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

C.P. Mandeville[1], E.B. Nilsen[2,3], A.G. Finstad[4]

[1]Centre for Biodiversity Dynamics, Department of Natural History, Norwegian University of Science and Technology. Erling Skakkes Gate 47B, 7021 Trondheim, Norway.

caitlin.mandeville@ntnu.no

[2]Norwegian Institute for Nature Research, Postboks 5685 Torgarden, 7485 Trondheim, Norway.

[3]Nord University, Kongens gate 42, 7713 Steinkjer, Norway.

[4]Centre for Biodiversity Dynamics, Department of Natural History, Norwegian University of Science and Technology. Erling Skakkes Gate 47B, 7021 Trondheim, Norway.

Abstract

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

Keywords

biodiversity; biodiversity data; citizen science; community science; protected areas; recreation ecology

Introduction

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

Biodiversity data from mass participation citizen science could play a greater role in filling a critical data gap for small protected areas, green spaces, and other multiple-use areas that contribute to other effective area-based conservation measures (OECMs [IUCN 2019]) (Schmeller et al. 2017, Maxwell et al. 2020, Adams et al. 2021), which are increasingly recognized as crucial for meeting biodiversity conservation targets (Kendal et al. 2017, Baldwin and Fouch 2018, Bonnet et al. 2020, Häkkilä et al. 2021, Rodríguez-Rodríguez et al.

2021). Such areas enhance connectivity, support ecosystem services, and play a key role in addressing environmental threats that manifest at a local scale (Oldekop et al. 2015, Volenec and Dobson 2019, Wintle et al. 2019, Hlásny et al. 2021, Gaget et al. 2021, Dreiss and Malcolm 2022). Still, small natural areas often have limited resources for biodiversity conservation, management, and monitoring, despite their high conservation value (Armsworth et al. 2011, Maxwell et al. 2020, Jansujwicz et al. 2020).

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

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

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

We aimed to respond to this call by investigating the spatial distribution of citizen science participation at a scale relevant to local area management. We took a case study approach, characterizing citizen science activity within a small, recreationally popular natural area in

Central Norway. The site was selected because it is a regional citizen science hotspot. Our objectives were to 1) test the hypothesis that the main predictors of citizen science activity at a broad spatial resolution—accessibility and natural interest—also drive citizen science at a local scale; 2) compare the distribution of citizen science activity throughout the study area with that of professional biodiversity data collection; 3) test the hypothesis that citizen science activity would primarily occur within a short distance of trails and roads; and 4) compare the distribution of citizen science activity along the area's trail network with that of other recreational trail use, represented by activity tracking data from Strava Metro.

Methods

2.1 Study site

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

2.2 Data

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

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

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

2.3 Analysis

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

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

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

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

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

Next, we conducted a small exploratory analysis intended to provide insight into whether citizen science participants tend to make observations from the trail or to leave the trail before making observations. We expected that if participants tend to make observations from the trail, then the distance between the recorded observation coordinates and the nearest trail would be greatest for taxonomic groups that are more often visible and identifiable from a distance (e.g., mammals, birds, plants). If participants tend to leave the trail to make observations, then we would not expect this relationship. We used a Kruskal-Wallis test to examine the relationship between distance to trail and the observed taxonomic group (grouped in the following categories: birds, fungi, invertebrates, mammals, plants, and reptiles/amphibians) and a Dunn’s post-hoc test for pairwise comparisons between taxonomic groups. This analysis was repeated with the professional dataset.

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

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

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

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

Results

3.1 Biodiversity data

The filtered citizen science data consisted of 44206 observations from seven citizen science platforms. The vast majority (91%) were contributed through the Norwegian Species Observation Service (https://www.biodiversity.no/), which is Norway's main biodiversity citizen science platform. Citizen science data were contributed by 560 participants. As is typical of digital citizen science datasets (Wood et al. 2011, Boakes et al. 2016, Rowley et al. 2019), a small number of highly active participants contributed the majority of the data; the most active five percent of participants contributed 79% of the total data, while the median participant contributed just six observations. The filtered professionally collected data available on GBIF consisted of 2059 observations from 31 data providers.

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

3.2 Environmental covariates of citizen science activity and professional data collection: grid-based analysis.

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

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

3.3 Relationship between citizen science and trail network.

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

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

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

The tested covariates had limited ability to explain variation in citizen science activity among trail segments; four models had a substantial level of support, totaling 24.2% of the weight of evidence (Table S3.3, Figure S3.4). All effect sizes were relatively small compared to the grid-based models (Figure 3, Figure 7, Figure S3.1, Table S3.2). Notably, most covariates that were important at the grid scale were not important to describe variation between trail segments; proximity to the nearest access point, the eastern area edge, and freshwater were important at the grid level but had only a small and uncertain relationship to citizen science activity along trail segments. The most important variable was forest cover, which had a small negative relationship to citizen science activity.

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

Discussion

We responded to calls for research on citizen science within protected and other natural areas by examining citizen science activity in a small natural area that serves as a regional citizen science hotspot. Our results illustrate that citizen science participation is spatially heterogeneous on a local scale. Ease of area access was the dominant landscape characteristic driving the distribution of citizen science in our study area, and a key component of accessibility is the use of a trail network to access observation sites. However, the distribution of citizen science activity along the trail network differed from that of other trail users. In general, citizen science activity was more evenly dispersed over a wider range of trail characteristics than other trail use; for example, citizen science participants were more likely than other trail users to spend time both in more developed parts of the natural area and also on less well-established paths that do not function as main travel routes.

The importance of area access is a notable result of our study. It is known that accessibility and natural interest are major regional determinants of citizen science activity, but our results are among the first to show that, within a small natural area, accessibility has a stronger relationship to citizen science than particular landscapes perceived as the most natural. To the contrary, citizen science activity was positively associated with cultivated and developed land within the area. This may be partially explained by the increased accessibility afforded by infrastructure in these areas, or by interest in the biodiversity of these land cover types. But it may also stem from an affinity for cultivated and developed land cover, as suggested by recent findings that the integration of biodiversity with cultural and agricultural heritage plays an important role in communities' relationship to natural areas (Cusens et al. 2021). Proximity to water was positively associated with citizen science activity in our study area, as has previously been shown at regional scales and within urban areas (Boakes et al. 2016, Tiago et al. 2017). This could be explained by trends in either participant behavior (e.g., participants might prefer spending time near water or observing species found near water) or in species availability (e.g., landscapes containing freshwater may be more species-rich or afford greater detectability for species that are present).

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

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

When comparing citizen science and professionally collected data, is important to note that the professional data available on GBIF for a natural area is almost certainly not a complete record of professional biodiversity data that has been collected in the area; while the value of openly sharing data is increasingly recognized, barriers still prevent much biodiversity data from being shared (Mandeville et al. 2021). Many small natural areas also support locally managed, place-based citizen science programs that are typically structured to address specific research and monitoring questions (Mandeville and Finstad 2021, Rosemartin et al. 2021). Such programs are highly valuable but are often resource-intensive to coordinate at a local level and therefore may not be feasible to implement in all settings (Tancoigne 2019, Rosemartin et al. 2021, Alfonso et al. 2022). Further, they often produce data that are not openly shared on GBIF (Mandeville et al. 2021). For this reason, the open biodiversity data collected through opportunistic citizen science platforms are particularly valuable for their relative ease of access, allowing them to fill gaps both when other data do not exist and when other data cannot be accessed. In parallel with increasing the utility of citizen science data for area management, it is critical to continue increasing area managers' access to other existing data sources; among other reasons, this is because opportunistic citizen science data are often most valuable when integrated with other data types (Kühl et al. 2020, Dobson et al. 2020, Mandeville et al. 2021).

In addition to informing more effective analysis of existing citizen science data, knowledge of citizen science activity patterns can be used by area managers to promote and guide future data collection. First, managers could use knowledge about citizen science trends to reach out to current participants to prompt collection of data to meet specific monitoring needs, for example by posting signs in areas regularly frequented by citizen science participants or communicating through customization features offered by citizen science platforms (Callaghan et al. 2021, Koen and Newton 2021, Gosal et al. 2021). Second, managers could identify areas of low citizen science activity to target for recruiting new participants (Weaver and Lawton 2017). For instance, recreational facilities were not closely associated with citizen science participation in our study area; collaboration with relevant recreational organizations and facilities to promote citizen science could more firmly ground local citizen science participation in a sense of place and engage recreational visitors who do not yet participate in citizen science (Newman et al. 2017, Allf et al. 2022). Finally, managers may be able to prioritize professional data collection to complement citizen science by emphasizing areas of low citizen science activity.

Knowledge of spatial trends in citizen science activity can further inform overall management strategies for natural areas and green spaces. The needs and preferences of area visitors are regularly used to make management decisions about natural areas and even to justify ongoing area protection, but because different subsets of visitors prioritize different types of area management, it can be challenging to identify the diverse needs of area visitors (Hornigold et al. 2016, Mancini et al. 2018, Muñoz et al. 2020, Komossa et al. 2021). Our results show that citizen science participants in our study area tend to use the area's trail network differently than other visitors, so their needs may be overlooked if not explicitly considered. Citizen science participants may even serve as a useful proxy to represent a broader group of nature-oriented visitors whose area use might differ in similar ways from the more activity-oriented visitors captured in the Strava Metro data (Havinga et al. 2020, Cambria et al. 2021). The Strava Metro dataset is itself biased towards visitors with a focus on athletic recreation, though a recent study elsewhere in Norway found a high correlation between Strava activities and absolute counts of segment users, suggesting that Strava is relatively representative of the dominant trends in segment use, particularly in areas of high activity (Venter et al. 2021).

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

Acknowledgements

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

Conflict of interest

The authors declare no conflicts of interest.

Author contributions

CM conceived of the idea, analyzed the data, and led the writing of the manuscript. EN and AF supported the conceptual development and writing of the manuscript. All authors contributed to the drafts and gave approval for publication.

Data availability

Biodiversity data are obtained from the Global Biodiversity Information Facility and are available here: https://doi.org/10.15468/dl.pd3tce (GBIF 2021). This report includes aggregated and de-identified data from Strava Metro. R scripts for the analyses are available from Open Science Framework: 10.17605/OSF.IO/BGF3D (Mandeville et al. 2022).

References

Adams VM, Visconti P, Graham V, Possingham HP. 2021. Indicators keep progress honest: A call to track both the quantity and quality of protected areas. One Earth 4: 901-906.

Ahlstrøm A., Bjørkelo K., Frydenlund J. 2014. AR5 klassifikasjonssyste: Klassifikasjon av Arealressurser. Rapport 6. Ås: Norsk institutt for skog og landskap.

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

Allf BC, Cooper CB, Larson LR, Dunn RR, Futch SE, Sharova M, Cavalier D. 2022. Citizen science as an ecosystem of engagement: implications for learning and broadening participation. BioScience 72: 651-663.

Armsworth PR, Cantú-Salazar L, Parnell M, Davies ZG, Stoneman R. 2011. Management costs for small protected areas and economies of scale in habitat conservation. Biological Conservation 144: 423–429.

Baldwin RF, Fouch NT. 2018. Understanding the biodiversity contributions of small protected areas presents many challenges. Land 7: 123.

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

Bardos DC, Guillera-Arroita G, Wintle BA. 2015. Valid auto-models for spatially autocorrelated occupancy and abundance data. Methods in Ecology and Evolution 6: 1137–1149.

Binley AD, Proctor CA, Pither R, Davis SA, Bennett JR. 2021. The unrealized potential of dommunity science to support research on the resilience of protected areas. Conservation Science and Practice 3: e376.

Boakes EH, Gliozzo G, Seymour V, Harvey M, Smith C, Roy DB, Haklay M. 2016. Patterns of contribution to citizen science biodiversity projects increase understanding of volunteers' recording behaviour. Scientific Reports 6: 33051.

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

Bonney R. 2021. Expanding the Impact of Citizen Science. BioScience 71: 448–451. Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: A practical information-theoretic approach. New York: Springer. 488 p.

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

Burnham KP and Anderson DR. 2002. Model Selection and Inference: A Practical Information-Theoretic Approach. 2nd Edition, Springer-Verlag, New York.

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

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

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

Callaghan CT, Rowley JJL, Cornwell WK, Poore AGB, Major RE. 2019. Improving big citizen science data: Moving beyond haphazard sampling. PLOS Biology 17: e3000357.

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

Cambria VE, Campagnaro T, Trentanovi G, Testolin R, Attorre F, Sitzia T. 2021. Citizen science data to measure human use of green areas and forests in European cities. Forests 12: 779.

Cheung SY, Leung Y-F, Larson LR. 2022. Citizen science as a tool for enhancing recreation research in protected areas: Applications and opportunities. Journal of Environmental Management 305: 114353.

Cooper CB, Shirk J, Zuckerberg B. 2014. The Invisible Prevalence of Citizen Science in Global Research: Migratory Birds and Climate Change. PLOS ONE 9: e106508.

Cretois B, Simmonds EG, Linnell JDC, van Moorter B, Rolandsen CM, Solberg EJ, Strand O, Gundersen V, Roer O, Rød JK. 2021. Identifying and correcting spatial bias in opportunistic citizen science data for wild ungulates in Norway. Ecology and Evolution 11: 15191–15204.

Cusens J, Barraclough AMD, Måren IE. 2021. Participatory mapping reveals biocultural and nature values in the shared landscape of a Nordic UNESCO Biosphere Reserve.

People and Nature 2021.

Di Cecco GJ, Barve V, Belitz MW, Stucky BJ, Guralnick RP, Hurlbert AH. 2021. Observing the observers: how participants contribute data to iNaturalist and implications for biodiversity science. BioScience 71: 1179–1188.

Dobson ADM, Milner-Gulland EJ, Aebischer NJ, Beale CM, Brozovic R, Coals P, Critchlow R, Dancer A, Greve M, Hinsley A, Ibbett H, Johnston A, Kuiper T, Le Comber S, Mahood SP, Moore JF, Nilsen EB, Pocock MJO, Quinn A, Travers H, Wilfred P, Wright J, Keane A. 2020. Making messy data work for conservation. One Earth 2: 455–465.

Dreiss LM, Malcom JW. Identifying key federal, state, and private lands strategies for achieving 30 × 30 in the United States. Conservation Letters 15: e12849.

Feldman MJ, Imbeau L, Marchand P, Mazerolle MJ, Darveau M, Fenton NJ. 2021. Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. PLOS ONE 16: e0234587.

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

Gaget E, Johnston A, Pavón-Jordán D, Lehikoinen AS, Sandercock BK, Soultan A, Božič L, Clausen P, Devos K, Domsa C, Encarnação V, Faragó S, Fitzgerald N, Frost T, Gaudard C, Gosztonyi L, Haas F, Hornman M, Langendoen T, Ieronymidou C, Luigujõe L, Meissner W, Mikuska T, Molina B, Musilová Z, Paquet J-Y, Petkov N, Portolou D, Ridzoň J, Sniauksta L, Stīpniece A, Teufelbauer N, Wahl J, Zenatello M, Brommer JE. Protected area characteristics that help waterbirds respond to climate warming. Conservation Biology: 10.1111/cobi.13877.

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

GBIF. 3 August 2021. GBIF Occurrence Download https://doi.org/10.15468/dl.pd3tce

Geldmann J, Heilmann-Clausen J, Holm TE, Levinsky I, Markussen B, Olsen K, Rahbek C, Tøttrup AP. 2016. What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. Diversity and Distributions 22: 1139–1149.

Gosal AS, McMahon JA, Bowgen KM, Hoppe CH, Ziv G. 2021. Identifying and mapping groups of protected area visitors by environmental awareness. Land 10: 560.

Gurney GG, Darling ES, Ahmadia GN, Agostini VN, Ban NC, Blythe J, Claudet J, Epstein G, Estradivari, Himes-Cornell A, Jonas HD, Armitage D, Campbell SJ, Cox C, Friedman WR, Gill D, Lestari P, Mangubhai S, McLeod E, Muthiga NA, Naggea J, Ranaivoson R, Wenger A, Yulianto I, Jupiter SD. 2021. Biodiversity needs every tool in the box: use OECMs. Nature 595: 646–649.

Hagen D, Vistad OI, Rød-Eriksen L. 2019. Sårbarhetsvurdering av utvalgte ferdselslokaliteter i Bymarka naturreservat. Kvistingstien, Gråkallen rundt, St. Olavsspranget. NINA Rapport 1680. Norsk institutt for naturforskning.

Häkkilä M. 2021. Are small protected habitat patches within boreal production forests effective in conserving species richness, abundance and community composition? A systematic review. Environmental Evidence 10: 2.

Halliwell P, Whipple S, Bowser G. 2021. Learning to love protected areas: Citizen science projects inspire place attachment for diverse students in United States National Parks. Journal of Geoscience Education 0: 1–9.

Havinga I, Bogaart PW, Hein L, Tuia D. 2020. Defining and spatially modelling cultural ecosystem services using crowdsourced data. Ecosystem Services 43: 101091.

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

Hornigold K, Lake I, Dolman P. 2016. Recreational use of the countryside: no evidence that high nature value enhances a key ecosystem service. PloS One 11: e0165043.

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

Jansujwicz JS, Calhoun AJK, Bieluch KH, McGreavy B, Silka L, Sponarski C. 2021. Localism “reimagined”: building a robust localist paradigm for overcoming emerging conservation challenges. Environmental Management 67: 91–108.

Johnston A, Moran N, Musgrove A, Fink D, Baillie SR. 2020. Estimating species distributions from spatially biased citizen science data. Ecological Modelling 422: 108927.

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

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

Kendal D, Zeeman BJ, Ikin K, Lunt ID, McDonnell MJ, Farrar A, Pearce LM, Morgan JW. 2017. The importance of small urban reserves for plant conservation. Biological Conservation 213: 146–153.

Knape J, Coulson SJ, van der Wal R, Arlt D. 2022. Temporal trends in opportunistic citizen science reports across multiple taxa. Ambio 51: 183–198.

Koen EL, Newton EJ. 2021. Outreach increases detections of an invasive species in a crowdsourced monitoring program. Biological Invasions 23: 2611–2620.

Komossa F, Wartmann FM, Verburg PH. 2021. Expanding the toolbox: Assessing methods for local outdoor recreation planning. Landscape and Urban Planning 212: 104105.

Kühl HS, Bowler DE, Bösch L, Bruelheide H, Dauber J, Eichenberg D, Eisenhauer N, Fernández N, Guerra CA, Henle K, Herbinger I, Isaac NJK, Jansen F, König-Ries B, Kühn I, Nilsen EB, Pe’er G, Richter A, Schulte R, Settele J, van Dam NM, Voigt M, Wägele WJ, Wirth C, Bonn A. 2020. Effective biodiversity monitoring needs a culture of integration. One Earth 3: 462-474.

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

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

Mair L, Ruete A. 2016. Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa. PLOS ONE 11: e0147796.

Mancini F, Coghill GM, Lusseau D. 2019. Quantifying wildlife watchers’ preferences to investigate the overlap between recreational and conservation value of natural areas. Journal of Applied Ecology 56: 387–397.

Mandeville CP, Finstad AG. 2021. Community science supports research on protected area resilience. Conservation Science and Practice 3: e442.

Mandeville CP, Koch W, Nilsen EB, Finstad AG. 2021. Open data practices among users of primary biodiversity data. BioScience 71: 1128–1147.

Mandeville CP, Nilsen EB, Finstad AG. 2022. Spatial distribution of biodiversity citizen science in a natural area. Open Science Framework 10.17605/OSF.IO/BGF3D.

Maund PR, Irvine KN, Lawson B, Steadman J, Risely K, Cunningham AA, Davies ZG. 2020.

What motivates the masses: Understanding why people contribute to conservation citizen science projects. Biological Conservation 246: 108587.

Maxwell SL, Cazalis V, Dudley N, Hoffmann M, Rodrigues ASL, Stolton S, Visconti P, Woodley S, Kingston N, Lewis E, Maron M, Strassburg BBN, Wenger A, Jonas HD, Venter O, Watson JEM. 2020. Area-based conservation in the twenty-first century. Nature 586: 217–227.

Millar EE, Hazell EC, Melles SJ. 2019. The ‘cottage effect’ in citizen science? Spatial bias in aquatic monitoring programs. International Journal of Geographical Information Science 33: 1612–1632.

Moen, A. 1998. Nasjonalatlas for Norge: Vegetasjon. Statens kartverk.

Mueller MA, Drake D, Allen ML. 2019. Using citizen science to inform urban canid management. Landscape and Urban Planning 189: 362–371.

Muñoz L, Hausner VH, Runge C, Brown G, Diagle R. 2020. Using crowdsourced spatial data from Flickr vs. PPGIS for understanding nature's contribution to people in Southern Norway. People and Nature 2: 437-449.

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

Oldekop JA, Holmes G, Harris WE, Evans KL. 2016. A global assessment of the social and conservation outcomes of protected areas. Conservation Biology 30: 133–141.

Pebesma E. 2018. Simple features for R: standardized support for spatial vector data.” The R Journal 10: 439–446.

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

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

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

Romo H, García-Barros E, Lobo JM. 2006. Identifying recorder-induced geographic bias in an Iberian butterfly database. Ecography 29: 873–885.

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

Rowley JJL, Callaghan CT, Cutajar T, Portway C, Potter K, Mahoney S, Trembath DF, Flemons P, Woods A. 2019. FrogID: Citizen scientists provide validated biodiversity data on frogs of Australia. Herpetological Conservation and Biology 14(1):155-170.

Salmon RA, Rammell S, Emeny MT, Hartley S. 2021. Citizens, scientists, and enablers: a tripartite model for citizen science projects. Diversity 13: 309.

Schmeller DS, Böhm M, Arvanitidis C, Barber-Meyer S, Brummitt N, Chandler M, Chatzinikolaou E, Costello M, Ding H, García-Moreno J, Gill M, Haase P, Jones M, Juillard R, Magnusson WE, Martin CS, McGeoch M, Mihoub J, Pettorelli N, Proença V, Peng C, Regan E, Schmiedel U, Simaika JP, Weatherdon L, Waterman C, Xu H, Belnap J. Building capacity in biodiversity monitoring at the global scale. Biodiversity and Conservation 26: 2765–2790.

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

Studd EK, Bates AE, Bramburger AJ, Fernandes T, Hayden B, Henry HAL, Humphries MM, Martin R, McMeans BC, Moise ERD, O'Sullivan AM, Sharma S, Sinclair BJ, Sutton AO, Templer PH, Cooke SJ. 2021. Nine maxims for the ecology of cold-climate winters. BioScience 71: 820–830.

Sutton AO, Studd EK, Fernandes T, Bates AE, Bramburger AJ, Cooke SJ, Hayden B, Henry HAL, Humphries MM, Martin R, McMeans B, Moise E, O'Sullivan AM, Sharma S, Templer PH. 2021. Frozen out: unanswered questions about winter biology. Environmental Reviews.

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

Tiago P, Ceia-Hasse A, Marques TA, Capinha C, Pereira HM. 2017. Spatial distribution of citizen science casuistic observations for different taxonomic groups. Scientific Reports 7: 12832.

Tredennick AT, Hooker G, Ellner SP, Adler PB. 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. Ecology 102: e03336.

Tulloch AIT, Mustin K, Possingham HP, Szabo JK, Wilson KA. 2013. To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale. Diversity and Distributions 19: 465–480.

Venter ZS, Barton DN, Gundersen V, Figari H, Nowell MS. 2021. Back to nature: Norwegians sustain increased recreational use of urban green space months after the COVID-19 outbreak. Landscape and Urban Planning 214: 104175.

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

Volenec ZM, Dobson AP. 2020. Conservation value of small reserves. Conservation Biology 34: 66-79.

Weaver DB, Lawton LJ. 2017. A new visitation paradigm for protected areas. Tourism Management 60: 140–146.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, Fraçois F, Gralemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. Journal of Open Source Software, 4: 1686.

Wintle BA, Kujala H, Whitehead A, Cameron A, Veloz S, Kukkala A, Moilanen A, Gordon A, Lentini PE, Cadenhead NCR, Bekessy SA. 2019. Global synthesis of conservation studies reveals the importance of small habitat patches for biodiversity. Proceedings of the National Academy of Sciences 116: 909–914.

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

Zulian V, Miller DAW, Ferraz G. 2021. Integrating citizen-science and planned-survey data Improves species distribution estimates. Diversity and Distributions 27: 2498–2509.

Tables

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

Environmental variables	Expected correlation with citizen science	Covariate structure in grid-based analysis	Covariate structure in trail-based analysis
Access variables			
Proximity to nearest access point*[1]	Positive	Negative distance (m) from grid centroid to nearest access point.	Negative distance (m) from trail segment centroid to nearest access point.
Proximity to recreational facilities (e.g., public tourist cabins, playgrounds, swimming beaches)[1]	Positive	Binary variable expressing whether grid cell contains a facility.	Negative distance (m) from trail segment centroid to nearest facility.
Elevation[2]	Negative	Maximum elevation (m) of grid cell.	Maximum elevation (m) of trail segment.
Longitude (eastness)†	Positive	Longitude (m) of grid centroid.	Longitude (m) of trail segment centroid.
Presence of recreational trails and access roads[1]	Positive	Total length (m) of trail within grid cell.	Function of segment as a main travel route, defined by the percentage of the trail segment characterized by the "transportation" land cover category.
Natural interest variables			
Cultivated land cover[3]	Negative	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Developed land cover[3]	Negative	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Forest land cover[3]	Positive	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Wetland land cover[3]	Positive	Area within grid cell covered by the land cover type.	Percentage of area in the trail segment corridor covered by the land cover type.
Proximity to a freshwater lake or stream[3]	Positive	Binary variable expressing whether each grid cell contains a freshwater body.	Negative distance (m) from trail segment centroid to nearest freshwater body.

*Access points were defined by intersections between a road or trail and the boundary of the natural area as well as public parking areas and public transit stops within or adjacent to the area.

† Longitude was used to represent distance from the nearest population center; the study area lies to the west of Trondheim's population center, so it was expected that eastern longitudes would be accessed more often.

[1] Trondheim Municipality (https://kart.trondheim.kommune.no)

[2] Norwegian Digital Elevation Model (https://www.kartverket.no)

[3] Norwegian Institute for Bioeconomics AR5 1:5000 land cover data (Ahlstrøm et al. 2014).

Figures

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

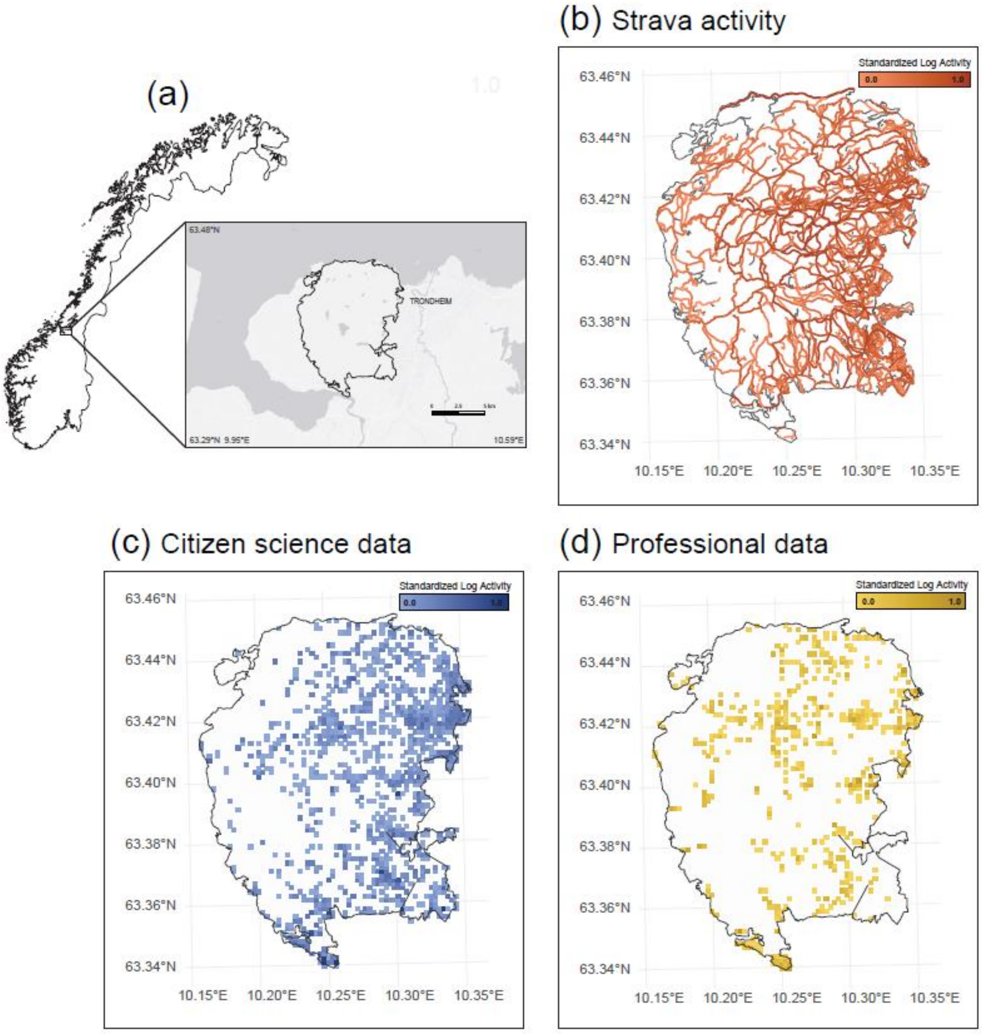


Figure 2. (a) Number of observations from each taxonomic group for citizen science and professional data; (b) number of species from each taxonomic group for citizen science and professional data; (c) month of observation for citizen science and professional data; (d) land cover type for citizen science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

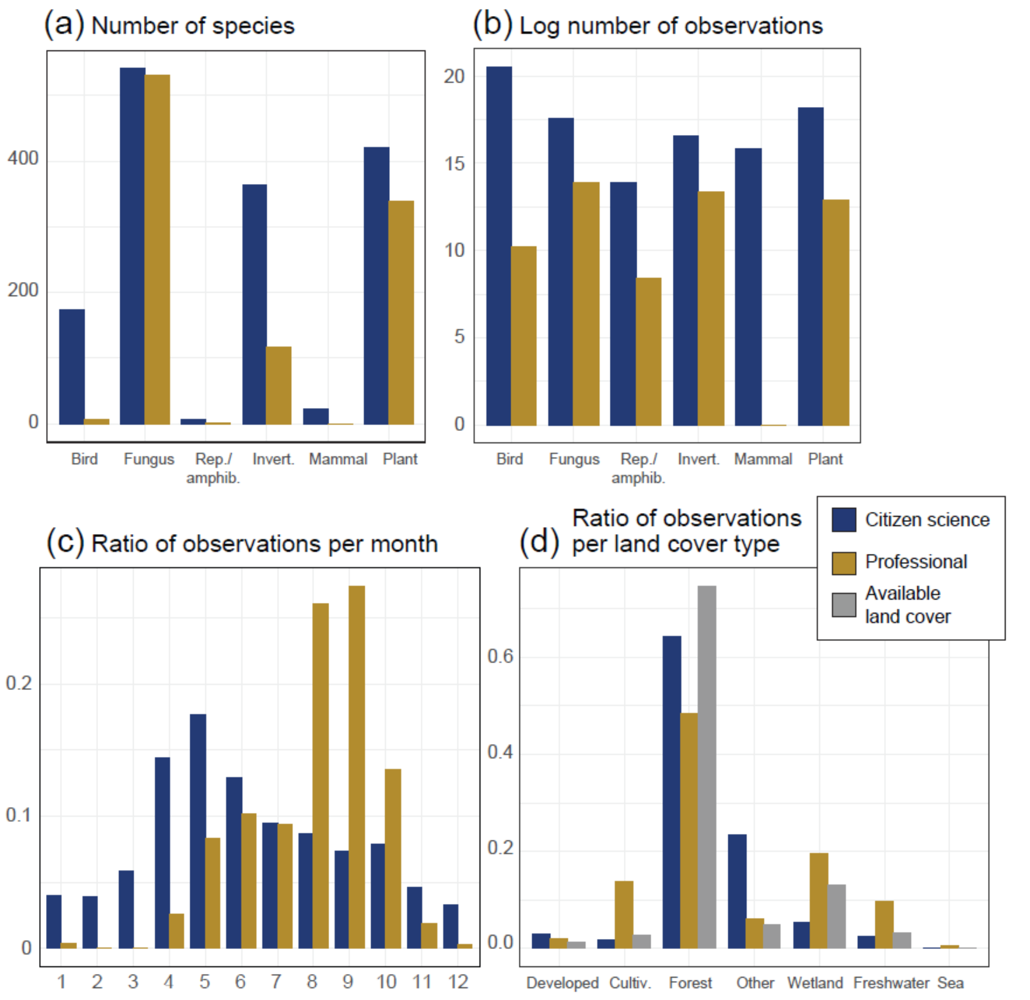


Figure 3. Effect of all covariates on the response variable of (a) the grid-based models of citizen science and professional data collection and (b) the trail-based models of citizen science and Strava activity. Decreasing color intensity indicates decreasing variable importance. All continuous variables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

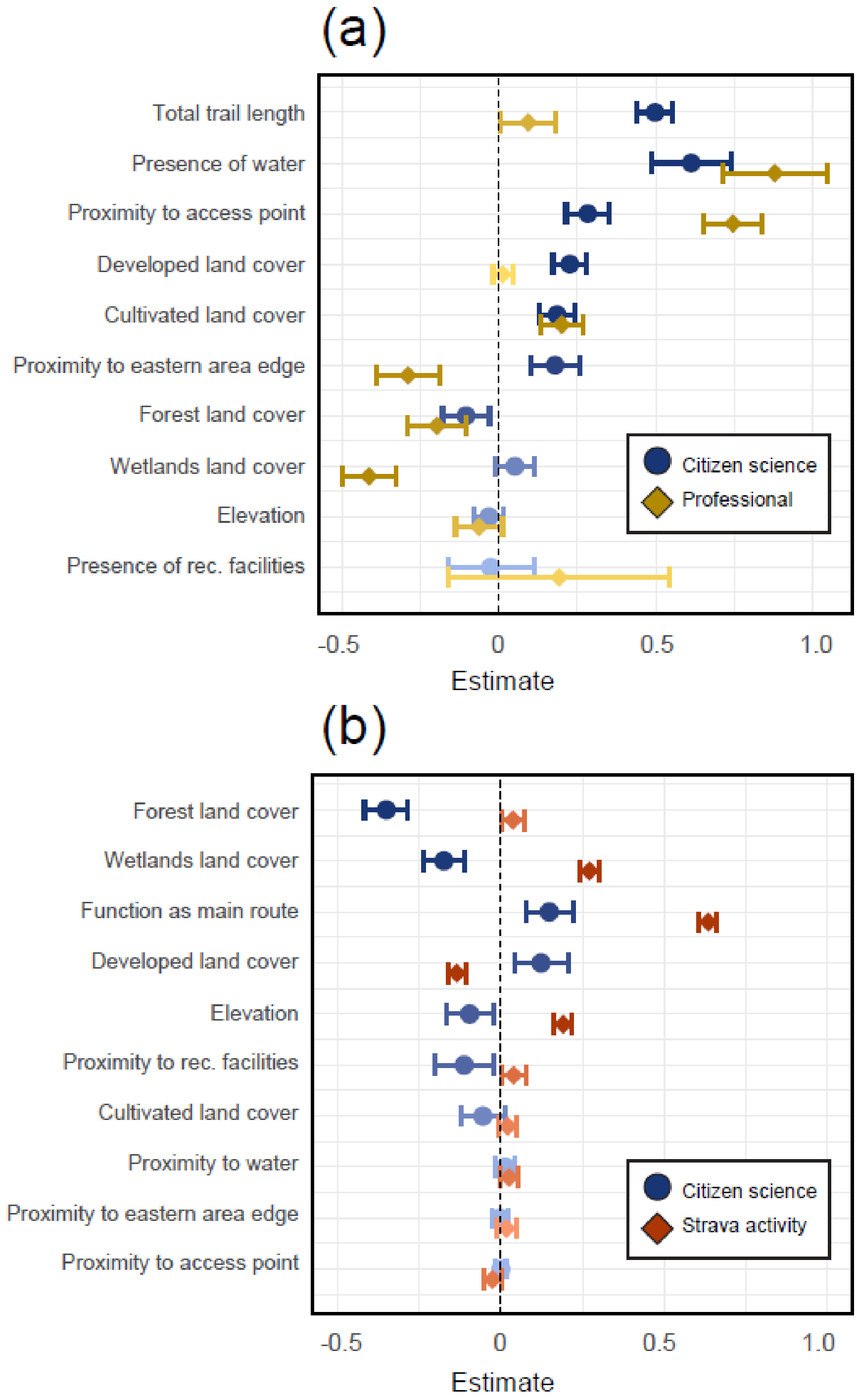


Figure 4. Predicted effect of each covariate on the number of citizen science observations per grid cell, modeled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All six models with substantial support ($\Delta AIC_c < 2$) are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

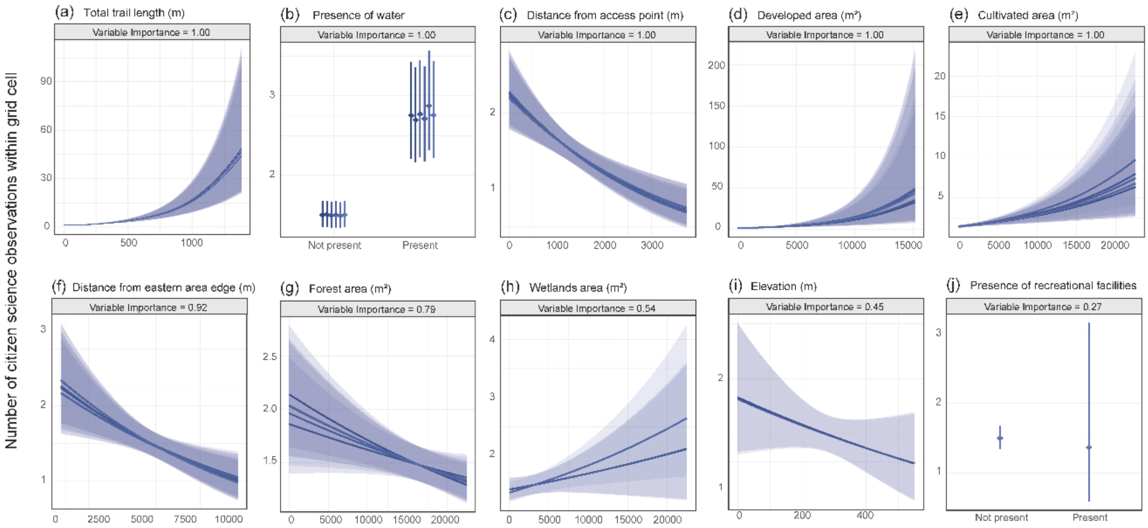


Figure 5. Predicted effect of each covariate on the number of professional biodiversity observations per grid cell, modeled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All nine models with substantial support ($\Delta AIC_c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

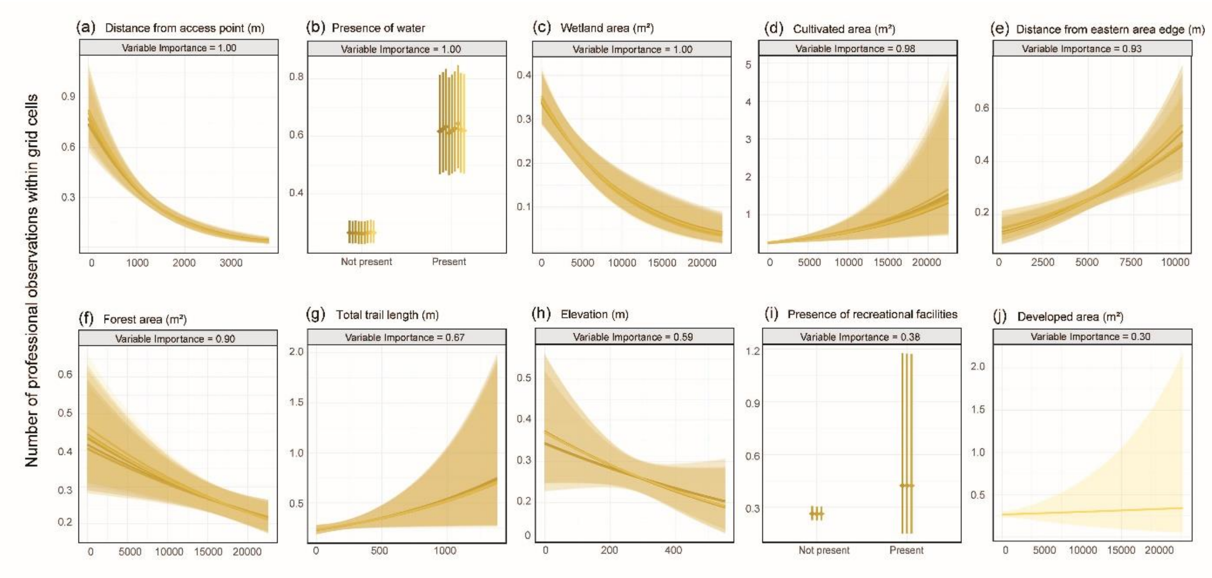


Figure 6. (a) Distance between reported observation coordinates and the nearest trail for the locations of citizen science data collection, professional data collection, and a random sample of locations in the study area. (b) Distance between reported observation coordinates and the nearest trail for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

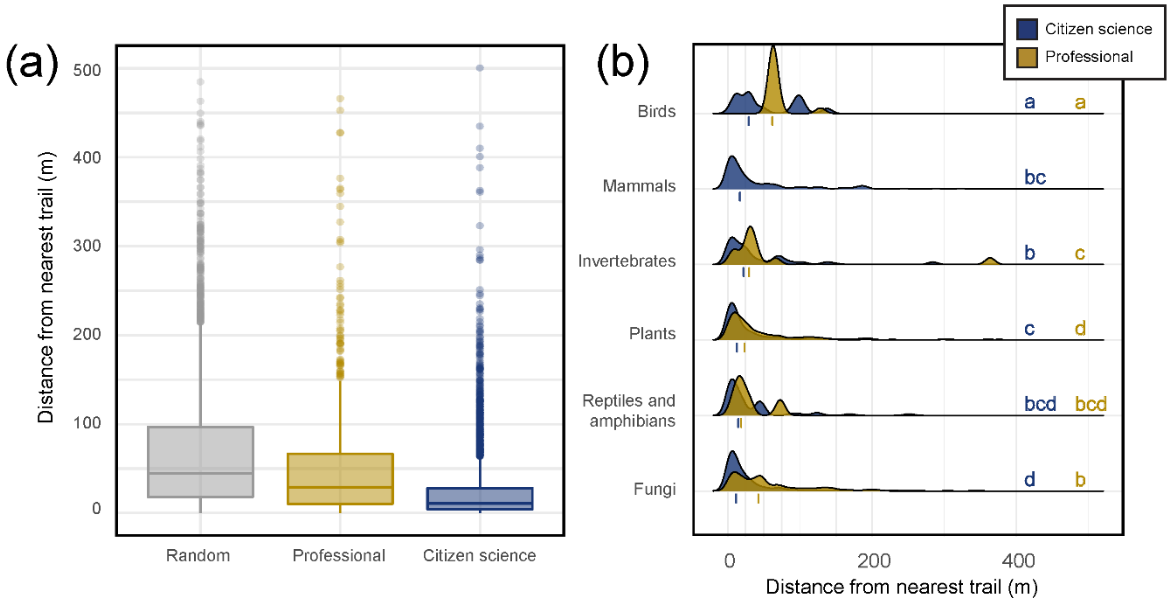


Figure 7. Predicted effect of each covariate on the number of citizen science observations per 300-meter-wide trail segment corridor, standardized by segment length, modeled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All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

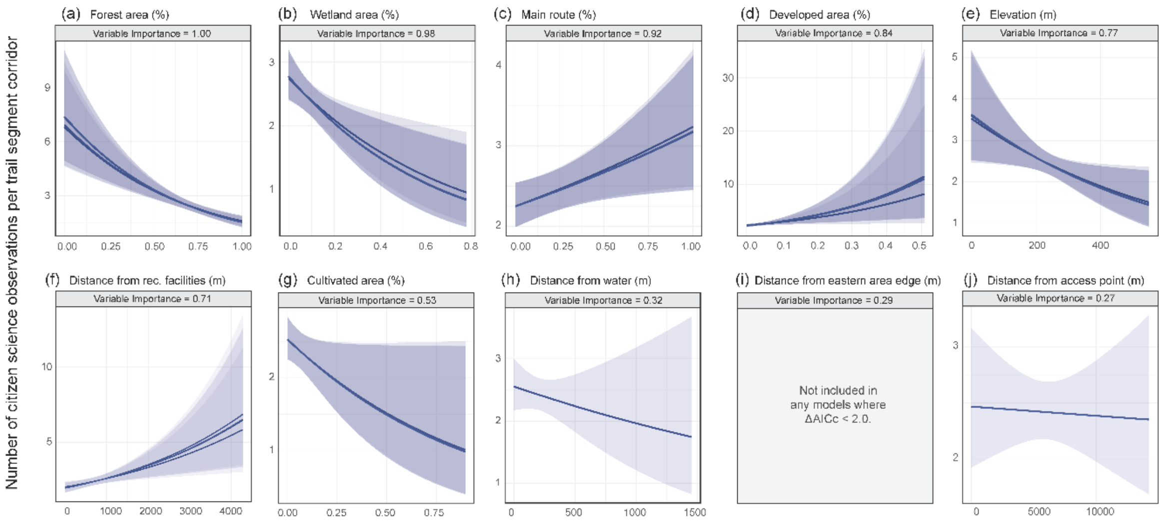
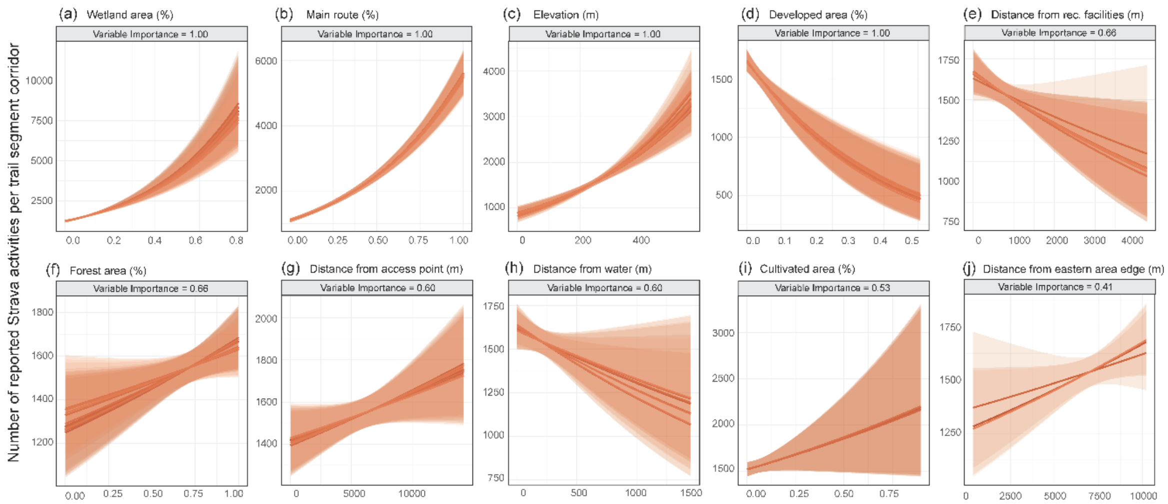


Figure 8. Predicted effect of each covariate on the number of reported Strava activities per trail segment corridor, modeled with a negative binomial generalized linear model structure. For predictions, other covariates are held at their mean value. All twelve models with substantial support ($\Delta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



Supporting Information

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

C.P. Mandeville, E.B. Nilsen, A.G. Finstad

Contents

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

Data source	**Type of source**	**n (included)**
Norwegian Species Observation Service	Citizen science - opportunistic	40376
eBird Observation Dataset	Citizen science - opportunistic	3450
iNaturalist Research-grade Observations	Citizen science - opportunistic	299
Pl@ntNet	Citizen science - opportunistic	50
Skandobs	Citizen science - opportunistic	11
Naturgucker	Citizen science - opportunistic	11
Observation.org	Citizen science - opportunistic	6
Vascular plant herbarium TRH, NTNU University Museum	Professional	492
Lichen herbarium TRH, NTNU University Museum	Professional	291
Terrestrial and limnic invertebrates systematic collection, NTNU University Museum	Professional	246
Mycology herbarium TRH, NTNU University Museum	Professional	218
Fungi field notes, Oslo (O)	Professional	170
NINA insect database	Professional	120
International Barcode of Life project (iBOL)	Professional	100
BioFokus	Professional	99
Geographically tagged INSDC sequences	Professional	67
Bryophyte herbarium TRH, NTNU University Museum	Professional	53
Lichen field notes, Oslo (O)	Professional	29
Royal Botanic Garden Edinburgh Living Plant Collections (E)	Professional	28

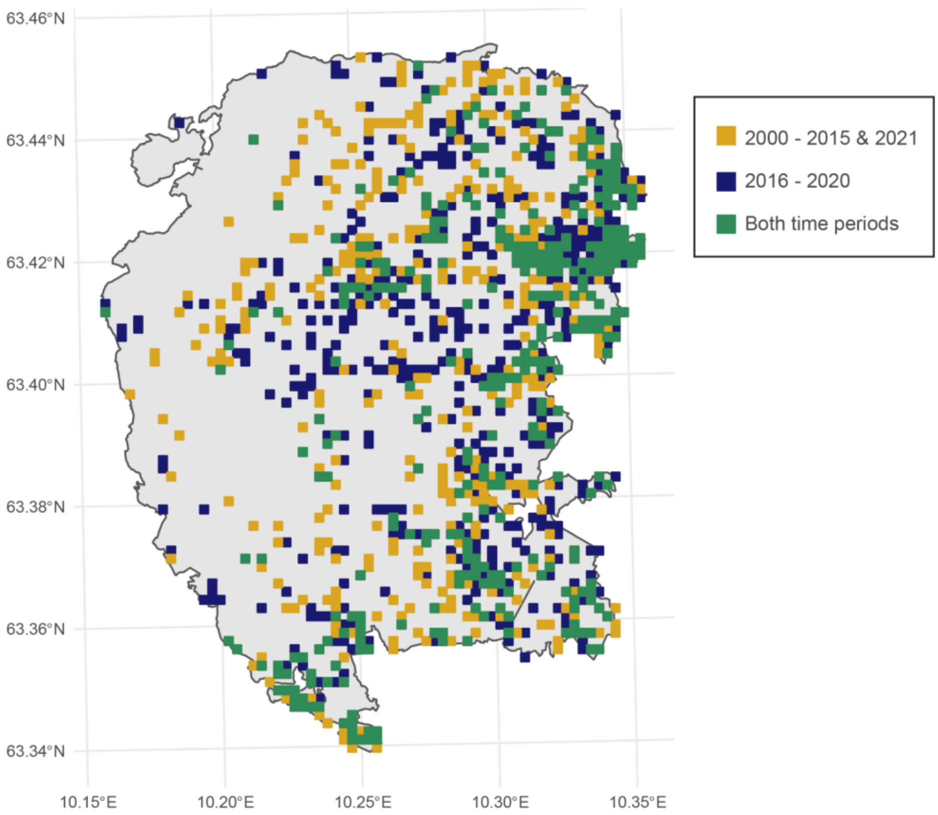
Bird collection NTNU University Museum	Professional	27
Lichen herbarium, Oslo (O) UiO	Professional	25
Mycology herbarium, Oslo (O) UiO	Professional	23
NHMO DNA Bank Vascular plants collection	Professional	19
Vascular Plant Herbarium, Oslo (O) UiO	Professional	15
NHMO DNA Bank Fungi and Lichens collection	Professional	7
Danish Mycological Society, fungal records database	Professional	6
Artsprosjekt: hypogeous_macrofungi	Professional	4
Bryophyte Herbarium, Oslo (O) UiO	Professional	4
Herpetile collection NTNU University Museum	Professional	4
Entomological collections, UiB	Professional	3
Lichen herbarium, UiB	Professional	2
Algae herbarium TRH, NTNU University Museum	Professional	1
Mycology collection, Norwegian Forest and Landscape Institute	Professional	1
Reptilia notes, NTNU University Museum	Professional	1
Seabirds in Norway - Estimated population sizes	Professional	1
The cryptogamy collection (PC) at the Herbarium of the Muséum national d'Histoire Naturelle (MNHN - Paris)	Professional	1
Tropicos Specimen Data	Professional	1
Vascular plant herbarium (KMN) UiA	Professional	1
Norwegian Biodiversity Information Centre - Other datasets	Citizen science - structured	0
Algae collection, Oslo (O) UiO	Professional	0
Algae, Norwegian College of Fishery Science	Professional	0

Birds ringed with Norwegian rings 1914-1960	Professional	0
Birds ringed with Norwegian rings 1961-1990	Professional	0
Bryophyte herbarium, UiT Tromsø Museum	Professional	0
Collembola collection of Arne Fjellberg, Norway	Professional	0
Entomology collection, UiT Tromsø Museum	Professional	0
Entomology Division, Yale Peabody Museum	Professional	0
Entomology, Natural History Museum, University of Oslo	Professional	0
Fish collection NTNU University Museum	Professional	0
Herbarium GB, University of Gothenburg	Professional	0
Huitfeldt Kaas: Freswhater fish distribution in Norway 1918	Professional	0
Ims fish tag database	Professional	0
Lichen herbarium, UiT Tromsø Museum	Professional	0
Limnic freshwater benthic invertebrates biogeographical mapping/inventory NTNU University Museum	Professional	0
Limnic freshwater pelagic invertebrates biogeographical mapping/inventory NTNU University Museum	Professional	0
Lund Botanical Museum (LD	Professional	0
Mammal collection NTNU University Museum	Professional	0
Marine invertebrate collection NTNU University Museum	Professional	0
Mycology herbarium, UiT Tromsø Museum	Professional	0
National fish tag database	Professional	0

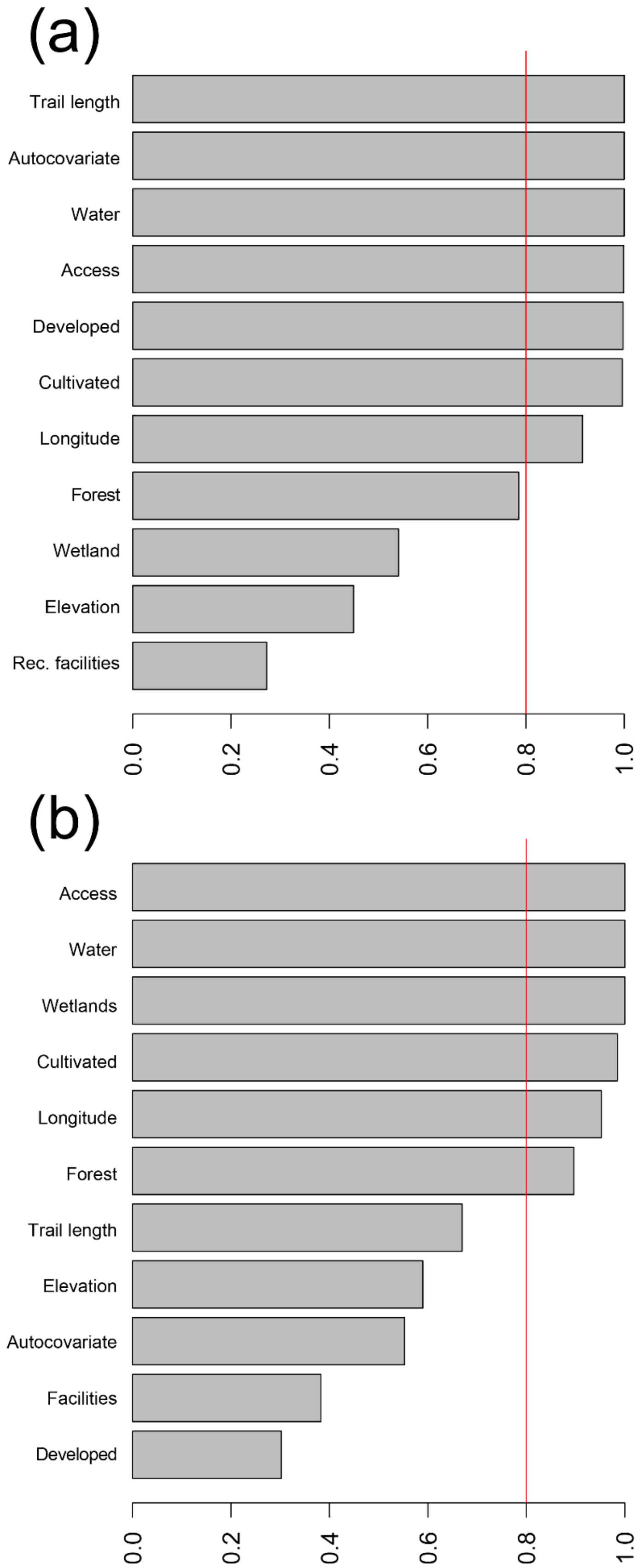
NHMO DNA Bank Fish and Herptile collection	Professional	0
NINA Vanndata fisk	Professional	0
NINA Vanndata øvrige arter	Professional	0
NMNH Extant Specimen Records	Professional	0
Notes from the Mycology Herbarium, Oslo (O)	Professional	0
NSW AVH data	Professional	0
Provincial Museum of Alberta, Edmonton,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

S1.2 Map indicates the year in which citizen science biodiversity data used in the study were collected. The 2016-2020 time period, which contains 41% of the citizen science data, corresponds with the dates of available Strava Metro data.

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



S2.1 Model-averaged relative variable importance of each covariate for the models of (a) citizen science and (b) professional biodiversity data observations among grid cells in the study area.



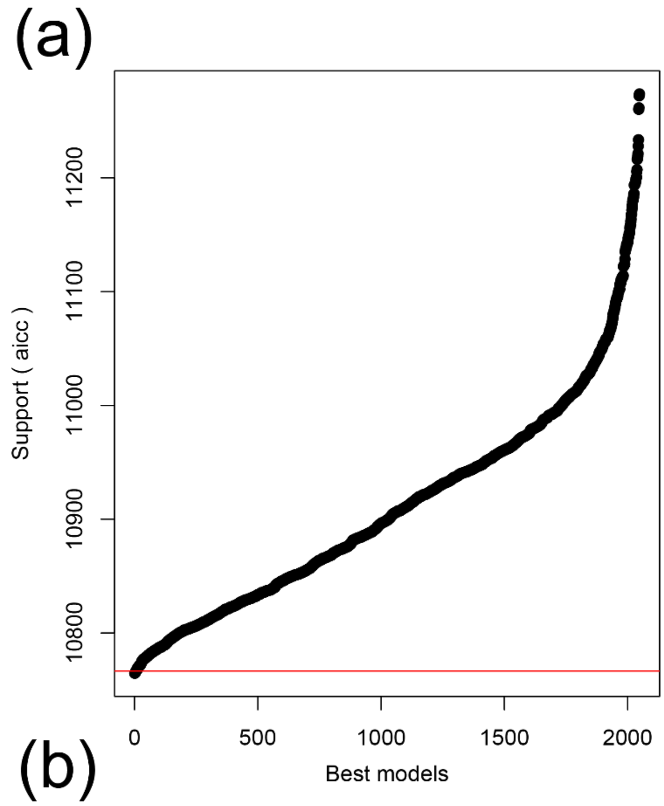
S2.2 Importance and model-averaged estimates and standard error for each covariate in the (a) citizen science and (b) professional model of biodiversity observations among grid cells in the study area.

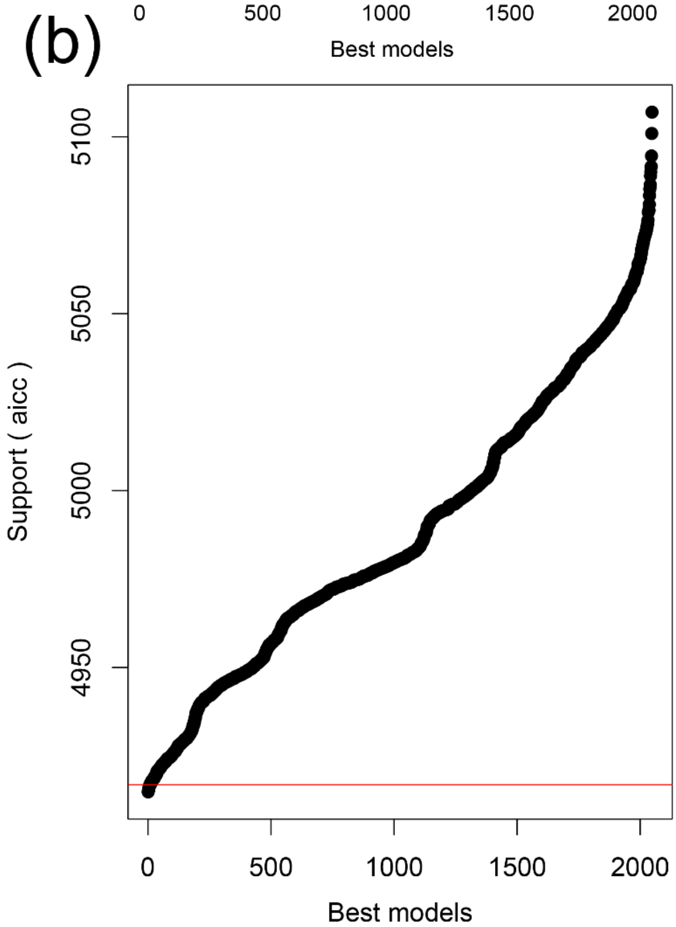
	(a) Citizen science			(b) Professional		
	Importance	Estimate	Std. error	Importance	Estimate	Std. error
Intercept	1.00	0.21	0.06	1.00	-1.34	0.01
Total trail length	1.00	0.50	0.06	0.67	0.09	0.09
Presence of water	1.00	0.61	0.13	1.00	0.88	0.17
Proximity to access	1.00	0.28	0.07	1.00	0.75	0.09
Developed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

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

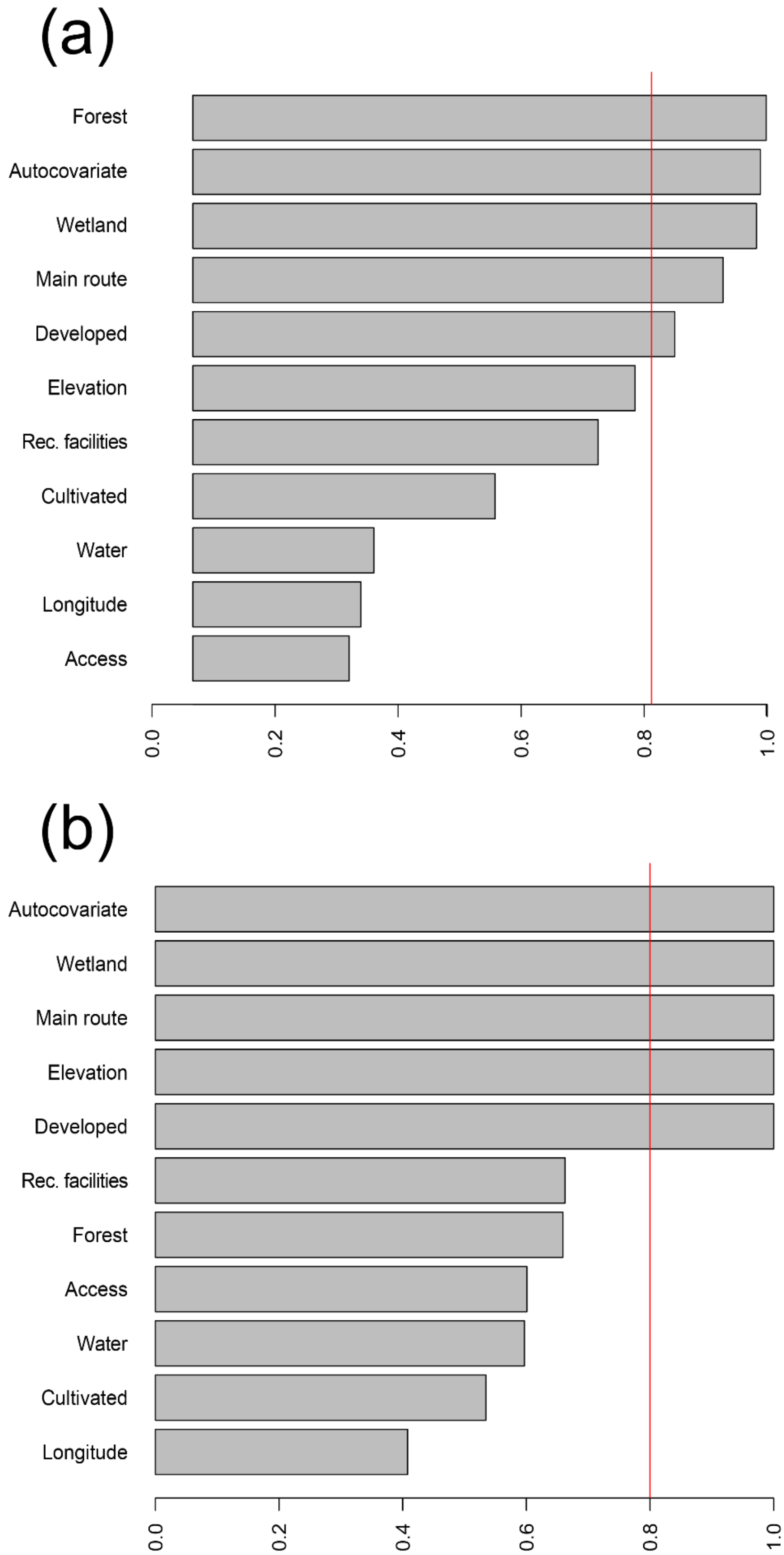
Model	AIC_c	ΔAIC_c	k	Evidence weight
(a) Citizen science				
water + access + trails + longitude + developed + cultivated + forest + ac	10764.62	0.00	9	0.15
water + access + trails + longitude + elevation + developed + cultivated + forest + ac	10764.93	0.14	10	0.14
water + access + trails + longitude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

S2.4 AIC_c weights of the 2000 highest rated negative binomial generalized linear models for the number of (a) citizen science and (b) professional observations per grid cell, out of a set consisting of all possible combinations of the ten covariates with no interactions. Models below the red line have substantial support ($\Delta AIC_c < 2$).





S3.1 Model-averaged relative variable importance of each covariate for the models of (a) citizen science observations, standardized by trail length, and (b) reported Strava activities among trail segment corridors in the study area.



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

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

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

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

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

S3.4 AIC_c weights of the 2000 highest rated negative binomial generalized linear models for the number of (a) citizen science observations, standardized by trail length, and (b) recorded Strava activities per trail segment, out of a set consisting of all possible combinations of the ten covariates with no interactions. Models below the red line have substantial support ($\Delta AIC_c < 2$).