To appear in the Journal of Industrial Ecology*, 2021*

This city is not a bin: Crowdmapping the distribution of urban litter

Andrea Ballatore[1], Stefano Cucurachi[2,*], Teun Johannes Verhagen[2], Zhije Li[2]

[1]Dept. of Geography, Birkbeck, University of London, London, UK

[2]CML - Institute of Environmental Sciences, Leiden University, Leiden, The Netherlands

[*]**Corresponding Author:** s.cucurachi@cml.leidenuniv.nl, Einsteinweg 2 2333CC, Leiden, The Netherlands

Conflict of Interest Statement: The authors declare no conflict of interest.

Keywords: spatial analysis; litter; waste; data analysis; Litterati; crowdmapping

Abstract

Urban litter, such as cans, packaging, and cigarettes, has significant impacts and yet little is known about its spatio-temporal distribution, with little available data. In contexts of data scarcity, crowdsourcing provides a low-cost approach to collecting a large amount of geo-referenced data. We consider 1.7 million litter observations in the Netherlands, collected by the crowdmapping project Litterati. First, we analyze the biases of this data at the province and municipality level. Second, in a local case study with high-quality data (the city of Purmerend), we investigate the spatial distribution of urban litter and the points of interest that attract it. This study's findings can support both the crowdmapping process, steering volunteers efforts, and policy-making to tackle litter at the urban level.

Article Type: Research and analysis

INTRODUCTION

The generation of solid waste has been steadily increasing in the past decades, outpacing the growth rates of global urbanization and industrialization. Worldwide urban waste generation is predicted to increase from 0.64 kg/capita/day in 2010 to an average of 1.2-1.4 kg/capita/day by 2025 (Hoornweg and Bhada-Tata 2012). A fraction of the anthropogenic waste generated is not treated in any facility. Mismanaged urban waste is comprised of urban litter, defined as waste that is not adequately disposed of (e.g. packaging left on the curb), and inadequately contained waste , which is transportable via runoff and wind from land to water bodies (Lebreton and Andrady 2019). There is increasing evidence that mismanaged waste on land, especially in coastal regions, is a major contributor to the accumulation of anthropogenic debris of all forms in the marine environment (Brennan et al. 2018; Hardesty et al. 2017).

The environmental consequences of the growing level of accumulation of macro- and micro-anthropogenic debris on land and global waters are still under assessment. Nevertheless, macro- and micro-litter types and plastics, in particular, have been associated with endpoints such as entanglement and smothering, and also with potential effects on toxicity, bioaccumulation, and reproduction, among others (Bergmann et al. 2015; Li et al. 2016). Litter not only presents a hazard to wildlife, but also affects human health, e.g. via ingestion of small particles that enter the foodweb (Carbery et al. 2018) and human wealth, as litter determines economic loss e.g. in the form of costly clean-up operations and decline in touristic attractiveness due to deteriorated landscapes (Kiessling et al. 2017).

Also, litter represents a missed opportunity for material recovery as part of ongoing efforts of urban mining (Di Maria et al. 2013), since materials such as plastics or aluminium could also enter remanufacturing streams, should litter be reduced or recovered from the natural environment.

Traditionally, because of the poor spatial and temporal quality of data, scientists have been making inferences on the spatial distribution of litter based on small samples of data, with wide margins of error and uncertainty (Lynch 2018). New data directly collected from citizens can shed new light on urban litter. Crowd-mapping, in particular, is a form of digitally-mediated collaborative work that empowers citizens to collect and share geographic knowledge for a variety of purposes (Quattrone et al. 2014; Senaratne et al. 2017). Using a smartphone, citizens and organizations can trace phenomena in space and time, providing data that can be processed cheaply by analysts.

Notably, applying this approach to litter mapping, the company Litterati (litterati.org) collects granular spatio-temporal data about various forms of litter through volunteers. The collection process of Litterati is organised around "challenges", i.e. mapping parties that gather volunteers in specific places or around individual efforts. The goal of Litterati is to build and maintain a global map of litter on land, supporting litter management and environmental activism (Litterati 2020). The Litterati dataset provides highly granular information about the presence of litter, more detailed than any other existing source, such as openlittermap.com (Lynch 2018).

The use and analysis of crowd-mapping data can increase the efficiency of urban sustainability governance (Certomà et al. 2015). Crowd-mapping data and other data from citizen science are considered a core new source of data that could be used to monitor and track progress, and to support decision-making on the sustainable development goals (SDG), as such data provide a finer spatial and temporal scale than traditional data sources (Fraisl et al. 2020; Fritz et al. 2019). Specifically, an improved understanding of the phenomenon of litter generation can contribute to better guiding policy-making for sustainability. The identification of the hotspots of litter generation and the coldspots of litter collection can better direct policy efforts towards improved collection, thus intervening to limit the environmental impacts of the sheer presence of litter in the natural environment, and can contribute to the recovery of potentially valuable materials. Such policy interventions could directly work as a fly-wheel for both litter generation and collection, as demonstrated by well-established findings from experimental research that suggest that individuals are more likely to litter in a littered environment, compared to a clean one, and are less likely to litter after observing someone pick up litter (Cialdini 2003; Keizer et al. 2008; Schultz et al. 2013; Hartley et al. 2018).

Litterati data, while abundant and detailed, may suffer from several biases intrinsic to the collection process. Volunteers are a self-selected population and tend to belong to specific demographic groups in terms of age, education, income, and ethnicity (Buil-Gil et al. 2020). The mapping activity is not spread evenly spatially and temporally, but tend to cluster in areas and periods that will be over-represented (Ballatore and De Sabbata 2020). While existing work investigated data representativeness in a variety of projects (Ballatore and Jokar Arsanjani 2019), crowdsourced litter data has not received any attention and little is known about its properties and limitations.

In order to understand the characteristics of urban litter, we consider 1.7 million crowdsourced litter observations for the Netherlands from Litterati, generated in the period 2016–2019. This dataset captures the presence and type of urban litter to an unprecedented scale and volume. First, we analyze the biases of litter data from Litterati at the national level for the Netherlands, addressing the following questions (***QA***): (1) What is the spatial and temporal distribution of litter observations at the national level? Are there hotspots? (2) What is the relationship between population density and litter observation density at multiple spatial scales? (3) What is the relationship between the geo-demographic characteristics (age, ethnicity, education, and income) and litter observations?

Second, we focus on a local case study (the city of Purmerend), in which the Litterati data is extremely dense and has been produced by a professional data collector and campaigner. To characterize the urban environment, we harvest a variety of points of interest (POIs) from Google Places. This analysis reveals previously unknown details of urban litter and allows answering (***QB***): (1) What is the temporal and spatial distribution of litter observations at the urban scale? Are there litter hotspots? (2) What types and volumes of litter are present? Are there visible brands? (3) What is the spatial clustering between litter observations? What is the spatial clustering between urban litter and points of interest (POIs), e.g. restaurants, cafés, schools? What types of POI tend to attract more litter?

Answering these research questions achieves two complementary goals. On the one hand, studying the characteristics of this crowdsourced data is the first step towards the production of better estimates of urban litter, mitigating the biases to extract more representative samples. On the other, the urban case study of Purmerend provides the first quantitative assessment of urban litter, enlightening multiple facets of this understudied phenomenon. These findings can inform both project owners and volunteers at initiatives such as Litterati, as well as urban policy-makers and waste managers.

To support reproducibility and replicability, compatibly with the terms of use of Litterati and Google, we make the study's data, results, and figures available as open data.[1] The next section covers the existing body of knowledge on litter and crowdmapping methods.

[1] https://github.com/andrea-ballatore/litter-dynamics (accessed in Feb 2021)

RELATED WORK

Litter estimation

In the field of industrial ecology, waste is classified as in use, post-consumer managed and mismanaged waste (Geyer et al. 2017). Waste statistics are notoriously scattered, fragmented, and present high variations by year, and country. As reported by Tisserant and co-authors (Tisserant et al. 2017), high-income countries usually have comprehensive waste accounts, while low- and middle-income countries have only a few waste types for which data are available. All of the waste accounts, no matter the country of origin, have poor coverage of unmanaged waste and litter even though these should be in principle covered by official statistics (Eurostat 2017). As a result, mismanaged waste including litter is typically either excluded due to lack of data [see e.g.(Brouwer et al. 2018)] or estimated based on the mass-balance, with considerable uncertainty on the effective quantity of mismanaged waste ending up in the natural environment.

An example of mass-balance estimation is the work of Jambeck and co-authors (Jambeck et al. 2015), which estimate mismanaged and littered plastics based on (i) the mass of waste generated per capita annually by the countries analyzed; (ii) the percentage of waste that is plastic; and (iii) the percentage of mismanaged plastic waste. In particular, the authors assume that the fraction of litter in mismanaged plastics corresponds to 2% of waste generated in all countries. The assumption is based on litter dynamics in the United States(Jambeck et al. 2015).

Similarly, Lebreton and Andrady (2019) conservatively consider a minimum of 1% as a threshold for mismanaged plastics. The authors acknowledge a significant degree of uncertainty associated with the value and include scenarios with upper and lower thresholds from 0.1 to 10% minimum mismanaged waste. Only for the city of Philadelphia, US, we found a study that explored the spatial and temporal elements of urban litter (Lockwood et al.

2020). No other approaches to map other urban litter types can be found to the best of our knowledge.

Estimates of urban litter by mass-balance and other model-based approximations are associated with large uncertainties and do not account for spatial heterogeneity within a country and between countries, i.e. the unevenness of litter distribution in different areas. More accurate alternatives require the sampling of litter on location, though such approaches are currently mostly limited to marine and freshwater systems [see e.g. Emmerik and co-authors (van Emmerik et al. 2018; Van Emmerik et al. 2019)]. Moreover, sampling approaches are expensive and time-consuming, as they require collection equipment and several people to perform the measurements (Kiessling et al. 2019; van Emmerik et al. 2018).

Crowdmapping and citizen science

Advances in information technology in the last decade have enabled new forms of data collection and citizen participation (See et al. 2016). In ecology, large projects involving citizen science have allowed keeping track of the ecological and social impacts of large-scale environmental change, e.g. monitoring phenology, relative abundance, distributions, survival, and reproductive success of organisms across time and space (Dickinson et al. 2010). Projects and methods of this kind hinge on contributions from non-expert, non-professional, typically unpaid users have been variously labelled as user-generated content, crowdsourcing, volunteered geographic information, as well as citizen science (Senaratne et al. 2017; See et al. 2016). Citizen science and citizen-generated data are being recognized as an important source of information for policy and actions on sustainable development (Fritz et al. 2019).

Litter has been mapped using citizen science data, e.g. in volunteer beach cleanup actions (Lynch 2018). These data sources are typical of lower quality than professionally-collected data but are often available in greater volumes and at a lower cost (Haarr et al. 2020).

Citizens, such as outdoor hobbyists, gather data over large geographic regions and enter them into a centralized system, which provides interactive graphs and maps (Cooper et al. 2007). When data have a strong geospatial component, collaborative mapping can produce large volumes of contributions at a high spatial and temporal granularity (See et al 2016). While non-profit projects such as OpenStreetMap uniquely rely on crowd-mapping, prominent commercial ventures such as Google Places crowdsource information from their users too to complement and enrich their informational assets.

Variable data quality poses a persistent challenge in this type of data production, limiting its applications in geographic data science (Ballatore and De Sabbata 2020). Crowdsourced data may be affected by different, interacting biases (Senaratne et al 2017). The demography of projects tends to be skewed towards relatively educated, affluent, urban contributors who have access to IT infrastructure and skills, although exceptions exist (Ballatore and Jokar Arsanjani 2019). Population density influences data density, making urban areas often better mapped than rural areas. Human-made features such as buildings are easier to categorize and demarcate than vague natural ones such as forests and rivers. Basic, clear mapping tasks are more likely to produce good quality data than complex ones.

Crowd-mapping projects also suffer from spatial heterogeneity that reflects the human geography of their managers and contributors. Data collection tends to happen in bursts that are highly localised in space and time (e.g., in mapping parties in cities), and popular locations attract more data than unpopular ones, neglecting low-density rural areas without particular attractions. The sustained, long-term, homogenous, systematic mapping performed by traditional mapping agencies is harder if not impossible to achieve (Senaratne et al. 2017). Some early work aims at developing bias-reduction methods, bootstrapping more representative models of phenomena from highly skewed samples (Buil-Gil et al. 2020). In the context of litter mapping, no prior work has been done on the characteristics of crowdsourced litter observations. Our study lies at the intersection between geographic data science and industrial ecology, showing how these novel data sources may contribute to managing crucial environmental challenges.

METHODS

Study overview

In this study, we use spatial analysis to investigate our research questions, starting from the distribution of urban litter observations to the association between litter observations and points of interest (POIs). As shown in Figure 1, we rely on two new datasets, containing litter observations from Litterati and POIs from Google Maps. The data is first analyzed at the national scale, using a clustering method (Gi*) to identify litter hotspots. A local case-study is then considered (Purmerend, NL). A temporal analysis is also performed at both national and local scale. At the local scale, we use Gi* to identify litter hotspots, and we perform a distance analysis, relying on the K and L functions. This spatial analysis method allows observing to what extent spatial objects or events tend to be associated. First, we observe the distance between litter observations of different types. Second, we consider the distance between POIs and litter, overall and by types, quantifying their spatial association. The remainder of this section will discuss the datasets and the methods for the temporal and spatial analyses in this study.

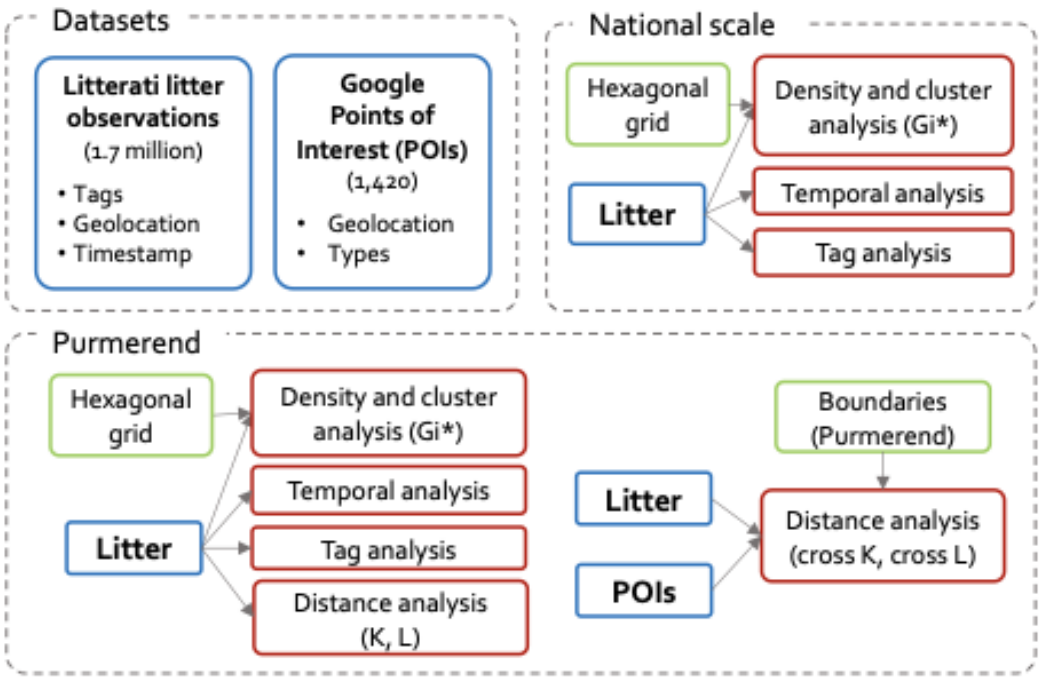

Figure 1. Overview of the study, including input datasets and methods

Litter dataset

Litterati has global coverage with 166,800 active members across 165 countries, with a total litter count currently at 5.8 million units as of January 2020 (Litterati 2020). Users from the Netherlands are the most active on the platform, followed by the United States and Austria. Users generate these observations with a dedicated smartphone app. When users see a piece of litter, they photograph it, select tags for it (e.g., "plastic", "wrapper"), and upload it, and are encouraged to dispose of it, combining data collection with activism and direct clean-up. This reduces the likelihood of double-counting the same piece of litter. In January 2020, we obtained from Litterati litter observations located in the Netherlands from 2012 to 2019, generated by about 7,000 users. As the data before 2016 is very sparse (687 data points in 2012-2014 and 5,278 in 2015), we included in the study litter data from 2016 to 2019.

This dataset comprises 1.7 million litter observations, with 3.5 million corresponding tags (with an average of 2.1 tags per litter observation), and the 2019 data also provides anonymized information about users (4,589 users). As discussed above, the coverage and sampling quality of these observations vary, and largely dependent on the practices and preferences of volunteers and local organizers. We distinguish here between litter *tags*, which are the labels attributed to items by the users of the Litterati platform, and litter *types*, which represent the litter items that can be identified from their tags.

Google points of interest

To study the relationship between urban POIs and litter in the Purmerend case study (*QB3*), we had to retrieve high-quality and up-to-date POI information. To the best of our knowledge, several data sources, including official government data, had limited coverage of POIs that could be associated with litter, and we could not identify a suitable specialist commercial provider. Hence, we concluded that Google Maps hosted the most appropriate data, including urban POIs at high coverage, collected and verified by the company with the contribution of the large Google user base (https://cloud.google.com/maps-platform/places).

Using a dedicated research tool that we developed (https://github.com/andrea-ballatore/google-places-api-r), we retrieved 1,420 POIs located in Purmerend from the Google Maps Places API in March 2020. This tool partitions the query area into small areas and generates API queries, re-splitting the areas as needed. This method ensures full coverage of POIs in the target area. Each resulting POI is associated with a location and with multiple "types" that describe its purposes. In particular, the dataset includes stores (850), transit stations (255), food (246), restaurants (128), schools (22), and places of worship (19) (note that a single POI can be categorised as "food", "meal takeaway", and "restaurant"). This kind of dataset provides very high coverage and accuracy of urban POIs relevant to littering behaviour. The analysis can be scaled up to any other geography covered by the Google API.

Hotspot analysis

To investigate the temporal dimension of the data collection process (QA1 and QB1), we structured the dataset in litter collected per year and the corresponding month to create a temporal overview of the data. We include a temporal analysis for all the available data in the Netherlands and our case study in Purmerend to gain insight into the possible seasonality of the data. To inspect spatial distribution, the litter observations are at first analyzed at the national scale using a density analysis based on a hexagonal grid. Subsequently, we carry out a hotspot analysis with Getis-Ord Gi* statistic, identifying hot and coldspots, i.e., areas of statistically significant concentration or dispersion of litter (Getis and Ord 2010).

Such density-based hotspot analysis is widespread in epidemiology and criminology, which are interested in identifying and predicting high-density areas. The core intuition is that, in spatial processes, it is hard to distinguish clusters of objects that occur stochastically (e.g., litter dispersed by the wind) and clusters that are non-random (e.g., litter clustered around a fast-food restaurant), and this method provides an interpretable test of statistical significance. A similar method is then applied at the urban scale for the urban environment of Purmerend, including density and hotspot analysis. By contrast, QA3, which focusses on the socio-demographic dimension of the data, is investigated through correlation analysis of socio-economic variables and litter counts at a fixed spatial scale.

Point pattern analysis

To study the clustering and dispersion of litter and POIs (QB3), we adopt methods from point pattern analysis, a branch of spatial analysis widely used in ecology and geography to study spatial processes in the natural and social sciences (Baddeley et al. 2015). Spatial association, also called clustering, can be observed between the litter points and between litter and POIs. Treating both litter observations and POIs as points in a projected metric space, we observe multi-distance clustering using point pattern analysis. The main function we use is Besag's L-function, a transformation of Ripley's K-function (O'Sullivan & Unwin, 2010). The K function summarizes the distance between points at a range of distances, for example between 0 and 1 km, dividing the mean of the sum of the number of points at different distance radii for each point by the area event density. More clustered phenomena will therefore produce higher values of K, having more events falling within the radius.

K can be applied to events of the same type (univariate) or to compare events of different type (cross K). K allows the detailed observation of the distance pattern, without having to choose an arbitrary radius, which is a major advantage in contexts where the exact scale of analysis is hard to determine (L is a more interpretable transformation of K). To ascertain to what extent K is likely to be the result of a random process or not, we use a Montecarlo simulation, as the direct calculation of a *p* value is difficult. When comparing K (or L) values, we scaled the values by considering their *z* scores to increase the interpretability of the results. Our model adopts Euclidian distance, as in an urban context it provides a good approximation of the street network distance (Apparicio et al. 2008). The next section outlines the results of these analyses.

RESULTS

Litter data biases at the national scale

Spatial and temporal distribution

Crowdsourced data may exhibit important spatial, temporal, and thematic biases. We start the analysis of national data (QA) by focusing on the spatio-temporal distribution of litter observations (*QA1*). Breaking down the 1.7 million litter observations per month reveals a relatively high collection rate during the summer and autumn period (June, July, August, September, October). The collection of litter in January and February is generally lower than the monthly collection during the summer or autumn. This trend can be explained by the wintry conditions in these months, which discourage outdoors activities, while the peak in the yearly collection of litter takes place during September as a result of the World Cleanup Day (15^{th} of September). 3.5 million litter tags provide information about the types of litter for 1.3 of the 1.7 million observations generated by participants (89% of observations have at least one tag). The most frequent tags are plastic (46%), can (12%), wrapper (10%), paper (9%), metal (8%), cigarette (5%), bottle (5%), with a tail of less frequent tags. Most litter types have a consistent collection rate throughout the year with little relative tag variation, with the exception of the seasonal "knetterball", a type of firework (see QB2 for more details on litter types and volumes). In terms of waste volume, estimating its weight based on material (KPLUSV 2015), the dominant types are glass (~5,500 kg), plastic bottles (~3,000 kg), metal cans (~2,500 kg), and plastic wraps (~1,300 kg). We refer the reader to the Supplementary Information for additional information.

Being an anthropic phenomenon, the distribution of litter observations is expected to be non-uniform, with variable density across the geographic space. Figure 2a illustrates the density of litter observations across the country (*QA1*). The hexagonal grid is fixed at 10 km^2, a spatial scale that was determined through a sensitivity analysis (varying the scale to higher and lower granularity by 100% had a negligible impact on the resulting hot and coldspots). The litter data appear to have national coverage, with all Dutch provinces being included, with higher density in some central areas of the country.

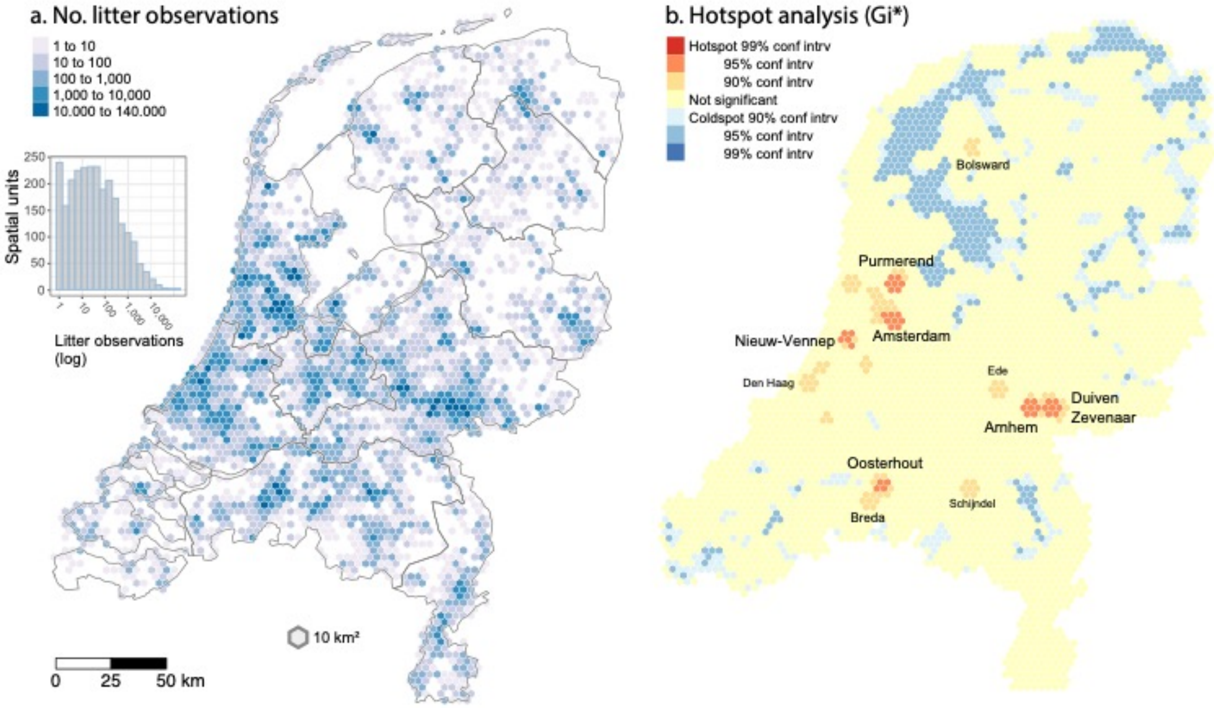

Figure 2. (a) Density of litter observations in the Netherlands (2016--2019) with province borders. Total litter observations: 1.71 million, grouped into five logarithmic bins, over 4,405 spatial units (each hexagon covers 10 km^2). (b) Hotspot analysis with Getis-Ord Gi* with significance level. Hotspots are labelled with municipality names. Most coldspots are bodies of water. The maps are north-oriented.

Assuming a homogeneous spatial Poisson process as a baseline, hotspot analysis allows detecting areas that have a higher or lower concentration of litter observations than expected with a random process. It is critical to note that these are hot and coldspots of litter observations generated by Litterati, and not necessarily of the actual litter distribution. As is possible to note in Figure 2b, six areas emerge as hotspots at 99% confidence, including Amsterdam and Purmerend, and other eight areas at 95% confidence (*QA1*). These hotspots indicate a higher level of activity by local Litterati volunteers and, rather surprisingly, do not correspond with the most populated cities. Hence, to explore this aspect more systematically, it is useful to observe the density of litter observations in proportion to the local population.

Litter observations by population

Observing the density of observations relative to the density of the location population allows studying the spatial distribution in the Litterati data, starting by observing the 12 Provinces of the Netherlands (*QA2*). Table 1 shows the summary of litter by 1,000 people for the provinces in descending order, including *z* scores for the divergence from average and inter-quartile range for distribution spread. Provinces range from those with an abundance of litter data ($z \sim 2$, more than 200 observations per 1,000 people, Gelderland, and Noord-Holland) to the least represented (Limburg, Overijssel, Drenthe, and Groningen, $z < -.5$). This indicator shows high regional variability, from 21 to 208, indicating that Litterati has uneven coverage with a few hotspots in the central part of the country dominating the dataset as discussed above. When considering population density, no clear pattern emerge, as litter observation density exhibits a weak correlation with population density at the province level (*rho* = .2) (Table 1).

Province	**Largest cities**	**Pop. (mill.)**	**Pop. Density (km^2)**	**Litter obs.**	**Litter obs. per 1,000 people**	***Z score***	***IQR***
Gelderland	Arnhem, Nijmegen, Apeldoorn	2.05	420	425,645	**208**	*2.1*	*152.3*
Noord-Holland	Amsterdam, Haarlem, Zaanstad	2.81	1,081	573,410	**204**	*2*	*134.5*
Noord-Brabant	Eindhoven, Tilburg, Breda	2.51	522	220,769	**88**	*.2*	*23.9*
Friesland	Leeuwarden, Sneek, Heerenveen	.65	195	53,109	**82**	*.1*	*77.9*
Zuid-Holland	Rotterdam, The Hague, Delft	3.62	215	215,647	**60**	*-.3*	*42.2*
Flevoland	Almere, Lelystad, Emmeloord	.41	300	23,476	**58**	*-.3*	*49.4*
Zeeland	Middelburg, Vlissingen, Goes	.38	912	19,490	**51**	*-.4*	*64.1*
Utrecht	Utrecht, Amersfoort, Houten	1.32	1,374	64,437	**49**	*-.4*	*35.9*

Limburg	Maastricht, Roermond, Heerlen	1.12	520	38,885	**35**	*-.6*	*23.1*
Overijssel	Zwolle, Almelo, Deventer	1.15	350	35,279	**31**	*-.7*	*22.1*
Drenthe	Emmen, Assen, Hoogeveen	.49	188	13,474	**27**	*-.8*	*9.5*
Groningen	Groningen, Oldambt, Stadskanaal	.58	252	12,493	**21**	*-.9*	*12.5*

Table 1. Provincial distribution of 1.7 million litter observations from Litterati 2016-2019 by Dutch Province, with litter observations per 1,000 people with Z scores and inter-quartile range. Ordered by litter per 1,000 people. Population estimates and density from CBS Statline 2017. Purmerend is located in Noord-Holland.

As shown in Figure 3a, the Eastern provinces are under-represented, while no clear spatial pattern emerges for the other parts of the country (QA2). The litter statistics calculated at the municipality level highlight the high variability within provinces (Figure 3b). The top municipalities include statistical outliers such as Duiven, Ouder-Amstel, Purmerend, and Oosterhout ($z > 3$, with values between 1,423 and 5,365 litter observations per 1,000 people). Surprisingly, major cities appear to be around the average or below average: Amsterdam and Almere are at the average (~100 litter observations per 1,000 people, $z = 0$), while Rotterdam, The Hague, Tilburg ($z = -.1$), Utrecht, Eindhoven, Groningen, Nijmegen, and Maastricht ($z = -.2$) are all slightly below average. The only notable exception is Breda, which is above average ($z = .3$). This confirms the absence of the urban bias that is usually expected in crowdmapping projects (QA2).

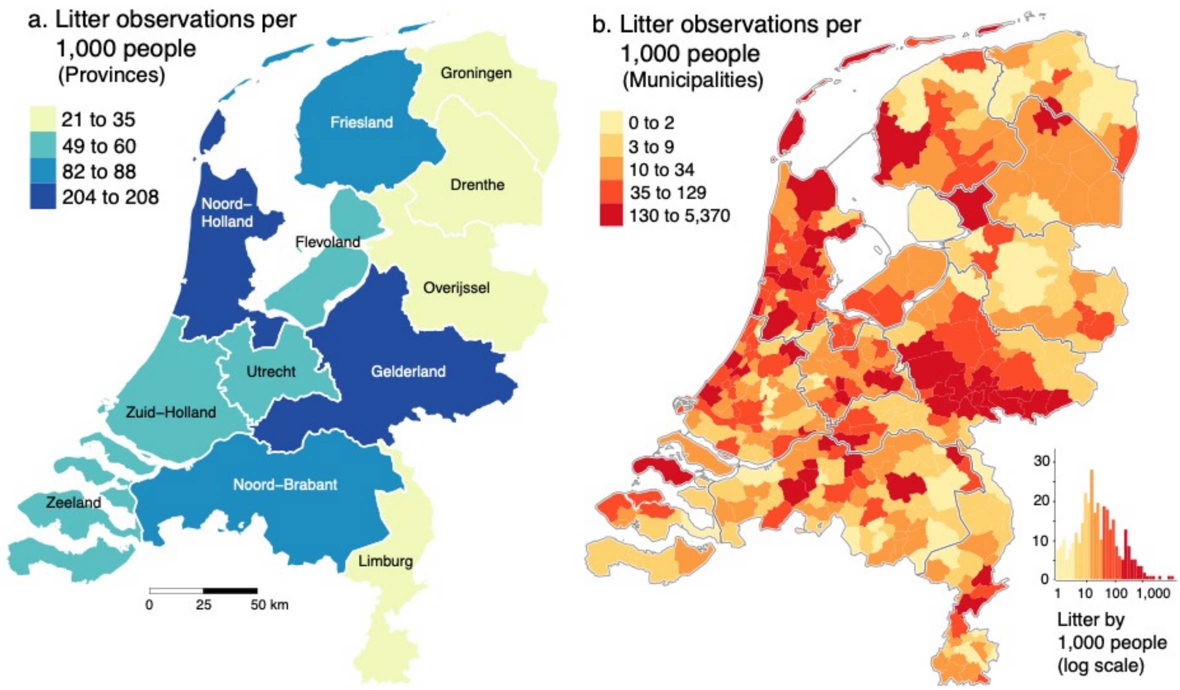

Figure 3. Litter observations from Litterati per 1,000 people in the Netherlands (2016--2019). (a) Province level, with manual bins (*provincies*, 12 units). (b) Municipality level, with Jenks bins (*gemeentes*, 388 units). Total litter observations: 1.71 million, total population: 17.2 million. Population data from Dutch Census (2017).

Geo-demographic analysis at the municipality scale

Considering the litter data at the national scale, socio-demographic factors of the resident population might influence the data distribution, highlighting geo-demographic biases (*QA3*). This type of analysis can reveal, for example, to what extent the collection tends to occur in older, more educated, wealthier areas of the country. In other studies, socio-economic factors are found to be of influence on the littering behaviour of residents within a city (Grimmer et al. 2016). Using data Dutch Census from 2017 (Kadaster 2017), we calculated a set of litter-related variables for 388 municipalities (*gemeente*), including the number of litter observations (overall and by litter type). Each municipality comprises between 941 and 844,947 residents with a median of 27,000 residents and is described in terms of total inhabitants, population density, and the percentage of residents by gender, age, and ethnicity. Average income (euro per capita) was also included in the analysis (CBS 2019).

When correlating litter observations and these socio-demographic factors, an expected correlation emerged between the population size and the number of litter observations (rho = .52, $p < .001$), and for this reason we only considered litter variables weighted by the local population (per 1,000 people). Using these variables, surprisingly, the correlations ranged from none to very weak. Higher population density weakly corresponds to more litter observations (rho = .17, $p < .001$), suggesting that denser urban areas have more active contributors. Most age groups showed no correlation, with a weak inverse correlation with the percentage of residents between 0 and 14 years of age (rho = -.15, $p < .01$). The percentage of different ethnic groups and average income also showed no correlation with litter. Overall, no clear and statistically significant trend was detected in the Litterati data at the municipality level (*QA3*). This indicates that the Litterati data collection does not present obvious geo-demographic biases, while volunteers are predominantly in their 30s and 40s.[2]

Urban litter in the Purmerend case study

Data collection and quality

In this second study (QB), we turn our attention to a local case study with high-quality data that can reveal insights about urban litter, with high confidence in the data's ability to reflect the actual distribution of litter. Purmerend emerged as the most promising case study for the quality and volume of its data. This municipality is located about 10 km north of Amsterdam, with 80,117 inhabitants as of 2019, occupying 23.39 km^2, and is an average urban area in the Dutch context from a socio-economic perspective. From 2016 to 2019, it received 178,557 litter observations with 456,140 tags (the third highest municipality). While the data was generated by 46 users, Purmerend exhibits an interesting case of extreme contribution inequality, with a single user generating 98% of the data.

[2] From Zoom interview with Dirk Groot (http://zwerfinator.nl), 4 November 2020.

This user, Dirk Groot - *Zwerfinator*, is well-known in the Netherlands as the top collector of litter data and has a public profile of advocacy and activism (Atlas Leefomgeving 2020). He is the only contributor that collects data professionally, generating litter data as a full-time paid job for local authorities and companies. This data is produced in a systematic way, focusing on high-litter areas for an average of 160 observations a day: Groot produces regular litter observations, following the same routes where he expects high litter density.[3] The spatial coverage, regularity, and volume of these litter observations make them suitable to answer research questions about urban litter (QB).

Spatial and temporal distribution

The collection of litter in Purmerend has the highest collection rate during the months of May, June and July, reaching 20,000–40,000 litter points. In these months Groot performs a yearly litter collection campaign for the city council of Purmerend. [2] The lowest collection rate (2,000–12,000) takes place in winter, reflecting the national trend. Given the systematic collection, this suggests that less litter is recorded in winter months (*QB1*). From personal communication with Groot,[2] we ascertained that the moment a piece of litter is collected (and recorded) is not a good predictor of the moment that piece of waste became litter. In particular, Groot could verify using personal data collected over the past three years that a higher quantity of beverage packaging could be found in the autumn months than in the summer months. Such a trend could be explained by the fact that shrubs lose their leaves and grass is mowed in the autumn, so that litter only becomes visible then.

Spatially, the litter observations in Purmerend exhibit high coverage, including the vast majority of the city's public spaces. Figure 4a shows the litter observation distribution, with hot- and coldspots highlighted in Figure 4b. Overall, the litter appears concentrated in a

---

[3] From Zoom interview with Dirk Groot (http://zwerfinator.nl), 4 November 2020.

North-West to South-East corridor, and absent only in some residential areas and a private golf course (*QB1*).

Most hotspots occur in central areas of the city, in the historic centre, train stations, and shopping malls. Hotspots in the outskirts are linked to fast-food restaurants and bus stations. Areas of social aggregation with litter hotspots include a laser tag hall, a shopping district, a church, a stadium, a school, and a water park. Another large hotspot in the same area is co-occurring with multiple supermarkets, bakeries and markets, which typically attract high foot traffic on a daily basis. The train stations are located centrally and intersect with several sidewalks. Most coldspots are located at the eastern edge of the municipality. The largest coldspot corresponds to a golf resort, only accessible to members and coldspots to the north and south of the golf course public parks, as well as residential areas, with a low density of litter and POIs. The coldspot at the northern edge of the city corresponds to a canal and a dual carriageway. Some smaller coldspots in the city center are located near canals.

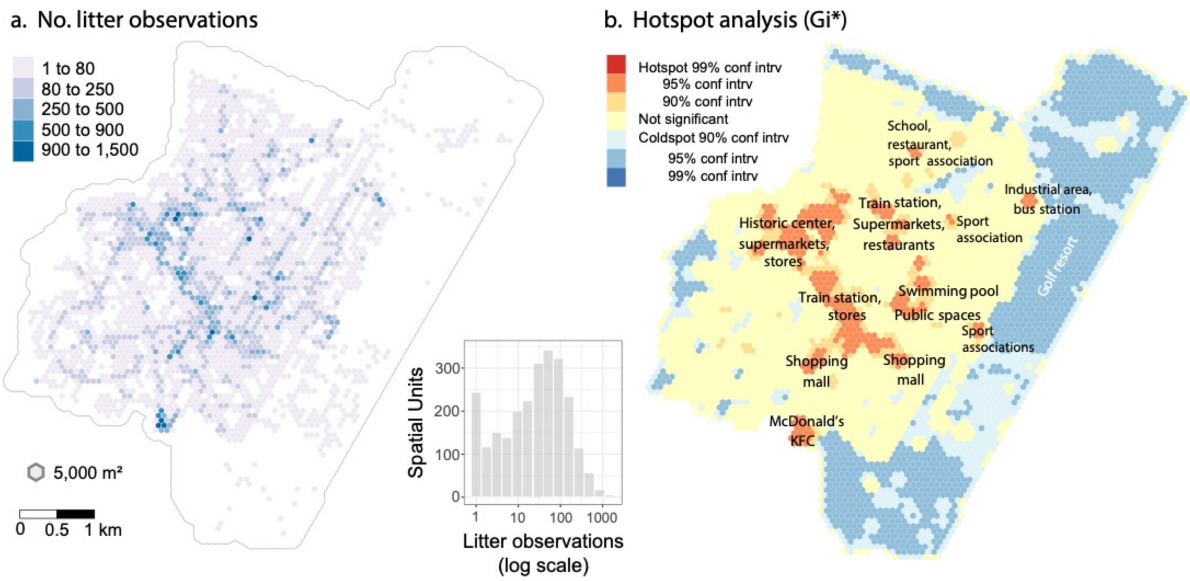

Figure 4. (a) The distribution of the 178,557 litter observations in Purmerend, Netherlands (2016–2019), grouped into 5 bins with Jenks. (b) Hotspot analysis with Getis-Ord Gi* with significance level. Hotspots and notable coldspots are labelled. The hexagonal grid includes 5,106 spatial units (each hexagon covers 5,000 m²). The maps are north-oriented.

Litter type analysis

This data allows quantifying the types of litter released in the urban environment. Most observations are tagged as plastic (50%), followed by wrapper (14%), can (12%), metal (11%), paper (11%), cigarette (7%), bag (5%), bottle (5%), energy drink (5%), beer (4%), and firework (4%), with other tags having lower prevalence (*QB2*).

Table 2 summarises the occurrence of litter types, which combine multiple tags, including the brands that emerge as most frequent. These findings are overall consistent with those at the national level: Urban litter is dominated by plastic wrappers, cans, cigarette butts, and plastic bags, with some brands being particular prominent (McDonald's, Knetterball, Anta-Flu, Red Bull, and Coca-Cola) (*QB2*). We also estimate the mass of litter. Glass bottles (30%), metal cans (17%), plastic bottles (14%), and plastic wrappers (13%) dominate the spectrum. By material category, both plastic and glass share around 30% of the total weight, in total taking up more than half of the total mass of litter.

Top litter type	*%*	*Mass(kg)*	*Weight per item(gram)* (KPLUSV 2015)	*Top brand*	*%*
Plastic wrapper	13	201.9	8.7	McDonald's	3.7
Metal can	10	267.8	15	Anta-Flu	2.6
Cigarette	7	6.2	0.5	Red Bull	2.5
Candy wrapper	6	21.4	2	Heineken	1.5
Piece of paper	5	8.9	1	Coca-Cola / Coke	1.0
Plastic bag	5	17.9~133.9	2.0-15	Marlboro	.9
Energy drink	5	133.9	15	Amstel	.8
Plastic bottle	4	212.8	29.8	Candyman Mac Bubble	.6
Beer can	3	80.4	15	Lipton	.5
Firework / Knetterball	3	46.6	8.7	Fanta	.5
Cup	2	16.1	4.5	Haribo	.3
Tissue	2	17.9	5	KFC	.3
Straw	1	0.7	0.42	Bavaria	.2
Glass bottle	1	491.0	275	Camel	.2

Table 2: Most prominent litter types with estimated mass and brands in the Litterati data for Purmerend (2016–2019). Total number of observations: 178,557.

Spatial clustering of litter and POIs

The various types of litter exhibit varying level of spatial clustering. Both litter observations and POIs can be modelled as points produced by a homogeneous Poisson point process. As stated in the methods section, the L-function allows studying whether points are clustered in a statistically significant way at different distances, with values higher than the complete spatial randomness baseline (CSR) indicating clustering, and lower values indicating dispersion.[4] To estimate the statistical significance of the L-function, we use 99 Monte Carlo simulations, resulting in an envelope at alpha level 0.02, derived as $2\ rank\ (n\ +\ 1)$, where *rank* is 1 and n is the number of simulations, which we consider acceptable for this analysis.

Litter clustering. We apply the L-function to litter observations of the most frequent types between 0 and 1,000 metres, a range suitable for this urban analysis (see also Supplementary Information). All litter groups are above the CSR and outside of their envelopes, indicating statistically significant clustering, as largely expected. These clustering patterns appear homogenous, with most types converging, particular at a short distance (0 to 200 meters). The only two types that stand out as being more clustered then the others are fireworks (also highly clustered temporally) and cigarettes, while bags are slightly less clustered than other types while having a similar overall pattern (*QB3*).

Clustering of Google POIs. To answer questions about the spatial association between litter and POIs, it is important to analyze the spatial structure of POIs to see how they cluster to one another in the urban space. Different types of POIs are expected to cluster at different levels (Krider and Putler 2013). When applying the L function to the most frequent 12 POI

---

[4] All functions were calculated with the R library *spatstat* https://cran.r-project.org/web/packages/spatstat (v1.64-1).

categories in Purmerend, the majority of Google POI types show, as expected, higher clustering than CSR, a common feature of urban structure, at almost all distances considered. The intensity and spatial structure of clustering vary considerably between POI types, with restaurants and cafes showing the highest level, exhibiting a tendency to agglomeration that is common in retail (Krider and Putler 2013). Most POIs are at an intermediate level (parking, meal takeaway, supermarkets, bakeries, and places of worship), while parks and transit stations show no clustering, confirming their planned nature, as they are deliberately not located too close to each other. Overall, the variability between POIs is much higher than that between litter types, which appear more spatially homogeneous (*QB3*).

Distance analysis. To study the spatial association between litter and POIs, we conduct a distance analysis, followed by a point pattern analysis. Assuming that litter observations are produced in a non-random spatial pattern, we consider POIs such as supermarkets, schools, and restaurants, which attract the presence of people for a variety of purposes. The analysis focuses on the spatial association between the 178,557 litter observations and the 1,420 Google POIs in Purmerend (*QB3*). The scale of the analysis is chosen to observe the co-location of litter and POIs, assuming that the presence of a POI can generate litter in the environment (e.g., fast-food packaging may be released near the restaurant that sold it).

According to Dirk Groot, litter is highly spatially structured and tends to appear in specific locations near specific POIs, for example in "candy&snack routes" in proximity to schools and shops that sell candies and snacks.[5] While spatial association does not imply causality, it is reasonable to assume that a substantial part of the data concerns litter coming from POIs, such as shops, supermarkets, and fast-food restaurants. Of 48 POI types defined by Google, we discarded generic ones (such as "food" and "health") in favor of those used as main POI

[5] From Zoom interview with Dirk Groot (http://zwerfinator.nl), 4 November 2020.

description (e.g. "restaurant" and "pharmacy"). Some types were discarded as too infrequent to provide useful information (e.g. "roofing contractors").

This selection resulted in 25 relevant POI types. When selecting which litter points fall within a given radius from any POI, a 50-meter radius comprises 44% of all litter, 100 meters 78%, while a radius of 200 meters reaches 98% of litter. Hence, we consider a distance between 0 and 100 meters as suitable to characterize the vast majority of litter in the dataset. Above this threshold, each litter point tends to fall within several POIs, making the association difficult to ascertain. As a first step, we observe the relative presence of litter in this distance range (0-100 meters) by POI type and the average litter by POI (*QB3*). These two indicators are calculated as:

$$c = \frac{|lit_{(poi,d)}|}{|lit|} \qquad c_p = \frac{|lit_{(poi,d)}|}{|poi_{cat}|}$$

where *lit* is the set of litter observations, *poi* is the set of POIs, *cat* is a POI category, and *d* is a distance threshold in metres. These two indicators are displayed in Figure 5, at five different distance thresholds, allowing a comparison between the overall presence of litter near POIs and their average, weighting very frequent and rare POIs. As some POIs are very frequent than others, it is beneficial to consider these ratios as opposed to an absolute count. 24% of all litter falls within 100 meters from Purmerend's 128 restaurants, while 3% of all litter is near the 4 department stores in Purmerend, but restaurants are co-located with litter below average (331 observations per POI) and department stores above (1,186). At 100 metres, the average percentage of litter by category is 4, while the average litter per POI is 670.

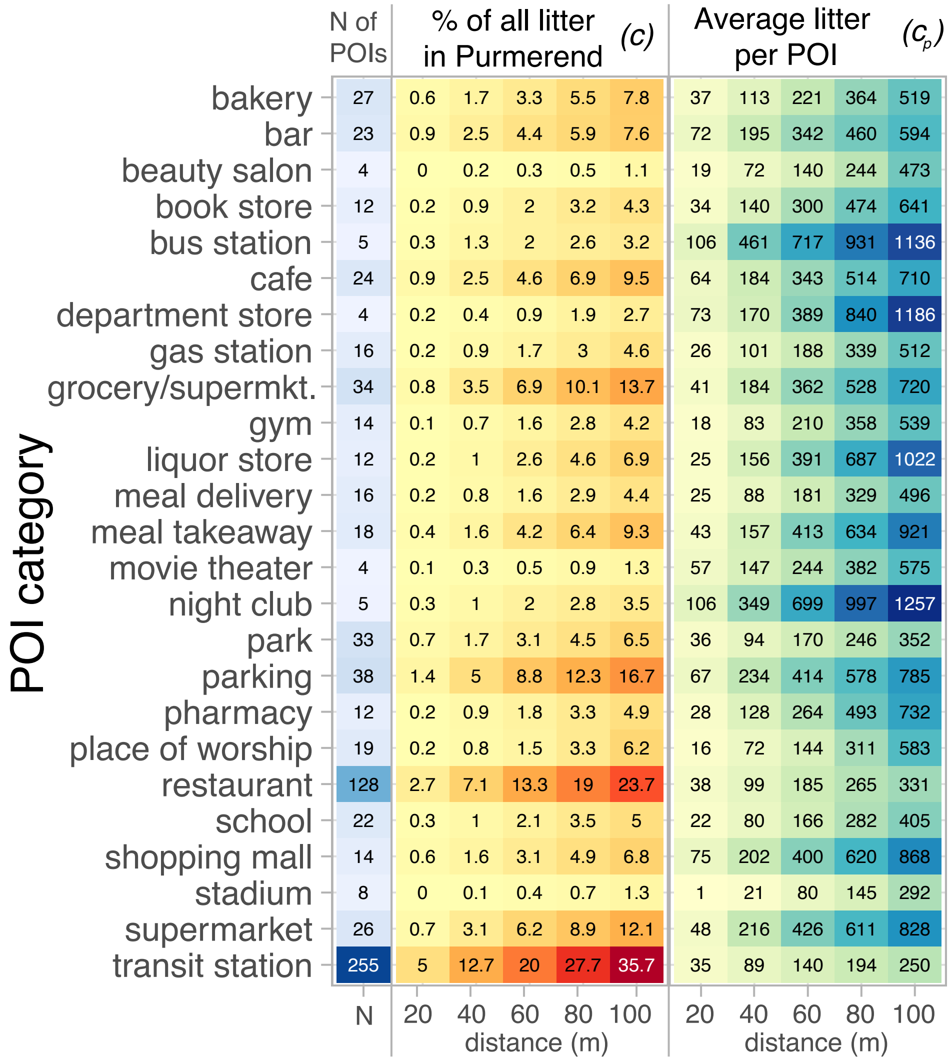

Figure 5. Co-occurrence of litter observations in Purmerend within a 100-meter radius from POIs for 25 Google POI types, including the number of POIs. For example, 13.3% of all litter in Purmerend is located within 60 metres from its 128 restaurants, corresponding to 185 litter observations per restaurant. Total number of observations: 178,557.

Figure 5 indicates how litter prevalence per POI (c_p) varies with distance: For instance, bars are above average at 20 metres, but closer to average at 100 metres. High-litter POIs provide products that can be opened and consumed in the street (department store, liquor store, and pharmacy). The below-average categories might be due to monitoring and cleaning that might reduce the presence of litter (transit stations), sales of goods that are consumed elsewhere (meal delivery), and consumption occurs primarily within the POI's premises (restaurants)

(QB3). Some litter types are found around specific POIs. Less litter tagged as “drink” than expected is located near transit stations and restaurants ($z < -2$). On the other hand, “plastic” litter is found in larger quantities than expected across the full spectrum of POIs, with stronger associations around liquor stores, night clubs, and pharmacies ($z > 2$). It is important to note that these co-occurrences are purely based on proximity and do not imply causation (*QB3*).

Clustering of litter to Google POIs. The previous analysis at fixed distances provides some insights, but it cannot estimate the statistical significance of the observed co-occurrences. Hence, we apply point pattern analysis to quantify the spatial association between litter and POIs by calculating a cross L-function for each *L(litter type, POI type)* pair. The clustering patterns between all litter observations and POIs grouped by type appear rather similar, with some types clustered more than average (bus station, cafe, night club, bar), some below average (stadium, school, park, beauty salon), and all other types are in between. The first column in Figure 6 summarises the clustering by POI type as average cross L, ranging from 58 to 220 (*QB3*).

To understand which POIs deserve particular attention as litter attractors, it is useful to observe the association of specific POI types with litter types. Using the average cross L-function between 0 and 100 meters, the matrix in Figure 6 provides an overview of the spatial relation between litter and POIs, both overall and in specific pairs. The “all POIs” and “all litter” values confirm that litter types tend to be rather homogenous, while the variation between POI categories is markedly higher (*QB3*). Surprisingly, litter observations are present between all POI categories and litter tags, confirming the high granularity and coverage of the Purmerend data. It is important to note that even cases with a negative z are not spatially dispersed, having L-function well above the CSR, with statistical significance at alpha level .05. The z scores allow the comparison between relative levels of clustering.

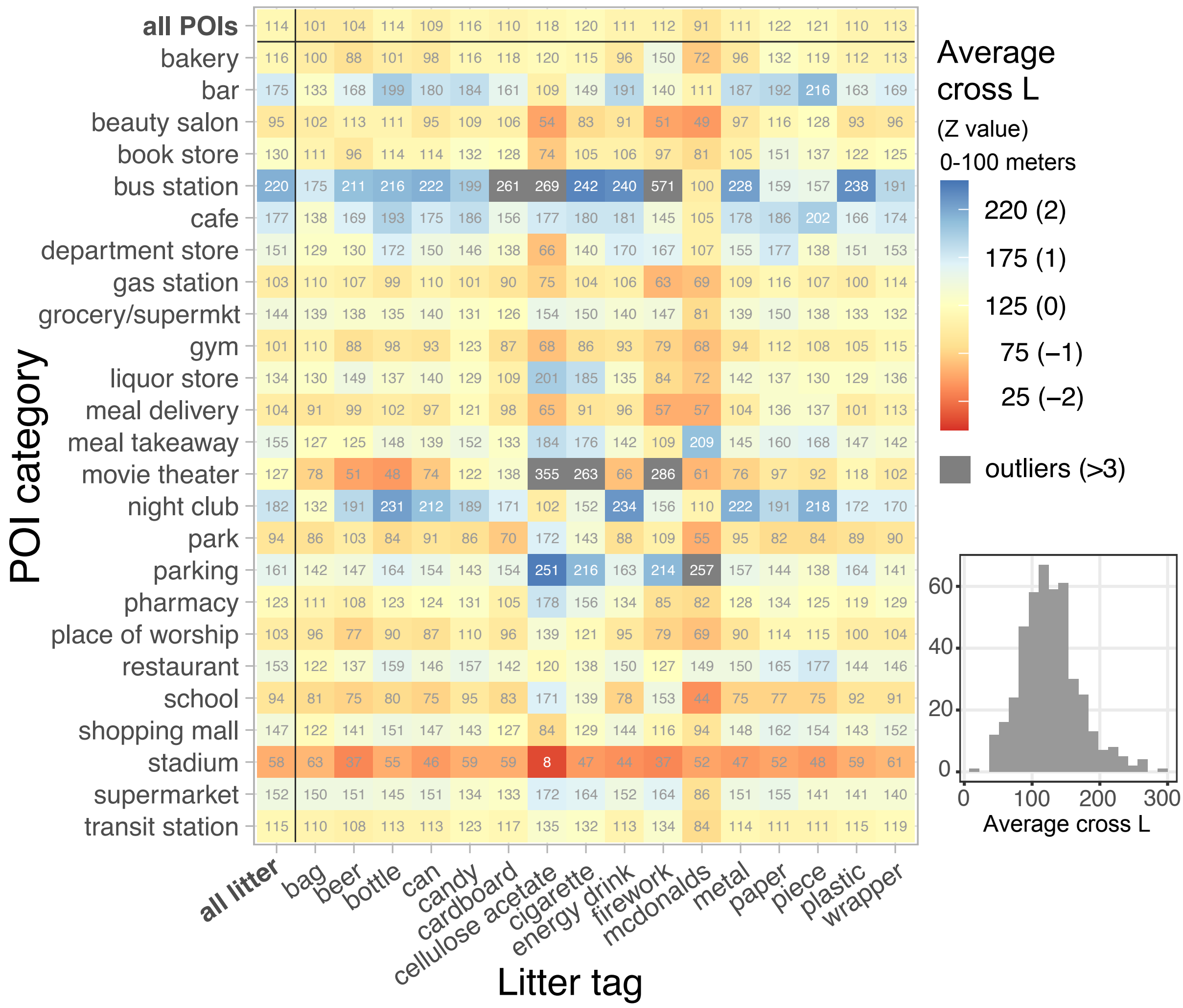

Figure 6. Average cross L-function (with *z* values) in range 0 to 100 meters between the most frequent 16 litter tags and 25 Google POI categories in Purmerend. Blue (red) values indicated higher (lower) clustering than average. The histogram shows the distribution of values omitting the outlier greater than 300. Gray cells are outliers with very high values ($z > 3$). For example, litter tagged as "cigarette" tends to cluster with bus stations, movie theatres, and parking, and little with gyms and beauty salons. Note that even the lowest values in this table are spatially clustered, well above the CSR. All results are significant at alpha level .05. Total number of observations: 178,557.

As the histogram in Figure 6 suggests, the distribution of the average cross L has a clear average at 125 with pairs below and above average, slightly skewed to the right. Outliers are present in the right tail of the distribution ($z > 3$) and include bus stations (highly clustered with cardboard, cellulose acetate, and firework litter), and movie theater (cellulose acetate, cigarette, and fireworks), as well as parking (McDonald's). High associations are observed around bars (piece of paper), bus stations (cigarettes, energy drinks, plastic bottles), night club (energy drinks, metal cans), and parking (cellulose acetate, cigarette, firework) ($z > 2$). The relative absence of litter can also be revealing ($z < -1$). McDonald's litter is highly clustered to a minority of POIs; beer bottles and cans tend not to be near bakeries nor movie theatres; cigarette occurs near parking lots and cafes; fireworks are present near parking lots, bus stations, movie theatres and, luckily, not in gas stations. At the opposite end of the distribution, extremely low L values are found for "stadium" POIs, which are sports facilities located in relatively peripheral areas, outside of the city centre (*QB3*).

DISCUSSION

This investigation on crowdsourced litter data has generated several findings, both concerning the biases of the Litterati data at the national level (*QA*) and on the local case study, in which high-quality data can provide insights into urban litter (*QB*), which can be summarized as follows. (*QA1*) The litter data has national coverage across all provinces, with higher density in some central areas of the country. Six areas are hot spots of litter observations at the national level, including Amsterdam, Purmerend, Nieuw-Vennep, Arnhem, Duiven, and Zevennar. Summer months have higher collection than winter months, with a peak at the World Cleanup Day. The most frequent tags are plastic (46%), cans (12%), wrapper (10%), and paper (9%). In terms of waste volume, the top types are glass (~5,500 kg), plastic bottles (~3,000 kg), metal cans (~2,500 kg), and plastic wraps (~1,300 kg). This corresponds to 1% of municipal waste officially collected in the Netherlands. (*QA2*) Considering litter observations per 1,000 residents, two Dutch provinces are largely over-represented (Gelderland and Noord-Holland), and four are under-represented (Limburg, Overijssel, Drenthe, and Groningen). At the municipality level, with the exception of Breda, all major Dutch cities are average or slightly under-represented. (*QA3*) The litter data does not exhibit geo-demographic biases. When considering age, income, and ethnicity at the municipality level, no correlation was found with litter observation density.

(*QB1*) In the Purmerend case study, the data covers the entire city, showing that litter is concentrated in a North-West to South-East corridor. Hotspots are located in the historic centre and other agglomerates of urban POIs, suggesting a strong relationship between urban flows and litter. Less litter is recorded in winter months, with the exception of fireworks. (*QB2*) Urban litter is dominated by plastic wrappers, cans, cigarette butts, and plastic bags. Some brands top the unenviable list of most present in litter (McDonald's, Anta-Flu, Red Bull, Heineken, Coca-Cola, Marlboro, Amstel, and Candyman Mac Bubble). (*QB3*) All litter types are clustered in a statistically significant way, with fireworks and cigarettes being more clustered than average and bags below average. 78% of litter is within 100 meters from POIs, and 98% are within 200 meters. The clustering between POIs varies more than the clustering between litter types, which appear more spatially homogeneous. Bus stations, cafes, bars, night clubs, parking lots, attract more litter than average; stadia (sports facilities), beauty salons, schools, parks attract the least litter, while all other POI types are in between.

Our national study of Litterati data indicates a strong spatial variability but without clear biases towards urban areas or to specific geo-demographic groups (*QA*). While a degree of heterogeneity is intrinsic to crowdsourced data, it is possible to mitigate it with explicit or implicit methods. Using our findings, project owners could explicitly nudge volunteers to map under-represented areas (e.g., the eastern Dutch provinces) and under-represented POIs and litter types. Implicit methods can deploy statistical methods such as bootstrapping to select more representative samples from the collected data (Buil-Gil et al. 2020). Efforts are needed to further improve the spatial, temporal and thematic quality of this data, estimating the distribution of litter at an increasingly small temporal and spatial scale. Crowdsourced data could more easily be integrated into monitoring infrastructure, and could support the creation of predictive models of urban litter.

Beyond litter, crowdsourcing tools, such as crowd mapping, are developing as a policy-tool that can increase the speed and ease of participation of citizens to the policy cycle (Taeihagh 2017). These data are also key in addressing challenging urban sustainability issues, such as energy consumption, transport efficiency, quality and availability of material resources, access to services, such as water and sanitation, and the availability and distribution of environmental services (Certomà et al. 2015; Afzalan and Muller 2018). Increased knowledge of such endpoints of urban sustainability could contribute both to the monitoring of sustainability goals (Fraisl et al. 2020; Lozano-Díaz and Fernández-Prados 2020), but also could provide a growing source of bottom-up data for industrial ecology analyses, such as material flow analysis. The system of assessment that we propose in this contribution can support also the study of other sustainability endpoints that can be crowdsourced, and for which traditional sampling strategies can be too costly.

Involving citizens in the process of analyzing and planning for sustainability has been expanding as part of ongoing initiatives of urban experimentation (Fuenfschilling et al. 2019). Urban, participatory, and people-centered approaches (Stevens and D'Hondt 2010) to which crowd-mapping efforts belong are a distributed-approach to addressing sustainability issues, which increase public engagement in planning for sustainability (Afzalan and Muller 2018), and which can empower collective action through everyday grassroots citizen science across blocks, neighborhoods, cities, and nations (Hasbrouck et al. 2007). Such approaches will have to be further integrated into the industrial ecology research agenda.

The case study we assessed (Purmerend) offers a first empirical quantification of urban litter (*QB*), which appear aligned with national trends in the Netherlands in terms of types of litter being collected. However, it might prove hard to generalise these findings to other contexts, without further field work. The geo-demography of the Netherlands does not show significant influence on the data, but this might not be the case at smaller scales. Moreover, the geographic space is not homogenous, as litter and POIs do not have the same likelihood of being located anywhere. Human mobility is constrained by infrastructure and obstacles, such as the canals commonly found in Dutch cities, and network distances would provide a more accurate model of the distances between litter observations and other geo-located objects.

The findings from crowdsourced data align with those based on traditional data, reinforcing the value of this data: smoking and food preparation and consumption are been seen as main litter sources in other countries (Hidalgo-Ruz et al. 2018; Kiessling et al. 2019; Zettler et al. 2017). Our study suggests that cigarettes are one of the most clustered litter types to many POIs, and some POIs deserve more attention than others for litter management (restaurants, cafes, takeaways). The qualitative experiences of litter collectors should inform litter research, and not only their data. Notably, Dirk Groot pointed out that unmanaged litter that remains permanently in the environment is located in peripheral, low-density areas and not in urban cores, where it tends to be collected by local services.[6] Research efforts in this direction might also help collection services design their routes more efficiently.

In this article, we analyzed crowdsourced data about urban litter to understand its characteristics and possible biases.Combining questions and methods from industrial ecology and geographic data science, this inter-disciplinary work can be extended to several promising directions, supporting policy-making, and enhancing our understanding of litter-related behaviour. The application of citizen science might have great potential in mapping

---

[6] From Zoom interview with Dirk de Groot (http://zwerfinator.nl), 4 November 2020.

mismanaged waste in the Global South, in which the impact of inadequate disposal is likely to become more severe in the near future (Lebreton and Andrady 2019).

The societal benefits of citizen participation in litter mapping cannot be understated and includes unmanaged and illegal waste dynamics (Geeraerts et al. 2017). Further work on the relationship between POIs and litter can help policy-makers and planners place more bins near litter-intensive POIs, or intensify the cleaning services, together with targeted campaigns. Specific policy actions can reduce the production of specific litter types can also play a role. For example, in January 2016, the Netherlands implemented a comprehensive ban on disposable plastic bags handed for free at shop counters (Ministerie van Infrastructuur en Milieu 2015; Rijksoverheid 2015; SAMR 2017). In the Litterati data, the percentage of observed plastic bags litter has slightly decreased by 4% yearly since 2016.

Comparative studies across urban systems in different contexts would be beneficial to identify generalizable trends and behaviours of humans and unmanaged litter. In this sense, crowdsourced data sources should be validated with ground-truth data, such as ad hoc multi-strata samples of litter observations, granular data from waste collection services, or information about urban behaviour, including transport flows and retail footfall.

ACKNOWLEDGEMENTS

The authors would like to thank Litterati.org for kindly providing access to their data. The raw data can be obtained upon request from Litterati.org and used under the CC-BY-SA 4.0 License. The authors wish to thank Dirk Groot (zwerfinator.nl) for his valuable insights into the interpretation of litter data and his great support in understanding how the volunteer communities collecting litter in the Netherlands operate. The authors also thank Google for making its points of interest data available through an API.

FUNDING INFORMATION

Stefano Cucurachi and Zhijie Li received support from the Packaging and Environment Foundation (SVM) and the Leiden University Fund (LUF).

REFERENCES

Afzalan, N. and B. Muller. 2018. Online Participatory Technologies: Opportunities and Challenges for Enriching Participatory Planning. *Journal of the American Planning Association* 84(2): 162–177.

Apparicio, P., M. Abdelmajid, M. Riva, and R. Shearmur. 2008. Comparing alternative approaches to measuring the geographical accessibility of urban health services: Distance types and aggregation-error issues. *International Journal of Health Geographics* 7(1): 7.

Atlas Leefomgeving. 2020. Zwerfinator Dirk Groot zet zwerfafval op de kaart.

Baddeley, A., E. Rubak, and R. Turner. 2015. *Spatial point patterns: methodology and applications with R*. CRC press.

Ballatore, A. and J. Jokar Arsanjani. 2019. Placing Wikimapia: an exploratory analysis. *International Journal of Geographical Information Science* 33(8): 1633–1650.

Ballatore, A. and S. De Sabbata. 2020. Los Angeles as a digital place: The geographies of user-generated content. *Transactions in GIS* 24(4): 880–902.

Bergmann, M., L. Gutow, and M. Klages. 2015. *Marine Anthropogenic Litter*. Ed. by Melanie Bergmann, Lars Gutow, and Michael Klages. *Environmental Science and Technology*. Vol. 53. Cham: Springer International Publishing.

Brennan, E., C. Wilcox, and B.D. Hardesty. 2018. Connecting flux, deposition and resuspension in coastal debris surveys. *Science of the Total Environment* 644: 1019–

1026.

Brouwer, M.T., E.U.T. van Velzen, A. Augustinus, H. Soethoudt, S. De Meester, and K. Ragaert. 2018. Predictive model for the Dutch post-consumer plastic packaging recycling system and implications for the circular economy. *Waste Management* 71: 62–85.

Buil-Gil, D., R. Solymosi, and A. Moretti. 2020. Nonparametric Bootstrap and Small Area Estimation to Mitigate Bias in Crowdsourced Data. In *Big Data Meets Survey Science*, 487–517. Wiley, October.

Carbery, M., W. O'Connor, and T. Palanisami. 2018. Trophic transfer of microplastics and mixed contaminants in the marine food web and implications for human health. *Environment International* 115: 400–409.

CBS. 2019. characterstics per zipcode. *Centraal Bureau Voor de Statistiek*.

Certomà, C., F. Corsini, and F. Rizzi. 2015. Crowdsourcing urban sustainability. Data, people and technologies in participatory governance. *Futures* 74: 93–106. http://www.sciencedirect.com/science/article/pii/S0016328714001918.

Cialdini, R.B. 2003. Crafting Normative Messages to Protect the Environment. *Current Directions in Psychological Science* 12(4): 105–109.

Cooper, C.B., J. Dickinson, T. Phillips, and R. Bonney. 2007. Citizen science as a tool for conservation in residential ecosystems. *Ecology and Society* 12(2).

Dickinson, J.L., B. Zuckerberg, and D.N. Bonter. 2010. Citizen science as an ecological research tool: challenges and benefits. *Annual Review of Ecology, Evolution, and Systematics* 41: 149–172.

Emmerik, T. van, T.-C.C. Kieu-Le, M. Loozen, K. van Oeveren, E. Strady, X.-T.T. Bui, M.

Egger, et al. 2018. A methodology to characterize riverine macroplastic emission into the ocean. *Frontiers in Marine Science* 5(OCT): 1–11.

Emmerik, T. Van, M. Loozen, K. Van Oeveren, F. Buschman, and G. Prinsen. 2019. Riverine plastic emission from Jakarta into the ocean. *Environmental Research Letters* 14(8): 84033.

Eurostat. 2017. *Guidance on municipal waste data collection*. https://ec.europa.eu/eurostat/documents/342366/351811/Municipal+Waste+guidance.

Fraisl, D., J. Campbell, L. See, U. Wehn, J. Wardlaw, M. Gold, I. Moorthy, et al. 2020. Mapping citizen science contributions to the UN sustainable development goals. *Sustainability Science* 15(6): 1735–1751.

Fritz, S., L. See, T. Carlson, M. Haklay, J.L. Oliver, D. Fraisl, R. Mondardini, et al. 2019. Citizen science and the United Nations Sustainable Development Goals. *Nature Sustainability* 2(10): 922–930.

Fuenfschilling, L., N. Frantzeskaki, and L. Coenen. 2019. Urban experimentation & sustainability transitions. *European Planning Studies* 27(2): 219–228.

Geeraerts, K., A. Illes, and J.-P. Schweizer. 2017. Illegal shipment of e-waste from the EU. A case study on illegal e-waste export from the EU to China.

Getis, A. and J.K. Ord. 2010. The analysis of spatial association by use of distance statistics. In *Perspectives on Spatial Data Analysis*, 127–145. Springer.

Geyer, R., J.R. Jambeck, and K.L. Law. 2017. Production, use, and fate of all plastics ever made. *Science Advances* 3(7): e1700782.

Grimmer, M., A.P. Kilburn, and M.P. Miles. 2016. The effect of purchase situation on realized pro-environmental consumer behavior. *Journal of Business Research* 69(5):

1582–1586.

Haarr, M.L., M. Pantalos, M.K. Hartviksen, and M. Gressetvold. 2020. Citizen science data indicate a reduction in beach litter in the Lofoten archipelago in the Norwegian Sea. *Marine Pollution Bulletin* 153: 111000.

Hardesty, B.D., J. Harari, A. Isobe, L. Lebreton, N. Maximenko, J. Potemra, E. van Sebille, A.D. Vethaak, and C. Wilcox. 2017. Using numerical model simulations to improve the understanding of micro-plastic distribution and pathways in the marine environment. *Frontiers in Marine Science* 4: 30.

Hartley, B.L., S. Pahl, J. Veiga, T. Vlachogianni, L. Vasconcelos, T. Maes, T. Doyle, et al. 2018. Exploring public views on marine litter in Europe: Perceived causes, consequences and pathways to change. *Marine Pollution Bulletin* 133: 945–955.

Hasbrouck, J., T. Igoe, J. Mankoff, and A. Woodruff. 2007. Ubiquitous Sustainability: Technologies for Green Values. In *In Proc. Ubicomp'07 Workshops. 567*.

Hidalgo-Ruz, V., D. Honorato-Zimmer, M. Gatta-Rosemary, P. Nuñez, I.A. Hinojosa, and M. Thiel. 2018. Spatio-temporal variation of anthropogenic marine debris on Chilean beaches. *Marine Pollution Bulletin* 126(November 2017): 516–524.

Hoornweg, D. and P. Bhada-Tata. 2012. *What a waste: a global review of solid waste management*. World Bank, Washington, DC.

Jambeck, J.R.J., R. Geyer, C. Wilcox, T.R.T. Siegler, M. Perryman, A. Andrady, R. Narayan, and K.K.L. Law. 2015. Plastic waste inputs from land into the ocean. *Science* 347(6223): 768–771.

Kadaster. 2017. Gemeentegrenzen 2017.

Keizer, K., S. Lindenberg, and L. Steg. 2008. The Spreading of Disorder. *Science* 322(5908):

1681–1685.

Kiessling, T., K. Knickmeier, K. Kruse, D. Brennecke, A. Nauendorf, and M. Thiel. 2019. Plastic Pirates sample litter at rivers in Germany--Riverside litter and litter sources estimated by schoolchildren. *Environmental Pollution* 245: 545–557.

Kiessling, T., S. Salas, K. Mutafoglu, and M. Thiel. 2017. Who cares about dirty beaches? Evaluating environmental awareness and action on coastal litter in Chile. *Ocean & Coastal Management* 137: 82–95.

KPLUSV. 2015. *Kosten en omvang zwerfafval*. https://kenniswijzerzwerfafval.nl/document/kosten-en-omvang-zwerfafval.

Krider, R.E. and D.S. Putler. 2013. Which Birds of a Feather Flock Together? Clustering and Avoidance Patterns of Similar Retail Outlets. *Geographical Analysis* 45(2): 123–149.

Lebreton, L. and A. Andrady. 2019. Future scenarios of global plastic waste generation and disposal. *Nature Communications* 5(1): 1–11.

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

Litterati. 2020. Litterati.

Lockwood, B., B.R. Wyant, and H.E. Grunwald. 2020. Locating Litter: An Exploratory Multilevel Analysis of the Spatial Patterns of Litter in Philadelphia. *Environment and Behavior*: 0013916520906834.

Lozano-Díaz, A. and J.S. Fernández-Prados. 2020. Educating Digital Citizens: An Opportunity to Critical and Activist Perspective of Sustainable Development Goals. *Sustainability* 12(18): 7260.

Lynch, S. 2018. OpenLitterMap. com--Open Data on Plastic Pollution with Blockchain

Rewards (Littercoin). *Open Geospatial Data, Software and Standards* 3(1): 1–10.

Maria, F. Di, C. Micale, A. Sordi, G. Cirulli, and M. Marionni. 2013. Urban Mining: Quality and quantity of recyclable and recoverable material mechanically and physically extractable from residual waste. *Waste Management* 33(12): 2594–2599.

Ministerie van Infrastructuur en Milieu. 2015. *Voorbereid op verbod gratis plastic tas*.

Quattrone, G., A. Mashhadi, and L. Capra. 2014. Mind the map: the impact of culture and economic affluence on crowd-mapping behaviours. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 934–944.

Rijksoverheid. 2015. *Verbod op gratis plastic tassen*. https://www.rijksoverheid.nl/onderwerpen/afval/verbod-op-gratis-plastic-tassen.

SAMR. 2017. *Rapport – impact verbod gratis plastic tassen onder retailers | april 2017*. https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/rapporten/2017/04/18/rapportage-verbod-plastic-tassen-retailers/Rapportage+verbod+plastic+tassen+retailers.pdf.

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

See, L., P. Mooney, G. Foody, L. Bastin, A. Comber, J. Estima, S. Fritz, et al. 2016. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information* 5(5).

Senaratne, H., A. Mobasheri, A.L. Ali, C. Capineri, and M. Haklay. 2017. A review of volunteered geographic information quality assessment methods. *International Journal* *of Geographical Information Science* 31(1): 139–167.

Stevens, M. and E. D'Hondt. 2010. Crowdsourcing of pollution data using smartphones. In *Workshop on Ubiquitous Crowdsourcing*, 1–4.

Taeihagh, A. 2017. Crowdsourcing: a new tool for policy-making? *Policy Sciences* 50(4): 629–647.

Tisserant, A., S. Pauliuk, S. Merciai, J. Schmidt, J. Fry, R. Wood, and A. Tukker. 2017. Solid waste and the circular economy: a global analysis of waste treatment and waste footprints. *Journal of Industrial Ecology* 21(3): 628–640.

Zettler, E.R., H. Takada, B. Monteleone, N. Mallos, M. Eriksen, and L.A. Amaral-Zettler. 2017. Incorporating citizen science to study plastics in the environment. *Analytical Methods* 9(9): 1392–1403.