Check for updates

Methods for monitoring urban street litter: a comparison of municipal audits and an app-based citizen science approach

Lisa Watkins, [ID] †*[a] David N. Bonter,[b] Patrick J. Sullivan[c] and M. Todd Walter[a]

Street litter and the plastic pollution associated with it is an economic and environmental health issue in municipalities worldwide. Most municipal litter data are derived from costly audits, performed by consultants at sparse intervals. Mobile phone apps have been developed to allow citizen scientists to participate in collecting litter data. Both municipal audits and citizen science datasets may be useful not only for informing municipal management decisions but also for increasing scientific understanding of litter dynamics in urban environments. In this analysis, we compare the spatial patterns and composition of litter in Vancouver, Canada, measured through professional municipal audits and with Litterati, a widely used citizen science app. While reported litter composition was consistent across methods, regression analysis shows that spatially, Litterati submissions were more highly correlated with human population patterns than with correlates of litter. We provide method recommendations to improve the utility of resulting data, such that these non-traditional, underutilized datasets may be more fully incorporated into scientific inquiry on litter.

Environmental significance

Street litter and associated plastic pollution have negative effects on aquatic and urban ecosystems both due to physical blockages they cause in infrastructure and organisms and due to associated chemical contaminants they transport. The quantification and fate of plastic litter in the environment is a prerequisite for successful pollution management, but sources, sinks, and transport patterns continue to be areas of high uncertainty. Currently there is a wealth of litter data being collected in urban environments by non-traditional sources that are yet to be included in scientific efforts to understand these patterns. We find that freely available, non-traditional datasets, specifically professional municipal audits and app-based citizen science data, provide valuable spatial and compositional insights into urban street litter trends and drivers.

1. Introduction

Street litter is a concern for cities. It is an eyesore with economic and infrastructural impacts.[1] Cleaning up litter is costly; estimates of litter cleanup costs for the US, for example, are US $11 billion annually,[2] and cities actively seek data to more efficiently address the issue.[3–5] Limits on litter loads in stormwater, which are increasingly common in coastal municipalities,[6,7] provide cities with an additional regulatory incentive to monitor and control litter, given its ability to enter local waterways *via* stormwater infrastructure.[8–10]

The plastic component of litter is of particular concern. Once integrated into soil and aquatic environments, small plastic items are known to have both physical effects, such as false satiation, and chemical effects, such as hormonal mimicry, on organism health.[11–13] These effects motivate scientists, as well as cities, to better understand litter as a source of plastic pollution to the environment. With the majority of marine litter originating on land,[14,15] understanding litter plays an important role in global efforts to mitigate plastic pollution.[16]

Litter ends up in the environment from the improper disposal of items, through littering or illegal dumping, or through fugitive municipal solid waste escaping proper disposal, through wind or other disturbances.[17,18] Most litter is found in areas with relatively high car and foot traffic, as well as areas exhibiting signs of disorder such as graffiti.[19,20] Litter transport across city landscapes is limited. For example a recent study found littered receipts, on average, 1.6 km from their point of origin.[21] Therefore, predicting the spatial distribution of litter likely requires city-specific mapping of sources, human activity, and transport dynamics.

Current methods of learning about litter patterns and sources are expensive and time intensive. Municipalities often rely on

[a]*Department of Biological & Environmental Engineering, Cornell University, Ithaca, New York, USA*

[b]*Cornell Lab of Ornithology, Ithaca, New York, USA*

[c]*Department of Natural Resources and the Environment, Cornell University, Ithaca, New York, USA*

† Present affiliation: Washington Sea Grant, Seattle, Washington, USA, watkinsl@uw.edu.

© 2024 The Author(s). Published by the Royal Society of Chemistry

Environ. Sci.: Adv., 2024, **3**, 885–896 | **885**

outside consulting firms for annual or one-time audits of street litter.[3,5,22–24] City managers use the resulting data on street litter for three main reasons: first, to optimize resources, such as manual cleanups and street sweeping; second, for tackling street cleanliness issues; and third, for receiving quantitative feedback on intervention strategies.[25] While these audits are performed regularly in many cities across the globe, we find that their results are not commonly incorporated into scientific research on litter or plastic pollution or otherwise synthesized across jurisdictions, leaving a gap between practitioner knowledge and a broader understanding of mechanics and trends in urban litter.[26]

Citizen science, also referred to as participatory science, is one approach that shows promise for aiding litter monitoring efforts. Several apps, including Litterati, Marine Debris Tracker, Marine LitterWatch, and Clean Swell, attempt to facilitate collection of litter data by citizen scientists through a dedicated mobile phone app.[27–30] Each app was developed with a slightly different context in mind, from coastal shorelines to urban streets and offers differing levels of guidance *via* the user interface. All are similar in that they aren't prescriptive about the methods followed; they offer the ability to submit data on litter in an opportunistic manner, *e.g.* without completing a timed or spatially-constrained survey, while being flexible enough to integrate with more robust survey methods if users desire.

This participatory approach to data collection has many benefits for learning about litter in a city. For one, by engaging community members in the process of contributing to data collection, these apps become a powerful avenue for education, both about the scientific process and about the prevalence and detrimental effects of litter.[31] While city governments have tools to react to litter, preventing litter in the first place requires some level of behavioral intervention from citizens.[32,33] These apps provide an avenue for dispensing knowledge, building ownership, and otherwise engaging citizens around litter concerns.

As is common with existing citizen science datasets, their scientific utilization, in this case for better understanding sources of litter, is limited.[34] For example, Litterati's 255 000 users and over 15 million observations across 185 countries have been utilized in 5 peer-reviewed publications from 2013–2021.[35] Engaging citizen scientists in data collection allows for a greater number of observations than researchers could traditionally achieve on their own. While cities typically have audits conducted once per year, citizen scientists on these apps voluntarily submit data year-round.

Both the citizen science dataset and the municipal audits have potential to assist managers and the scientific community's understanding of urban litter. Differences in the methods currently followed by citizen science app-users and professional auditors require some assessment to understand the best uses for and any shortcomings in their resulting datasets. We use this paper both to highlight the existence of these data sources and to assess their utility and robustness for investigating litter patterns in urban areas.

2. Methods

To identify opportunities where citizen science data may be used to complement or supplement municipal audit data, we focus on litter in Vancouver, B. C., Canada, and compare Litterati app submissions from 2017–2019 to the three litter audits commissioned by the City of Vancouver over the same years. First, we compare composition of street litter, as determined from professional municipal audits and citizen science data, *i.e. via* the Litterati app. Second, we compare the spatial coverage and limitations of each of these two data sources. Third, we compare spatial, temporal, and human-behavior predictors of litter, as determined from municipal audits and citizen science data. Finally, we make recommendations for citizen science app improvements that would allow the volunteers' data to be better leveraged by scientists and municipalities alike for improving understanding of litter patterns in urban environments. Fig. 1 provides a comprehensive overview of these steps.

2.1 Municipal audit dataset description

The City of Vancouver hired Dillon Consulting Limited[36] to perform standardized repeated audits of street litter across the urban region. In 2017, 2018, and 2019 the same 108 sites were visited across a 3-day period in September of each year. Sites were public street sections selected randomly with GIS software and paired down to optimize for accessibility, adjacent-site proximity, and diversity of represented land-use and street types. Sampled sites were 61 m long and 5.5 m wide, including 0.5 m of roadway width. In cases where a narrow public-right-of-way limited full sampling, the size of the sampling area was noted and litter densities calculated accordingly. All litter items larger than 25.8 cm^2 (4 in^2) were counted and categorized according to material and item type and labeled as "large litter". Smaller litter items, labeled "small litter", were fully counted and categorized within three 0.55 m^2 (or 1.5% of the total site area) randomly selected subsets at each site.[4,22,37] We refer to these efforts as "municipal audits" in this analysis. Municipal audits include only city-wide breakdowns of litter compositions; we could not freely access site-by-site data on litter composition. We hand-transcribed the site locations and audit data from the audit reports shared with us in pdf form by the City of Vancouver.

2.2 Litterati dataset description

Litterati is an app-based citizen science tool developed for recording observations of litter throughout the world. Users are encouraged to submit photos of littered items before properly disposing of the litter. Users may submit as frequently as desired from any outdoor location. Submissions are geotagged photographs of the litter, plus associated metadata, which includes time of submission and a categorizing label generated by the user, referred to in this paper as a "user-generated tag". We assigned material and item types to user-generated tags when labels were unambiguous, without consulting the photographs. While Litterati has developed additional methodology deployed to specific users for more rigorous and systematic surveys, our analysis includes only their most basic user submission-types, which are presence-only and include no additional information on area sampled or time spent actively looking for litter. We received all available user submissions for the Vancouver area spanning June 2013 through April 2020,
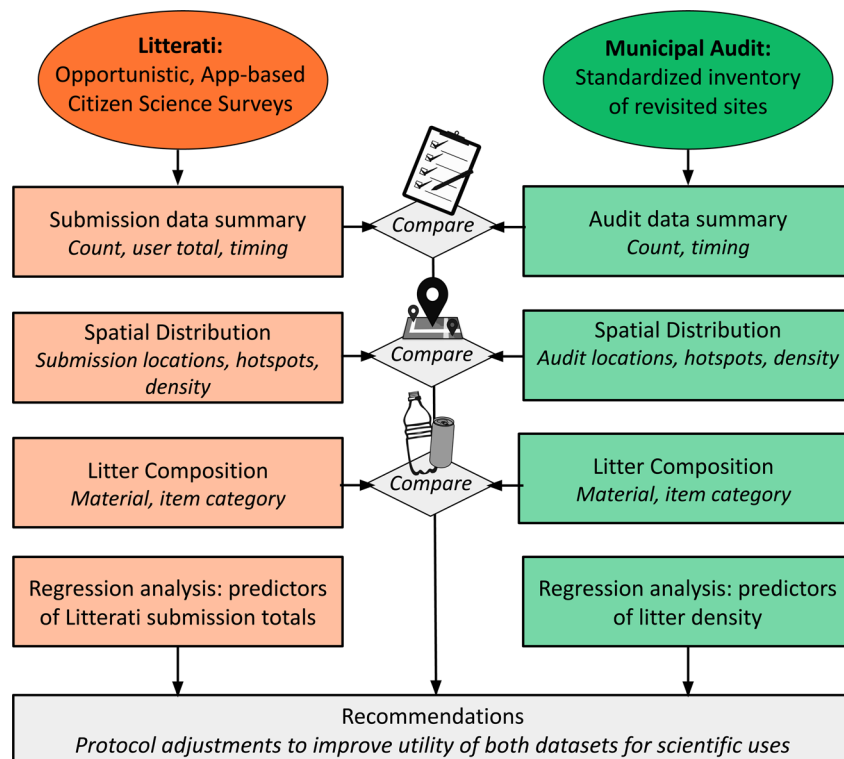
© 2024 The Author(s). Published by the Royal Society of Chemistry

Fig. 1 Flowchart summarizing the analyses used in this study of two non-traditional litter datasets: Litterati (orange) and municipal audits (green).

which we subsequently narrowed down to include only submissions made in the Vancouver metro area between January 2017 and December 2019, to overlap in time and space with the municipal audit data.

2.3 Comparison of spatial coverage

All spatial manipulation and associated calculations were performed in QGIS version 3.18.2.[38] To determine the spatial distribution of sites specifically, we rely on a nearest neighbor index calculated with the built-in nearest neighbor algorithm.

2.4 Predictors of litter density and submission totals

All statistical analyses were performed in R version 4.0.3.[39] Our benchmark for significance was a p-value < 0.05. Variables were selected based on availability across all neighborhoods, sites, and dates, as well as literature- or observational-relevancy to litter. Before use in the model, we checked for multicollinearity

Table 1 Variables included in litter density regression measured through the municipal audit

Factor	Factor levels
Litter bin within the site[a]	Yes; No
Bus stop within or near the site[a]	Yes; No
Fast food or convenience store within or near the site[a]	Yes; No
Grass height at the site[a]	None (0 cm); short (<8 cm); mid-length (8–15 cm); tall (>15 cm)
Street type at the site[a]	Major; minor
Zoning category near the site[a]	Residential; commercial; park; developed, other (industrial, mixed, essential, and institutional)
Year of collection[ab]	2017, 2018, and 2019
Population density, by neighborhood[c]	Numeric (people per km^2)
iNaturalist users per person, by neighborhood[d]	Numeric (users per person)
Calls to '311' per person, by neighborhood[e]	Numeric (calls per person)
Litterati item count, by neighborhood[b]	Quantiles: 0; 1–2; 3–13; 14–315

[a] From the Vancouver street litter audit.[4,22,37] [b] Litterati submissions, aggregated by neighborhood for the years 2017, 2018, and 2019. [c] Calculated with 2016 census data and Vancouver neighborhood boundaries.[42,43] [d] All observations on iNaturalist from 2017 to 2019, aggregated by neighborhood and normalized by neighborhood population. [e] All calls to 311 made during September 2017, 2018, and 2019, aggregated by neighborhood and normalized by neighborhood population.

© 2024 The Author(s). Published by the Royal Society of Chemistry

Environ. Sci.: Adv., 2024, **3**, 885–896 | **887**

by calculating both Spearman correlation and variance-inflation factors (VIF) for all variables. No redundant variables (VIF > 5) were included in the presented model (Table 1).

To test which of the available factors were most predictive of litter density, as measured through the municipal audits, we fit a linear regression, including all variables in Table 1, to predict litter density, in items per m^2. Residuals were found to be normally distributed, confirming that our Gaussian assumption was reasonable for these data. We included variables in the regression to test local site effects (presence of a litter bin, bus stop or convenience store; grass height; street type; zoning category), neighborhood effects (population density), and human activity effects (calls to '311'; iNaturalist users).

The population-normalized total calls to '311', Vancouver's municipal non-emergency hotline for complaints and requests regarding infrastructure and services, was selected as a measure of overall human activity because we speculate that more hotline-reported disorder would correspond to more litter on the streets.[40] The municipality already collects this dataset, which would make it a valuable proxy, if correlated with litter. We also chose to include the population-normalized number of users submitting to iNaturalist, a popular app-based citizen science dataset that encourages observations of any form of "nature", as another measure of human activity.[41] Given iNaturalist's large user-base and broad spatial coverage, even within urban areas, we hypothesized that "nature" observations may be a proxy for humans on foot. We also included the year of collection to look for city-wide changes in litter volume by year and Litterati observations in the neighborhood to test whether the count of items from the citizen scientists correlated with the item densities found by the auditors.

We use linear regression a second time to identify predictors of the total number of Litterati submissions in a given neighborhood. "Neighborhoods" referenced in this study are the official, named delineations, as shown in Fig. 2, used by the City of Vancouver for delivering city services and resources (average area = 5.4 ± 1.9 km^2). We use a natural log-corrected, Laplace-smoothed version of this value as our independent variable, ln(submissions + 1), to allow our model to reasonably meet all linear regression assumptions. Regression predictors included the population-normalized iNaturalist submission count, calls to '311', and local Litterati user total, as well as the year of submission and population density. We utilized calls to '311' and iNaturalist (as introduced in Table 1), as freely available proxies to see whether other citizen science app-users or populations already observing and submitting information on their surroundings may be good proxies for Litterati submissions. We include population density and municipal audits, hypothesizing that the model will indicate that Litterati submissions are more reflective of the number of people in an area, rather than the amount of litter there.

3. Results and discussion

3.1 Summary of recorded litter

The three municipal audits (from 2017, 2018, and 2019) detected 5536 items at the 108 sites resampled for each audit. This amounts to an average of 17 items per audit-site. The Litterati dataset included 3917 items observed from 2017–2019 by 223 unique users with an average of 10.6 items per user per day. Two thirds of these submissions also include a user-generated tag indicating the item category.
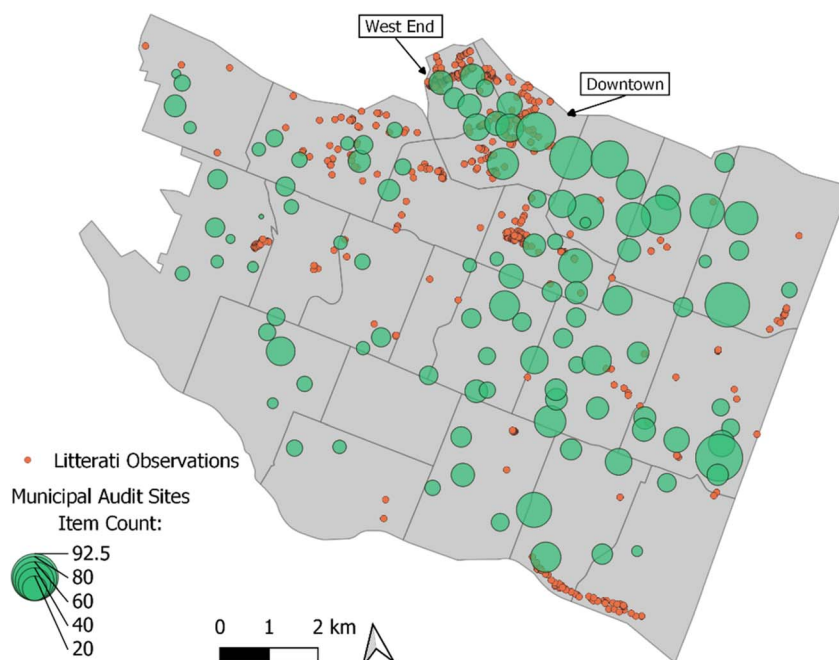

Fig. 2 Litterati observations from 2019 (orange) in Vancouver B. C. neighborhoods (gray regions with dark boundaries) where municipal audits (green dots sized by the 2019 item count) were also performed.

Municipal audits demonstrate an increasing presence of litter each subsequent September, and Litterati submission totals, too, show increasing annual numbers. Unlike municipal audits, which occurred in September, however, Litterati submissions occurred year-round. Most active monitoring through Litterati, in terms of the number of unique users, occurred in summer months, specifically June & July. Because of this decoupled timing, we did not investigate temporal trends of litter in this analysis.

3.2 Comparison of spatial coverage

We confirmed the random spatial distribution of municipal audit locations across the Vancouver metro area (nearest neighbor index = 1.01). This is in contrast to the patchy coverage of Litterati submissions, which tend to cluster near parks and more populated neighborhoods. Because the Litterati dataset does not contain submissions of zero-litter sites, we cannot be certain whether Litterati submissions are missing from large portions of the city for lack of volunteers looking *vs.* lack of litter. In contrast, municipal audits fully report site contents, even when no litter is present: on average, 3 sites per year with no small litter present, 14 sites per year without large litter, and 1 site in 2018 with no litter at all.

The locations of 2019 Litterati submissions and municipal audit sites are shown in Fig. 2. The highest prevalence of Litterati observations took place in the West End neighborhood of Vancouver (Fig. 2) while the highest litter densities as measured through municipal audits took place in neighborhoods east of downtown.

Litterati users have the opportunity to link submissions by labeling them as part of a "challenge" when logging observations. These linked observations provide a possible mechanism for combining otherwise discrete and random observations in the dataset in order to determine local litter density. We did not find sufficient details in the current metadata to fully utilize these linked sampling events, but by assuming a width of sampling observation and a fully searched sampling length between observations, we were able to calculate a rough litter density from these sampling events. As an example, one event involved 2 users sampling approximately 2.25 linear km of downtown Vancouver streets over 3 days in 2019. Assuming a search width similar to municipal audits (5.5 m), their 318 submissions indicate a litter density of 0.03 items per m^2. In contrast, 2019 municipal audits from downtown Vancouver report an average of 5.2 items per m^2 or, if only including items larger than 25.8 cm^2, 0.06 items per m^2. While these linked events provide opportunity for leveraging Litterati observations to calculate litter density, it is clear that missing metadata about the search area and completeness, as well as further understanding about the detection biases of untrained volunteers, currently limits spatial density calculations from Litterati data.

3.3 Vancouver litter composition

Municipal audits indicate that plastic has steadily become more prevalent on Vancouver streets, from 34% of all large litter in 2017, to 42% in 2018, and 46% in 2019. Paper has become less prevalent, falling from 39% in 2017 to 33% in 2019 (Fig. 3A). These relative trends are also seen in the Litterati data. Plastic was the most common material type submitted, comprising 33% of submissions in 2017, 44% in 2018 and 49% in 2019. The prevalence of paper litter also declined through the years in Litterati submissions from 15% of submission in 2017 to 12% in 2019 (Fig. 3B).

Municipal audits indicate that the most common item types found on Vancouver streets change little over time. The most common six item types consistently comprised ~40% of the large items found. These include cup lids, napkins, tobacco products, miscellaneous plastic items, receipts, and snack food packaging. Small items found in municipal audits were similarly consistent in type: tobacco products, paper and chewing gum comprised over 2/3 of all small items found each year (67–74%).
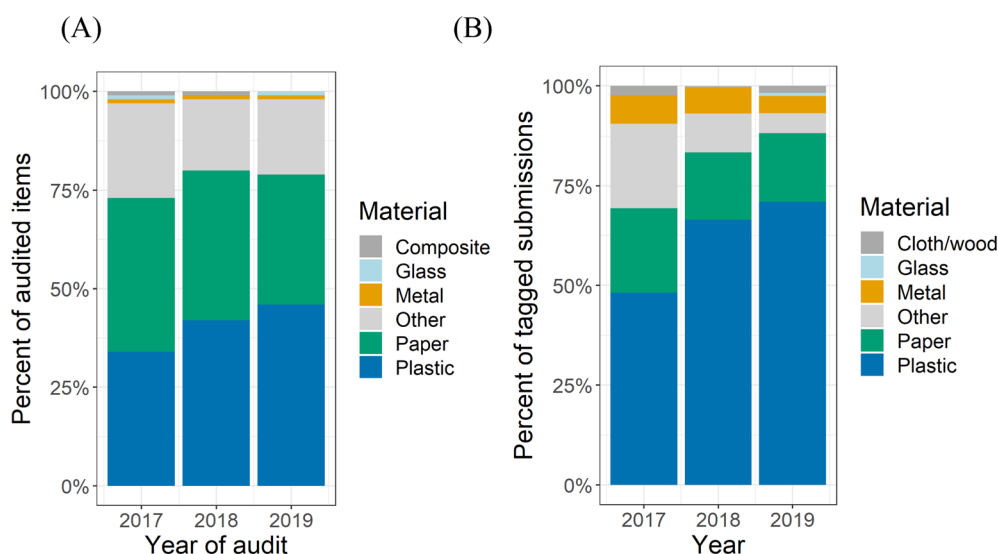

Fig. 3 Composition of litter on Vancouver streets, as determined through (A) municipal audit and (B) aggregated Litterati submissions.

Litterati user-generated tags frequently lacked specific enough information to identify an item beyond its material. For those that were able to be classified, the most commonly identified item types were beverage containers, paper, bags, cups, tobacco products, and wrappers. Without having more complete tag information for Litterati submissions, we cannot confirm that Litterati users are without bias in terms of the kinds of items they perceive as litter, but it appears that the two data collection methods do a comparable job at highlighting the most common litter types on Vancouver streets.

3.4 Predictors of litter density, as measured through the municipal audit

We present the results of a linear regression to highlight factors that appear to be predictive of litter density (items per m^2), as measured through the municipal audit (Table 2).

Some of the site-specific characteristics measured through the municipal audit are significant predictors of litter density, including positive correlation with the presence of litter bins and grass height (Table 2). Further study would be needed to understand whether litter bin presence is predictive of litter density due to bins of properly disposed garbage becoming litter sources from wind or animal disturbance or due to successful bin placement by the city, in areas of frequent trash generation. Studies show that people are less likely to litter when a bin is convenient, supporting the idea of this correlation being due to unintentional or "fugitive" litter.[33] We hypothesize that taller grass is correlated with increased litter due to vegetation trapping wind-transported litter, as well as litter levels remaining higher in areas that receive less property upkeep such as mowing.

We suspect that bus stops and fast food or convenience stores were not predictive of litter density in this study due to the limited distance over which urban litter travels. For example, Lockwood et al.[20] found litter in Philadelphia, USA, was greater within 61 m of convenience stores and fast-food restaurants. The Vancouver municipal audits, however, recorded these stores and restaurants whenever they were "within sight" of the sampling location. The lack of correlation between bus stops, convenience stores, fast food restaurants and litter in these municipal audits is further evidence, therefore, that the presence of litter is driven most significantly by sources only in very close proximity to where litter is observed.[21]

We find that land-use and -pressures are also predictive of litter density. When sites are located on major roads, such as arterial streets, higher litter density is observed (Fig. 4A). We also find that the zoning category is a significant predictor of litter density, with higher litter densities in developed zones including institutional, essential, and industrial areas and lower densities in parkland and residential areas (Fig. 4B). This aligns with the expectation that higher pedestrian and car traffic is associated with more litter.[2] Similarly, higher neighborhood population density, colinear with the percent of residents who commute to work by biking or walking, is predictive of higher litter density (Fig. 4A).

The number of calls to Vancouver's non-emergency complaint and request line '311' is predictive of litter density (Fig. 4B). Note that the number of calls has been normalized by the population, making this signal independent of population density. Calls to '311' span report types, including graffiti, illegal dumping, potholes, leaks, and broken signs. These various reports of "disorder" correlating with litter density is consistent with previous work (e.g. Lockwood et al.,[20]) which suggests that the presence of litter increases perception of crime and is correlated with other indicators of urban "disorder".

We hypothesized that iNaturalist users could be used as a proxy of foot-traffic and would therefore be positively correlated with litter. We found, however, that iNaturalist user abundance, normalized by the neighborhood population, is negatively correlated with litter density. This unexpected finding could be a behavior indicator that areas where people are looking for nature tend to have fewer people leaving litter behind. It could also be a reflection of where people participating in nature-observation citizen science choose to submit; a simple visual analysis of iNaturalist submissions indicate iNaturalist users tend to submit from parklands, which, in Vancouver, are less prevalent in the neighborhoods where municipal audits found the highest litter counts.

Table 2 Summary of linear regression^a used to predict litter density (items per m^2), as measured through the municipal audit

Factor	Estimate	Std. error	t value	p-value
Intercept	−0.20	0.79	−0.26	0.80
Litter bin within the site	0.73	0.24	3.07	<0.005
Bus stop near the site	0.02	0.24	0.07	0.95
Fast food or convenience store near the site	0.18	0.24	0.75	0.45
Grass height at the site	0.23	0.12	1.96	0.05
Street type at the site^b: minor	−0.88	0.28	−3.17	<0.005
Zoning category^c: residential	−0.36	0.29	−1.22	0.22
Zoning category^c: park	−0.53	0.42	−1.27	0.21
Zoning category^c: developed, other	0.46	0.28	1.65	0.10
Year of collection	0.46	0.13	3.56	<0.0005
Population density of the neighborhood^d	78.60	22.44	3.50	<0.005
iNaturalist users per neighborhood population	−237.90	71.86	−3.31	<0.005
Calls to '311' per neighborhood population	71.01	30.31	2.34	<0.05
Litterati item count, by neighborhood^e	−0.03	0.07	−0.45	0.65

^{a} Adjusted R-squared for this model is 0.37, with an F-statistic of 15.48 on 13 and 306 degrees of freedom and a p-value of $<2.2 \times 10^{-16}$. ^{b} Compared to 'street type: major'. ^{c} Compared to 'zoning category: commercial'. ^{d} Values displayed in units of people per m^2. ^{e} Binned using quantile.
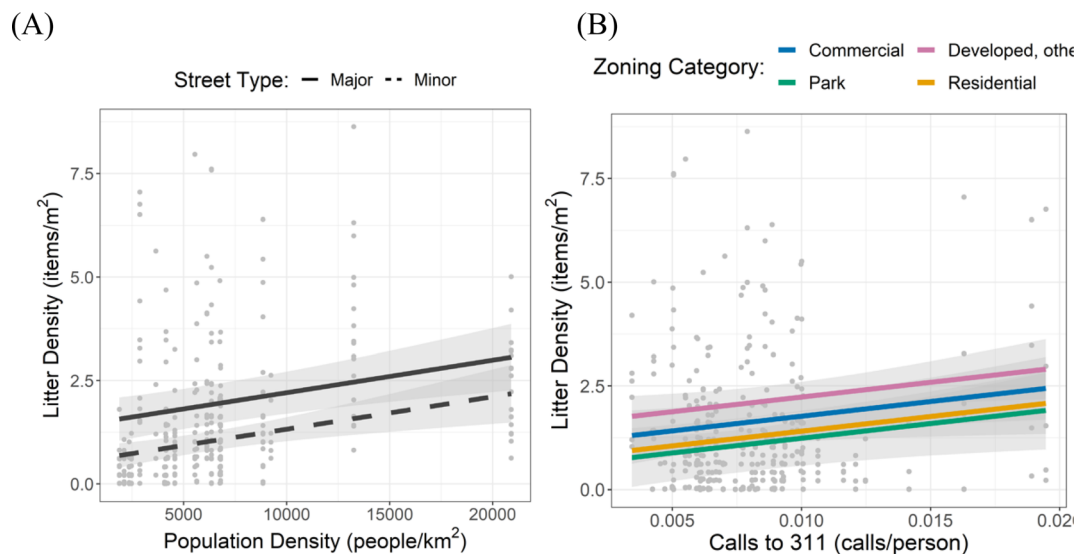
(A)

(B)

Fig. 4 Predicted influence of (A) population density on measured litter, by street type and (B) calls to 311, by the zoning category overlayed on observed litter counts (gray points) from 2017–2019 municipal litter audits in Vancouver, BC. Shading indicates 95% confidence interval.

Table 3 Summary of linear regression[a] used to predict the natural log of Litterati submissions, aggregated by neighborhood

Factor	Estimate	Std. error	t value	p-value
Intercept	0.10	0.21	0.47	0.64
Year of submission	0.37	0.08	4.57	<0.0005
Population density of the neighborhood[b]	0.01	5.81×10^{-4}	11.38	<0.0005
iNaturalist users per neighborhood population	−81.65	45.40	−1.80	0.07
Calls to '311' per neighborhood population	−44.70	18.45	−2.42	<0.05
Litterati users per neighborhood population[c]	12.75	0.70	18.17	<0.0005

[a] Adjusted R-squared for this model is 0.70, with an F-statistic of 153 on 5 and 318 degrees of freedom and a p-value of $<2.2 \times 10^{-16}$. [b] Values displayed in units of people per m². [c] Values displayed in units of users per 1000 residents.

3.5 Predictors of submission totals to Litterati

We had fewer site-specific variables to pair with Litterati observations, given that users do not collect these kinds of metadata when submitting. We, however, do find that Litterati submissions, aggregated by neighborhood, are positively correlated with population density (Table 3 and Fig. 5). The submission total is negatively correlated with calls to '311', likely reflecting a Litterati user bias toward sampling in areas with lower signs of disorder. The Litterati submission count was found to be only weakly correlated with litter densities measured through the municipal audit (Spearman's $\rho = 0.24$), indicating that correlations with Litterati submission are likely more a reflection of user behavior than of litter trends.

We also test whether normalizing Litterati submissions by the number of active users changes these results, hypothesizing that areas where users submit more items may indicate areas of more visible litter abundance. We find that submissions per user shows similar patterns as total submissions; it is correlated with population density ($\rho = 0.47$) and year of submission ($\rho = 0.15$) and is negatively correlated with calls to '311' per person ($\rho = -0.22$).
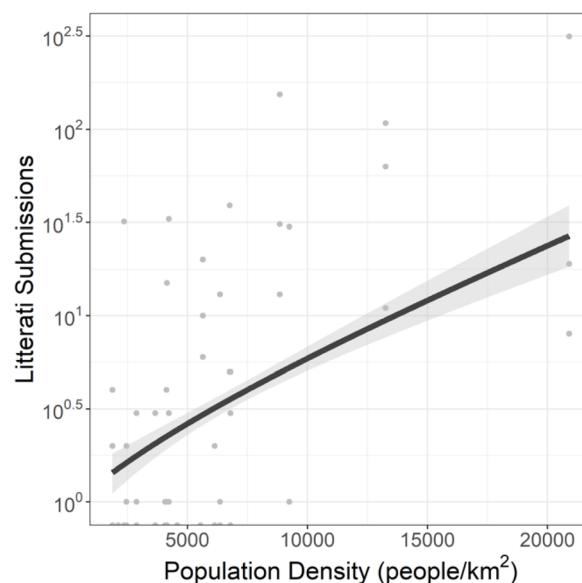

Fig. 5 Predicted relationship between population density and measured litter, from the linear regression of Litterati submissions/user, overlayed on Litterati submissions (gray points) from all 2017–2019 observations in Vancouver, BC. Gray shading indicates 95% confidence interval.

© 2024 The Author(s). Published by the Royal Society of Chemistry

Environ. Sci.: Adv., 2024, **3**, 885–896 | **891**

Litterati submissions were largely made along residential streets (65%), as opposed to major roadways, including arterials and collectors. The municipal audit results had indicated that litter was more abundant on major roadways, which again supports the suggestion that Litterati submission patterns are more linked to volunteer behavioral preferences (e.g., collecting data on quieter streets), which may not represent true litter patterns.

Together, this lack of alignment between Litterati and municipal audit data indicates that submissions from Litterati are not a reliable way of quantifying city-wide litter density or spatial patterns. The result suggesting that Litterati submissions are reflective of the daily movement patterns of their users rather than of underlying spatial patterns in litter distribution is not unique to the Litterati platform. Unstructured citizen science data sets often confront the "recorder effort problem".[44]

4. Recommendations

To meet the current needs of municipalities, data on the distribution of litter need to identify spatial hotspots and allow for year-to-year comparisons such that interventions can be monitored and assessed.[25] Similarly, the scientific community largely seeks to understand spatial and temporal trends in urban litter distribution and quantity, as well as composition.[21,37,45,46] From this analysis we find that citizen science litter data from Litterati's basic platform provides some helpful information about sources. Litter composition, for instance, is one step towards understanding what types of litter need to be better controlled. The data in their current form are, however, inadequate for quantifying spatial hotspots. Presence-only data, as is being collected by Litterati users, presents a challenge to hotspot identification. Inconsistent observations, particularly those fueled by one-time events, obscure temporal trends, and inevitable inconsistencies in volunteer submissions should be accounted for by data analysts. These issues can be overcome with some methodological changes. Below we elaborate on these issues and suggest methods for[1] improving spatial coverage and quantification of litter,[2] improving utility of user-submitted images through computer vision, and[3] accounting for detection biases.

4.1 Improving spatial coverage and quantification of litter

Given Litterati's voluntary-participation model, areas of the city without Litterati submissions could be either litter-free or user-free. This is not an uncommon uncertainty for uncontrolled-effort citizen science models.[47] Common methods for using presence-only observations to understand spatial patterns or abundance rely on an understanding of underlying patterns.[48] For instance, understanding a plant species' habitat needs allows environmental data to supplement and weigh volunteer observations. Further study of litter patterns through regular audits across time and urban landscapes could similarly inform Litterati observations.

One strategy for enhancing Litterati observations to provide context to litter counts would be to collect additional metadata that could be used to normalize counts and provide a measure of relative litter abundance. Unlike municipal audits, where every item within a set area is quantified, many citizen science quantification schemes allow users to engage for however much area or time they choose. A measure of effort, whether that be time spent looking for litter or distance traveled while counting or area surveyed, would allow counts to be normalized by time or area searched, providing a comparable metric between surveyed regions (e.g. items per min, items per m or, preferably, items per m^2). This additional information could be recorded in the background by the user's phone, for instance. Litterati has developed protocols that allow for this enhanced segment-style observation for special projects, but their basic app interface does not collect this additional information.

To enhance this normalized metric, one simple question could be asked of the user, a self-evaluation about their own efforts, "Is this submission inclusive of all litter present?".[49] This information may similarly be gathered through more advanced analysis of their submission photo. Given that the most polluted sites may have a smaller ratio of submitted litter to total litter present than a pristine area, this additional "absence" information would allow for more appropriate conclusions to be drawn from observation densities.

Spatial biases are also introduced into the dataset by the user's tendencies to collect data in convenient areas, namely ones near their home.[50] Ecological research also suggests that volunteers may be motivated to submit data in areas with higher diversity;[51] for litter, this perhaps translates to areas where volunteers anticipate more litter to be found. To encourage broader spatial coverage of the city, project managers could predetermine, through random sampling, a network of promoted road segments across the city, perhaps weighted by street type and zoning designation. Through gamification, users may be enticed to regularly submit from those preferred sites, if reasonably convenient to them.[52] This could achieve repeat sampling, which would be beneficial for detecting temporal changes in composition and density of litter, as well as increased submissions from under-sampled regions of the city.

4.2 Implement computer vision algorithms for data validation

Making further use of the submitted photo is one way that Litterati could enhance the quality of submitted data, which is an important concern of any citizen science project. Computer vision uses algorithms to extract meaningful information from images. When applied to submitted, geotagged photographs, a computer vision program could learn to classify items according to type and material.[53,54] This likely would also allow for contributions of even finer-scaled data based on logos or branding, which would enable rudimentary source-tracing for items such as fast-food containers and wrappers. In the process, this same tool could be used to flag submissions based on improper images.

Though our data-use agreement did not permit us to test a computer vision algorithm in this way, we inspected a subset of images visually. The 50 images (1%) selected by a random

number generator included one taken indoors, 11 that were too blurry or far away to determine the material, item type, or count, 3 microplastic items, and 5 that included more than one item. Thirty-three of the inspected submissions included user-generated tags, making them seem otherwise complete and reliable, but of those seemingly complete submissions, 20% had images containing these noted flaws. Through further analysis, we incidentally encountered additional seemingly complete submissions that contained images of humans, full but not overflowing garbage bins, and previously submitted items, which suggests that not all submitted items are true litter or subsequently disposed of as requested. These few examples indicate that relying only on data field entries for data cleaning is not sufficient, especially when user instructions imply that images are the central data being collected. We were unable to determine any unifying characteristics between users who submitted such images; for example, omitting submissions from first-time users would not successfully remove these types of out-of-scope observations.

The substantial amount of missing and inconsistent user-generated tags is another concern for data quality in litter-related citizen science apps. Replacing free-response user-generated tags with check-boxes of known categories or auto-fill options is one low-tech method of enforcing consistency between observers. Ideally, all incomplete data could be omitted before analyses, unless being used to inform questions of where users are making litter observations. With increasing user and engagement numbers, omitting large portions of suboptimal data while retaining enough for analysis becomes easier.

4.3 Account for biases in detection

One consideration in using volunteer data is anticipating and compensating for the inevitable bias of user sampling decisions. Temporal and detection bias are both present in Litterati data. Submissions are not made consistently over time, by day of week or by month (Fig. 6), unlike municipal audit data, which is a measurement of the same sites on the same dates each year.

This temporal bias is likely driven in part by when people are inclined to participate; for instance, during Vancouver's rainy and dark winter months, outdoor data collection is likely less attractive for users. Similarly, weekends and outside of work hours are times when volunteering is more possible. We observe large volumes of weekend observations from June labeled with the same 'challenge' name, indicating that users were participating together in a litter cleanup and encouraged to log all items with Litterati. In some ways, these clustered events mimic the municipal audit results, which are derived from the single collection event in Septembers. Identifying and understanding the goals and design of these events is necessary for properly correcting spatial and temporal hotspots biased by coordinated collection events.

Detection bias occurs in these citizen science data when a volunteer fails to recognize or submit all litter types. While we do not observe this in the Vancouver data, applying similar methods to San Francisco municipal audits and Litterati data shows that Litterati users there are much less likely to submit observations of broken glass and gum than municipal audits quantify.[3,55] This is one example of users not associating certain kinds of prevalent items as "litter" that can be submitted to Litterati. Instances like this may affect the diversity of items reported from Litterati data but may also provide opportunities for social and behavioral analyses that could utilize Litterati data to provide insights on how community members perceive litter. Size distribution biases would also affect the comparability of results between volunteers. For example, if volunteers tend to ignore smaller items, the composition, abundance, and spatial distribution of litter would all be affected.

Users' selection biases could be identified or eliminated by adapting protocols from other fields, such the pebble count methodology developed by fluvial geomorphologists to remove visual bias when collecting representative estimates of heterogeneous streambed composition.[56,57] In the pebble count method, the scientist selects a stream-length and without looking at their feet, walks a random pattern, picking up one pebble at regular intervals, for instance, at the end of their toe at
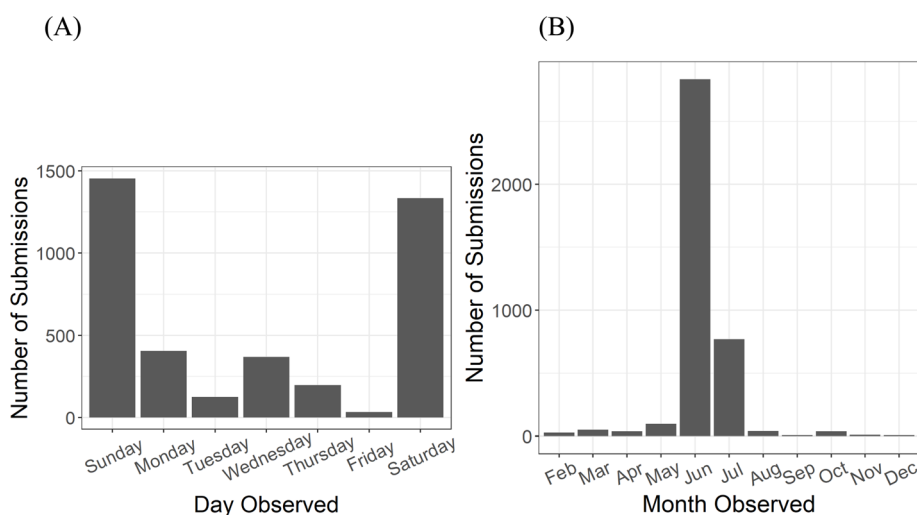
(A)   (B)

Fig. 6 Distribution of Litterati observations made between 2017 and 2019 by (A) day of week and (B) month.

each stride length and recording its size, repeating for 100 strides to record 100 pebbles. An adaptation of this method for litter could involve volunteers recording all items (or lack of items), every fifth sidewalk square or five minutes, for instance. That would provide absence data, a full record of all sizes present within the area, and potentially a density metric.

5. Summary

In their current state, both municipal audits and Litterati are useful data sources for understanding the composition of urban litter, including for answering manager questions such as, "What trash categories need to be better controlled to reduce litter management costs" or scientific questions such as, "Is urban litter a major source of secondary microplastics to the environment?" To answer questions related to spatial distribution of litter, of interest to managers for prioritizing street sweeping routes and scientists looking to better understand sources of pollution, municipal audits are a more robust information source than Litterati data. To improve their usefulness to the broader research community, municipalities should release digitized versions of audit data, in addition to the reports of audit results currently available. If citizen scientists begin collecting metadata related to distance or time searched, as well as observing over broader spatial areas as we outline above, these citizen science apps can also be a source of spatial information on litter. Temporal questions, such as whether street sweeping schedules can be optimized by day or season or whether temporal trends in litter reflect temporal trends in plastic pollution elsewhere, cannot currently be addressed through either method. Increasing the frequency of municipal audits is likely cost-prohibitive, but citizen science could be used to fill these data gaps, if observations can be normalized to account for organized event submission surges and additional metadata on effort are collected to allow litter density to be compared through time.

In summary, both citizen science data and municipal audit data require some improvements to become adequate substitutes for scientific research on urban litter, but both even in the current form, are valuable resources for complementing existing research efforts. These non-traditional data sources can be utilized in scientific inquiry for understanding litter composition and in the case of municipal audits, quantifying litter densities and distributions.

Author contributions

Lisa Watkins led the project from conceptualization, method development, and data curation through to analysis, writing, and submission. David Bonter, Patrick Sullivan, and Todd Walter provided support and review of the methodology, analysis, visualization, and writing.

Conflicts of interest

There are no conflicts to declare.

References

1 N. Armitage and A. Rooseboom, The removal of urban litter from stormwater conduits and streams: Paper 1 – The quantities involved and catchment litter management options, *Water SA*, 2000, **26**(2), 181–188.

2 Keep America Beautiful, *2020 National Litter Study, Summary Report*, 2020.

3 HDR, San Francisco 2014 Litter Study, San Francisco, 2014, p. 7, Available from: **http://www.sfexaminer.com/wp-content/uploads/2015/12/San-Francisco-2014-Litter-Study_final.pdf**.

4 Dillon Consulting Limited, City of Vancouver Street Litter Audit, 2017 Results, Vancouver, 2017, p. 59. Report No.: 17–6318, Available from: **https://vancouver.ca/files/cov/street-litter-audit-survey-report-2017.pdf**.

5 K. Blair, 2016 Toronto Litter Audit, AET Group Inc., 2016, p. 90, Report No.: TOR_WAC1516_110, Available from: **https://www.toronto.ca/wp-content/uploads/2017/10/8ed5-Toronto-Litter-2016-Final-Report_App_Final.pdf**.

6 California Regional Water Quality Control Board, San Francisco Bay Region Municipal Regional Stormwater Permit, 2015, p. 10, Report No.: R2-2015–0049, Available from: **https://www.cleanwaterprogram.org/images/uploads/R2-2015-0049.pdf**.

7 Maryland Department of the Environment, National Pollutant Discharge Elimination System Municipal Separate Storm Sewer System Discharge Permit, 2010, p. 4, Report No.: 06-DP-3320, Available from: **https://www.montgomerycountymd.gov/DEP/Resources/Files/downloads/water-reports/npdes/MOCO_MS4_Permit.pdf**.

8 E. A. Weideman, V. Perold, G. Arnold and P. G. Ryan, Quantifying changes in litter loads in urban stormwater run-off from Cape Town, South Africa, over the last two decades, *Sci. Total Environ.*, 2020, **724**, 138310.

9 K. Mann, *Motivations Regarding Street Litter Data*, San Jose, CA, 2020.

10 P. E. Duckett, V. Repaci, P. E. Duckett and V. Repaci, Marine plastic pollution: using community science to address a global problem, *Mar. Freshwater Res.*, 2015, **66**(8), 665–673.

11 K. Critchell and M. O. Hoogenboom, Effects of microplastic exposure on the body condition and behaviour of planktivorous reef fish (Acanthochromis polyacanthus), *PLoS One*, 2018, **13**(3), e0193308.

12 C. J. Foley, Z. S. Feiner, T. D. Malinich and T. O. Höök, A meta-analysis of the effects of exposure to microplastics on fish and aquatic invertebrates, *Sci. Total Environ.*, 2018, **631–632**, 550–559.

13 M. D. Prokić, T. B. Radovanović, J. P. Gavrić and C. Faggio, Ecotoxicological effects of microplastics: Examination of biomarkers, current state and future perspectives, *TrAC, Trends Anal. Chem.*, 2019, **111**, 37–46.

14 J. B. Ross, R. Parker and M. Strickland, A survey of shoreline litter in Halifax Harbour 1989, *Mar. Pollut. Bull.*, 1991, **22**(5), 245–248.

15 W. C. Li, H. F. Tse and L. Fok, Plastic waste in the marine environment: A review of sources, occurrence and effects, *Sci. Total Environ.*, 2016, **566–567**, 333–349.

16 C. Wallis, Terrestrial fugitive plastic packaging: the blind spot in resolving plastic pollution, *Green Mater.*, 2020, **8**(1), 3–5.

17 E. R. Zylstra, Accumulation of wind-dispersed trash in desert environments, *J. Arid Environ.*, 2013, **89**, 13–15.

18 J. R. Jambeck, R. Geyer, C. Wilcox, T. R. Siegler, M. Perryman, A. Andrady, *et al.*, Plastic waste inputs from land into the ocean, *Science*, 2015, **347**(6223), 768–771.

19 S. Robinson, Littering behavior in public places, *Environ. Behav.*, 1976, **8**(3), 363–384.

20 B. Lockwood, B. R. Wyant and H. E. Grunwald, Locating Litter: An Exploratory Multilevel Analysis of the Spatial Patterns of Litter in Philadelphia, *Environ. Behav.*, 2020, **53**(6), 601–635.

21 W. Cowger, A. Gray, H. Hapich, J. Osei-Enin, S. Olguin, B. Huynh, *et al.*, Litter origins, accumulation rates, and hierarchical composition on urban roadsides of the Inland Empire, California, *Environ. Res. Lett.*, 2022, **17**(1), 015007.

22 Dillon Consulting Limited, City of Vancouver Street Litter Audit, 2019 Results, Vancouver, 2019, p. 71, Report No.: 19–1393, Available from: https://vancouver.ca/files/cov/street-litter-audit-survey-report-2019.pdf.

23 J. Maracle, *2017 City of Edmonton Litter Audit Report*, AET Group Inc., 2017, p. 67, Report No.: EDM_WAW1617_112.

24 S. R. Stein, *Anacostia Watershed Litter Survey*, Environmental resources Planning, LLC, 2015, p. 97.

25 P. Chauo, *Vancouver Municipal Litter Audits*, 2020.

26 National Academies of Sciences, Engineering, and Medicine, *Reducing Litter on Roadsides*, The National Academies Press, Washington, DC, 2009, DOI: 10.17226/14250.

27 European Environment Agency, *Marine LitterWatch*, Available from: https://marinelitterwatch.discomap.eea.europa.eu/Index.html.

28 Litterati, Vancouver litter submissions: June 2, 2013 to April 22, 2020, Global Litter Database, Available from: https://opendata.litterati.org/.

29 NOAA Marine Debris Program, University of Georgia College of Engineering, Marine Debris Tracker, Available from: https://debristracker.org/.

30 Ocean Conservancy, Clean Swell, Available from: https://oceanconservancy.org/trash-free-seas/international-coastal-cleanup/cleanswell/.

31 J. R. Jambeck and K. Johnsen, Citizen-Based Litter and Marine Debris Data Collection and Mapping, *Comput. Sci. Eng.*, 2015, **17**(4), 20–26.

32 R. J. Bator, A. D. Bryan and P. Wesley Schultz, Who Gives a Hoot?: Intercept Surveys of Litterers and Disposers, *Environ. Behav.*, 2011, **43**(3), 295–315.

33 P. W. Schultz, R. J. Bator, L. B. Large, C. M. Bruni and J. J. Tabanico, Littering in Context: Personal and Environmental Predictors of Littering Behavior, *Environ. Behav.*, 2013, **45**(1), 35–59.

34 H. K. Burgess, L. B. DeBey, H. E. Froehlich, N. Schmidt, E. J. Theobald, A. K. Ettinger, *et al.*, The science of citizen science: Exploring barriers to use as a primary research tool, *Comput. Sci. Eng.*, 2017, **208**, 113–120.

35 Litterati, Data for a litter free world, Available from: https://www.litterati.org.

36 Dillon Consulting Limited, Available from: https://www.dillon.ca/.

37 Dillon Consulting Limited, *City of Vancouver Street Litter Audit, 2018 Results*, Vancouver, 2018, p. 101, Report No.: 18–8281.

38 QGIS Development Team, QGIS Geographic Information System, Open Source Geospatial Foundation Project, 2021, Available from: https://qgis.org/en/site/.

39 R Core Team, *R: A language and environment for statistical computing*, Vienna, Austria, R Foundation for Statistical Computing, 2020, Available from: https://www.R-project.org/.

40 City of Vancouver. 311 Contact center Services, City of Vancouver Open Data Portal, 2022, Available from: https://opendata.vancouver.ca/explore/dataset/3-1-1-contact-centre-metrics/information/.

41 iNaturalist, Available from: https://www.inaturalist.org.

42 City of Vancouver. Local Area Boundary, City of Vancouver Open Data Portal, 2019, Available from: https://opendata.vancouver.ca/explore/dataset/local-area-boundary/information/?disjunctive.name.

43 City of Vancouver, Statistics Canada. Census local area profiles 2016, City of Vancouver Open Data Portal; 2018, Available from: https://opendata.vancouver.ca/explore/dataset/local-area-boundary/information/?disjunctive.name.

44 J. R. Prendergast, S. N. Wood, J. H. Lawton and B. C. Eversham, Correcting for Variation in Recording Effort in Analyses of Diversity Hotspots, *Biodivers. Lett.*, 1993, **1**(2), 39–53.

45 J. Ammendolia, J. Saturno, A. L. Brooks, S. Jacobs and J. R. Jambeck, An emerging source of plastic pollution: Environmental presence of plastic personal protective equipment (PPE) debris related to COVID-19 in a metropolitan city, *Environ. Pollut.*, 2021, **269**, 116160.

46 O. Pietz, M. Augenstein, C. B. Georgakakos, K. Singh, M. McDonald and M. T. Walter, Macroplastic accumulation in roadside ditches of New York State's Finger Lakes region (USA) across land uses and the COVID-19 pandemic, *J. Environ. Manage.*, 2021, **298**, 113524.

47 L. Mair and A. Ruete, Explaining Spatial Variation in the Recording Effort of Citizen Science Data across Multiple Taxa, *PLoS One*, 2016, **11**(1), e0147796.

© 2024 The Author(s). Published by the Royal Society of Chemistry

Environ. Sci.: Adv., 2024, **3**, 885–896 | **895**

48 J. L. Pearce and M. S. Boyce, Modelling distribution and abundance with presence-only data, *J. Appl. Ecol.*, 2006, **43**(3), 405–412.

49 B. L. Sullivan, C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink and S. Kelling, eBird: A citizen-based bird observation network in the biological sciences, *Comput. Sci. Eng.*, 2009, **142**(10), 2282–2292.

50 R. L. H. Dennis and C. D. Thomas, Bias in Butterfly Distribution Maps: The Influence of Hot Spots and Recorder's Home Range, *J. Insect Conserv.*, 2000, **4**(2), 73–77.

51 A. I. T. Tulloch, K. Mustin, H. P. Possingham, J. K. Szabo and K. A. Wilson, To boldly go where no volunteer has gone before: predicting volunteer activity to prioritize surveys at the landscape scale, *Divers. Distrib.*, 2013, **19**(4), 465–480.

52 A. Bowser, D. Hansen, Y. He, C. Boston, M. Reid, L. Gunnell, *et al.*, Using gamification to inspire new citizen science volunteers, in *Proceedings of the First International Conference on Gameful Design, Research, and Applications*, Association for Computing Machinery, New York, NY, USA, 2013, pp. 18–25, DOI: **10.1145/2583008.2583011**.

53 M. S. Rad, A. von Kaenel, A. Droux, F. Tieche, N. Ouerhani, H. K. Ekenel, *et al.*, A Computer Vision System to Localize and Classify Wastes on the Streets, in *Computer Vision Systems*, ed. Liu M., Chen H. and Vincze M., Cham, Springer International Publishing, 2017. pp. 195–204.

54 A. Kankane and D. Kang, Detection of Seashore Debris with Fixed Camera Images using Computer Vision and Deep learning, in *2021 6th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, 2021, pp. 34–38.

55 Litterati. San Francisco litter submissions, Global Litter Database, 2020, Available from: **https://opendata.litterati.org/**.

56 G. M. Kondolf and S. Li, The pebble count technique for quanitfying surface bed material size in instream flow studies, *Rivers*, 1992, **3**(2), 80–87.

57 M. G. Wolman, A method of sampling coarse river-bed material. Transactions, *Am. Geophys. Union*, 1954, **35**(6), 951–956.