# Turning observations into biodiversity data: Broadscale spatial biases in community science

Ellyne M. Geurts, John D. Reynolds, and Brian M. Starzomski

2023

Received: 17 December 2022 | Revised: 16 March 2023 | Accepted: 23 March 2023

DOI: 10.1002/ecs2.4582

ECOSPHERE
AN ESA OPEN ACCESS JOURNAL

ARTICLE

Methods, Tools, and Technologies

# Turning observations into biodiversity data: Broadscale spatial biases in community science

**Ellyne M. Geurts**[1] | **John D. Reynolds**[2] | **Brian M. Starzomski**[1]

[1]School of Environmental Studies, University of Victoria, Victoria, British Columbia, Canada

[2]Earth to Ocean Research Group, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada

**Correspondence**
Brian M. Starzomski
Email: starzom@uvic.ca

**Funding information**
BC Parks; Ministry of Land, Water and Resource Stewardship; Natural Sciences and Engineering Research Council of Canada; Pacific Wildlife Foundation; Sitka Foundation

**Handling Editor:** John Humphreys

## Abstract

Biodiversity community science projects are growing rapidly in popularity. The enormous amounts of data generated by these programs are transforming how we conduct ecological research and conservation management. However, as with other biodiversity surveys, community science datasets suffer from biases in time and locations of observations. To better use these data, we modeled the spatial biases present in the popular community science platform, iNaturalist. iNaturalist uses crowdsourcing to collect georeferenced and time-stamped observations of all taxa worldwide. With its wealth of biodiversity data, iNaturalist is now being used to answer a broad range of questions in ecology and conservation, but little is known about the platform's spatial biases. We focus on the more than 1.75 million iNaturalist observations available (as of December 2021) from British Columbia, Canada, a region with a strong community science presence and diversity of ecosystems. Using machine learning and species distribution modeling, we examined which landscape factors (e.g., protected areas, roads, human population density, habitat zones, elevation) were most important in determining where observations are taken, and we created a predicted probability map revealing how likely different regions are to be sampled by community scientists. We found strong road biases for observations in iNaturalist, with over 94% of observations within 1 km of roads. In addition, human population density and broad habitat ecosystem zones played a large role in predicting where iNaturalist observations occur across the landscape. These methods demonstrate tools for modeling the effects of spatial biases in large opportunistic datasets that can then be used to produce more accurate species distribution and biodiversity models from community science data.

**KEYWORDS**

citizen science, community science, geographical bias, iNaturalist, Maxent, opportunistic data, presence-only, sampling bias, species distribution modeling, unstructured data

## INTRODUCTION

The use of community science to collect data on biodiversity is growing rapidly with advances in technology and online platforms like eBird and iNaturalist (Loarie, 2020; Miller-Rushing et al., 2012; Pocock et al., 2018; Sullivan et al., 2009). These community science (also known as citizen science) platforms range from user-driven opportunistic data collections (e.g., iNaturalist) to standardized surveys with volunteers helping collect specific data (e.g., Breeding Bird Surveys). Community science platforms are producing massive amounts of biodiversity and ecological data across large geographical and temporal scales (Loarie, 2022; Pocock et al., 2018; Zhang, 2020). These data are being used to answer questions of phenology (Barve et al., 2020; Nowak et al., 2020), species distributions (Johnston et al., 2021), population trends (Neate-Clegg et al., 2020), phenotypic variation (Drury et al., 2019; Lehtinen et al., 2020), health (Hamilton et al., 2021), and species interactions (Saldivar & Romero, 2022). In addition, these community science platforms are contributing to the discovery of new species and monitoring and management of exotic and rare species (Hausdorf et al., 2021; Jain et al., 2022; Roberts et al., 2022; Werenkraut et al., 2020). The contribution of these datasets to our biodiversity knowledge hinges on whether we understand a variety of biases, especially related to the spatial distribution of observers (Brown & Williams, 2019; Dickinson et al., 2010; Isaac et al., 2014; Johnston et al., 2018).

Spatial biases, taxonomic biases, and variability in observer sampling effort are common limitations in community science biodiversity datasets. Observations are often concentrated in regions with high human population densities (Ballesteros-Mejia et al., 2013; Ruete, 2015; Speed et al., 2018) and in areas that are easily accessible such as roads and tourist locations (Kadmon et al., 2004; Oliveira et al., 2016). The data often also exhibit taxonomic biases, favoring large charismatic taxa such as birds and mammals over cryptic taxa like spiders (Isaac & Pocock, 2015; Troudet et al., 2017). Furthermore, variability in sampling effort is common in crowdsourced community science, where there are little to no sampling guidelines, resulting in large variability in distances surveyed, duration of surveys, and intensity of observations, which may (e.g., eBird) or may not (e.g., iNaturalist) be accounted for (Isaac et al., 2014; Ruete, 2015). Note that biases exist in all datasets, even professional surveys (Kosmala et al., 2016), and there are methods to explicitly account for spatial biases in species distributions (Fithian et al., 2015; Zizka et al., 2020), observer variability in species occupancy-detection and biodiversity modeling (Isaac et al., 2014; Johnston et al., 2018; Kelling et al., 2015; Meyer et al., 2016), and temporal biases in phenology studies (Courter et al., 2013). However, these methods require knowledge of the extent and strength of the biases and the factors affecting them before use.

Despite the importance of understanding the extent and strength of biases in community science diversity datasets, many projects have limited information on biases and errors, with the exception of bird-focused programs, such as eBird and breeding bird surveys (La Sorte & Somveille, 2020; van Wilgenburg et al., 2015; Zhang, 2020). The popular platform iNaturalist has the largest number of participants with over 2.6 million users and the broadest taxonomic coverage in the world with more than 135 million observations (Callaghan et al., 2020; iNaturalist, 2023). iNaturalist also displays strong evidence of spatial biases and observer variability due to its opportunistic data collection (Di Cecco et al., 2021; iNaturalist, 2023; Loarie, 2020). Current research has only begun to scratch the surface of the biases within iNaturalist data (Callaghan et al., 2020; Di Cecco et al., 2021; Mesaglio & Callaghan, 2021). Studies so far have focused on describing the broad taxonomic, temporal, and spatial biases within iNaturalist such as taxonomic specialization of iNaturalist observers, weekend temporal biases, and spatial bias toward developed land (Di Cecco et al., 2021; Mesaglio & Callaghan, 2021). However, there is currently no example framework available for investigators interested in visualizing and modeling the spatial sampling biases on iNaturalist to show where predicted community science activity will be high versus low. In addition, we do not know which landscape features are the most influential for where iNaturalist users make observations and how these are related to the probability of an observation being made at a particular location.

Our study addresses this knowledge gap on biases in community science diversity datasets by modeling spatial biases in the iNaturalist database using the large and geographically diverse province of British Columbia (BC), Canada as a case study. We use Maxent, a popular machine learning software that produces species distribution models with presence-only data (Phillips et al., 2020). The software can also be used to model sampling effort (i.e., spatial sampling biases) and test which variables such as habitats and distance to roads influence observer behavior (Barber et al., 2022; Merow et al., 2013). We developed a workflow that could be applied anywhere to: (1) model spatial biases of iNaturalist observations across BC and predict where observations are likely to occur across the landscape; and (2) determine the strength and direction of relationships between environmental variables and probability of iNaturalist observations. Given the opportunistic nature of iNaturalist observations and previous studies of community science

projects, we predicted that distance to roads and human population density will be the most important environmental variables biasing the distribution of observations, with land cover type such as urban and agricultural regions and tourist locations (e.g., parks) having a lesser effect. We predicted an exponential negative relationship between distance to roads and probability of observation, and a positive linear relationship between human population density and probability of observation.

# METHODS

## Background: iNaturalist and study area

iNaturalist is a social network platform where users upload their own georeferenced and time-stamped photos and audio recordings of organisms for community identification (iNaturalist, 2023). The platform is designed for users of all skill levels with the primary goals of education and connecting people with nature. iNaturalist has no sampling guidelines; people select what, when, and where they want to observe nature. While this lack of sampling guidelines has some distinct advantages, it leads to large variability in sampling effort and spatial, temporal, and taxonomic coverage of observations (Di Cecco et al., 2021). As a result, observations on iNaturalist are considered presence-only data, which require specialized statistics for analysis (Dickinson et al., 2010).

BC is an excellent study region with its wide range of habitats and population densities (Environmental Reporting BC, 2018; Meidinger & Pojar, 1991), strong iNaturalist community (iNaturalist, 2023), and publicly available fine-resolution spatial data (BC Data Catalogue: https://catalogue.data.gov.bc.ca/). The BC iNaturalist database currently contains over 2 million georeferenced observations across the taxonomic spectrum from 40,000 observers from 1937 to 2022, with the number of observations growing exponentially (iNaturalist, 2023).

## Study datasets

We downloaded iNaturalist observations ($n = 1{,}769{,}501$) directly from the iNaturalist website (https://www.inaturalist.org/) on December 2, 2021. We selected distance to roads as a landscape feature that we expected to influence the spatial coverage of iNaturalist observations (Kadmon et al., 2004; Reddy & Dávalos, 2003; Stolar & Nielsen, 2015; Tye et al., 2017). We included provincial and national parks because many parks are popular tourist spots in BC (BC Parks, 2018) and tourist spots can bias number of community science records in a region (Boakes et al., 2010). In addition, there are increasing initiatives by government agencies and nonprofit societies encouraging the use of the iNaturalist platform in parks in BC (BC iNaturalist Program, 2021; Parks Canada, 2022; Strathcona Wilderness Institute, 2022). We expected human population density to be related to the spatial distribution of iNaturalist observations (Ballesteros-Mejia et al., 2013; Ruete, 2015; Speed et al., 2018). We selected the fine-scale land cover type (e.g., cropland, urban, and mixed forest) from MODIS (International Geosphere-Biosphere Programme [IGBP] global vegetation classification scheme; Appendix S1: Table S1) because land cover type can cause spatial bias in volunteer community science (Di Cecco et al., 2021; Geldmann et al., 2016; Petersen et al., 2021). We also included data from the broader scale Biogeoclimatic Ecosystem Classification (BEC) system, used in BC to classify landscape-level ecosystems (Geldmann et al., 2016; Meidinger & Pojar, 1991; Tulloch & Szabo, 2012). We also analyzed elevation as mountainous regions are less accessible than lower elevation sites, which likely causes spatial bias in observations (Fernández & Nakamura, 2015; Mair & Ruete, 2016). See Table S2 in Appendix S1 for further information on the spatial datasets.

## Spatial data preparation

We removed marine observations ($n = 141{,}147$), resulting in 1,628,354 terrestrial observations for analysis. We further refined the data for Maxent analysis by spatially filtering the observations following Kass et al. (2022) where we retained only one observation per grid cell (277-m resolution) to reduce spatial autocorrelation. The final cleaned occurrence dataset contained 152,785 terrestrial observations. We cleaned and processed environmental spatial layers to ensure identical spatial projection, cell size, extent, and origin. We clipped the roads dataset to the BC terrestrial boundary to remove boat routes and Yukon highways. We created a Euclidean distance to the road raster layer with a cell size of 25 m. We rasterized the national and provincial park polygon layers, converted them to binary surfaces, and combined them to create a raster layer of parkland versus non-parkland. We took the log of human population density following Mair and Ruete (2016) and Barber et al. (2022) to better examine the effect of population density changes at very low densities. We projected all environmental layers in BC Albers projection, cropped and masked using the BC terrestrial polygon, then resampled to cell size of 277 m to match the coarsest raster layer (MODIS land cover). This resolution is comparable or even finer than other studies of spatial sampling bias in

community science (El-Gabbas & Dormann, 2018; Mair & Ruete, 2016; Stolar & Nielsen, 2015). We used bilinear resampling for continuous raster layers (distance to road, human population size, and elevation). We conducted spatial data preparations in R (version 4.1.3), RStudio (R Core Team, 2021; RStudio Team, 2021), and ArcMap 10.6.1 (Redlands, 2017). We used the following R packages: bcmaps (Teucher et al., 2021), MODIStsp (Busetto & Ranghetti, 2016), sf (Pebesma, 2018), dplyr (Wickham et al., 2021), raster (Hijmans, 2021a), fasterize (Ross, 2020), and Terra (Hijmans, 2021b).

## Statistical analyses

We quantified the number of terrestrial iNaturalist observations near roads using an empirical cumulative distribution function, which we compared with a random null model. We selected 1 million random points across BC, extracted Euclidean distance to road for each random spatial point, then took the mean of the million random points. We bootstrapped those million distances 10,000 times. Mean and SE were measured for each bootstrap sample. We then selected 1 million data points randomly from the observed terrestrial iNaturalist observations and took bootstrapped samples in the same manner. We chose a sample size that was similar to the observed number of observations in BC. We compared the observed mean distances between the two distributions using a Welch two-sample *t* test. We conducted these analyses in RStudio using the stats, sf, raster, and dplyr packages (Hijmans, 2021a; Pebesma, 2018; R Core Team, 2021; RStudio Team, 2021; Wickham, 2016; Wickham et al., 2021).

We used the species distribution modeling software, Maxent, via the ENMeval and dismo packages to investigate which environmental variables are strong predictors of where iNaturalist observations are made across the province (Hijmans et al., 2021; Kass et al., 2021; Phillips et al., 2020). We selected Maxent because it handles presence-only data (Elith et al., 2011), is widely used (Fourcade et al., 2014), and is ranked as one of the top species distribution models for presence-only data for predictiveness (Valavi et al., 2022). In addition to modeling species distributions, Maxent can also be used to create a bias file of sampling effort that can then be fed into a species distribution model to correct for sampling and geographical bias (Barber et al., 2022; Elith et al., 2011; Phillips et al., 2020). Appendix S1 shows how a bias file can be incorporated in the Maxent GUI to account for spatial sampling bias (Appendix S1: Figure S1). Creating a bias file is usually based on target group sampling, where occurrence data of species within the same taxonomic category of the focal species with similar sampling biases are pooled together and modeled with different covariates related to observer behavior, for example, distance to roads and urban centers (Merow et al., 2013; Phillips et al., 2009). We adapted Maxent to model the distribution of iNaturalist observers across BC. People who make an observation are thus our "species." We pooled all terrestrial observations to model the probability of occurrence of an observation (i.e., sampling spatial bias) across the province and to determine which environmental variables best explain where observations are made (Elith et al., 2011; Phillips, 2017). We selected the R package ENMeval to run Maxent as it allows multiple tuning parameters to be tested at once, produces reproducible code, provides metrics such as corrected Akaike information criterion ($AIC_c$) to allow comparison of different maxent models, and has the function to test Maxent models against a null model (Kass et al., 2021). The null model analysis for species distribution modeling is based on Bohl et al. (2019).

We conducted a Pearson correlation matrix analysis for continuous raster layers using the ENMTools package (Warren & Dinnage, 2022) and Cramer's *V* for similarity association measurements of categorical layers using the rcompanion R package (Mangiafico, 2022) to ensure the environmental variables were not highly correlated (correlation metric <0.50) (Fourcade et al., 2014; Merow et al., 2013). We produced similarity matrices using Schoener's *D* (i.e., niche overlap) of the predicted Maxent values among the different Maxent models in geographic space (Kass et al., 2022; Schoener, 1968). The code and workflow we used can be found on Zenodo (https://doi.org/10.5281/zenodo.7710337).

## Maxent settings

We ran Maxent models with the following tuning arguments and settings in ENMeval. We used the Randomkfold partition method with $k = 5$ and the algorithm argument = "maxent.jar" following the default Maxent GUI settings (Kass et al., 2021). We used 300,000 background points and made the model randomly select the background points from across BC (Kass et al., 2021; Phillips, 2017; Valavi et al., 2022). We did not restrict background area to buffer zones around presence points because we are interested in making inferences for the entire province (Fourcade et al., 2014; Kass et al., 2021). Maxent removed 4664 occurrence points with NA predictor variable values. We tested regularization multiplier values from 0.5 to 2 by increments of 0.5. The feature classes we included were linear (L), quadratic (Q), hinge (H), and product (P), with the following combinations tested: L, Q, H, P, LQ, LQH, LQP, and LQHP. We tested 32 different Maxent models. A prediction map for the top

model was produced using the cloglog output format. The top Maxent model was selected using $AIC_c$ and area under the curve (AUC) validation (Kass et al., 2021). It was then compared with a null model using the "ENMnulls" function from ENMeval package with 100 iterations implemented (Kass et al., 2021). We used the tuning arguments from the top model as inputs for the Maxent GUI to produce the response curves and variable importance graphs (Phillips, 2017).

We examined variable importance of the top maxent model using percent contribution, permutation importance, and a jackknife test of regularized training gain (Phillips, 2017). We analyzed the relationships of the environmental variables with the Maxent predicted probability of iNaturalist observations using marginal and isolated variable response curves. Marginal response curves show the relationship between predicted probability of observation with an environmental variable across its range while all other variables are held at their average sample value. Isolated variable response curves consider only one variable at a time.

# RESULTS

## Where are iNaturalist observations likely to occur?

The highest probabilities of observations in BC are in the south and along highways (Figure 1). The lowest predicted probabilities are in the north, particularly in the northwest (Figure 1). These probabilities are based on the top Maxent model in terms of both $AIC_c$ and AUC validation values ($AUC_{val} = 0.906$, $\Delta AIC_c = 0$, number of coefficients = 176; Appendix S1: Table S3). This was also the most complex model with all four feature classes included (=linear, quadratic, hinge, and product) and a regularization multiplier value of 0.5 (i.e., low penalty toward complexity). There was high "niche overlap" among the 32 Maxent model predictions, suggesting the different Maxent settings produced similar maps. The similarity matrices using Schoener's $D$ of the predicted Maxent model values in geographic space ranged from $D = 0.818$ to $D = 0.998$. Lastly, the top Maxent model was significantly different from the null model using the AUC validation metric ($Z = 15.25$, $p < 0.00001$; Appendix S1: Table S4 and Figure S2).

## Which environmental features best predict where iNaturalist observations occur?

For the top Maxent model, distance to road was the most influential variable with a permutation importance value of 51.6%, while BEC was the second most important variable at 28.7% (Table 1). Human population density was much less important (permutation importance value 9.2%), though its percent contribution was similar to distance to roads (34%; Table 1). Since percent contribution values are pathway (i.e., algorithm) dependent and permutation importance values are derived from the final

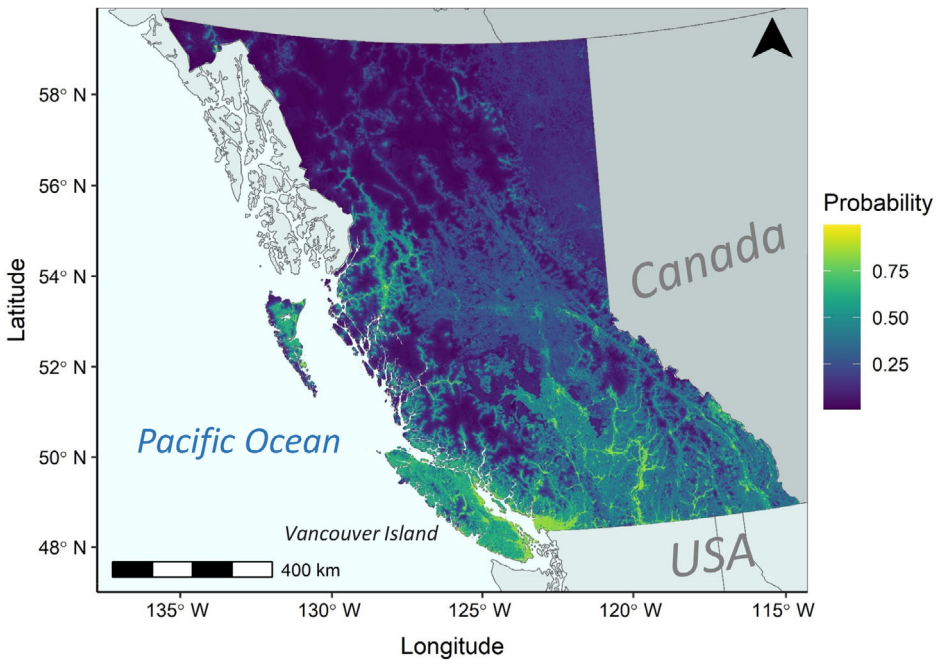



**FIGURE 1** Maxent predicted probability of presence for iNaturalist observers making observations in British Columbia, Canada. Predictions are based on the cloglog output format. Cell resolution = 277 m.

**TABLE 1** Two measures of variable importance for the top Maxent model (feature classes = LQHP, regularization multiplier = 0.5) selected by the ENMeval R package for the distribution of iNaturalist observations in British Columbia, Canada.

| Variable | Permutation importance (%) | Percent contribution (%) |
|---|---|---|
| Distance to roads (m) | 51.6 | 34.2 |
| Biogeoclimatic zones | 28.7 | 19.3 |
| Human population density | 9.2 | 34.1 |
| Park versus non-parkland | 7.0 | 7.2 |
| Elevation (m) | 2.3 | 4.6 |
| Land cover type | 1.2 | 0.6 |

*Note*: Permutation importance values are derived from the final Maxent model. Percent contribution values are algorithm (i.e., Maxent) dependent.

Maxent model (Phillips, 2017), this drop in relative importance suggests that population density was important for the Maxent algorithm to create the model, but ultimately distance to roads and BEC play a larger role for predicting where iNaturalist observations occur (Table 1). Land cover type was the least important variable with permutation importance of 1.2% and percent contribution of 0.6% (Table 1). See Appendix S1 for jackknife test results of variable importance (Appendix S1: Figure S3).

The strong bias of observations toward roads can be seen in Figure 2. A total of 75% of observations were within 160 m and 94% of observations were within 1 km of roads (Figure 2). The mean distance from roads was 309 ± 0.01 m (SE) from roads, whereas random points were 5277 ± 0.09 m from roads (Welch two-sample $t$ test: $t = 55{,}695$, df = 10,422, $p$ value <0.001; Figure 2c,d).

### How do the environment variables affect predicted probability of iNaturalist observation?

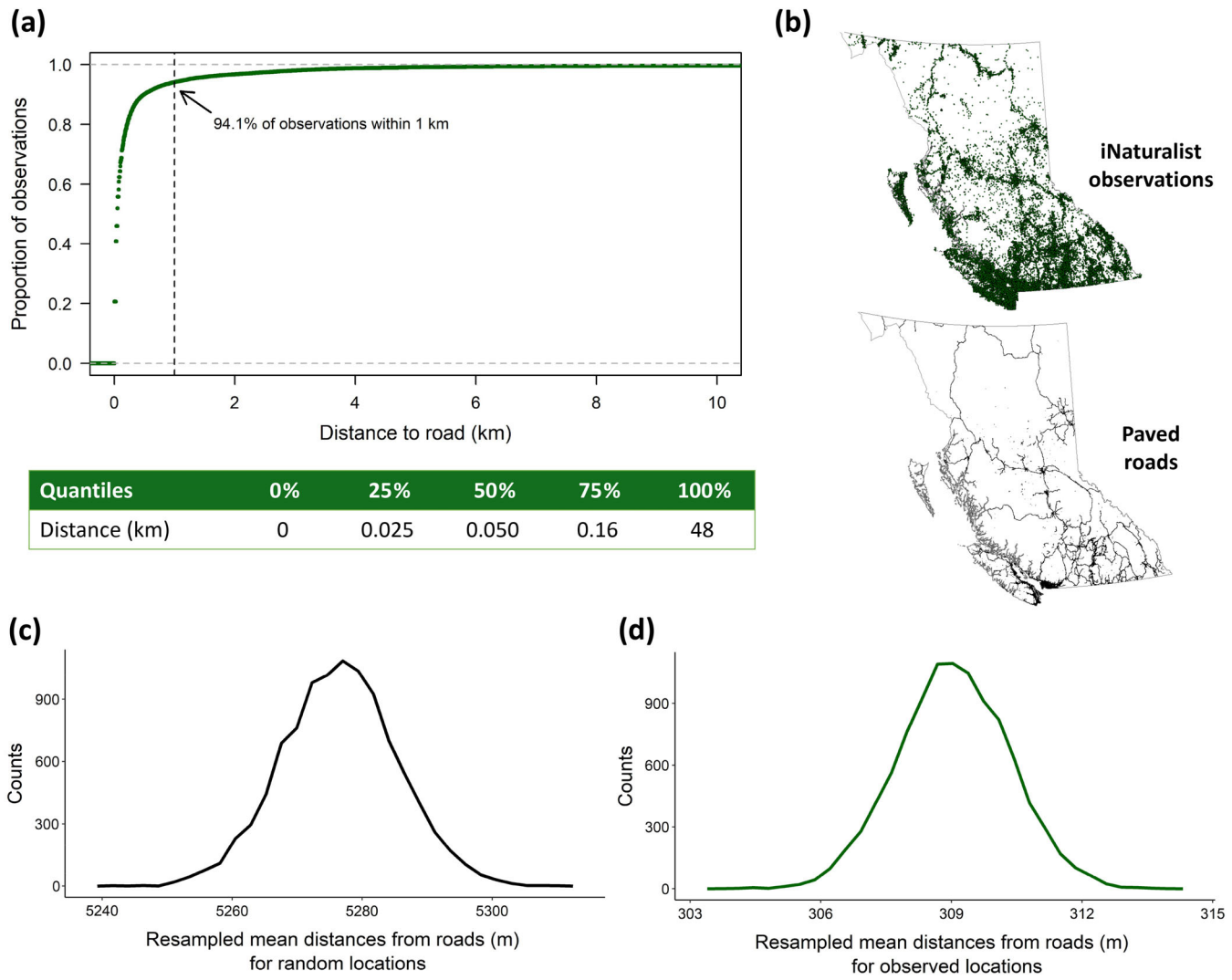
Marginal response curves provide another way of assessing and visualizing the roles of individual predictor variables in biasing the locations of observations by holding all other predictors at their average sample value. As expected, there was higher predicted probability of an iNaturalist observation for locations close to roads (Figure 3). When controlling for other variables, for example, human population density, the biogeoclimatic zones with the highest predicted probability of observation were the Coastal Mountain-heather Alpine and Mountain Hemlock zones, and the lowest probabilities were the Boreal White and Black Spruce and Sub-Boreal Pine–Spruce zones (Figure 3). When the other variables are not controlled for, the highest predicted BEC zones are Coastal Douglas-Fir and Ponderosa Pine (Appendix S1: Figure S4). As predicted, there was a positive relationship between predicted probability of observation and human population density, and a higher predicted probability of iNaturalist observations within parks than outside parks (Figure 3). See Appendix S1 for marginal and isolated variable response curves for all six environmental variables (Appendix S1: Figure S4).

## DISCUSSION

Our results provide a workflow to visualize the spatial sampling biases present within iNaturalist, the world's largest community science biodiversity platform. Using this method, we were able to determine the important environmental drivers behind those spatial biases. Our workflow demonstrates a method to identify spatial sampling bias in presence-only data (e.g., iNaturalist) that could then be subsequently incorporated into species distribution models using these data, as well as other community science summary analyses (Geldmann et al., 2016). As predicted, distance to roads played the largest part in influencing where people make iNaturalist observations (Table 1). Unexpectedly, the broad habitat variable (i.e., BEC) was more important than human population density for predicting where iNaturalist observations occur (Table 1).

These results align with other studies examining variable importance in presence-only datasets (El-Gabbas & Dormann, 2018; Geldmann et al., 2016; Mair & Ruete, 2016). Geldmann et al. (2016) also found distance to roads, population density, and land cover type (e.g., urban) are important factors influencing spatial biases in four different community science projects in Denmark; however, they did not look at any additional variables. Mair and Ruete (2016) found road access and population density were consistently the most important variables in the Swedish LifeWatch platform. Lastly, El-Gabbas and Dormann (2018) found accessibility covariates (e.g., distances to roads, cities, and protected areas) better accounted for spatial biases than other environmental and effort variables for opportunistically collected data on bats in Egypt.

Although elevation did not appear to have a large influence on where iNaturalist observations occur (Table 1), this may be because distance to road was the better measure of accessibility. Mair and Ruete (2016) also found that roads were more important than




| Quantiles | 0% | 25% | 50% | 75% | 100% |
|---|---|---|---|---|---|
| Distance (km) | 0 | 0.025 | 0.050 | 0.16 | 48 |

**FIGURE 2** (a) Empirical cumulative distribution function and quantiles of observed distance to roads for terrestrial iNaturalist observations in British Columbia, Canada. Analysis includes all road types: paved, maintained, and unmaintained. (b) Maps of iNaturalist observations and paved roads. (c) Frequency polygon plot of the mean distances from roads for random locations ($n = 10{,}000$ bootstrapped samples). (d) Frequency polygon plot of the mean distances from roads for observed locations ($n = 10{,}000$ bootstrapped samples). Mean distance was calculated for each bootstrapped sample. Each sample contained 1 million resampled data points.

elevation for accessibility. There are other potential metrics of accessibility such as steepness (Mair & Ruete, 2016), terrain ruggedness (Stolar & Nielsen, 2015), and travel time to major cities (Barber et al., 2022). Lastly, protected areas (i.e., national and provincial protected parks) and land cover types explained relatively little of the spatial biases (Table 1). This was unexpected, considering they have been important in other opportunistically collected datasets (El-Gabbas & Dormann, 2018; Geldmann et al., 2016; Petersen et al., 2021; Rocchini et al., 2011; Stolar & Nielsen, 2015). In particular, it was interesting to see the land cover type variable ranking so low when Di Cecco et al. (2021) found evidence of iNaturalist observations being biased across different land cover categories in the United States. However, they did not look at multiple spatial variables simultaneously. Thus, iNaturalist observations are likely biased by land cover type, but accessibility (i.e., distance to roads) is a more important factor for predicting where observations will occur. The small effect of parks (7%; Table 1) may be due to many of the parks, in particular large ones, being in remote northern regions with no year-round road access. Thus, we recommend including distance to roads, human population density, and broad habitat classification when accounting for spatial biases when using iNaturalist data and consider including protected areas and land cover land type if available.

The negative exponential relationship in predicted probabilities with distance from roads (Figure 3) mirrors the sharp cumulative curve with 94% of iNaturalist

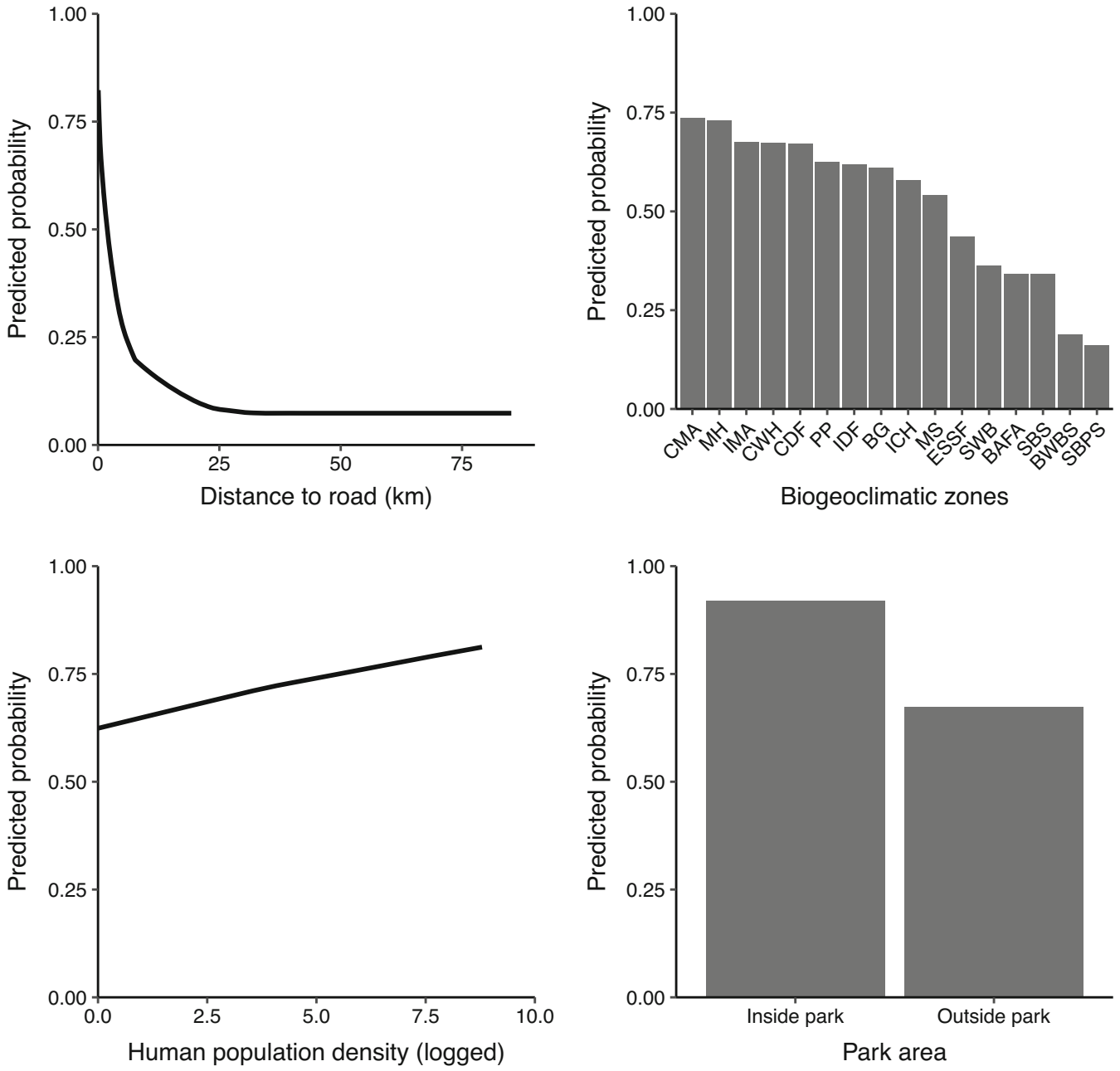



**FIGURE 3** Marginal response curves for the top four ranked environmental variables in the Maxent model. These show the relationship between predicted probability of observation with an environmental variable while all other variables are held at their average sample value. Biogeoclimatic zones: BAFA, Boreal Altai Fescue Alpine; BG, Bunchgrass; BWBS, Boreal White and Black Spruce; CDF, Coastal Douglas-fir; CMA, Coastal Mountain-heather Alpine; CWH, Coastal Western Hemlock; ESSF, Engelmann Spruce–Subalpine Fir; ICH, Interior Cedar–Hemlock; IDF, Interior Douglas-fir; IMA, Interior Mountain-heather Alpine; MH, Mountain Hemlock; MS, Montane Spruce; PP, Ponderosa Pine; SBPS, Sub-Boreal Pine–Spruce; SBS, Sub-Boreal Spruce; and SWB, Spruce–Willow–Birch.

observations within 1 km of roads in BC (Figure 2). This relationship was similarly found in Kadmon et al. (2004) with 61% of plant observations in Israel within 500 m of roads and 97% of observations within 4 km. Predicted probability increasing with human population density (Figure 3) is in line with other studies of opportunistically collected datasets (Mair & Ruete, 2016). Although the effect of parks on observation distribution was smaller than the other variables we tested, there was higher predicted probability of observations in parks than outside parks (Figure 3). The highest predicted zones for the BEC variable were Coastal Mountain-heather Alpine and Mountain Hemlock (Figure 3). This was likely due to popular provincial parks occurring within these zones that have intense

community science activity (BC iNaturalist Program, 2021; Egan, 1997; Strathcona Wilderness Institute, 2022).

## Applicability and future directions

We restricted our analyses to a single province as proof of concept, but our results may be broadly applicable beyond BC given the large land area (947,800 $km^2$), diversity of ecosystems, and large ranges in human population and road densities we tested (Environmental Reporting BC, 2018; Government of Canada, 2017; Meidinger & Pojar, 1991). To test for spatial bias in other regions, we recommend that people start by incorporating broad habitat classification, human population density, and distance to roads. This should likely sufficiently account for most of the spatial bias in iNaturalist data given that our results broadly agree with other community science studies (El-Gabbas & Dormann, 2018; Geldmann et al., 2016; Mair & Ruete, 2016).

For application of broad habitat classification outside of BC, we suggest using the most accurate habitat dataset available (Geldmann et al., 2016). We selected the BEC dataset as it is the most accurate broad habitat dataset for BC, and Geldmann et al. (2016) used a national habitat and land cover dataset of Denmark for modeling spatial biases. If modeling on an international scale, global habitat datasets are becoming more available and accurate with improvements in remote sensing (Jung et al., 2020).

Spatial sampling bias can be accounted for in species distribution models in a variety of ways (Fourcade et al., 2014). One method is to use a raster file of predicted probabilities of a sampling event (e.g., iNaturalist observations) like Figure 1 and upload it to the Maxent GUI as a bias file (Appendix S1: Figure S1) (Fourcade et al., 2014). In addition to species distribution models, quantification of spatial sampling bias and identification of under- and oversampled regions (Figure 2) can be combined with measures of species richness to determine where richness has been over- or underestimated due to variation in observer effort (Geldmann et al., 2016).

It would be helpful for future research to explore different taxonomic groups such as fungi, birds, and arthropods in relation to spatial bias (Geldmann et al., 2016; Mair & Ruete, 2016). This would aid conservation management in revealing which areas and taxa are under- and oversampled.

## CONCLUSIONS

With its exponential growth, community science continues to be of increasing importance in supplying data for analyses of biodiversity patterns and processes. This work shows how researchers can identify and account for spatial biases in such data. There are additional biases that need attention, including taxonomy of species and interobserver variability in where people go and what they record. We feel it is important to remember that no dataset is without bias, from community science to professionally collected data, and that it would be a disservice to our pursuit of ecological knowledge to view community science as unusable due to strong biases. We hope that further studies building on our findings can improve the scientific value of community science platforms, including testing the limits of inference that are possible.

### ACKNOWLEDGMENTS

We are thankful to the amazing iNaturalist community that volunteer their time to help document and identify the biodiversity that is under increasing pressure globally. Thank you to the members of the Reynolds and Starzomski labs, as well as Macgregor Aubertin-Young, Laura Eliuk, and Maxwell Borden for their support, input, and statistical advice. Thank you especially to the BC Parks team of Sharilynn Wardrop, James Quayle, and Jennifer Grant; and the former Ministry of Forests, Lands, and Natural Resource Operations and Rural Development (Jennifer Psyllakis). This work was supported by BC Parks, the Ministry of Land, Water and Resource Stewardship, the Sitka Foundation, the Pacific Wildlife Foundation, and NSERC Discovery grants to Brian M. Starzomski and John D. Reynolds, and a CGS-M NSERC scholarship to Ellyne M. Geurts.

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

### DATA AVAILABILITY STATEMENT

Datasets used for this research were retrieved from iNaturalist.org (https://www.inaturalist.org/observations/export), BC Data Catalogue (https://catalogue.data.gov.bc.ca/dataset/digital-road-atlas-dra-master-partially-attributed-roads), ArcGIS hub (https://www.arcgis.com/home/item.html?id=bd852a2c42924732a69b37632c25c585), Government of Canada (https://open.canada.ca/data/en/dataset/9e1507cd-f25c-4c64-995b-6563bf9d65bd), and Oak Ridge National Laboratory (https://landscan.ornl.gov/). Query details used for Oak Ridge National Laboratory: Year = 2020 and Product = LandScan global. Digital Elevation Model and Biogeoclimatic Ecosystem Classification data were accessed via the bcmaps package in R. Land cover type data from MODIS (product: MCD12Q1) was downloaded via MODIStsp package in R. Query details for MODIS under "Product and Layers"

were the following: Product category = Land Cover Characteristics – Land Cover, Platform = Terra, Product name = LandCover_Type_Yearly_500m (MCD12Q1), and MODIS layers = Land Cover Type 1 (IGBP). In addition, under "Spatial/Temporal options," we selected: Temporal range = 2020.01.01 to 2022.03.31, Data range type = Full, Output projection = User Defined, MODIS Sinusoidal = 3005, Output resolution = Native, Resampling method = Near, and Spatial extent – selection method = Load From Spatial File (used BC boundary shapefile). Query details for retrieving iNaturalist data were as follows: Place = British Columbia, Canada; Date range – end = December 2, 2021. As csv file exports are limited to 200,000 entries per request, requests were broken up by taxa until observations were downloaded. All other query settings were set to default.

## ORCID

*Ellyne M. Geurts* https://orcid.org/0000-0002-4440-8273

## REFERENCES

Ballesteros-Mejia, L., I. J. Kitching, W. Jetz, P. Nagel, and J. Beck. 2013. "Mapping the Biodiversity of Tropical Insects: Species Richness and Inventory Completeness of African Sphingid Moths." *Global Ecology and Biogeography* 22(5): 586–95. https://doi.org/10.1111/geb.12039.

Barber, R. A., S. G. Ball, R. K. A. Morris, and F. Gilbert. 2022. "Target-Group Backgrounds Prove Effective at Correcting Sampling Bias in Maxent Models." *Diversity and Distributions* 28(1): 128–41. https://doi.org/10.1111/ddi.13442.

Barve, V. V., L. Brenskelle, D. Li, B. J. Stucky, N. V. Barve, M. M. Hantak, B. S. McLean, et al. 2020. "Methods for Broad-Scale Plant Phenology Assessments Using Citizen Scientists' Photographs." *Applications in Plant Sciences* 8(1): 1–10. https://doi.org/10.1002/aps3.11315.

BC iNaturalist Program. 2021. "BC iNaturalist." https://www.bcinat.com/.

BC Parks. 2018. "BC Parks 2017/18 Statistics Report." https://bcparks.ca/about/news-publications/reports-surveys/.

Boakes, E. H., P. J. K. McGowan, R. A. Fuller, D. Chang-Qing, N. E. Clark, K. O'Connor, and G. M. Mace. 2010. "Distorted Views of Biodiversity: Spatial and Temporal Bias in Species Occurrence Data." *PLoS Biology* 8(6): e1000385. https://doi.org/10.1371/journal.pbio.1000385.

Bohl, C. L., J. M. Kass, and R. P. Anderson. 2019. "A New Null Model Approach to Quantify Performance and Significance for Ecological Niche Models of Species Distributions." *Journal of Biogeography* 46(6): 1101–11. https://doi.org/10.1111/jbi.13573.

Brown, E. D., and B. K. Williams. 2019. "The Potential for Citizen Science to Produce Reliable and Useful Information in Ecology." *Conservation Biology* 33(3): 561–9. https://doi.org/10.1111/cobi.13223.

Busetto, L., and L. Ranghetti. 2016. "MODIStsp: An R Package for Preprocessing of MODIS Land Products Time Series." *Computers & Geosciences* 97: 40–8. https://doi.org/10.1016/j.cageo.2016.08.020.

Callaghan, C. T., I. Ozeroff, C. Hitchcock, and M. Chandler. 2020. "Capitalizing on Opportunistic Citizen Science Data to Monitor Urban Biodiversity: A Multi-Taxa Framework." *Biological Conservation* 251: 108753. https://doi.org/10.1016/j.biocon.2020.108753.

Courter, J. R., R. J. Johnson, C. M. Stuyck, B. A. Lang, and E. W. Kaiser. 2013. "Weekend Bias in Citizen Science Data Reporting: Implications for Phenology Studies." *International Journal of Biometeorology* 57: 715–20. https://doi.org/10.1007/s00484-012-0598-7.

Di Cecco, G. J., V. Barve, M. W. Belitz, B. J. Stucky, R. P. Guralnick, and A. H. Hurlbert. 2021. "Observing the Observers: How Participants Contribute Data to iNaturalist and Implications for Biodiversity Science." *Bioscience* 71(11): 1179–88. https://doi.org/10.1093/biosci/biab093.

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter. 2010. "Citizen Science as an Ecological Research Tool: Challenges and Benefits." *Annual Review of Ecology, Evolution, and Systematics* 41: 149–72. https://doi.org/10.1146/annurev-ecolsys-102209-144636.

Drury, J. P., M. Barnes, A. E. Finneran, M. Harris, and G. F. Grether. 2019. "Continent-Scale Phenotype Mapping Using Citizen Scientists' Photographs." *Ecography* 42(8): 1436–45. https://doi.org/10.1111/ecog.04469.

Egan, B. 1997. "The Ecology of the Mountain Hemlock Zone." https://www.for.gov.bc.ca/hfd/pubs/docs/bro/bro51.pdf.

El-Gabbas, A., and C. F. Dormann. 2018. "Improved Species-Occurrence Predictions in Data-Poor Regions: Using Large-Scale Data and Bias Correction with Down-Weighted Poisson Regression and Maxent." *Ecography* 41(7): 1161–72. https://doi.org/10.1111/ecog.03149.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. "A Statistical Explanation of MaxEnt for Ecologists." *Diversity and Distributions* 17(1): 43–57. https://doi.org/10.1111/j.1472-4642.2010.00725.x.

Environmental Reporting BC. 2018. "Trends in B.C.'s Population Size & Distribution. State of Environment Reporting." British Columbia, Canada. https://www.env.gov.bc.ca/soe/indicators/sustainability/bc-population.

Fernández, D., and M. Nakamura. 2015. "Estimation of Spatial Sampling Effort Based on Presence-Only Data and Accessibility." *Ecological Modelling* 299: 147–55. https://doi.org/10.1016/j.ecolmodel.2014.12.017.

Fithian, W., J. Elith, T. Hastie, and D. A. Keith. 2015. "Bias Correction in Species Distribution Models: Pooling Survey and Collection Data for Multiple Species." *Methods in Ecology and Evolution* 6(4): 424–38. https://doi.org/10.1111/2041-210X.12242.

Fourcade, Y., J. O. Engler, D. Rödder, and J. Secondi. 2014. "Mapping Species Distributions with MAXENT Using a Geographically Biased Sample of Presence Data: A Performance Assessment of Methods for Correcting Sampling Bias." *PLoS One* 9(5): e97122. https://doi.org/10.1371/journal.pone.0097122.

Geldmann, J., J. Heilmann-Clausen, T. E. Holm, I. Levinsky, B. Markussen, K. Olsen, C. Rahbek, and A. P. Tøttrup. 2016. "What Determines Spatial Bias in Citizen Science? Exploring Four Recording Schemes with Different Proficiency Requirements." *Diversity and Distributions* 22(11): 1139–49. https://doi.org/10.1111/ddi.12477.

Government of Canada. 2017. "British Columbia's Provincial Symbols." https://www.canada.ca/en/canadian-heritage/services/provincial-territorial-symbols-canada/british-columbia.

Hamilton, S. L., V. R. Saccomanno, W. N. Heady, A. L. Gehman, S. I. Lonhart, R. Beas-Luna, F. T. Francis, et al. 2021. "Disease-Driven Mass Mortality Event Leads to Widespread Extirpation and Variable Recovery Potential of a Marine Predator across the Eastern Pacific." *Proceedings of the Royal Society B: Biological Sciences*. 288: 20211195. https://doi.org/10.1098/rspb.2021.1195.

Hausdorf, B., M. Parr, L. J. Shappell, J. Oldeland, and D. G. Robinson. 2021. "The Introduction of the European *Caucasotachea vindobonensis* (Gastropoda: Helicidae) in North America, Its Origin and Its Potential Range." *Biological Invasions* 23(11): 3281–9. https://doi.org/10.1007/s10530-021-02579-4.

Hijmans, R. J. 2021a. "Raster: Geographic Data Analysis and Modeling." R Package Version 3.5-2. https://cran.r-project.org/package=raster.

Hijmans, R. J. 2021b. "Terra: Spatial Data Analysis." R Package Version 1.4-11. https://cran.r-project.org/package=terra.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith. 2021. "Dismo: Species Distribution Modeling." R Package Version 1.3-5. https://cran.r-project.org/package=dismo.

iNaturalist. 2023. "iNaturalist." https://www.inaturalist.org.

Isaac, N. J. B., and M. J. O. Pocock. 2015. "Bias and Information in Biological Records." *Biological Journal of the Linnean Society* 115(3): 522–31. https://doi.org/10.1111/bij.12532.

Isaac, N. J. B., A. J. van Strien, T. A. August, M. P. de Zeeuw, and D. B. Roy. 2014. "Statistics for Citizen Science: Extracting Signals of Change from Noisy Ecological Data." *Methods in Ecology and Evolution* 5(10): 1052–60. https://doi.org/10.1111/2041-210X.12254.

Jain, P., H. Forbes, and L. A. Esposito. 2022. "Two New Alkali-Sink Specialist Species of *Paruroctonus* Werner 1934 (Scorpiones, Vaejovidae) from Central California." *ZooKeys* 1117: 139–88. https://doi.org/10.3897/zookeys.1117.76872.

Johnston, A., D. Fink, W. M. Hochachka, and S. Kelling. 2018. "Estimates of Observer Expertise Improve Species Distributions from Citizen Science Data." *Methods in Ecology and Evolution* 9(1): 88–97. https://doi.org/10.1111/2041-210X.12838.

Johnston, A., W. M. Hochachka, M. E. Strimas-Mackey, V. R. Gutierrez, O. J. Robinson, E. T. Miller, T. Auer, S. T. Kelling, and D. Fink. 2021. "Analytical Guidelines to Increase the Value of Community Science Data: An Example Using eBird Data to Estimate Species Distributions." *Diversity and Distributions* 27(7): 1265–77. https://doi.org/10.1111/ddi.13271.

Jung, M., P. R. Dahal, S. H. M. Butchart, P. F. Donald, X. De Lamo, M. Lesiv, V. Kapos, C. Rondinini, and P. Visconti. 2020. "A Global Map of Terrestrial Habitat Types." *Scientific Data* 7: 256. https://doi.org/10.1038/s41597-020-00599-8.

Kadmon, R., O. Farber, and A. Danin. 2004. "Effect of Roadside Bias on the Accuracy of Predictive Maps Produced by Bioclimatic Models." *Ecological Applications* 14(2): 401–13. https://doi.org/10.1890/02-5364.

Kass, J. M., R. Muscarella, P. J. Galante, C. L. Bohl, G. E. Pinilla-Buitrago, R. A. Boria, M. Soley-Guardia, and R. P. Anderson. 2021. "ENMeval 2.0: Redesigned for Customizable and Reproducible Modeling of Species' Niches and Distributions." *Methods in Ecology and Evolution* 12(9): 1602–8. https://doi.org/10.1111/2041-210X.13628.

Kass, J. M., R. Muscarella, G. E. Pinilla-Buitrago, and P. J. Galante. 2022. "ENMeval 2.0 Vignette." ENMeval 2.0.4. https://jamiemkass.github.io/ENMeval/articles/ENMeval-2.0-vignette.

Kelling, S., A. Johnston, W. M. Hochachka, M. Iliff, D. Fink, J. Gerbracht, C. Lagoze, et al. 2015. "Can Observation Skills of Citizen Scientists Be Estimated Using Species Accumulation Curves?" *PLoS One* 10(10): e0139600. https://doi.org/10.1371/journal.pone.0139600.

Kosmala, M., A. Wiggins, A. Swanson, and B. Simmons. 2016. "Assessing Data Quality in Citizen Science." *Frontiers in Ecology and the Environment* 14(10): 551–60. https://doi.org/10.1002/fee.1436.

La Sorte, F. A., and M. Somveille. 2020. "Survey Completeness of a Global Citizen-Science Database of Bird Occurrence." *Ecography* 43(1): 34–43. https://doi.org/10.1111/ecog.04632.

Lehtinen, R. M., B. M. Carlson, A. R. Hamm, A. G. Riley, M. M. Mullin, and W. J. Gray. 2020. "Dispatches from the Neighborhood Watch: Using Citizen Science and Field Survey Data to Document Color Morph Frequency in Space and Time." *Ecology and Evolution* 10(3): 1526–38. https://doi.org/10.1002/ece3.6006.

Loarie, S. 2020. "We've Reached 1,000,000 Observers!" iNaturalist. https://www.inaturalist.org/blog/35758-we-ve-reached-1-000-000-observers.

Loarie, S. 2022. "We've Passed 100,000,000 Verifiable Observations on iNaturalist!" iNaturalist. https://www.inaturalist.org/blog/66531-we-ve-passed-100-000-000-verifiable-observations-on-inaturalist.

Mair, L., and A. Ruete. 2016. "Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa." *PLoS One* 11(1): e0147796. https://doi.org/10.1371/journal.pone.0147796.

Mangiafico, S. 2022. "Rcompanion: Functions to Support Extension Education Program Evaluation." R Package Version 2.4.15. https://cran.r-project.org/package=rcompanion.

Meidinger, D., and J. Pojar. 1991. *Ecosystems of British Columbia*. Victoria, BC: Ministry of Forests and Range Research Branch.

Merow, C., M. J. Smith, and J. A. Silander. 2013. "A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36(10): 1058–69. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Mesaglio, T., and C. T. Callaghan. 2021. "An Overview of the History, Current Contributions and Future Outlook of iNaturalist in Australia." *Wildlife Research* 48(4): 289–303. https://doi.org/10.1071/WR20154.

Meyer, C., P. Weigelt, and H. Kreft. 2016. "Multidimensional Biases, Gaps and Uncertainties in Global Plant Occurrence Information." *Ecology Letters* 19(8): 992–1006. https://doi.org/10.1111/ele.12624.

Miller-Rushing, A., R. Primack, and R. Bonney. 2012. "The History of Public Participation in Ecological Research." *Frontiers in Ecology and the Environment* 10(6): 285–90. https://doi.org/10.1890/110278.

Neate-Clegg, M. H. C., J. J. Horns, F. R. Adler, M. Ç. K. Aytekin, and Ç. H. Şekercioğlu. 2020. "Monitoring the World's Bird Populations with Community Science Data." *Biological Conservation* 248: 108653. https://doi.org/10.1016/j.biocon.2020.108653.

Nowak, K., J. Berger, A. Panikowski, D. G. Reid, A. L. Jacob, G. Newman, N. E. Young, J. P. Beckmann, and S. A. Richards. 2020. "Using Community Photography to Investigate Phenology: A Case Study of Coat Molt in the Mountain Goat (*Oreamnos americanus*) with Missing Data." *Ecology and Evolution* 10(23): 13488–99. https://doi.org/10.1002/ece3.6954.

Oliveira, U., A. P. Paglia, A. D. Brescovit, C. J. B. de Carvalho, D. P. Silva, D. T. Rezende, F. S. F. Leite, et al. 2016. "The Strong Influence of Collection Bias on Biodiversity Knowledge Shortfalls of Brazilian Terrestrial Biodiversity." *Diversity and Distributions* 22(12): 1232–44. https://doi.org/10.1111/ddi.12489.

Parks Canada. 2022. "Citizen Science – Science and Conservation." https://www.pc.gc.ca/en/nature/science/impliquez-involved/science.

Pebesma, E. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *The R Journal* 10(1): 439–46. https://doi.org/10.32614/RJ-2018-009.

Petersen, T. K., J. D. M. Speed, V. Grøtan, and G. Austrheim. 2021. "Species Data for Understanding Biodiversity Dynamics: The What, Where and When of Species Occurrence Data Collection." *Ecological Solutions and Evidence* 2(1): e12048. https://doi.org/10.1002/2688-8319.12048.

Phillips, S. 2017. "A Brief Tutorial on Maxent." http://biodiversityinformatics.amnh.org/open_source/maxent/.

Phillips, S. J., M. Dudík, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. "Sample Selection Bias and Presence-Only Distribution Models: Implications for Background and Pseudo-Absence Data." *Ecological Applications* 19(1): 181–97. https://doi.org/10.1890/07-2153.1.

Phillips, S. J., M. Dudík, and R. E. Schapire. 2020. "Maxent Software for Modeling Species Niches and Distributions." Version 3.4.4. https://biodiversityinformatics.amnh.org/open_source/maxent/.

Pocock, M. J. O., M. Chandler, R. Bonney, I. Thornhill, A. Albin, T. August, S. Bachman, et al. 2018. "A Vision for Global Biodiversity Monitoring with Citizen Science." *Advances in Ecological Research* 59: 169–223. https://doi.org/10.1016/bs.aecr.2018.06.003.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/.

Reddy, S., and L. M. Dávalos. 2003. "Geographical Sampling Bias and Its Implications for Conservation Priorities in Africa." *Journal of Biogeography* 30(11): 1719–27. https://doi.org/10.1046/j.1365-2699.2003.00946.x.

Redlands, C.E.S.R.I. 2017. "ArcMap 10.6.1." ArcGIS Desktop: Release 10. https://desktop.arcgis.com/en/quick-start-guides/10.6/arcgis-desktop-quick-start-guide.htm.

Roberts, C. J., A. Vergés, C. T. Callaghan, and A. G. B. Poore. 2022. "Many Cameras Make Light Work: Opportunistic Photographs of Rare Species in iNaturalist Complement Structured Surveys of Reef Fish to Better Understand Species Richness." *Biodiversity and Conservation* 31(4): 1407–25. https://doi.org/10.1007/s10531-022-02398-6.

Rocchini, D., J. Hortal, S. Lengyel, J. M. Lobo, A. Jiménez-Valverde, C. Ricotta, G. Bacaro, and A. Chiarucci. 2011. "Accounting for Uncertainty When Mapping Species Distributions: The Need for Maps of Ignorance." *Progress in Physical Geography* 35(2): 211–26. https://doi.org/10.1177/0309133311399491.

Ross, N. 2020. "Fasterize: Fast Polygon to Raster Conversion." R Package Version 1.0.3. https://cran.r-project.org/package=fasterize.

RStudio Team. 2021. *RStudio: Integrated Development Environment for R*. Boston, MA: RSudio, PBC. http://www.rstudio.com/.

Ruete, A. 2015. "Displaying Bias in Sampling Effort of Data Accessed from Biodiversity Databases Using Ignorance Maps." *Biodiversity Data Journal* 3: e5361. https://doi.org/10.3897/BDJ.3.e5361.

Saldivar, J.'l. A., A. N. Romero, and E. E. Wilson Rankin. 2022. "Community Science Reveals High Diversity of Nectaring Plants Visited by Painted Lady Butterflies (Lepidoptera: Nymphalidae) in California Sage Scrub." *Environmental Entomology* 51(6): 1141–9. https://doi.org/10.1093/ee/nvac073.

Schoener, T. W. 1968. "The Anolis Lizards of Bimini: Resource Partitioning in a Complex Fauna." *Ecology* 49(4): 704–26.

Speed, J. D. M., M. Bendiksby, A. G. Finstad, K. Hassel, A. L. Kolstad, and T. Prestø. 2018. "Contrasting Spatial, Temporal and Environmental Patterns in Observation and Specimen Based Species Occurrence Data." *PLoS One* 13(4): e0196417. https://doi.org/10.1371/journal.pone.0196417.

Stolar, J., and S. E. Nielsen. 2015. "Accounting for Spatially Biased Sampling Effort in Presence-Only Species Distribution Modelling." *Diversity and Distributions* 21(5): 595–608. https://doi.org/10.1111/ddi.12279.

Strathcona Wilderness Institute. 2022. "iNaturalist." https://strathconapark.org/swi-research/inaturalist-data/.

Sullivan, B. L., C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling. 2009. "eBird: A Citizen-Based Bird Observation Network in the Biological Sciences." *Biological Conservation* 142(10): 2282–92. https://doi.org/10.1016/j.biocon.2009.05.006.

Teucher, A., S. Hazlitt, and S. Albers. 2021. "Bcmaps: Map Layers and Spatial Utilities for British Columbia." R Package Version 1.0.2. https://cran.r-project.org/package=bcmaps.

Troudet, J., P. Grandcolas, A. Blin, R. Vignes-Lebbe, and F. Legendre. 2017. "Taxonomic Bias in Biodiversity Data and Societal Preferences." *Scientific Reports* 7(1): 9132. https://doi.org/10.1038/s41598-017-09084-6.

Tulloch, A. I. T., and J. K. Szabo. 2012. "A Behavioural Ecology Approach to Understand Volunteer Surveying for Citizen Science Datasets." *Emu – Austral Ornithology* 112(4): 313–25. https://doi.org/10.1071/MU12009.

Tye, C. A., R. A. McCleery, R. J. Fletcher, D. U. Greene, and R. S. Butryn. 2017. "Evaluating Citizen vs. Professional Data for Modelling Distributions of a Rare Squirrel." *Journal of Applied Ecology* 54(2): 628–37. https://doi.org/10.1111/1365-2664.12682.

Valavi, R., G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Elith. 2022. "Predictive Performance of Presence-Only Species Distribution Models: A Benchmark Study with Reproducible Code." *Ecological Monographs* 92(1): e01486. https://doi.org/10.1002/ecm.1486.

van Wilgenburg, S. L., E. M. Beck, B. Obermayer, T. Joyce, and B. Weddle. 2015. "Biased Representation of Disturbance Rates in the Roadside Sampling Frame in Boreal Forests: Implications for Monitoring Design." *Avian Conservation and Ecology* 10(2): 5. https://doi.org/10.5751/ACE-00777-100205.

Warren, D., and R. Dinnage. 2022. "ENMTools: Analysis of Niche Evolution Using Niche and Distribution Models."

R Package Version 1.0.6. https://cran.r-project.org/package=ENMTools.

Werenkraut, V., F. Baudino, and H. E. Roy. 2020. "Citizen Science Reveals the Distribution of the Invasive Harlequin Ladybird (*Harmonia axyridis* Pallas) in Argentina." *Biological Invasions* 22: 2915–21. https://doi.org/10.1007/s10530-020-02312-7.

Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Wickham, H., R. François, L. Henry, and K. Müller. 2021. "Dplyr: A Grammar of Data Manipulation." R Package Version 1.0.7. https://cran.r-project.org/package=dplyr.

Zhang, G. 2020. "Spatial and Temporal Patterns in Volunteer Data Contribution Activities: A Case Study of eBird." *ISPRS International Journal of Geo-Information* 9(10): 597. https://doi.org/10.3390/ijgi9100597.

Zizka, A., A. Antonelli, and D. Silvestro. 2020. "Sampbias, a Method for Quantifying Geographic Sampling Biases in Species Distribution Data." *Ecography* 44(1): 25–32. https://doi.org/10.1111/ecog.05102.

## SUPPORTING INFORMATION

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

**How to cite this article:** Geurts, Ellyne M., John D. Reynolds, and Brian M. Starzomski. 2023. "Turning Observations into Biodiversity Data: Broadscale Spatial Biases in Community Science." *Ecosphere* 14(6): e4582. https://doi.org/10.1002/ecs2.4582