

From Point Events to Continuous Surfaces: A Framework for Multi-scale Spatial Analysis

Sachit Desa(✉), Jonathan J. Huck, Olatunji Johnson, Matthew Dennis, Matt D. Tomkins, and Angela Harris

The University of Manchester, Oxford Road, Manchester M13 9PL, UK
{sachit.desa,jonathan.huck,olatunji.johnson,matthew.dennis,
matthew.tomkins,angela.harris}@manchester.ac.uk

Abstract. Existing approaches for analysing point-based data either model stochastic intensity or operate on fixed representations, limiting the ability to explicitly examine how relationships vary across space. This study presents a methodological framework that transforms point-event data into a continuous, mass-preserving spatial surface, enabling flexible, user-defined multi-scale analysis of spatial processes. The framework is demonstrated using unmanaged waste sites across three sub-Saharan cities. Results highlight distinct scale-dependent mechanisms across cities: in Lagos, population density structures city-scale patterns while accessibility shapes local accumulation; in Freetown, accessibility dominates at the city scale but local patterns reflect economic gradients; and in Nairobi, urban morphology governs city-scale patterns while informality drives local variation. By linking discrete events to continuous representations, the framework provides a flexible and generalisable approach for analysing point-based phenomena in systems where understanding spatial heterogeneity is critical for interpretation.

Keywords: Point processes · Exposure surface modelling · Random Forest · Urban analytics · Log-Gaussian Cox Process

1 Introduction

Many spatial phenomena are observed as discrete events or point locations, yet are shaped by continuous processes operating across space [21]. Such events are common across ecology, epidemiology, urban systems, and climate-related hazards, often reflecting the interaction of overlapping spatial dynamics [2]. Two broad classes of methods commonly used to analyse these processes are: local spatial regression approaches, such as geographically weighted regression (GWR) [4,6,8], and spatial point-process models (PPM), which provide a probabilistic framework for analysing event occurrence [2,16].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16765, pp. 624–642, 2027.
https://doi.org/10.1007/978-3-032-30536-7_40

Local spatial regression methods, including GWR and its extensions, are widely used to investigate spatially varying predictor–outcome relationships. [6,8]. These models assume that a sufficiently coherent response surface exists over which local patterns can be meaningfully specified. Whilst this assumption may hold for some phenomena (i.e. disease spread, species presence, agricultural yield), it is problematic for phenomena observed as spatially fragmented point patterns (i.e. crime events, waste site presence). Extensions such as Multiscale GWR (MGWR), and spatial variants of Random Forest, improve flexibility but remain reliant on the assumption that the response can be represented as a sufficiently dense or continuous surface for local inference [6].

Spatial point-process models offer a complementary perspective by modelling event patterns directly. Gibbs processes capture interaction between nearby events, while Cox processes model spatially varying intensity through latent stochastic fields. Log-Gaussian Cox Processes (LGCPs), in particular, are widely used to represent clustered event patterns and account for spatial dependence [2,16]. These models provide a rigorous framework for estimating event intensity and identifying process scales; however, they lack user-defined scale-setting, limiting analysis of heterogeneity across spatial scales. The primary inferential target of these models is stochastic intensity rather than an interpretable response surface for downstream analysis. In complex settings, they can become computationally demanding and are often restricted to additive formulations, limiting their ability to capture highly non-linear and locally heterogeneous relationships [2,7].

In settings where point events are spatially discontinuous or clustered, and drivers vary across space and scale, current methods do not provide a continuous, interpretable surface for explanatory analysis. In these approaches, the scale of inference is typically dictated by implicit bandwidth optimisation or covariance structure, rather than aligned with the spatial support of the phenomenon [6,13]. This limits the incorporation of domain-specific scale knowledge and constrains analysis to a single scale. As a result, researchers lack a flexible framework to construct response surfaces that preserve the original event pattern while supporting analysis of broad trends and local heterogeneity.

This study introduces an analyst-controlled, deterministic, mass-preserving, and scale-explicit event surface that links discrete events to the continuous processes that shape them. Rather than replacing point-process models, the framework provides an interpretable response surface through which domain knowledge can be directly incorporated, supporting analysis of spatial heterogeneity while remaining faithful to the original event pattern. As an example of a downstream analysis, we demonstrate the utility of this framework by exploring the drivers of unmanaged waste sites. Random Forest models are applied at a global and local scale to identify city-scale drivers and local variation. To situate these results within established spatial modelling paradigms, we compare the predictors identified by the framework with those derived from a traditional LGCP intensity surface.

2 The Study

Across sub-Saharan Africa (SSA), rapid urbanization has caused waste generation to outpace formal management capacity [19,26]. Consequently, the region now records the lowest formal collection rates and some of the highest volumes of unmanaged waste globally. This has resulted in cascading impacts on public health, local ecosystems, and climate systems [12,24]. Whilst waste sites are typically represented by polygons or point locations, their emergence and persistence is shaped by continuous, and interacting, spatial gradients in population density, accessibility, neighbourhood morphology, and socio-economic conditions. These drivers interact through complex, non-linear dependencies that vary geographically, producing heterogeneous patterns within a single metropolitan area. Analysing such phenomena requires methods capable of linking discrete event data to the continuous urban processes that generate them. Urban waste presents an ideal case for evaluating the framework, as it reflects the interaction of multiple spatial gradients operating across scales. By exploring the drivers of waste at different scales, this framework demonstrates how it can be used on spatial point-event data to generate meaningful insights and uncover (layered) patterns.

2.1 Data Used for Study

The framework was implemented using mapped waste sites across three urban centres: Lagos, Freetown, and Nairobi. These cities were selected for their contrasting built morphologies, infrastructure coverage, and stages of urbanisation, enabling assessment across diverse contexts. Waste sites were represented as spatial point locations with associated footprint extents derived from manual mapping. Five predictor groups were considered for each city, selected based on existing literature and known drivers of uncontrolled waste deposition.

The groups of predictors considered in this study are: (i) demographic measures (population density and U-18 population share), capturing population pressure and household structure associated with waste generation [9]; (ii) economic proxies (relative wealth index and night-time lights), closely linked with goods manufacturing and consumption [15]; (iii) settlement informality (slum probability), associated with poor service access and separation from urban infrastructure [1,3,15]; (iv) accessibility (distance to tertiary & secondary roads and an accessibility index), capturing network constraints on service provision [17,23]; and (v) urban morphology (building footprint and spacing, orientation dispersion, distance to open space, Clark–Evans R, and a built-environment index), reflecting city structure and configuration [15,17]. Predictors were harmonised and interpolated onto a common 30 m grid across each city.

3 Methodology

Figure 1 summarises the proposed framework, in which discrete waste-site locations are transformed into a continuous, mass-preserving surface for downstream

analysis. Random Forest (RF) models are then applied to examine drivers of waste presence, with global models identifying city-scale patterns and local models capturing neighbourhood-level variation. LGCP models provide a complementary comparison by assessing spatial dependence and process scale. The surface is constructed through five steps that preserve relative site importance, ensure bounded influence, and maintain comparability across the urban area.

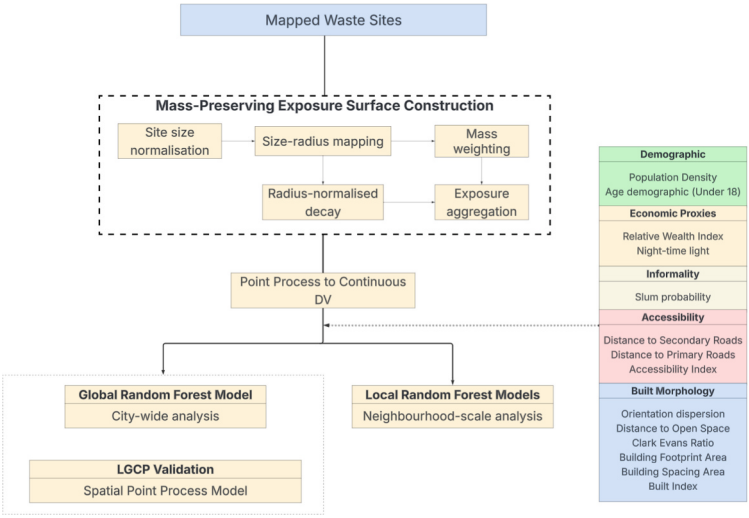


Fig. 1. Framework for generating a continuous surface from point-process data

Site Size Normalisation

Mapped events can vary substantially in their spatial extent or magnitude (e.g., footprint area, intensity). Without adjustment, larger events can disproportionately influence the resulting exposure surface and obscure meaningful spatial variation. To preserve relative differences in event magnitude while limiting the influence of extreme outliers, each event j was assigned a bounded size score $s_j \in [0, 1]$ derived from its footprint area A_j. Site areas were log-transformed to reduce right-skew, then Winsorised by replacing values below and above the 5th and 95th percentiles of the log-area distribution, respectively. Let

$$\ell_j = \ln(A_j), \tag{1}$$

where A_j is the footprint area of event j, and let L_5 and L_{95} denote the 5th and 95th percentiles of the set $\{\ell_j\}$. The Winsorised log-area, $\ell_j^{(w)}$, was defined as

$$\ell_j^{(w)} = \begin{cases} L_5, & \text{if } \ell_j < L_5, \\ \ell_j, & \text{if } L_5 \le \ell_j \le L_{95}, \\ L_{95}, & \text{if } \ell_j > L_{95}. \end{cases} \tag{2}$$

The normalised size score was then computed as

$$s_j = \frac{\ell_j^{(w)} - L_5}{L_{95} - L_5}, \tag{3}$$

which ensures $s_j \in [0, 1]$.

This transformation preserves the relative ordering of site sizes while limiting the influence of extreme values on the resulting event surface.

Radius Mapping Based on Site Size

Spatial phenomena often exhibit multi-scalar influence, with the drivers and impacts of point-represented events extending beyond their immediate location and attenuating with distance, consistent with Tobler's first law [8,21]. The spatial extent of this influence varies with site characteristics and local context. Empirical studies describe impacts ranging from highly localised effects, such as odour, poor air quality, and immediate health risks [11,20], to broader regional-scale influences linked to environmental conditions [25]. Collectively, this literature suggests that contributing factors and impacts operate within a plausible influence zone in the order of 100–800 m [15,22].

Guided by these findings, a sensitivity analysis was conducted to identify plausible bounds of site influence. Candidate radius mappings were evaluated separately for each city using a fixed lower bound of 75 m and upper bounds between 200 m and 400 m. The evaluation assessed whether the resulting exposure surfaces provided sufficient spatial support for downstream analysis while preserving local structure. Candidate settings were compared based on surface coverage, site overlap, and the stability of waste–predictor relationships under regression diagnostics. Results indicated that:

- Influence distances below 75 m produced minimal variation in regression diagnostics, indicating highly localised effects for smaller sites;
- Increasing distances beyond 75 m led to systematic increases in surface coverage and site overlap, consistent with local-level interactions; and
- Beyond 400 m, surfaces became overly smooth and less locally interpretable, with unstable diagnostics and divergence from established waste–predictor relationships.

Each site, j, was assigned an effective radius of influence, r_j, representing the spatial extent that contributes to site presence and size. Radii were mapped from normalised size scores, s_j, as:

$$r_j = R_{\min} + (R_{\max} - R_{\min})\, s_j^{\gamma}, \tag{4}$$

where s_j denotes the scaled size of site j, $R_{\min} = 75\,\text{m}$ (approximately three grid cells), $R_{\max} \in [200, 400]\,\text{m}$, and $\gamma = 1$ specifies a linear mapping (alternative γ values allow non-linear scaling if different size–influence relationships are desired). This imposes a monotonic relationship between site size and spatial influence, such that larger sites affect wider surrounding areas while maintaining bounded, comparable influence across the urban footprint.

Quantity (Mass) Weighting
While the radius controls spatial extent, a separate mass term governs each event's total contribution to the surface. If $\beta = 1$, site contribution scales linearly with site area. However, given substantial variation in site sizes, very large sites, often formal landfills or sanctioned disposal areas, may not relate to the predictors in the same way as smaller, uncontrolled sites. To account for this, a sublinear scaling exponent β was introduced so that larger sites contribute more mass, but with diminishing influence relative to size. The exponent was defined by anchoring the relative contribution of large sites, such that those at the 95th percentile contribute a fixed multiple R of the mass of a median-sized site:

$$\beta = \frac{\ln(R)}{\ln\left(A_{95}/A_{50}\right)}, \tag{5}$$

where $R = 3$ was selected to provide moderate weighting of larger sites without allowing them to dominate the exposure surface. This choice was supported by sensitivity analysis, which indicated that stronger scaling led to reduced interpretability and increased dominance of a small number of large sites. The resulting mass weight for each site was then computed as:

$$q_j = A_j^{\beta}. \tag{6}$$

This formulation preserves a monotonic relationship between site size and contribution while ensuring that the event field reflects a balanced combination of localised and neighbourhood-scale influences. Mass-weighting is flexible, and alternative values of β (including $\beta = 1$ for linear scaling) may be adopted depending on the application context and underlying data characteristics.

Radius-Normalised Distance Decay
Spatial influence is assumed to decay smoothly with distance, reflecting the diminishing relevance of an event's effects with increasing separation [21]. A radius-normalised exponential decay kernel was adopted to ensure continuous attenuation while remaining interpretable and consistent across sites of different sizes. This provides a simple, monotonic, and parameter-light formulation, avoiding strong assumptions about spatial form while maintaining consistency across heterogeneous urban settings. The radius-normalised exponential kernel selected for the study is defined as:

$$w\left(d_{pj}; r_j\right) = \exp\left(-\frac{d_{pj}}{r_j}\right), \tag{7}$$

where $w(\cdot)$ denotes the contribution weight, d_{pj} is the distance from grid point p to site j, and r_j is the radius of influence. This ensures larger sites influence over longer distances, while smaller sites decay more rapidly. To define an effective spatial support, the kernel was truncated using a small threshold (ε), with contributions below this level treated as negligible. This yields finite support, ensuring computational tractability while preserving continuous decay.

Exposure Surface Construction
The final event surface was constructed by aggregating the kernel-weighted contributions, of all sites, at each grid point. For grid location, p, exposure was:

$$\mathrm{DV}(p) = \sum_{j:\, d_{pj} \le \tau_j} q_j \, \exp\left(-\frac{d_{pj}}{r_j}\right), \tag{8}$$

where q_j is the site mass, r_j is the site-specific radius of influence, and τ_j denotes the limit of the spatial support beyond which influence is assumed to be negligible ($\varepsilon = 0.05$).

A hexagonal tessellation was used as the finite-element representation of the exposure surface to support spatial aggregation and local analysis. Hex sizes of 80–100 m were selected to stabilise local exposure estimates while remaining comparable to the lower bound of influence ($R_{\min}$) and well below the upper bound ($R_{\max}$). This ensured that each cell aggregates multiple 30 m grid points, reducing noise while preserving local variation.

Parameter Selection
The surface is constructed through a parameterised aggregation of event-level contributions, allowing adaptation to different phenomena, spatial scales, and modelling objectives. Key components are controlled through user-defined parameters:

1. **Spatial influence bounds** ($R_{\min}, R_{\max}$): define the lower and upper limits of event influence.
2. **Radius–size mapping exponent** (γ): controls how the radius of influence r_j scales with normalised event size s_j.
3. **Mass weighting factor** (R, β): governs the relative contribution of larger versus smaller events; here, β is defined such that upper-percentile events contribute a fixed multiple R of a median event.
4. **Distance decay function** $w(d_{pj}; r_j)$: specifies how influence attenuates with distance, implemented as a radius-normalised exponential decay.
5. **Minimum influence threshold** (ε): defines the cutoff below which contributions are negligible.
6. **Spatial aggregation scheme** (e.g., hexagon size): determines the resolution and structure of analysis units.

These parameters enable application-specific tuning within a consistent framework. For this study, values were selected based on prior literature, sensitivity analysis, and the spatial characteristics of waste. Table 1 summarises the parameters used for each city.

This aggregation captures cumulative influence from single large sites, multiple smaller sites, or overlapping influence zones. The resulting surface is a mathematical representation of waste influence across the city and is used as the dependent variable (DV) in subsequent global and local Random Forest analyses. Figure 2 illustrates the constructed surface for a ward within a local government area (LGA) in Lagos.

Table 1. Parameter settings used for exposure surface construction across cities.

City	$R_{\min}$	$R_{\max}$	Decay rate	β	Hex-grid size
Lagos	75	250	0.8	0.57	100
Freetown	75	300	0.8	0.40	80
Nairobi	75	350	0.8	0.42	80

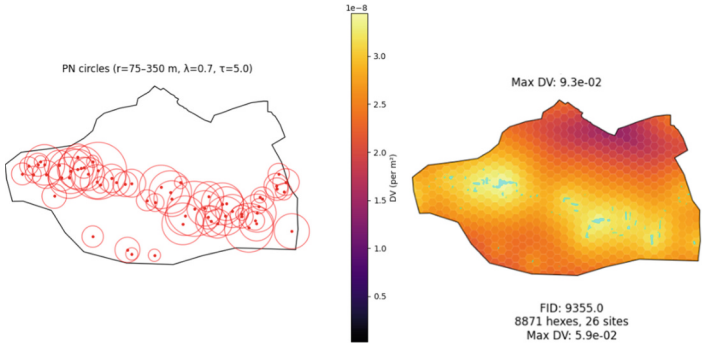


Fig. 2. Surface for mapped waste sites in Sabo-Oniba (ward), Ojo (LGA), Lagos. Red circles show site radii (75–350 m), illustrating bounded influence (left); the corresponding event surface is shown on the right

3.1 Global Model

Following construction of the continuous, non-sparse event surface, relationships between the response variable and spatial predictors can be analysed directly. The parameterised surface enables analysis across scales. Here, we examine both city-wide and local relationships to assess how drivers of waste deposition vary across the urban landscape. A global Random Forest (RF) model is first used to identify predictors explaining variation in waste presence at the city scale, providing a baseline for assessing spatial heterogeneity.

To ensure computational efficiency while preserving the full range of DV values, the global RF was trained on a stratified and spatially balanced sample of approximately 20,000 grid points per city. Sampling followed two principles. First, dependent-variable stratification ensured representation across DV levels by including a subset of zero-exposure cells alongside samples drawn from each decile of the positive DV distribution. Second, spatial balancing was applied to mitigate the effects of spatial autocorrelation. Grid points were partitioned into spatial blocks using k-means clustering in projected (UTM) coordinate space, with these blocks serving as sampling strata and cross-validation folds. This limited localised overfitting and improved the robustness of variable-importance estimates.

Blocked cross-validation was used to assess model performance (R^2, MAE), with Moran's I computed on test residuals as a spatial robustness check. Fol-

lowing validation, RF was refit on the sampled dataset (20,000 points). Permutation importance was then computed from this fitted model, with the mean importance and associated standard deviation (SD) derived from repeated permutations. RF models were implemented using 600 decision trees, a maximum tree depth of 20, and a minimum node split of 5 observations. These parameters were selected to balance three competing objectives: sufficient model flexibility to capture non-linear relationships and interactions; controlled tree depth to limit overfitting in complex spatial settings; and stable estimation of variable importance across cross-validation folds.

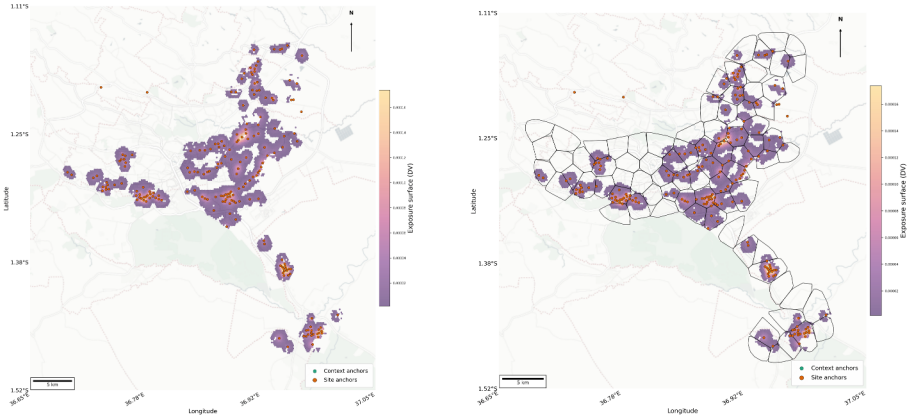
3.2 Localised RF Modelling

To characterise spatial heterogeneity in the drivers of waste, a region-stratified RF framework was implemented in which separate RF models were fitted within spatially contiguous sub-regions of each city. Unlike the global RF, which estimates an average relationship across the entire urban domain, this approach allows predictor importance to vary spatially, thereby capturing non-stationarity in the relationships between waste presence and urban characteristics.

Spatial Partitioning of the Urban Domain
The derived event surface was partitioned into K spatial regions using k-means clustering applied to projected UTM coordinates of hexagon centroids. This produced a set of non-overlapping, spatially contiguous regions that collectively covered the full surface. Following partitioning, regions were ranked by total DV value (exposure mass), defined as the sum of the exposure hexagons within each region. To focus local modelling on areas that either contain waste or that contribute to waste sites, regions with the highest total DV value were retained for subsequent analysis. An elbow and cumulative-share diagnostic was used to assess whether the selected number of regions captured a substantial fraction of total mass, providing a principled basis for the choice of K regions. Figure 3 illustrates the continuous waste surface derived from mapped waste sites and its partitioning into spatial regions used for local Random Forest analysis.

For each region, an independent RF model was fitted using all assigned grid cells. Locality is defined by region membership: each cell belongs to one region, and each region yields a single RF model. Predictors were standardised within regions using region-specific scalers prior to fitting, ensuring numerical stability for permutation importance while allowing adaptation to local predictor distributions. RF parameters were fixed across regions to enable comparison of relative importance patterns. Permutation importance was computed for each model using repeated feature shuffling on the training data. As models are fitted on distinct spatial subsets and independently standardised, importance magnitudes are interpreted within regions rather than as globally comparable effects.

Spatial heterogeneity in predictor relevance was quantified by aggregating permutation importance across regions. For each predictor, the coefficient of variation (CV) of importance scores was computed. Low CV indicates spatially stable influence, while high CV reflects strong spatial differentiation. This



(a) Continuous waste exposure surface. (b) Spatial regions for local RF analysis.

Fig. 3. Waste exposure in Nairobi: (a) continuous exposure surface from mapped sites; (b) k-means partitioning into regions capturing local intensity and extent

regional RF framework provides a structured representation of spatial heterogeneity in waste drivers, complementing the global RF by showing how local importance and stability compare with city-scale influence.

3.3 Log-Gaussian Cox Process for Spatial Comparison

LGCP models were used to characterise the underlying spatial point process governing event locations. LGCP models intensity as:

$$\Lambda(\mathbf{p}) = \exp\left(\mathbf{X}(\mathbf{p})^\top \boldsymbol{\beta} + Z(\mathbf{p})\right),$$

where $\mathbf{X}(\mathbf{p})$ are covariates, $\boldsymbol{\beta}$ are fixed effects, and $Z(\mathbf{p})$ is a Gaussian process capturing residual spatial dependence, parameterised by variance σ^2 and range ϕ. Parameters were estimated using Bayesian Inference in R-INLA [18].

While Random Forest models operate on the DV surface, LGCP models analyse event locations directly, explicitly accounting for spatial dependence. LGCP therefore provides a complementary, process-based perspective. Comparing LGCP covariate effects with RF-derived predictor importance allows assessment of whether identified relationships align with the underlying spatial structure.

4 Results

This section presents results from the global RF, local RF, and LGCP analyses for each city. Global RF results summarise dominant predictors at the city scale, while local RF models characterise spatial heterogeneity in predictor importance.

LGCP results provide an independent, point-process perspective by explicitly modelling spatial dependence, enabling comparison with RF-derived patterns. In the LGCP, coefficient sign indicates the direction of association with waste occurrence (positive: increased likelihood; negative: decreased likelihood), while magnitude reflects the strength of this relationship on the log-intensity scale.

4.1 Global and Local RF Modelling Results for Lagos

Tables 2a to 3 present the Global RF model, Local RF model, and LGCP results for Lagos. The global RF model identifies population density, accessibility, night-time lights (NTL), and relative wealth index (RWI) as the strongest city-scale predictors of waste exposure (Table 2a). These predictors exhibit low variability across cross-validation folds, indicating stable global importance. Results from the LGCP model corroborate these findings, with the same predictors showing statistically robust effects (Table 3), providing independent confirmation that the RF-derived patterns align with the underlying spatial process.

Local RF models provide additional insight, revealing that distance to tertiary and secondary roads exhibits both high local importance and substantial spatial heterogeneity across the city. Population density, despite being the dominant global predictor, displays the highest coefficient of variation locally, indicating marked spatial variability in its local influence. Built morphology predictors, including footprint area and spacing, exhibit moderate global importance but heterogeneous local behaviour (Table 2b).

Table 2. RF predictor importance for Lagos: (a) Global and (b) Local models.

(a) Global Random Forest

Predictor	Mean Imp.	SD	Rank
Population Density	1.249	0.027	1
Accessibility Index	0.482	0.011	2
Night-Time Lights	0.377	0.009	3
Relative Wealth Index	0.224	0.007	4
Footprint (Mean)	0.192	0.005	5
Dist. to Secondary Roads	0.164	0.004	6
Under-18 Population	0.138	0.011	7
Dist. to Tertiary Roads	0.135	0.003	8
Footprint Spacing	0.112	0.004	9
Slum Likelihood	0.056	0.002	10
Built Environment Index	0.033	0.001	11
Dist. to Open Space	0.030	0.001	12
Orientation Dispersion	0.025	0.000	13
Clark–Evans R	0.019	0.001	14

(b) Local Random Forest

Predictor	Mean Imp.	SD	CV
Population Density	0.119	0.455	3.822
Dist. to Open Space	0.052	0.127	2.451
Slum Likelihood	0.052	0.108	2.086
Built Environment Index	0.075	0.139	1.861
Under-18 Population	0.020	0.036	1.773
Dist. to Tertiary Roads	0.425	0.751	1.766
Orientation Dispersion	0.029	0.045	1.526
Footprint Spacing	0.044	0.061	1.404
Footprint (Mean)	0.118	0.166	1.401
Clark–Evans R	0.025	0.033	1.339
Dist. to Secondary Roads	0.459	0.534	1.163
Night-Time Lights	0.530	0.513	0.967
Accessibility Index	0.558	0.507	0.909
Relative Wealth Index	0.559	0.474	0.847

Table 3. LGCP covariate effects for Lagos arranged by absolute coefficient value.

Predictor	Coef.	SE	CI	Sig.
Footprint (Mean)	−1.800	0.198	[−2.19, −1.41]	TRUE
Accessibility Index	−1.015	0.127	[−1.26, −0.77]	TRUE
Under-18 Population	−0.783	0.060	[−0.90, −0.66]	TRUE
Night-Time Lights	−0.729	0.090	[−0.91, −0.55]	TRUE
Relative Wealth Index	0.432	0.100	[0.24, 0.63]	TRUE
Slum Likelihood	0.381	0.063	[0.26, 0.50]	TRUE
Dist. to Tertiary Roads	−0.351	0.080	[−0.51, −0.19]	TRUE
Built Environment Index	−0.331	0.057	[−0.44, −0.22]	TRUE
Footprint Spacing	−0.326	0.048	[−0.42, −0.23]	TRUE
Dist. to Secondary Roads	−0.269	0.163	[−0.59, 0.05]	FALSE
Population Density	0.241	0.041	[0.16, 0.32]	TRUE
Clark–Evans R	0.077	0.026	[0.03, 0.13]	TRUE
Orientation Dispersion	0.024	0.031	[−0.04, 0.08]	FALSE

4.2 Global and Local RF Modelling Results for Freetown

Tables 4a–5 present the Global RF, Local RF, and LGCP results for Freetown. The global RF model identifies accessibility, road connectivity (secondary and tertiary roads), and built morphology as the dominant city-scale predictors of waste exposure. Economic proxies, NTL and RWI, also exhibit strong influence (Table 4a). LGCP results broadly corroborate these findings, with distance to secondary and tertiary roads, NTL, and RWI exhibiting statistically robust effects (Table 5).

Local RF models indicate that road connectivity remains important at a local scale but exhibits spatial heterogeneity across the city. In contrast, NTL and RWI retain relatively stable local importance (Table 4b). Together, the results indicate that waste exposure in Freetown is structured by fragmented urban morphology, with road accessibility and economic gradients shaping the spatial distribution of waste presence.

4.3 Global and Local RF Modelling Results for Nairobi

Tables 6a–7 present the Global RF, Local RF, and LGCP results for Nairobi. The global RF model identifies distance to secondary roads, RWI, accessibility, and BEI as the dominant city-scale predictors of waste presence (Table 6a). LGCP results identify a consistent set of structural drivers, with built morphology (footprint spacing, footprint size, and BEI), alongside road proximity and RWI, emerging as significant covariates (Table 7).

Local RF models indicate substantial heterogeneity, particularly for built morphology predictors, including BEI, footprint size, and spacing. Road connectivity and accessibility remain locally important, whilst RWI exhibits com-

Table 4. RF predictor importance for Freetown: (a) Global and (b) Local models.

(a) Global Random Forest

Predictor	Mean Imp.	SD	Rank
Dist. to Secondary Roads	7.161	0.109	1
Accessibility Index	1.628	0.030	2
Relative Wealth Index	0.416	0.010	3
Footprint Spacing	0.377	0.007	4
Dist. to Tertiary Roads	0.295	0.005	5
Night-Time Lights	0.232	0.005	6
Built Environment Index	0.151	0.009	7
Dist. to Open Space	0.088	0.002	8
Under-18 Population	0.076	0.002	9
Footprint (Mean)	0.071	0.001	10
Slum Likelihood	0.040	0.001	11
Clark–Evans R	0.029	0.001	12
Orientation Dispersion	0.019	0.000	13
Population Density	0.013	0.000	14

(b) Local Random Forest

Predictor	Mean Imp.	SD	CV
Under-18 Population	0.133	0.341	2.558
Population Density	0.000	0.000	2.462
Clark–Evans R	0.061	0.150	2.452
Orientation Dispersion	0.052	0.123	2.390
Built Environment Index	0.110	0.246	2.243
Dist. to Open Space	0.043	0.092	2.127
Footprint Spacing	0.104	0.195	1.877
Dist. to Secondary Roads	0.211	0.384	1.816
Dist. to Tertiary Roads	0.120	0.216	1.806
Footprint (Mean)	0.122	0.201	1.641
Accessibility Index	0.113	0.158	1.392
Night-Time Lights	0.417	0.548	1.313
Slum Likelihood	0.032	0.042	1.302
Relative Wealth Index	0.364	0.441	1.213

Table 5. LGCP covariate effects for Freetown arranged by absolute coefficient value.

Predictor	Coef.	SE	CI	Sig.
Dist. to Sec. Roads	−1.379	0.284	[−1.94, −0.82]	TRUE
Relative Wealth Index	0.713	0.185	[0.35, 1.08]	TRUE
Footprint Spacing	−0.623	0.158	[−0.93, −0.31]	TRUE
Night-Time Lights	−0.592	0.131	[−0.85, −0.34]	TRUE
Dist. to Tert. Roads	0.561	0.217	[0.14, 0.99]	TRUE
Dist. to Open Space	−0.466	0.112	[−0.68, −0.25]	TRUE
Under-18 Population	−0.263	0.080	[−0.42, −0.11]	TRUE
Footprint (Mean)	−0.211	0.079	[−0.37, −0.06]	TRUE
Population Density	0.145	0.137	[−0.12, 0.41]	FALSE
Accessibility Index	0.134	0.268	[−0.39, 0.66]	FALSE
Slum Likelihood	0.080	0.054	[−0.03, 0.19]	FALSE
Orientation Dispersion	−0.077	0.053	[−0.18, 0.03]	FALSE
Clark–Evans R	−0.059	0.085	[−0.23, 0.11]	FALSE
Built Env. Index	0.027	0.106	[−0.18, 0.23]	FALSE

paratively stable influence across local areas (Table 6b). Slum likelihood also increases in local importance relative to the global model. Together, the results indicate that waste exposure in Nairobi is structured by interacting accessibility, economic, and built morphology gradients, with strong spatial variation in local-scale influence.

Table 6. RF predictor importance for Nairobi: (a) Global and (b) Local models.

(a) Global Random Forest			
Predictor	Mean Imp.	SD	Rank
Dist. to Secondary Roads	0.560	0.014	1
Relative Wealth Index	0.459	0.010	2
Accessibility Index	0.329	0.008	3
Built Environment Index	0.320	0.010	4
Footprint Spacing	0.215	0.002	5
Population Density	0.205	0.004	6
Dist. to Tertiary Roads	0.138	0.003	7
Night-Time Lights	0.072	0.002	8
Footprint (Mean)	0.033	0.001	9
Slum Likelihood	0.021	0.000	10
Under-18 Population	0.019	0.000	11
Clark–Evans R	0.017	0.000	12
Orientation Dispersion	0.014	0.000	13
Dist. to Open Space	0.012	0.000	14

(b) Local Random Forest			
Predictor	Mean Imp.	SD	CV
Built Environment Index	0.151	0.378	2.506
Under-18 Population	0.034	0.078	2.284
Population Density	0.142	0.311	2.190
Footprint (Mean)	0.062	0.122	1.974
Slum Likelihood	0.047	0.092	1.947
Orientation Dispersion	0.020	0.038	1.895
Footprint Spacing	0.081	0.129	1.604
Night-Time Lights	0.136	0.213	1.559
Accessibility Index	0.268	0.416	1.552
Dist. to Open Space	0.009	0.013	1.389
Dist. to Tertiary Roads	0.258	0.340	1.319
Dist. to Secondary Roads	0.226	0.293	1.293
Relative Wealth Index	0.316	0.381	1.208
Clark–Evans R	0.021	0.025	1.170

Table 7. LGCP covariate effects for Nairobi arranged by absolute coefficient value.

Predictor	Coef.	SE	CI	Sig.
Footprint Spacing	−1.477	0.306	[−2.08, −0.88]	TRUE
Footprint (Mean)	−0.894	0.462	[−1.80, 0.01]	FALSE
Dist. to Open Space	−0.752	0.123	[−0.99, −0.51]	TRUE
Built Env. Index	0.634	0.103	[0.43, 0.84]	TRUE
Dist. to Sec. Roads	0.475	0.079	[0.32, 0.63]	TRUE
Night-Time Lights	−0.396	0.140	[−0.67, −0.12]	TRUE
Relative Wealth Index	−0.374	0.090	[−0.55, −0.20]	TRUE
Slum Likelihood	−0.287	0.102	[−0.49, −0.09]	TRUE
Under-18 Population	−0.267	0.077	[−0.42, −0.11]	TRUE
Accessibility Index	−0.209	0.142	[−0.49, 0.07]	FALSE
Dist. to Tert. Roads	−0.180	0.113	[−0.40, 0.04]	FALSE
Orientation Dispersion	−0.046	0.075	[−0.19, 0.10]	FALSE
Clark–Evans R	−0.040	0.114	[−0.26, 0.18]	FALSE
Population Density	−0.015	0.062	[−0.14, 0.11]	FALSE

5 Discussion

Results between the global RF and LGCP models display close alignment in the dominant domains associated with waste across cities. In Lagos, both models consistently prioritise economic conditions and accessibility-related factors as primary drivers, highlighting the combined influence of wealth gradients, night-time activity, and access to basic infrastructure. In Freetown, accessibil-

ity emerges as the dominant mechanism, with road connectivity and associated income proxies jointly shaping waste presence. Nairobi exhibits a more balanced configuration, in which built-form characteristics, secondary road connectivity, and economic indicators jointly govern waste presence. Across all cities, this convergence at the domain level demonstrates that the RF-derived exposure surface captures meaningful spatial processes, with LGCP results providing independent confirmation of the same underlying mechanisms. The predictor-level discrepancy in results between Global RF and LGCP may be attributed to both approaches characterizing spatial relationships in different ways. RF variable importance provides a descriptive ranking of predictor relevance based on contribution to predictive accuracy, but does not indicate the direction, magnitude, or functional form of associations. By contrast, LGCP explicitly models the intensity of the point process and estimates directional covariate effects alongside residual spatial dependence through mathematically optimised spatial structures. Furthermore, the LGCP operates on the observed point pattern itself, whereas the RF models are fitted to a user-informed, deterministic surface designed to represent the spatial influence of the studied phenomena. The framework proposed here is not intended to reproduce LGCP results exactly, but rather to provide a complementary perspective against which broader spatial patterns can be compared.

Global RF and LGCP models provide a first-order characterisation of waste exposure patterns, but their aggregated nature obscures locally distinct mechanisms within heterogeneous urban systems. This is particularly relevant in cities with fragmented land use, uneven infrastructure provision, and mixed socio-economic conditions. In such contexts, city-scale drivers associated with waste presence differ from the predictors and interactions operating at a sub-city scale.

Figure 4 illustrates how local RF models extend city-scale insights by revealing shifts in the relative importance of predictor domains within urban sub-regions. In Lagos, local RF results show a marked re-weighting of drivers, with economic and accessibility indicators increasing in importance relative to the global model, while population density becomes substantially less influential. This suggests that city-wide demographic pressure masks more localised constraints linked to infrastructure access and economic conditions. In Freetown, divergence between global and local RF results is more pronounced. While accessibility remains important at both scales, local models indicate a more even balance between urban structure, demographic pressure, and accessibility, alongside a strengthened role for economic proxies. This pattern highlights strong within-city heterogeneity that is not apparent in global summaries. Nairobi exhibits the greatest consistency between global and local RF results, indicating comparatively stable drivers across scales. The primary local-level deviation is an increased importance of informality, suggesting that slum probability acts as a locally intensified modifier rather than a dominant city-wide driver. Figure 5 provides mapped examples linking the Global and Local RF results to spatial context across Lagos, Freetown, and Nairobi.

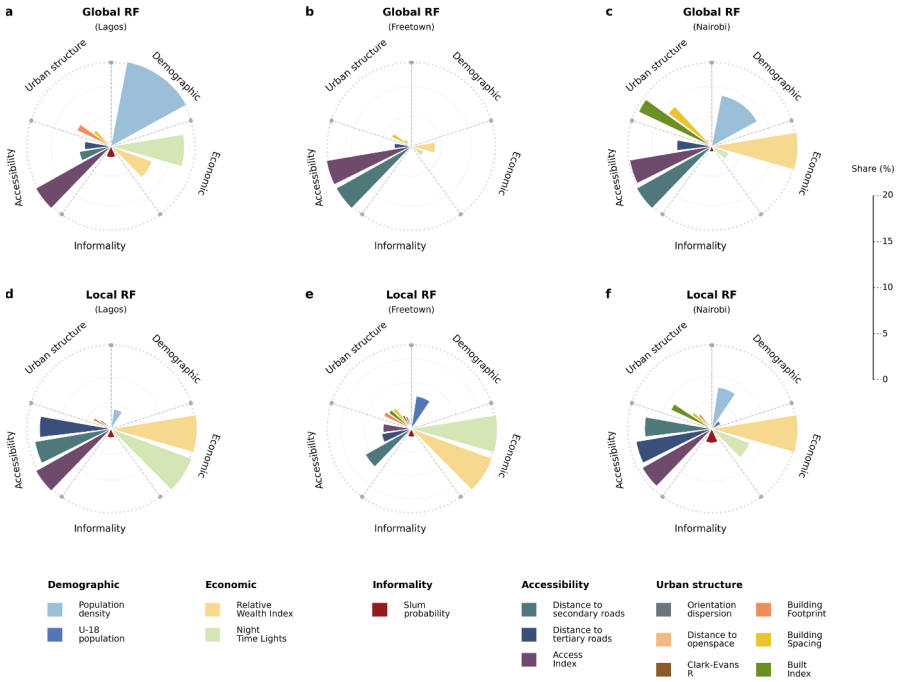


Fig. 4. Figures a-c present the relative predictor share at a city-scale for Lagos, Freetown, and Nairobi, respectively. Figures d-f present the relative predictor importance share at a local-level for the same cities

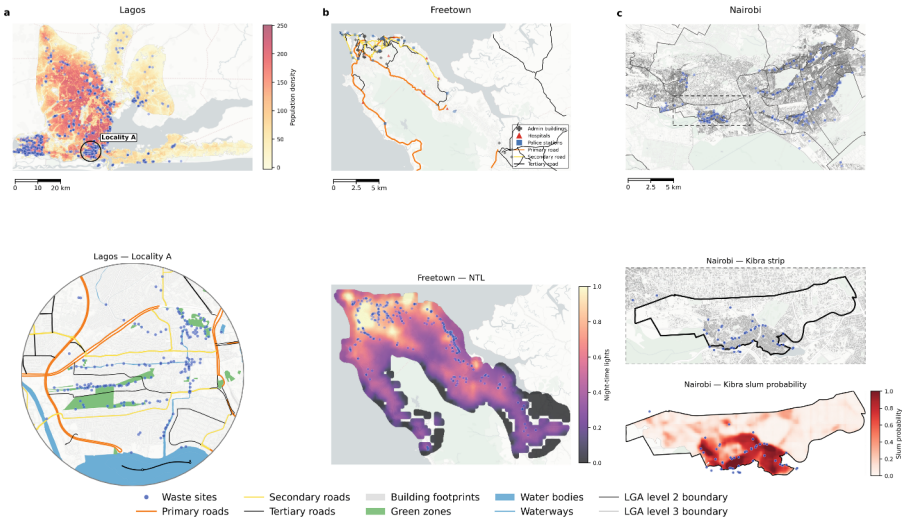


Fig. 5. Examples of the city-scale and local-level drivers of waste that were predicted by Global and Local RF models

Lagos, Freetown, and Nairobi represent contrasting urban configurations: rapid urban expansion, severe infrastructure deficit, and moderate infrastructure development, respectively. In Lagos (column a), rapid population growth and urban expansion (5.7%) [14] have outpaced infrastructure provision, resulting in population density acting as a city-wide envelope within which waste sites emerge. Simultaneously, secondary and tertiary road access concentrates waste within disconnected peripheral areas, consistent with the importance of accessibility in the Local RF model (Fig. 4). In Freetown (column b), which exhibits the highest infrastructure deficit among all cities ($\leq$46% of the population in severe or extreme access constraints) [3], road infrastructure and services are concentrated within the urban core and remain limited in peri-urban areas. Consequently, accessibility governs waste presence at the city scale, while economic inequality shapes waste accumulation across peri-urban and lower-income areas. Nairobi (column c) represents an intermediate system, where demographic, economic, accessibility, and infrastructure pressures jointly structure waste exposure. Although urban structure displays a broad city-scale association with waste, local clustering closely aligns with areas of high slum probability, such as Kibra ward, reinforcing the increased local importance of informality in systems where infrastructure provision has not fully kept pace with urban growth.

As demonstrated in this case study, the drivers of point-based phenomena are inherently scale-dependent, with distinct relationships emerging at different spatial extents. This behaviour is not unique to urban systems. Evidence from ecology and epidemiology shows that processes governing spatial distributions vary across scales, with different drivers becoming dominant under different spatial supports. For example, biotic factors have been shown to dominate pathogen spread at local scales ($\sim 10^2$–$10^3\,\mathrm{km}^2$), while climatic and demographic drivers become significant only at broader regional scales ($> 10^4\,\mathrm{km}^2$) [5]. Similarly, invasive species distributions are often shaped by climate at large scales, but by infrastructure and human activity at local scales [10]. More broadly, spatial epidemiological studies demonstrate how reliance on global trends can obscure heterogeneous transmission dynamics and locally specific drivers. These findings highlight the importance of frameworks that explicitly incorporate spatial variation into analysis. Within this context, the proposed framework allows domain knowledge to guide the generation of a continuous, mass-preserving response surface. The resulting surface can then be interrogated at user-defined resolutions, supporting the identification of predictors and mechanisms operating across spatial scales.

6 Conclusion

This study introduces a framework for analysing point-based phenomena by transforming discrete event locations into a continuous, mass-preserving influence surface. The results demonstrate that the drivers of waste are strongly scale-dependent, with global models capturing dominant patterns while local analyses reveal substantial heterogeneity within urban systems. By enabling analysis

at user-defined spatial resolutions, the framework provides a flexible means of exploring how relationships vary across space while preserving the spatial structure of the original data. Beyond waste, this approach offers a generalisable and computationally practical framework for analysing point-based phenomena in complex systems, where understanding spatial heterogeneity is critical for interpretation and targeted intervention.

References

1. Anbazu, T., et al.: Slum expansion and municipal solid waste management in urban Ghana. Environ. Dev. Sustain. (2022)
2. Baddeley, A., Rubak, E., Turner, R.: Spatial Point Patterns: Methodology and Applications with R. Interdisciplinary Statistics, Chapman and Hall/CRC, Boca Raton (2015). https://doi.org/10.1201/b19708
3. Bettencourt, L.M.A., Marchio, N.: Infrastructure deficits and informal settlements in sub-Saharan Africa. Nature **645**(8080), 399–406 (2025)
4. Brunsdon, C., Fotheringham, A.S., Charlton, M.E.: Geographically weighted regression: a method for exploring spatial nonstationarity. Geogr. Anal. **28**(4), 281–298 (1996). https://doi.org/10.1111/j.1538-4632.1996.tb00936.x
5. Cohen, J.M., et al.: Spatial scale modulates the strength of ecological processes driving disease distributions. Proc. Natl. Acad. Sci. **113**(24), E3359–E3367 (2016)
6. Comber, A., et al.: A route map for successful applications of geographically weighted regression. Geogr. Anal. **55**(1), 155–178 (2023)
7. Diggle, P.J., Moraga, P., Rowlingson, B., Taylor, B.M.: Spatial and spatio-temporal log-gaussian cox processes: extending the geostatistical paradigm. Stat. Sci. **28**(4) (2013). https://doi.org/10.1214/13-sts441
8. Fotheringham, A.S., Brunsdon, C., Charlton, M.: Geographically Weighted Regression: The Analysis of Spatially Varying Relationships. Wiley, Chichester (2002)
9. Kodua, T., et al.: Estimating the willingness to pay for solid waste management services in a Ghanaian municipality. Waste Manag. Res. (2020)
10. Lindestad, O., Ehrlén, J., Neri, M., Hylander, K.: Distribution drivers in a globally invasive plant: climate dominates at large scales, human presence at local scales. Ecosphere **17**(1), e70552 (2026). https://doi.org/10.1002/ecs2.70552
11. Maquart, P.O., Froehlich, Y., Boyer, S.: Plastic pollution and infectious diseases. Lancet Planet. Health **6**(10), e842–e845 (2022)
12. Mebratu, D., Mbandi, A.: Engineering X (Royal Academy of Engineering) and the UN High-Level Climate Champions, UN-HLCC (2022). Technical report Open burning of waste in Africa: Challenges and opportunities
13. Møller, J., Waagepetersen, R.P.: Modern Statistics for Spatial Point Processes. CRC Press, Boca Raton (2003). https://doi.org/10.1201/9780203496398
14. OECD/AfDB/UNOPS-Cities Alliance/UCLG Africa: Africa's Urbanisation Dynamics 2025: Planning for Urban Expansion. West African Studies, OECD Publishing, Paris (2025). https://doi.org/10.1787/2a47845c-en
15. Parrot, L., Sotamenou, J., Dia, B.K.: Municipal solid waste management in Africa: strategies and livelihoods in Yaoundé, Cameroon. Waste Manag. **29**(2), 950–959 (2009). https://doi.org/10.1016/j.wasman.2008.05.005
16. Renner, I.W., et al.: Point process models for presence-only analysis. Methods Ecol. Evol. **6**(4), 366–379 (2015)

17. Rodseth, C., Notten, P., Von Blottnitz, H.: A numerical analysis of the determinants of household solid waste service provision in South Africa. Waste Manag. Res. **38**(10), 1143–1152 (2020). https://doi.org/10.1177/0734242X20902868
18. Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent gaussian models by using integrated nested laplace approximations. J. Roy. Stat. Soc. Ser. B Stat. Methodol. **71**(2), 319–392 (2009). https://doi.org/10.1111/j.1467-9868.2008.00700.x
19. Scheinberg, A., Wilson, D.C., Rodic-Wiersma, L.: Solid Waste Management in the World's Cities: Water and Sanitation in the World's Cities 2010. UN-HABITAT/Earthscan, London; Washington, DC (2010)
20. Singhal, A., Goel, S.: Spatio-temporal analysis of open waste dumping sites using google earth: a case study of Kharagpur City, India. In: Shit, P.K., Adhikary, P.P., Sengupta, D. (eds.) Spatial Modeling and Assessment of Environmental Contaminants. ECS, pp. 137–151. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-63422-3_9
21. Tobler, W.R.: A computer movie simulating urban growth in the detroit region. Econ. Geogr. **46**(sup1), 234–240 (1970)
22. Tomita, A., Cuadros, D.F., Burns, J.K., Tanser, F., Slotow, R.: Exposure to waste sites and their impact on health: a panel and geospatial analysis of nationally representative data from south africa, 2008–2015. Lancet Planet. Health **4**(6), e223–e234 (2020)
23. Velis, C.A., et al.: Socio-economic development drives solid waste management performance in cities: a global analysis using machine learning. Sci. Total Environ. **872**, 161913 (2023)
24. Wang, Y., et al.: Methane emissions from landfills differentially underestimated worldwide. Nat. Sustain. **7**(4), 496–507 (2024). https://doi.org/10.1038/s41893-024-01307-9
25. Wekisa, E., Majale, C.: Spatial distribution of waste collection points and their implications on quality of life in Mombasa County, Kenya. J. Urban Manag. **9**(2), 250–260 (2020). https://doi.org/10.1016/j.jum.2020.02.003
26. Wilson, D.C., Velis, C.A.: Waste management-still a global challenge in the 21st century: an evidence-based call for action. Waste Manag. Res. **33**(12), 1049–1051 (2015). https://doi.org/10.1177/0734242X15616055