Effects of Accessibility on the Records of Biodiversity Observation by the Citizens of Urban Forests

Shoma Jingu[1], Yui Ogawa[2]

[1] Dept. of Forest Management, Forestry and Forest Products Research Institute, Forest Research and Management Organization, Tsukuba 305-8687, Japan - jingu_shoma250@ffpri.go.jp

[2] Degree Programs in Life and Earth Sciences, Univ. of Tsukuba, Tsukuba 305-8577, Japan - ogawa.yui.ta@alumni.tsukuba.ac.jp

Keywords: Community Science, Biodiversity Monitoring, Urban Forests, Hotspot Analysis, Wildlife Watching, Human–Nature Interaction

Abstract

The leveraging of social media for citizen science is a powerful tool for generating large-scale records of wildlife observations and engaging the public in biodiversity conservation. However, the utility of observation records is adversely affected by their biases. In particular, understanding the drivers behind observation hotspots—locations where observation records are concentrated—is vital for interpreting data and optimizing project design. This study investigated the factors forming observation hotspots in the urban forests of Tsukuba Science City, Japan, assuming that accessibility by humans is as important as ecological factors. We analyzed 17,224 observation records from a social media platform for citizen science across a 54-km^2 study area and classified location accessibility into three levels (public wayside, public inland, or remote) based on road proximity and public access. Subsequently, models for five taxonomic groups were compared using (model 1) basic land-cover categories and (model 2) land-cover and accessibility categories. Model 2 consistently outperformed Model 1 in predicting hotspots and specific land-cover factors connected to species distributions. These findings highlighted the critical role of identifying biases that drive data patterns in citizen science, which is essential for interpreting data, designing effective engagement strategies, and planning biodiversity-friendly and accessible urban green spaces.

1. Introduction

Biodiversity conservation contributes to human well-being and sustainable development (Cardinale et al., 2012; Díaz et al., 2018); however, biodiversity continues to be lost in the 2020s (Stephens, 2023). In December 2022, the Kunming–Montreal Global Biodiversity Framework (KM-GBF) was established to promote the Nature Positive concept, which aims to halt and reverse global biodiversity loss by 2030 (Stephens, 2023). KM-GBF promotes engagement in biodiversity conservation–related activities and invites various stakeholders to support community-based monitoring and citizen science (United Nations Conservation on Biological Diversity, 2022). Citizen science is defined as "scientific work undertaken by members of the general public" (Oxford English Dictionary, 2014) and is considered a positive force for biodiversity conservation (Bonney et al., 2009). For example, detailed information on local biodiversity is required to inform conservation policies (Kindsvater et al., 2018; Stephens, 2023). Researchers have generally relied on time-consuming and costly surveys conducted by experts to collect information, but citizen science enables data collection at a scale that is inaccessible to scientists alone and helps citizens enhance their understanding of local nature and biodiversity (Amano et al., 2016; Cohn, 2008; Kindsvater et al., 2018; Kobori et al., 2016). Scientists, public institutions, and companies can organize citizen science platforms to collect biodiversity information on a large scale in various regions and help societies understand the value of biodiversity. Promoting citizen science can help individuals consider biodiversity in their daily lives, resulting in social change through a virtuous cycle (Kobori et al., 2016) and is therefore critical for achieving a Nature Positive society (Costanza et al., 2017; United Nations Conservation on Biological Diversity, 2022).

However, using citizen science to collect information for biodiversity conservation requires the participation of diverse members of the society (Schröter et al., 2017). Traditional citizen science participants include minorities, such as bird-watchers (Carlen et al., 2024; Cohn, 2008). Given that social media are essential for increasing the number of participants in citizen science (Ghermandi and Sinclair, 2019), multiple social media platforms have been developed to collect biodiversity information, e.g., eBird (Sullivan et al., 2009), iNaturalist (Nugent, 2018), and Biome (Atsumi et al., 2024). These platforms function as information-sharing websites where users are able to generate the necessary big data for biodiversity conservation (Jacobs and Zipf, 2017; Schröter et al., 2017). To motivate data collection, some platforms attract users using applications that incorporate gamification elements (Atsumi et al., 2024; Costanza et al., 2017; Morschheuser et al., 2017). For example, Biome—a mobile application available in Japanese—uses artificial intelligence–based species identification algorithms and gamification elements, which has collected more than six million observation records of organisms in the form of uploaded photos since its launch in 2019 (Atsumi et al., 2024).

Effectively using such voluntary observation records requires understanding the biases inherent to the collection location (Atsumi et al., 2024; Carlen et al., 2024). Numerous previous studies have used observation records as the indicators of the presence of a species and built models of its distribution based on biological and environmental factors, such as land cover. However, biological and environmental factors on their own are insufficient for building species distribution models (Fourcade et

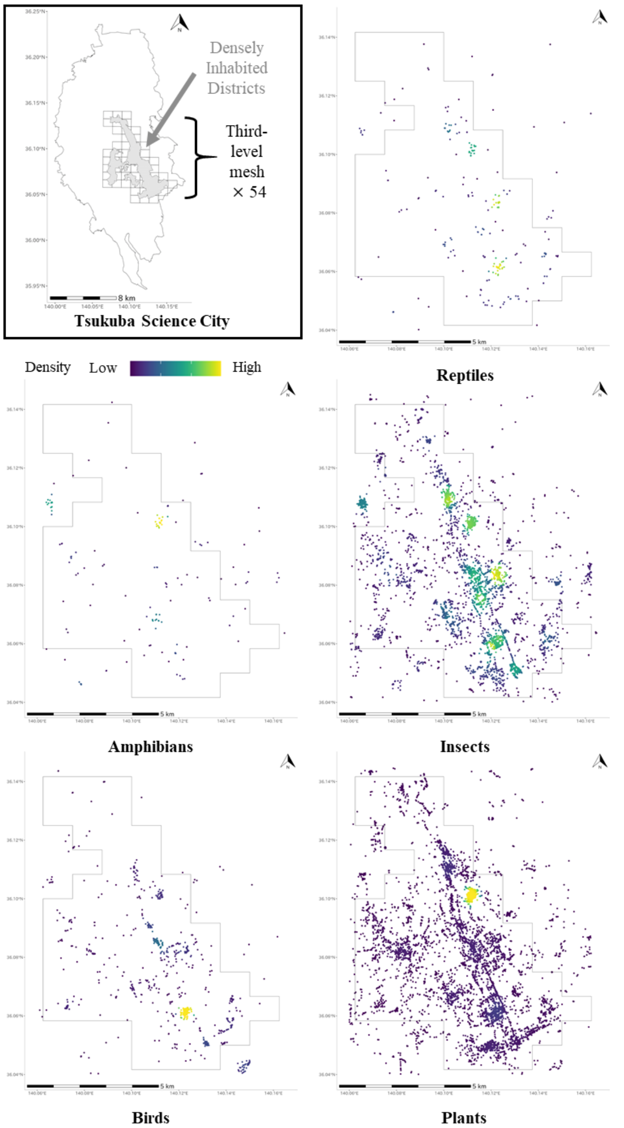

Figure 1. Study area and spatial distribution of citizen science data in Tsukuba Science City. Black points show the photos taken spots for each species. The heat map displays the hotspots of contribution. The more photos posted, the darker the red.

al., 2014) as an observation record is generated only when the organism and observer are present (Carlen et al., 2024). This requirement leads to spatial (e.g., sampling related) and observation effort–related biases, in which the observation frequency of an organism is strongly influenced by its population density and its accessibility to human observers. For example, despite the excellent environmental conditions of an area, the observation records may be scarce if it is difficult to access or suitable paths are not maintained (You et al., 2022). Therefore, the effective utilization of citizen science for data collection requires the consideration of the ecological factors related to the habitats of local species and human factors that regulate the ability of observers to generate records (Carlen et al., 2024). In addition, observation records are influenced by areas visited by people and opportunities on detecting an organism (Carlen et al., 2024). Herein, we focused on the first factor, namely, the spatial biases in observation records owing to accessibility to observers.

This study aimed to clarify human factors shaping the formation of observation hotspots (i.e., locations where wildlife observation records are concentrated) in urban forests. The formation of these hotspots was hypothesized to depend on the ecological factors making urban forests a suitable habitat for target organisms (e.g., land cover, abundance of vegetation, and presence of water bodies) and the human factors facilitating access by potential observes (e.g., proximity to roads and openness of land to the general public). In particular, we investigated the influence of accessibility on observation hotspots to determine whether observation patterns are primarily determined by accessibility or the enhanced biodiversity of ecologically rich areas.

2. Materials and Methods

2.1 Study Area

In Tsukuba Science City (Ibaraki Prefecture, Japan), forests have been maintained for nature conservation and use by urban residents through policy interventions. Tsukuba Science City, which lies ~50 km northeast of Tokyo and had a population of ~260,000 in 2023 (Tsukuba City Government, 2024), was planned and developed as a science city in the late 1960s. In the design phase, policies mandating the preservation of parts of the original forest and grassland through integration into urban parks and green spaces accessible to residents were implemented (Grabska-Szwagrzyk et al., 2024; Jingu, 2020). This history makes Tsukuba Science City an appropriate area to test our hypothesis.

The study area (Figure 1) was defined based on densely inhabited districts (DID), which are statistical areal units reflecting a high residential population and high daytime population density. Forests within or near DIDs are likely to be accessible to residents and visitors and were included in the study area when present within a walkable distance from a given DID. Owing to the irregular shapes of DIDs, we used the third-level mesh of the standard regional grid system for Japan, which has a resolution of approximately 1 km × 1 km, as a reference unit (e-Stat, 2024). All third-level meshes overlapping with DIDs of Tsukuba Science City were selected, which resulted in a study area of 54 km^2 (i.e., 54 meshes).

2.2 Data Collection

2.2.1 Citizen Science Data: We used anonymized wildlife observation records submitted by citizen scientists via Biome (Atsumi et al., 2024) within the study area from April 26, 2019 (i.e., when Biome was launched) to May 24, 2024 (Figure 1). Each record contained a contributor-taken photo with sufficient characteristics for species identification, geographic coordinates (latitude and longitude), observation date, and taxonomic information (often to the species level, although higher taxonomic levels were also included). The data were filtered to include only observations resulting from wildlife watching, which involved manually removing records taken indoors, such as the photographs of organisms in indoor botanical gardens or pets.

2.2.2 Ecological Factors: Geospatial data representing potential ecological factors for the presence of a species were compiled primarily from land-cover data. We used the high-resolution land-use land-cover (LULC) map provided by the Japan Aerospace Exploration Agency (JAXA, 2023), combining satellite imagery with multiple information sources and machine learning to ensure classification accuracy (Takahashi et al., 2013;

Tsutsumida et al., 2019). The latest available version (LULC v23.12), reflecting land-cover conditions as of 2022, was used. The map was provided in a raster format with a resolution of approximately 10 m × 10 m per cell. Each cell was classified into 1 of 14 land-cover types. These types were aggregated into seven broader categories: built-up areas (i.e., urban areas, buildings, and roads), croplands (i.e., cultivated fields excluding rice paddies), forests of various types, grassland (including bare land), rice paddies, water bodies (including wetlands), and others (i.e., solar panels and greenhouses).

2.2.3 Human Factors: Geospatial data representing accessibility were compiled with a focus on the forest land-cover category. The following factors majorly influence the distribution of species: water bodies, cultivated land, grasslands, and forests. In urban and suburban settings, forests are often subjected to diverse management approaches ranging from active recreational use to strict protection with restricted entry. Therefore, forest accessibility depends on regulations (i.e., land ownership and protection status) and proximity to access routes (paths and roads). This variation in accessibility could have a stronger influence on the distribution of observation records within forests than within other land-cover categories. For example, croplands are largely private with limited variation in accessibility. Water bodies and grasslands are often open to the public; however, their accessibility often depends solely on the availability of means of access (e.g., boats for water bodies and free movement across managed grasslands). Thus, we established three additional categories for the forest land-cover category based on accessibility.

1. A public wayside forest (PWF) is open to the public and adjacent to publicly maintained and accessible roads or paths. This category was delineated by identifying overlaps between forest land-cover cells and the buffer zone around officially maintained roads, which we defined using road edge data from the Fundamental Geospatial Data (FGD) provided by the Geospatial Information Authority of Japan (GSI, 2024a). A road edge was defined as the outer boundary of a road with a width of ≥1.0 m, which allowed us to identify roads used by pedestrians, even along wide multilane highways. A forest within a 5-m road edge was defined as adjacent to a road because landscape elements within a 5-m road edge are subjected to ecological factors influenced by edge effects (Watkins et al., 2003) and perceptual factors related to the human scale (Simpson et al., 2019). Thus, we used spatial analysis functions in the GIS software (R package sf, st_buffer function) to generate an ~5-m buffer around road edges and identified forest cells overlapping with this buffer.
2. A public inland forest (PIF) is open to the public but not adjacent to officially maintained roads. Forests open to the public were identified based on land ownership, lacking fencing, or fencing coupled with permissive entry policies. PIFs include forests within urban parks as defined under the Urban Green Space Act, university campuses without access restrictions, government facility grounds open to the public, and the precincts of Shinto shrines and Buddhist temples. The delineation involved visual inspection and cross-referencing of the forest land-cover category with FGD (GSI, 2017), aerial photographs (GSI, 2024b), OpenStreetMap (OSM) (OpenStreetMap Foundation, 2024), and cadastral maps (G Space Information Center, 2023). In some areas, public access restrictions were confirmed through publicly available information from facility managers or site visits.
3. A remote forest (RF) is neither open to the public (i.e., private or restricted access) nor adjacent to officially maintained roads. RFs include forests within restricted public facilities (e.g., water management areas), private facilities with limited user access (e.g., golf courses), backyard forests associated with private residences, and commercial timber plantations.

2.3 Data Analysis

The Maximum Entropy Modeling (MaxEnt) software (version 3.4.4) was used to build models for explaining and predicting the distribution of observation records from citizen scientists. MaxEnt, widely used for modeling species distributions and habitat suitability (Phillips et al., 2006), was used to estimate the probability distribution of suitable observing points using presence-only data and environmental variables. The modeling process was as follows.

The observation records from Biome were used to identify presence points for each of the five target taxonomic groups (reptiles, amphibians, insects, plants, and birds). To prevent over-prediction in MaxEnt models, multiple occurrence records falling within the same analysis cell for a given taxonomic group were treated as a single data point, with duplicate records removed prior to analysis. For each taxonomic group, the environmental conditions across the study area were sampled by the random selection of 10,000 background points and the occurrence data were randomly partitioned into a 75% training set and a 25% test set. LULC data were analyzed at a resolution of 10 m, which is consistent with the prediction unit of MaxEnt. The focal function in the R raster package (v. 3.6.20) was used to calculate the number of cells of each land-cover category within its neighborhood, thus accounting for the influence of the surrounding environment on the presence and observation probabilities for a given cell. Two neighborhood sizes were considered, each centered on the target cell: a 3 × 3 cell neighborhood with 9 cells and a ~15-m radius and a 9 × 9 cell neighborhood with 81 cells and a ~45-m radius. The smaller neighborhood was selected to account for potential global positioning system inaccuracies in the Biome records observed in previous studies (Asari and Fujiki 2020) and represented the immediate vicinity. The larger neighborhood was selected to represent a broad environmental context and the spatial scales often considered in habitat buffer analyses. The obtained focal statistics (e.g., count of built-up cells in a 3 × 3 cell neighborhood or count of PWF cells in a 9 × 9 cell neighborhood) were used as environmental variables. To adjust the model complexity and prevent overfitting, we set the regularization multiplier (betamultiplier) to 2.0. The model included linear, product, and hinge features. The algorithm was terminated after 500 iterations or convergence.

To test our hypothesis, we compared two models for each of the five taxonomic groups. Model 1 considered only environmental variables derived from the seven land-cover categories (see Section 2.2.2). Model 2 considered nine categories where the forest land-cover category was replaced by the three accessibility categories (PWF, PIF, and RF). Model performances were evaluated using the area under the receiver operating characteristics curve (AUC) by test data, measuring the ability of a model to discriminate between presence and background points. Higher values indicate improved performance. We compared the average AUCs of Models 1 and 2 across the five taxonomic groups and performed a variable importance analysis for Model 2 to determine the contribution (%) of each environmental variable. Permutation importance was used to measure the AUC drop owing to a random change in the value of a specific variable, and variables causing large drops were classified as more important. The jackknife test using AUC on test data was used to evaluate the model gain when each variable was used in isolation versus when each variable was excluded from the full model, which indicated whether a variable was redundant or made unique contributions. Finally, response curves were plotted to visualize the change in the predicted probability of a species' presence (i.e., observation probability) along the range of each environmental variable when other variables were fixed to their average values. Data preparation and spatial analysis were performed using R (v. 4.3.1, The R Foundation for Statistical Computing, 2023) with the raster (v. 3.6.20) and sf (v. 1.0.13) packages and QGIS (v. 3.28.11). Models were constructed using the standalone MaxEnt software (Phillips et al., 2020).

3. Results

3.1 Observation Hotspots

The Biome data were manually filtered to obtain a dataset with 17,224 valid observation records across various taxonomic groups. The records were classified according to species into nine initial groups: reptiles (*n* = 225), amphibians (*n* = 111), mammals (*n* = 30), insects and spiders (*n* = 5,328), plants (*n* = 10,161), crustaceans (*n* = 89), fish (*n* = 63), birds (*n* = 1,029), and others (*n* = 188). Owing to insufficient sample size, the mammal, crustacean, fish, and "other" groups were excluded, and five groups, namely, reptiles, amphibians, insects, plants, and birds therefore remained. The spatial distributions of the observation records for each taxonomic group were highly heterogeneous (Figure 1), with observation hotspots primarily occurring within the specific zones (e.g., inside the large public parks) and lines (e.g., pedestrian decks connecting the north and south parts) of DIDs. Visually, these hotspots were confirmed to be strongly associated with high human activity areas.

3.2 Model Comparison

Model 2 outperformed Model 1 in predicting the distributions of all taxonomic groups (Table 1). The largest and smallest differences were observed for amphibians (AUC 0.05 higher for Model 2) and reptiles (AUC 0.01 higher for Model 2). These results demonstrate that incorporating accessibility improved the explanatory power of the model.

Taxon	AUC for Model 1	AUC for Model 2
Reptiles	0.76	0.77
Amphibians	0.58	0.63
Insects	0.68	0.71
Plants	0.66	0.68
Birds	0.78	0.82

Table 1. Performance comparison between Models 1 and 2.

3.3 Variable Importance Analysis

Table 2 presents the results of variable importance analysis for Model 2, revealing the importance of human and ecological factors. Although ecological factors were important, human factors (in terms of accessibility) were crucial for explaining the observation hotspots.

3.3.1 Percent Contribution: Across the five taxonomic groups, environmental variables reflecting human-dominated landscapes or accessibility (e.g., cropland and forest-related categories) often had high percent contributions. The taxonomic groups differed in the percent contributions of the accessibility categories.

3.3.2 Permutation Importance: PWF consistently ranked among the six most important variables for all taxonomic groups, which indicated the importance of accessibility. The predictive performance substantially degraded when PWF was scrambled.

3.3.3 Jackknife Test: The jackknife test was used to compare the learning gains when each variable was excluded or used alone to evaluate the uniqueness of its information. The uniqueness of the variables depended on the taxonomic group. Interestingly, the forest-related variables were consistently one of the three most important variables for all taxonomic groups. This suggests that accessibility was an important factor for predicting the distributions of all taxonomic groups.

3.4 Response Curves

Figure 2 shows the response curves generated to visualize the effect of each environmental variable on the observation probability predicted by Model 2, revealing that changing a single environmental variable while keeping all other environmental variables fixed at their average sample values had a marginal effect on the predicted probability. The response curves for the accessibility categories generally remained similar across taxonomic groups. Increasing the PWF proportion steeply increased the observation probability of all taxonomic groups, strongly suggesting that observation records were more likely to occur along forests adjacent to public roads, which offered the easiest access. Increasing the proportion of PIF also generally increased the observation probability, although the slope was less steep or saturated at lower values for some taxonomic groups. This phenomenon may reflect the more restricted access. Increasing the RF proportion decreased the observation probability or kept it low for most taxonomic groups. These results support the hypothesis that observation records are less likely to be generated within forests more difficult to access.

The response curves for the other land-cover categories showed highly pronounced differences among the taxonomic groups. Increasing the water related proportion increased the observation probability of reptiles, amphibians, and birds. Insects and plants generally exhibited weaker responses to the various natural land-cover categories. However, all taxonomic groups showed interesting responses to built-up areas and croplands. Increasing the built-up area proportion tended to increase the observation probability for all taxonomic groups, whereas the opposite trend was observed for cropland. These results suggested that human factors (i.e., accessibility) dominated the observation probability of a given species.

4. Discussion

Model 2 consistently outperformed Model 1 in predicting the observations of all five taxonomic groups, indicating that accounting for human accessibility remarkably enhances our ability to explain the localization of observation records. This result supported our recognition that when MaxEnt models are applied to citizen science data, they predict the environment where observers are present rather than the true distribution of species. This behavior also reflects a distinct spatial bias (sampling bias and observation efforts) that is inherent to citizen

Taxon Group	Variable Importance	Environments	PIF	PWF	RF	Built-up	Cropland	Grassland	Rice paddies	Water bodies	Others
Reptiles	Percent Contribution	taken point	0.45	1.34	3.02	4.62	3.80	2.24	2.35	10.06	0.47
		surrounding	12.04	8.27	8.71	1.51	19.93	1.80	6.46	12.80	0.12
	Permutation Importance	taken point	0.08	2.26	0.00	21.45	0.63	5.38	0.00	3.02	0.08
		surrounding	1.68	14.01	9.49	5.22	8.12	3.65	8.07	16.73	0.15
	Jackknife Test: without	taken point	0.76	0.76	0.77	0.76	0.76	0.77	0.76	0.77	0.76
		surrounding	0.77	0.75	0.78	0.76	0.77	0.74	0.77	0.76	0.77
	Jackknife Test: only	taken point	0.53	0.52	0.55	0.58	0.57	0.55	0.54	0.60	0.51
		surrounding	0.59	0.61	0.56	0.57	0.63	0.50	0.56	0.62	0.52
Amphibians	Percent Contribution	taken point	2.19	16.84	5.22	14.66	0.13	2.05	6.24	1.38	0.44
		surrounding	1.27	13.88	2.17	8.34	4.59	3.86	13.92	2.62	0.21
	Permutation Importance	taken point	0.91	2.51	7.29	3.93	0.35	8.24	16.54	4.00	4.55
		surrounding	1.32	6.01	4.15	1.75	8.71	5.47	17.69	5.35	1.24
	Jackknife Test: without	taken point	0.65	0.63	0.59	0.64	0.61	0.63	0.63	0.63	0.63
		surrounding	0.63	0.60	0.61	0.64	0.62	0.62	0.58	0.62	0.63
	Jackknife Test: only	taken point	0.54	0.61	0.51	0.45	0.53	0.52	0.53	0.42	0.47
		surrounding	0.62	0.67	0.51	0.47	0.54	0.51	0.55	0.62	0.46
Insects	Percent Contribution	taken point	0.01	1.53	4.93	2.21	1.66	0.11	7.54	0.30	0.29
		surrounding	9.39	6.37	14.23	7.80	40.16	1.92	1.30	0.24	0.01
	Permutation Importance	taken point	0.19	2.23	2.94	2.72	2.95	4.52	0.83	2.17	0.23
		surrounding	6.51	25.91	9.87	10.80	13.18	5.83	3.35	5.77	0.00
	Jackknife Test: without	taken point	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
		surrounding	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71	0.71
	Jackknife Test: only	taken point	0.53	0.54	0.54	0.59	0.56	0.53	0.53	0.52	0.51
		surrounding	0.58	0.55	0.57	0.61	0.61	0.54	0.55	0.53	0.51
Plants	Percent Contribution	taken point	0.00	2.31	5.91	9.54	2.68	2.39	5.26	0.41	0.00
		surrounding	8.04	2.08	9.67	3.22	42.65	1.85	2.72	0.66	0.60
	Permutation Importance	taken point	0.00	3.17	2.96	0.00	0.97	0.00	1.60	2.54	0.00
		surrounding	6.65	31.80	4.62	19.61	7.76	3.31	2.69	11.86	0.44
	Jackknife Test: without	taken point	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68
		surrounding	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68	0.68
	Jackknife Test: only	taken point	0.53	0.54	0.53	0.57	0.56	0.52	0.53	0.52	0.51
		surrounding	0.57	0.55	0.56	0.59	0.60	0.53	0.54	0.53	0.51
Birds	Percent Contribution	taken point	0.00	0.84	0.53	1.50	0.00	0.06	1.30	3.09	0.05
		surrounding	19.42	4.60	11.29	0.96	16.08	2.84	0.83	36.59	0.00
	Permutation Importance	taken point	0.01	1.26	4.09	7.62	0.00	0.67	0.73	0.09	0.07
		surrounding	0.09	10.41	16.14	3.22	22.59	2.58	2.36	28.06	0.00
	Jackknife Test: without	taken point	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82
		surrounding	0.82	0.82	0.81	0.82	0.81	0.81	0.82	0.81	0.82
	Jackknife Test: only	taken point	0.53	0.58	0.55	0.52	0.58	0.53	0.53	0.62	0.51
		surrounding	0.66	0.66	0.60	0.54	0.65	0.57	0.56	0.69	0.52

Table 2. Variable importance metrics for Model 2 across taxonomic groups. In the Environment column, "taken point" represents the 3 × 3 cell neighborhood while "surrounding" represents the 9 × 9 cell neighborhood. Darker colors indicate that the selected environmental variable has a higher importance value according to the selected metric for the selected taxonomic group (rows).

science data (Carlen et al., 2024) and primarily related to where observers visit (You et al., 2022), which should be distinguished from species detectability itself. The dominant influence of human factors was most evident in the responses to the different accessibility categories. The steep positive response curves for PWF demonstrate that observation records are concentrated along easily accessible roads and probably reflect records taken opportunistically during walks or commutes. The high permutation importance of PWF further highlights the reliance of the model on this category for accurate predictions. PIF also had a positive association with observation records albeit generally weaker than PWF. This suggests that PIF serves as a destination for wildlife observation but may experience less incidental traffic. Conversely, the negative or flat response curves for RF strongly agree with the hypothesis that areas with difficult or restricted access receive far fewer observations regardless of their ecological value (Sarmento and Berger, 2017) and reflect the biases inherent to observation records, which are primarily obtained in locations that people can and do visit (Carlen et al., 2024). This result strongly suggested that numerous observations

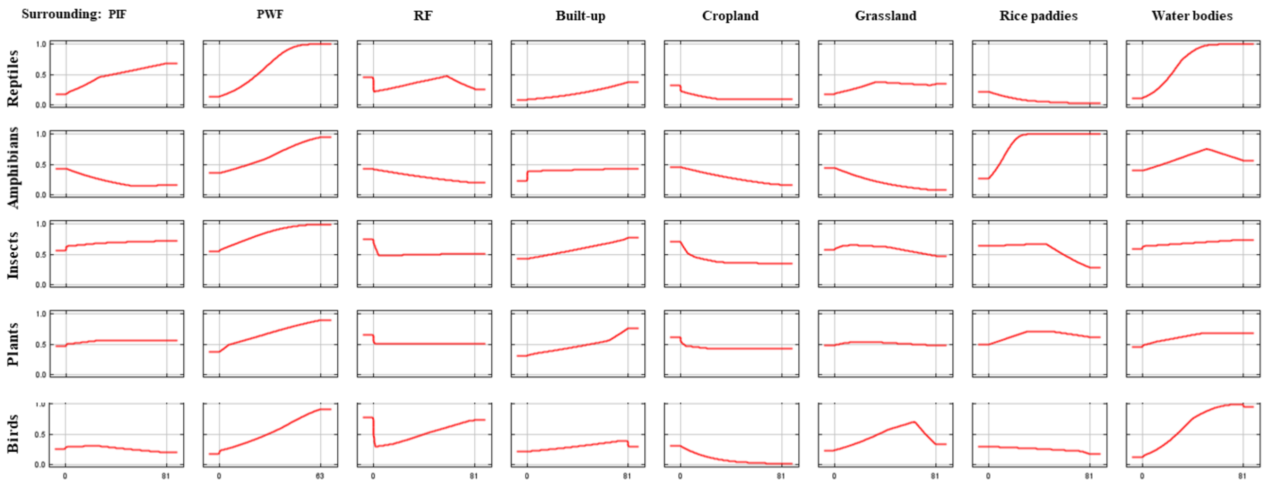

Figure 2. Response curves for Model 2 showing the marginal effect of environmental variables on the observation probability. The x-axis represents the cell count for each land-cover category within the 9 × 9 cell neighborhood whereas the y-axis shows the predicted observation probability (logistic output). The curves illustrate the relationship when only the selected variable is varied while the other variables are fixed at their average values.

are likely to be reported in accessible areas, even if they are not ecologically rich, whereas fewer observations may be reported in ecologically rich but difficult-to-access areas.

Despite the dominance of accessibility, the ecological factors also demonstrated relevance. Water related environments were important for the observation of reptiles, amphibians, and birds, aligning with their known habitat requirements and the ecological importance of riparian zones within urban landscapes (Riis et al., 2020). This finding indicated that ecologically suitable environments can form more prominent observation hotspots when combined with accessibility. However, the responses to human-dominated land-cover categories revealed intriguing patterns that potentially indicate human factors overriding simple ecological constraints. The positive responses of all taxonomic groups to an increase in built-up areas contradicted the typical expectation of urbanization limiting biodiversity. This trend does not necessarily imply that built-up areas are superior habitats for diverse species but rather reflects a high observer density near residential areas, the increased use of small green patches within built-up zones, or specific features associated with built-up edges that attract observers. Similarly, the negative response to cropland across all taxonomic groups may be related more to restricted access or observers preferring forests and parks rather than cropland being an inherently poor habitat for the taxonomic groups. These findings indicated that calibrating models with citizen science data strongly reflects human presence and observation opportunities (Carlen et al., 2024; Sarmento and Berger, 2017; You et al., 2022).

These results have several implications for the use of citizen science data. First, interpreting the raw data as a direct proxy for biodiversity richness in urban areas is clearly problematic, and biases owing to accessibility must be accounted for. Second, the accessibility categories (i.e., PWF, PIF, and RF) offer a practical approach for incorporating human factors into distribution models or survey designs based on LULC data. Third, for urban planning aimed at fostering human–nature connections and supporting Nature Positive goals, our findings advocated the network creation of highly accessible green spaces (PWFs and PIFs) and the strategic management of less accessible areas (RFs) for core habitat functions. The proposed forest classification provides urban planners and managers with concrete design guidelines for securing green spaces and encouraging citizens to regularly engage with nature, recognize biodiversity value, and ultimately foster behavioral changes toward nature conservation. Furthermore, as the societal awareness of contributing to data expansion grows, our methodology, leveraging open data and citizen science, enables the quantification of the relative value of urban forests.

The limitations of this study include its focus on a single city, reliance on a single social media platform (Biome), the possible misidentification of observation records (Bonney et al., 2009; Dickinson et al., 2010), and the proxy nature of the selected accessibility categories. Further limitations inherent to citizen science data, such as the potential biases stemming from the demographics and varying behavioral patterns of the observers, seasonality in observations and corresponding target species/frequencies, visual attractiveness of certain species, and interface design of citizen science applications (Atsumi et al., 2024; Carlen et al., 2024; Jacobs and Zipf, 2017), should be considered. Future work may involve comparing observation records across cities and platforms and integrating finer-scale data on human movement. In particular, the model can be improved by considering the biases estimated from participation behavior using citizen science applications.

5. Conclusion

Human accessibility was demonstrated to be a critical and often dominant factor shaping the distribution of biodiversity observations within the urban forests of Tsukuba Science City. By explicitly dividing the forest land cover into categories based on accessibility (PWF, PIF, and RF), we improved the ability of the developed model to predict observation locations compared with using basic land-cover categories alone. The observations of five diverse taxonomic groups were concentrated in easily accessible areas, highlighting the bias of accessibility inherent to data obtained by citizen science. Although ecological factors, such as proximity to water, remained important for specific taxonomic groups (Riis et al., 2020), the influence of accessibility underscores the need to integrate human behavioral patterns when interpreting citizen science data derived from urban environments. We elucidated the drivers of data submission patterns for citizen science and thus facilitated its application to biodiversity conservation. These findings are crucial for accurately assessing biodiversity patterns, designing effective citizen science projects that account for spatial biases, and informing urban planning strategies that aim to enhance biodiversity and human engagement with nature. Ultimately, recognizing and incorporating the human dimension is essential for effectively leveraging citizen science to achieve conservation objectives and foster Nature Positive societies in an increasingly urbanized world.

Acknowledgements

This study was partially supported by the Support program of FFPRI for researchers having family obligations and JSPS KAKENHI Grant Number 24K20980. We would like to thank Biome Inc., Keisuke Atsumi, Shotaro Matsuda, Saya Sunada and Tomoro Mukai for data contribution.

References

Amano, T., Lamming, J.D.L., Sutherland, W.J., 2016: Spatial gaps in global biodiversity information and the role of citizen science. *BioScience*, 66(5), 393-400. doi.org/10.1093/biosci/biw022.

Asari, Y., and Fujiki. S., 2020: Report on a preliminary test of roadkill observation system in Amami-Oshima Island, Japan. *Wildlife and Human Society*, 8, 43-45. 10.20798/awhswhs.8.0_43.

Atsumi, K., Nishida, Y., Ushio, M., Nishi, H., Genroku, T., Fujiki, S., 2024: Boosting biodiversity monitoring using smartphone-driven, rapidly accumulating community-sourced data. *eLife*, 13, RP93694. doi.org/10.7554/eLife.93694.3.

Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V., Shirk, J., 2009: Citizen science: A developing tool for expanding science knowledge and scientific literacy. *Bioscience,* 59(11), 977-984. doi.org/10.1525/bio.2009.59.11.9.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., Narwani, A., Mace, G.M., Tilman, D., Wardle, D.A., Kinzig, A.P., Daily, G.C., Loreau, M., Grace, J.B., Larigauderie, A., Srivastava, D.S., Naeem, S., 2012: Biodiversity loss and its impact on humanity. *Nature*, 486(7401), 59-67. doi.org/10.1038/nature11148.

Carlen, E.J., Estien, C.O., Caspi, T., Perkins, D., Goldstein, B.R., Kreling, S.E.S., Hentati, Y., Williams, T.D., Stanton, L.A., Des Roches, S., Johnson, R.F., Young, A.N., Cooper, C.B., Schell, C.J., 2024: A framework for contextualizing social-ecological biases in contributory science data. *People and Nature*, 6(2), 377-390. doi.org/10.1002/pan3.10592.

Cohn, J.P., 2008: Citizen science: Can volunteers do real research? *Bioscience*, 58(3), 192-197. doi.org/10.1641/B580303.

Costanza, R., de Groot, R., Braat, L., Kubiszewski, I., Fioramonti, L., Sutton, P., Farber, S., Grasso, M., 2017: Twenty years of ecosystem services: How far have we come and how far do we still need to go? *Ecosystem Services*, 28, 1-16. doi.org/10.1016/j.ecoser.2017.09.008.

Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R.T., Molnár, Z., Hill, R., Chan, K.M.A., Baste, I.A., Brauman, K.A., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P.W., van Oudenhoven, A.P.E., van der Plaat, F., Schröter, M., Lavorel, S., Aumeeruddy-Thomas, Y., Bukvareva, E., Davies, K., Demissew, S., Erpul, G., Failler, P., Guerra, C.A., Hewitt, C.L., Keune, H., Lindley, S., Shirayama, Y., 2018: Assessing nature's contributions to people. *Science*, 359(6373), 270-272. doi.org/10.1126/science.aap8826.

Dickinson, J.L., Zuckerberg, B., Bonter, D.N., 2010: Citizen science as an ecological research tool: Challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics*, 41(1), 149-172. doi.org/10.1146/annurev-ecolsys-102209-144636.

e-Stat, 2024. Government Statistics Portal of Japan: Statistical GIS. e-Stat. https://www.e-stat.go.jp/gis (25 March 2022).

Fourcade, Y., Engler, J.O., Rödder, D., Secondi, J., 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. doi.org/10.1371/journal.pone.0097122.

Geospatial Information Authority of Japan, 2024a. FG-GSI: Fundamental Geospatial Data Download Service - Road Edge. FG-GSI: Fundamental Geospatial Data Download Service. https://fgd.gsi.go.jp/download/menu.php (14 August 2024).

Geospatial Information Authority of Japan, 2024b. GitHub - The source of GSI Maps. GitHub - The Source of GSI Maps. https://www.google.com/url?q=https://github.com/gsi-cyberjapan/gsimaps&sa=D&source=docs&ust=1745554714478160&usg=AOvVaw1FwNL3GeBUfxwxkTe9QVvn (12 September 2024).

Geospatial Information Authority of Japan, 2017. GitHub - GSI-CyberJapan/Vector-Tile-Experiment. https://github.com/gsi-cyberjapan/vector-tile-experiment (12 September 2024).

Ghermandi, A., Sinclair, M., 2019: Passive crowdsourcing of social media in environmental research: A systematic map. *Global Environmental Change*, 55, 36-47. doi.org/10.1016/j.gloenvcha.2019.02.003.

Gould, R.K., Morse, J.W., Adams, A.B., 2019: Cultural ecosystem services and decision-making: How researchers describe the applications of their work. *People and Nature*, 1(4), 457-475. doi.org/10.1002/pan3.10044.

Grabska-Szwagrzyk, E., Hashemvand Khiabani, P., Pesoa-Marcilla, M., Chaturvedi, V., de Vries, W.T., 2024: Exploring land use dynamics in rural areas. An analysis of eight cases in the Global North. *Land Use Policy*, 144, 107246. doi.org/10.1016/j.landusepol.2024.107246.

G Space Information Center, 2023. Registered maps at the Legal Affairs Bureau (cadastral maps): Tsukuba City. Registered maps at the Legal Affairs Bureau (cadastral maps). https://front.geospatial.jp/moj-chizu-shp-download/ (1 August 2024).

Jacobs, C., Zipf, A., 2017: Completeness of citizen science biodiversity data from a volunteered geographic information perspective. *Geo-spatial Information Science*, 20(1), 3-13. doi.org/10.1080/10095020.2017.1288424.

Japan Aerospace Exploration Agency, 2023. High Resolution Land Use and Land Cover Map of Japan [2022] (Version 23.12). Homepage of High-Resolution Land Use and Land Cover Map Products.

https://www.eorc.jaxa.jp/ALOS/jp/dataset/lulc/lulc_v2312_j.htm (22 July 2024).

Jingu, S., 2020: Temporal continuities of grasslands and forests as patches of natural land in urban landscapes: A case study of the Tsukuba science city. *Land*, 9(11), 425. doi.org/10.3390/land9110425.

Kindsvater, H.K., Dulvy, N.K., Horswill, C., Juan-Jordá, M.-J., Mangel, M., Matthiopoulos, J., 2018: Overcoming the data crisis in biodiversity conservation. *Trends in Ecology & Evolution*, 33(9), 676-688. doi.org/10.1016/j.tree.2018.06.004.

Kobori, H., Dickinson, J.L., Washitani, I., Sakurai, R., Amano, T., Komatsu, N., Kitamura, W., Takagama, S., Koyama, K., Ogawara, T., Miller-Rushing, A.J., 2016: Citizen science: A new approach to advance ecology, education, and conservation. *Ecological Research*, 31(1), 1-19. doi.org/10.1007/s11284-015-1314-y.

Morschheuser, B., Hamari, J., Koivisto, J., Maedche, A., 2017: Gamified crowdsourcing: Conceptualization, literature review, and future agenda. *International Journal of Human-Computer Studies*, 106, 26-43. doi.org/10.1016/j.ijhcs.2017.04.005.

Nugent, J., 2018: iNaturalist: Citizen science for 21st-century naturalists. *Science Scope*, 41(7), 12-15. doi.org/10.2505/4/ss18_041_07_12.

OpenStreetMap Foundation, 2024. OpenStreetMap. OpenStreetMap. http://openstreetmap.org (12 September 2024).

Oxford English Dictionary, 2014. *Citizen Science*. Oxford University Press, Oxford.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006: Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190(3-4), 231-259. doi.org/10.1016/j.ecolmodel.2005.03.026.

Phillips, S.J., Dudík, M., Schapire, R.E., 2020. Maxent Software for Modeling Species Niches and Distributions (Version 3.4.4). Maxent. http://biodiversityinformatics.amnh.org/open_source/maxent/ (24 April 2025).

Riis, T., Kelly-Quinn, M., Aguiar, F.C., Manolaki, P., Bruno, D., Bejarano, M.D., Clerici, N., Fernandes, M.R., Franco, J.C., Pettit, N., Portela, A.P., Tammeorg, O., Tammeorg, P., Rodríguez-González, P.M., Dufour, S., 2020: Global overview of ecosystem services provided by riparian vegetation. *Bioscience*, 70(6), 501-514. doi.org/10.1093/biosci/biaa041.

Sarmento, W.M., Berger, J., 2017: Human visitation limits the utility of protected areas as ecological baselines. *Biological Conservation*, 212, 316-326. doi.org/10.1016/j.biocon.2017.06.032.

Schröter, M., Kraemer, R., Mantel, M., Kabisch, N., Hecker, S., Richter, A., Neumeier, V., Bonn, A., 2017: Citizen science for assessing ecosystem services: Status, challenges and opportunities. *Ecosystem Service*, 28, 80-94. doi.org/10.1016/j.ecoser.2017.09.017.

Simpson, J., Thwaites, K., Freeth, M., 2019: Understanding visual engagement with urban street edges along non-pedestrianised and pedestrianised streets using mobile eye-tracking. *Sustainability*, 11(15), 4251. doi.org/10.3390/su11154251.

Stephens, T., 2023: The Kunming–montreal global biodiversity framework. *International Legal Materials*, 62(5), 868-887. doi.org/10.1017/ilm.2023.16.

Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009: eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282-2292. doi.org/10.1016/j.biocon.2009.05.006.

Takahashi, M., Nasahara, K.N., Tadono, T., Watanabe, T., Dotsu, M., Sugimura, T., Tomiyama, N., 2013: JAXA high resolution land-use and land-cover map of Japan. 2013 *IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, 2384-2387. doi.org/10.1109/igarss.2013.6723299.

Tsukuba City Government, 2024. *Tsukuba Statistical Yearbook* 2023. Tsukuba City Government, TsukubaJapan.

Tsutsumida, N., Nagai, S., Rodríguez-Veiga, P., Katagi, J., Nasahara, K., Tadono, T., 2019: Mapping spatial accuracy. *of* forest type classification in jaxa's high–resolution land use and land cover map, in: joint workshop "multidisciplinary *remote sensing* for environmental monitoring." *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-3-W1, 57-63. doi.org/10.5194/isprs-annals-IV-3-W1-57-2019.

United Nations Conservation on Biological Diversity, 2022: Monitoring framework for the Kunming-Montreal global biodiversity framework. *Fifteenth Meeting of the Conference of the Parties to the Convention on Biological Diversity (Part Two)*, 20-45.

Watkins, R.Z., Chen, J., Pickens, J., Brosofske, K.D., 2003: Effects of forest roads on understory plants in a managed hardwood landscape. *Conservation Biology*, 17(2), 411–419. doi.org/10.1046/j.1523-1739.2003.01285.x.

You, S., Zheng, Q., Chen, B., Xu, Z., Lin, Y., Gan, M., Zhu, C., Deng, J., Wang, K., 2022: Identifying the spatiotemporal dynamics of forest ecotourism values with remotely sensed images and social media data: A perspective of public preferences. *Journal of Cleaner Production*, 341, 130715. doi.org/10.1016/j.jclepro.2022.130715.