, there was little difference in the propensity for individuals from different samples to map preferences. And yet there were a few important preference differences for *wanting more koalas* and for *residential* and *tourism* development. In identifying the location of koala conservation opportunities, our judgment is to keep the preference data separated by group for analysis and reporting.

A missing sampling component in this project was a random household sample of residents in the study region. A random household sample would be a useful addition to the sampling design to increase data completeness and increase the geographic representativeness of participants in the study region. And yet, the response rate would likely be low (i.e., around 10%) and the sample would likely be biased toward older, more formally educated male participants with higher incomes (Brown & Kyttä, 2014).

Based on our findings, the sampling design for geographic citizen science to inform conservation policy should be differentiated. For citizen science, data completeness for geographic coverage appears most important, while general population and stakeholder representativeness appears more important for participatory mapping intended for land use planning and decision support. With careful design, sampling for citizen science and planning-driven participatory mapping can be combined in the same project. However, there are trade-offs when including both observation and preference attributes in the same spatial survey. Including both elements increases the complexity of the survey for participants and, given that participant mapping effort appears finite regardless of the number of markers, response data would be spread out over more markers rather than focused on key variables. On the contrary, including both elements allows participants to contribute preferences about koala conservation even if the participant has not observed many koalas. A further rationale for including both elements is that the relationship between the presence of the subject wildlife species and potential actions that support or reduce its viability is made explicit in the mapping process, thus enhancing public education about conservation efforts.

4.5 | Approach and limitations

The use of citizen science for koala surveys has been a successful venture in Australia, and NSW in particular (Harris & Goldingay, 2003; Lunney et al., 2009, 2016; Lunney, Stalenberg, Santika, & Rhodes, 2017; Predavec et al., 2016, 2018; Santika, McAlpine, Lunney, Wilson, & Rhodes, 2014). These studies have been able to identify trends in koala populations over time, reveal spatial preferences for planning, and present options for management that were not possible without the spatial reach and wider public participation where citizens share their knowledge and contribute to the koala conservation debate. With geographic citizen science, there are limitations. The first is spatial: with a clustering of most of the population near the coast, the hinterland is less well represented. The hinterland supports more landowners with 20 acres or more land. Without explicit sampling for *landowners*, this important perspective would be missed. For example, of all threats, *landowners* perceived the logging of native forests to be the least and wild dogs to be the greatest. Dogs, rather than the conspicuous koala roadkill, were found to be the silent killers during a 3-year radio-tracking study of koalas in the NSW coastal shire of Port Stephens (Lunney, Gresser, O'Neill, Matthews, & Rhodes, 2007). This finding lends support to the *landowners'* view. The *panel* perceived bushfire to be a significantly greater threat than either *volunteers* or *landowners*, but the

Port Stephens koala population recovered quickly from a major fire, while being susceptible to prescribed, low-intensity fire used as a fire management technique.

A second limitation is the proportion of state forests in the shires, and their location. This is a weak zone for citizen science, including *volunteer* sampling, as there are simply too few citizens spending extended amounts of time in forests, and people are more often there by day and miss the night-time movements of koalas and the bellowing calls of adult males in spring (Law et al., 2018). In this regard, well-designed field surveys are most important. A third limitation is that caution must be taken when interpreting key threats identified by citizens which can potentially mislead managers to address visible threats while missing primary causes such as land clearing. The key result of habitat loss is a reduction in population size, but the public manifestation of the loss and fragmentation of habitat is a sick or injured koala, a roadkill, or a dog attack. Notwithstanding, there is sufficient capacity for ecologists and planners to see the connections and manage the land and the planning provisions accordingly.

Finally, we are well aware that the koala, being such an iconic species, is readily co-opted to a cause and geographic citizen science may, intentionally or not, play a political role. One of the local causes is an anti-logging stance, both logging on private native forests and state forests (i.e., commercial forests owned and managed by the state, in this case NSW). We are also aware of the political divisions in the communities on the far north coast of NSW, with the major political parties, and the Green Party, all vying for local votes in these marginal electorates. Unsurprisingly, the koala is being used as the flagship species to prevent logging, restore habitat, or slow vehicles on shire roads. Given such wide social divisions, the concept of the “general public” may not be so relevant in the far north coast of NSW. Sampling in this project likely identified the competing interest groups, the relative size of which is changing with the years—less logging, more conservation—but with solid support remaining for a sustained agricultural industry.

5 | CONCLUSIONS

Previous citizen science wildlife research has not examined data quality from multiple samples having different origins. In this study, three sample groups produced spatially accurate observation data despite originating from different sources (*volunteer* recruitment, *Internet panel*, or *landowner* interviews). A second important finding was that the sample groups differed in some mapped preferences for land use, confirming previous findings that participant characteristics (e.g., experience, knowledge, livelihood, values, and attitudes) can influence subjective spatial preferences that are linked to land use decisions. Conservation strategies and actions typically require commitment of public resources and support and therefore, obtaining participant representativeness of the general population is important for assessing the social acceptability and political viability of potential conservation actions.

A key conclusion from our assessment of sampling effects on data quality in geographic citizen science is that sampling design plays an important role in implementing and interpreting citizen science studies of wildlife. Sampling for wildlife distribution and sampling for the acceptability of various land uses have different, but related, purposes. Whereas citizen observations seek to augment scientific assessments of wildlife populations, identifying land use preferences seeks to identify conservation opportunities and constraints that need to be reconciled with wildlife habitat. If these two research purposes are included in the same geographic citizen science project, the sampling design should reflect the different criteria used to assess data quality. A sound sampling design for observation data would target a large, geographically distributed population within the study area, while sampling for conservation opportunity would target study area population subgroups that are representative and important in providing evidence of public support for conservation action.

These study findings serve to inform future geographic citizen science methods wherein: (a) sample size, not sample group per se, is most important to data quality for observational data; and (b) sample group based on population representativeness is important for data quality when mapping subjective spatial attributes such as

land use preferences. Both types of data—the location of koalas and preferences for land use—are important to effective koala conservation outcomes.

From a socio-political perspective, our study findings are consistent with the way that an iconic species is viewed in a community of diverse interests where the species becomes a visible symbol of the success (or failure) of government efforts at all levels (local, state, and commonwealth) to be proactive in addressing threats. The approach adopted here, namely identifying sampling effects, enables ecologists and planners to acknowledge the values and knowledge of different segments of the general population to tailor a conservation approach that incorporates the full range of social perspectives and land tenures that provide a greater chance for the effectiveness of any plan or action. We advocate incorporating an assessment of sampling effects on data quality in future geographic citizen science projects.

ORCID

Greg Brown  <https://orcid.org/0000-0003-3283-4619>

REFERENCES

- Adams-Hosking, C., McBride, M. F., Baxter, G., Burgman, M., de Villiers, D., Kavanagh, R., ... McAlpine, C. A. (2016). Use of expert knowledge to elicit population trends for the koala (*Phascolarctos cinereus*). *Diversity & Distributions*, 22(3), 249–262.
- Antoniou, V., & Skopeliti, A. (2015). Measures and indicators of VGI quality: An overview. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2, 345.
- Australian Bureau of Statistics. (2016). 2016 Census. QuickStats. Retrieved from <http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20QuickStats>
- Brown, G. (2013). Relationships between spatial and non-spatial preferences and place-based values in national forests. *Applied Geography*, 44, 1–11.
- Brown, G. (2017). A review of sampling effects and response bias in internet participatory mapping (PPGIS/PGIS/VGI). *Transactions in GIS*, 21(1), 39–56.
- Brown, G., Kelly, M., & Whittall, D. (2014). Which “public”? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. *Journal of Environmental Planning & Management*, 57(2), 190–214.
- Brown, G., & Kytä, M. (2014). Key issues and research priorities for public participation GIS (PPGIS): A synthesis based on empirical research. *Applied Geography*, 46, 122–136.
- Brown, G., & Kytä, M. (2018). Key issues and priorities in participatory mapping: Toward integration or increased specialization? *Applied Geography*, 95, 1–8.
- Brown, G., McAlpine, C., Rhodes, J., Lunney, D., Goldingay, R., Fielding, K., ... Vass, L. (2018). Assessing the validity of crowdsourced wildlife observations for conservation using public participatory mapping methods. *Biological Conservation*, 227, 141–151.
- Brown, G., Strickland-Munro, J., Kobryn, H., & Moore, S. (2016). Stakeholder analysis for marine conservation planning using public participation GIS. *Applied Geography*, 67, 77–93.
- Brown, G., Weber, D., Zanon, D., & de Bie, K. (2012). Evaluation of an online (opt-in) panel for public participation geographic information systems (PPGIS) surveys. *International Journal of Public Opinion Research*, 24(4), 534–545.
- Bubalo, M., van Zanten, B. T., & Verburg, P. H. (2019). Crowdsourcing geo-information on landscape perceptions and preferences: A review. *Landscape & Urban Planning*, 184, 101–111.
- Fitz-Gibbon, C., & Morris, L. (1987). *How to analyze data*. Newbury Park, CA: Sage.
- Girres, J., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS*, 12(4), 435–459.
- Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and ordinance survey datasets. *Environmental Planning B*, 37, 682–703.
- Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge* (pp. 105–122). Berlin, Germany: Springer.
- Harris, J. M., & Goldingay, R. L. (2003). A community-based survey of the koala *Phascolarctos cinereus* in the Lismore region of northeastern New South Wales. *Australian Mammalogy*, 25(2), 155–167.

- Hollow, B., Roetman, P. E., Walter, M., & Daniels, C. B. (2015). Citizen science for policy development: The case of koala management in South Australia. *Environmental Science & Policy*, 47, 126–136.
- Law, B. S., Brassil, T., Gonsalves, L., Roe, P., Trusking, A., & McConville, A. (2018). Passive acoustics and sound recognition provide new insights on status and resilience of an iconic endangered marsupial (koala *Phascolarctos cinereus*) to timber harvesting. *PLoS One*, 13(10), e0205075.
- Lunney, D., Crowther, M. S., Shannon, I., & Bryant, J. V. (2009). Combining a map-based public survey with an estimation of site occupancy to determine the recent and changing distribution of the koala in New South Wales. *Wildlife Research*, 36(3), 262–273.
- Lunney, D., Gresser, S., O'Neill, L. E., Matthews, A., & Rhodes, J. (2007). The impact of fire and dogs on koalas at Port Stephens, New South Wales, using population viability analysis. *Pacific Conservation Biology*, 13(3), 189–201.
- Lunney, D., O'Neill, L., Matthews, A., & Coburn, D. (2000). Contribution of community knowledge of vertebrate fauna to management and planning. *Ecological Management & Restoration*, 1, 175–184.
- Lunney, D., Predavec, M., Miller, I., Shannon, I., Fisher, M., Moon, C., ... Rhodes, J. R. (2016). Interpreting patterns of population change in koalas from long-term datasets in Coffs Harbour on the north coast of New South Wales. *Australian Mammalogy*, 38(1), 29–43.
- Lunney, D., Stalenberg, E., Santika, T., & Rhodes, J. (2017). A rebuttal to 'Mooted extinction of koalas at Eden. Improving the information base' – A misleading, even dangerous, polemic. *Wildlife Research*, 44, 453–457.
- McAlpine, C., Lunney, D., Melzer, A., Menkhorst, P., Phillips, S., Phalen, D., ... Close, R. (2015). Conserving koalas: A review of the contrasting regional trends, outlooks and policy challenges. *Biological Conservation*, 192, 226–236.
- Predavec, M., Lunney, D., Hope, B., Stalenberg, E., Shannon, I., Crowther, M. S., & Miller, I. (2016). The contribution of community wisdom to conservation ecology. *Conservation Biology*, 30, 496–505.
- Predavec, M., Lunney, D., Shannon, I., Lemon, J., Sonawane, I., & Crowther, M. (2018). Using repeat citizen science surveys of koalas to assess their population trend in the north-west of New South Wales: Scale matters. *Australian Mammalogy*, 40(1), 47–57.
- Predavec, M., Lunney, D., Shannon, I., Scotts, D., Turbill, J., & Faulkner, B. (2015). Mapping the likelihood of koalas across New South Wales for use in private native forestry: Developing a simple, species distribution model that deals with opportunistic data. *Australian Mammalogy*, 37, 182–193. (An updated version available at <http://www.epa.nsw.gov.au/your-environment/native539forestry/mapping-research/koala-mapping-program/mapping-koala-occurrence> was used in this study.)
- Reed, P., Lunney, D., & Walker, P. (1990). Survey of the koala *Phascolarctos cinereus* (Goldfuss) in New South Wales (1986–87), with an ecological interpretation of its distribution. In A. K. Lee, K. A. Handasyde, & G. D. Sanson (Eds.), *Biology of the koala* (pp. 55–74). Chipping Norton, NSW: Surrey Beatty & Sons.
- Robertson, C., & Feick, R. (2018). Inference and analysis across spatial supports in the big data era: Uncertain point observations and geographic contexts. *Transactions in GIS*, 22(2), 455–476.
- Santika, T., McAlpine, C. A., Lunney, D., Wilson, K. A., & Rhodes, J. R. (2014). Modelling species distributional shifts across broad spatial extents by linking dynamic occupancy models with public-based surveys. *Diversity & Distributions*, 20(7), 786–796.
- Spielman, S. E. (2014). Spatial collective intelligence? Credibility, accuracy, and volunteered geographic information. *Cartography & Geographic Information Science*, 41(2), 115–124.
- Zolkafli, A., Brown, G., & Liu, Y. (2017). An evaluation of the capacity-building effects of participatory GIS (PGIS) for public participation in land use planning. *Planning Practice & Research*, 32(4), 385–401.

How to cite this article: Brown G, Rhodes J, Lunney D, et al. The influence of sampling design on spatial data quality in a geographic citizen science project. *Transactions in GIS*. 2019;00:1–20. <https://doi.org/10.1111/tgis.12568>