www.nature.com/scientificdata

scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A global dataset of microplastics in urban stormwater runoff (2018–2024)

Abdul Mobin Ibna Hafiz[1,2], Ebrahim Ahmadisharaf[1,2] ✉, Maryam Salehi[3], Tim H. M. van Emmerik[4], Tamara Galloway[5], Jeffrey M. Farner[1,2], Jason C. White[6], Naipeng Han[7], Wen Zongguo[7] & Yanxu Zhang[8]

Microplastic (<5 mm) pollution is a growing environmental concern, with urban stormwater systems serving as critical transport pathways. While microplastics predominantly originate and circulate within urban environments, transported via stormwater runoff to receiving water bodies, existing datasets focus almost exclusively on marine and coastal waters. This mismatch between contamination sources and data infrastructure has left urban stormwater systems without dedicated data resources to support research and management. In this research, we present the Dataset of Urban RUnoff Microplastics (DURUM), a global dataset dedicated to microplastics in urban stormwater runoff. The objective of DURUM is to compile, standardize, and serve urban runoff microplastic data in a consistent format to enable data-driven approaches for developing predictive transport models. DURUM systematically compiles 180 sampling efforts from peer-reviewed studies published between 2018 and 2024, encompassing surface runoff from streets, residential catchments, commercial districts, industrial areas, and drainage infrastructure across 15 countries. This paper describes DURUM's structure, compilation methodology, standardization protocols, and future research directions.

Background & Context

Urban stormwater runoff represents a critical but poorly understood transport pathway connecting terrestrial microplastic (MP) sources to aquatic environments[1–3]. Despite the recognition that stormwater runoff carries MP from urban catchment surfaces generating from numerous sources, with substantial fractions originating from tire wear particles, road dust, and atmospheric deposition of textile fibers[2–5], data gaps persist in systematically characterizing which sources dominate, how land use patterns affect MP transport via stormwater, and what environmental factors control their spatiotemporal distribution[6]. The episodic, event-driven nature of urban runoff creates transport dynamics fundamentally different from the relatively stable conditions studied in marine surface waters or depositional sediment environments, yet no dedicated dataset exists to support research on these dynamic urban transport processes.

Understanding MP behavior in urban runoff requires detailed knowledge of sources, transport pathways, and the environmental variables that control their mobilization and fate. MPs in urban systems originate from diverse sources including primary particles intentionally manufactured at small sizes (<5 mm) and secondary particles generated through fragmentation of larger plastic debris via physical, chemical, and biological degradation[7–9]. Tire wear particles, now recognized as a dominant urban MP source, are released directly onto road surfaces through vehicle-pavement friction and subsequently mobilized during rainfall events[5]. Textile fibers enter urban runoff through atmospheric deposition following release during laundering or from outdoor textiles, while plastic fragments accumulate on urban surfaces through littering, building materials degradation,

[1]Department of Civil and Environmental Engineering, FAMU-FSU College of Engineering, Tallahassee, FL, USA. [2]Resilient Infrastructure and Disaster Response Center, FAMU-FSU College of Engineering, Tallahassee, FL, USA. [3]Civil and Environmental Engineering, University of Missouri, Columbia, MO, USA. [4]Hydrology and Environmental Hydraulics Group, Wageningen University, Wageningen, Netherlands. [5]Department of Biosciences, Faculty of Health and Life Sciences, University of Exeter, Exeter, UK. [6]Connecticut Agricultural Experiment Station, New Haven, CT, USA. [7]School of Environment, Tsinghua University, Beijing, China. [8]Department of Earth and Environmental Sciences, Tulane University, New Orleans, LA, USA. ✉e-mail: eahmadisharaf@eng.famu.fsu.edu; eascesharif@gmail.com

and waste mismanagement. Once deposited on urban surfaces, these particles undergo accumulation during dry periods and episodic mobilization during storm events, with concentrations often exhibiting pronounced "first flush" effects where initial runoff carries significantly elevated MP loads[10]. Critical environmental variables governing these processes include rainfall intensity and duration, antecedent dry period length, land use type (e.g., residential, commercial, industrial, or transportation corridors), traffic density, population patterns, and waste management infrastructure quality. However, systematic documentation of these relationships remains limited, with only a few studies reviewing clear source-to-runoff connections or enabling accurate flux estimations from urban environments to receiving waters[2,6,11,12]. Key knowledge gaps include: the relative contribution of different source types across varying urban contexts, how particle characteristics (e.g., size, morphology, and polymer type) vary by source and land use, what environmental thresholds trigger mobilization, how particles age and fragment during surface residence, and what retention mechanisms operate within stormwater infrastructure.

The persistent focus of existing MP datasets on marine and sedimentary environments creates a critical mismatch between available data and the requirements for understanding urban contamination patterns and developing predictive transport models[13]. Major datasets, including the NOAA Global Marine MPs Database[14], the Marine Plastic Database (MPDB)[15], and the DOER MPs Database for Sediments[16], provide valuable resources for studying aquatic systems and long-term depositional records, yet they fundamentally cannot capture the event-scale dynamics, rapid concentration fluctuations, and source variability characteristic of urban runoff. Regional initiatives such as the Southern California Coastal Water Research Project (https://www.sccwrp.org/) and crowd-sourced platforms like the Atlas of Ocean Microplastics (https://aomi.env.go.jp/) focus primarily on receiving water contamination rather than the terrestrial-to-aquatic transport processes that urban runoff represents. River-focused datasets[17] address freshwater systems but typically characterize baseflow conditions or long-term contamination trends rather than storm-driven transport that dominates urban MP loading[5,18]. This gap in data infrastructure limits the ability to quantify urban MP contributions to downstream environments, identify priority sources for intervention, evaluate the effectiveness of stormwater management practices for MP retention, and predict how urban development patterns and climate change will affect future loading patterns. Furthermore, unique challenges of urban runoff sampling, including contamination control during storm events, flow-weighted composite sampling across variable hydrographs, and characterization of particles spanning several orders of magnitude in size, require specialized methodological documentation often absent from datasets designed for marine or sediment sampling contexts[19,20].

Current sampling and analytical methods face fundamental limitations that constrain data availability and comparability across urban runoff studies. Urban runoff MP sampling must address contamination from airborne fibers, equipment fragments, and laboratory environments while capturing temporal concentration gradients during storm events[21]. Flow-weighted composite sampling across entire storm durations, which is essential for accurate load estimation and event mean concentration characterization, is resource-intensive, leading many studies to rely on grab samples that provide only instantaneous snapshots of highly dynamic conditions[19,22,23]. Laboratory identification presents equally formidable challenges, as visual identification methods that work reasonably well for particles $> 100\ \mu m$ become progressively unreliable for smaller size fractions where morphological features diminish and distinction from natural materials fails, requiring transition to time-intensive spectroscopic methods[24,25]. The absence of standardized protocols for sampling design, extraction procedures, analytical techniques, and data reporting prevents direct comparison of findings across studies and hinders aggregation of datasets into larger, statistically powerful compilations. These methodological inconsistencies propagate through the literature, creating datasets with variable size detection limits, inconsistent polymer identification criteria, and incompatible concentration units that must be carefully harmonized before meaningful synthesis becomes possible[26–29].

Given these challenges and the dual imperatives of understanding urban MP contamination patterns and enabling predictive modeling capabilities, we developed the first dedicated, standardized dataset specifically for source-level MP concentration in direct urban stormwater runoff. The Dataset of Urban RUnoff Microplastics (DURUM) serves two complementary research priorities: (1) characterizing exposure and contamination patterns by systematically documenting MP concentrations, size distributions, morphological characteristics, and polymer compositions across diverse urban contexts, land uses, and environmental conditions, thereby advancing fundamental understanding of urban MP pollution severity and variability; and (2) enabling development, calibration, and validation of predictive transport models by providing the quantitative benchmarks, process relationships, and environmental correlations under varying conditions. DURUM distinguishes itself from existing marine and sediment repositories through explicit focus on the dynamic, event-driven transport processes of urban runoff systems, capturing relationships between meteorological forcing (rainfall intensity and antecedent dry periods), source characteristics (land use and traffic patterns), and observed MP loadings that are essential for process-based model development. By systematically compiling data from 180 sampling events across 15 countries-territories and diverse urban contexts, implementing rigorous harmonization of concentration units and particle characterization methods, and documenting critical metadata including sampling protocols, environmental conditions, and analytical detection limits, DURUM provides the empirical foundation necessary for both understanding current urban MP contamination patterns and developing the predictive tools needed to forecast how future urban development, climate change, and management interventions will affect MP loading to aquatic ecosystems.

By shifting research focus from marine accumulation zones to urban mobilization pathways, DURUM enables a fundamental transition in MP management philosophy, from downstream symptom treatment to upstream source control. While beach cleanups and ocean surface sampling remain important for assessing environmental contamination, complementary source-level data from urban catchment surfaces provides the foundation for pollution prevention through targeted stormwater management interventions[30]. In contrast,

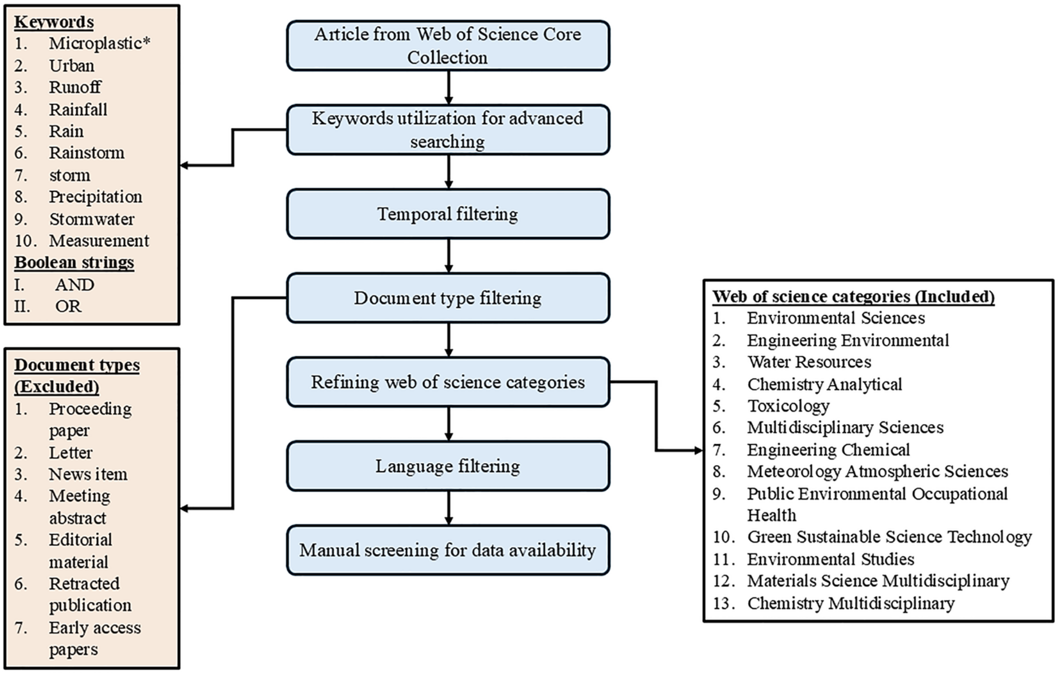

Fig. 1 Schematic diagram of the literature screening and study selection process.

characterizing urban runoff processes illuminates opportunities for stormwater infrastructure modification, treatment technology deployment, and source reduction strategies that intercept MPs before they disperse into aquatic ecosystems[31]. This dataset, thus, addresses a critical gap in the plastic pollution research, where substantial progress in characterizing MP transport through rivers, oceans, and wastewater systems has not been matched by equivalent source-level data from urban catchment surface runoff.

All dataset contents, data harmonization protocols, and analytical code are made freely available through Open Science Framework (OSF) repository to ensure transparency, reproducibility, and broad research accessibility. Statistical analyses were applied to characterize the distribution of MP concentration, assess variability across geographic regions and quantification methods, and evaluate relationships with land use and meteorological variables. This open-access framework enables researchers worldwide to validate our standardization procedures, apply the dataset to diverse research questions ranging from source apportionment to risk assessment, integrate additional datasets as urban monitoring programs expand, and develop novel analytical approaches that advance understanding of MP pollution dynamics in urban environments.

Methods

Dataset compilation and study selection. This dataset compiles empirical data on MP concentrations in urban stormwater runoff from peer-reviewed scientific literature. Studies were identified through Web of Science Core Collection using the search string "Microplastic* AND (urban OR runoff OR rain OR rainfall OR rainstorm OR storm OR stormwater OR precipitation OR measurement)" in the Topic field, covering publications from 2000 to 2024. Studies were included if they reported empirical measurements of MP in urban surface runoff, stormwater drainage systems, or precipitation-driven transport pathways within urban environments. Eligible studies provided quantitative data on MP concentrations, particle characteristics including size distributions, morphological properties, or polymer composition, reported in either main text or supplementary materials. The dataset encompasses measurements from diverse urban settings including residential catchments, commercial districts, industrial areas, transportation corridors, and stormwater infrastructure across all geographic regions and climatic zones. To align with the emergence of MP research as a distinct scientific field, the temporal scope was restricted to studies published between 2000 and 2024. Only peer-reviewed articles published in English were considered, excluding review papers, conference proceedings, editorials, and meeting abstracts.

Studies were excluded if they focused exclusively on marine environments, riverine systems, or wastewater treatment plant effluents without corresponding urban runoff measurements. This scope ensures that DURUM captures source-level MP loading from urban catchment surfaces, prior to any removal or attenuation by stormwater treatment structure. Publications lacking quantitative concentration data or reporting only presence-absence information were not included. Theoretical modeling studies and laboratory experiments without field measurements were similarly excluded, as were studies lacking sufficient methodological details for reliable data extraction. Following our systematic screening (Fig. 1) of titles, abstracts, and full texts, 28 studies spanning 15 countries and territories met the inclusion criteria and were incorporated into the dataset, yielding 180 individual measurement records. Data were extracted directly from published articles and supplementary materials without contacting authors for unpublished data, ensuring reproducibility and accessibility of the dataset compilation process.

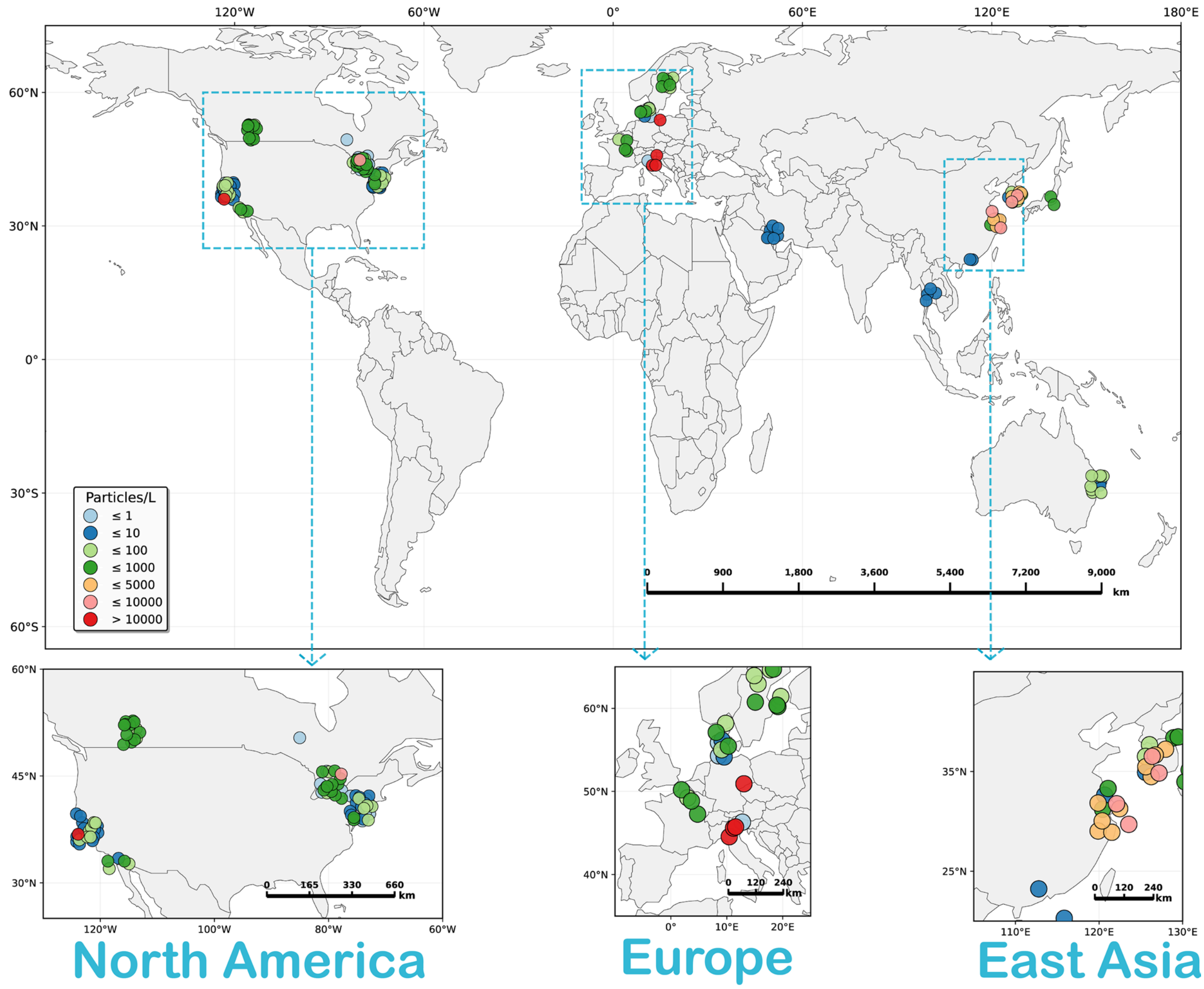

Fig. 2 Global distribution of sampling locations included in the Dataset of Urban RUnoff Microplastics (DURUM).

Dataset scope and compilation. The compiled dataset encompasses MP concentration measurements from urban runoff systems across 15 countries and territories (Fig. 2), including North America (Canada, United States, and Mexico), Europe (Denmark, France, and Italy), Asia (China, Japan, Iran, Hong Kong, and South Korea), and Oceania (Australia). This geographic distribution reflects established research infrastructure in high-income nations and emerging monitoring capacity in upper-middle income regions, while notable under-representation of lower middle- and low-income countries, particularly in Africa, South America, and tropical regions.

Sampling locations were classified according to dominant land use at the watershed scale, as reported in the original studies. Categories included residential areas, commercial districts, industrial zones, transportation corridors, green infrastructure, engineered infrastructure, urban watersheds, and mixed land use areas. Land use classifications were adopted directly from the source publications to maintain consistency with the original designs. Environmental metadata, including antecedent dry days, total rainfall, air temperature, and wind speed, were extracted when reported in the publications, though reporting of these variables was highly inconsistent across the studies.

Data standardization protocols. A fundamental challenge in synthesizing MP data from diverse studies is the substantial heterogeneity in measurement approaches, reporting units, and analytical protocols. To enable quantitative comparisons and support predictive modeling applications, comprehensive standardization protocols were developed and applied across all DURUM variables. These protocols are designed to be replicable, allowing other researchers to apply similar harmonization approaches to MP datasets in different environmental compartments. The standardization process involved systematic review of methodology sections from all source publications, with ambiguous cases resolved through examination of supplementary materials.

Sampling method standardization. Original studies employed diverse terminology to describe sampling approaches, often with inconsistent or overlapping definitions. To enable methodological comparisons, sampling methods were standardized into five primary categories through systematic keyword matching and

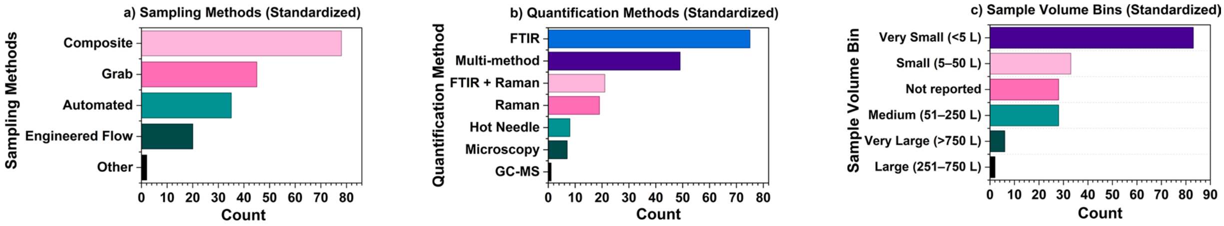

Fig. 3 Distribution of (**a**) standardized sampling methods, (**b**) standardized quantification techniques, and (**c**) sample volume categories across Dataset of Urban RUnoff Microplastics (DURUM; 180 records).

manual verification: composite, grab, automated, engineered flow or other sampling. Composite sampling was identified when studies explicitly mentioned "composite," "integrated," "combined subsamples," or described collection of multiple samples over time that were subsequently mixed or analyzed together. This category captures both time-weighted and flow-weighted composite approaches. Grab sampling was assigned when studies reported "grab," "discrete," "manual collection," or "single point" sampling, representing instantaneous collection at a specific time and location. Automated sampling was identified through keywords including "automated," "automatic," "ISCO sampler," "programmable sampler," or "flow-proportional sampler," indicating use of mechanical sampling equipment with timing or flow-triggered collection. Engineered flow sampling encompassed specialized infrastructure described as "continuous flow," "flow-through system," or "inline sampling," typically associated with permanent monitoring installations. Methods not fitting these categories or combining multiple approaches were classified as "Other."

Quantification method standardization. Analytical identification methods showed even greater terminological descriptions. Quantification methods were standardized based on the primary technique used for particle identification and confirmation, not preliminary screening steps. The main quantification methods included Fourier Transform Infrared Spectroscopy (FTIR), Raman spectroscopy, FTIR plus Raman, microscopy-based, and multi-method. Studies were classified as FTIR when reporting "FTIR," "ATR-FTIR," "micro-FTIR," or "μFTIR" as the primary identification method. Raman spectroscopy was identified through keywords "Raman," "micro-Raman," or "μRaman." Studies employing both spectroscopic techniques were classified as "FTIR plus Raman" only when both methods were applied systematically to the same sample set, not when one method was used only for verification of ambiguous particles. Microscopy-based identification was assigned when studies relied primarily on "stereomicroscopy," "visual identification," "optical microscopy," or "morphological analysis" without systematic spectroscopic confirmation. Multi-method approaches combining three or more techniques (e.g., FTIR, Raman, and SEM) were classified separately to recognize comprehensive analytical protocols. Methods not fitting these categories, including hot needle tests, staining techniques, or density-based presumptive identification, were classified as "Other."

The standardization of the quantification methods enables assessments of how analytical detection capabilities influence reported concentrations, particularly regarding size detection limits and polymer identification success rates. The classification prioritizes the definitive identification technique over preliminary separation or screening steps.

Sample volume standardization and binning. Sample volumes varied across four orders of magnitude, from 0.3 liters to over 1000 liters. To enable analyses of how sample volume influences detection and reported concentrations, volumes were classified into standardized bins. First, reported volumes required unit standardization. Most studies reported volumes in liters, while some studies did not specify the volume of the samples collected. The arithmetic mean was calculated and used for bin assignments. Volume bins were established based on natural breaks in the distribution and operational considerations. Very small volumes (less than 5 liters) represent grab samples and small-volume collections typical of manual sampling and analytical throughput constraints. Small volumes (5 to 50 liters) capture moderate composite samples and bucket-based collections. Medium volumes (51 to 250 liters) represent larger composite sampling campaigns and automated sampler capacities. Large volumes (251 to 750 liters) and very large volumes (greater than 750 liters) indicate specialized high-volume sampling systems or extended monitoring periods. Records lacking sample volume documentation (28 records, 15.6% of the 180 total records) were assigned to a "Not reported" category and flagged for exclusion from volume-dependent analyses. This reporting gap was documented as it represents a significant metadata limitation affecting reproducibility and cross-study comparisons. Figure 3 shows the count of sampling methods, quantification methods, and sample volume bins after applying the standardization methods.

Concentration unit standardization. Concentration reporting exhibited the greatest diversity across studies, with units such as particles per liter, particles per cubic meter, and mass-based measurement (mg/g). Standardization to a common unit (particles per liter) was essential for quantitative synthesis. The standardization protocol followed a hierarchical conversion approach. For studies reporting particles per cubic meter, direct conversion was applied using the factor of 1,000 liters/m^3. One study that reported mass-based measurement (mg/g) presented significant challenge. Conversion from mass to particle number requires assumptions about average particle mass, which depends on particle size distribution, polymer density, and particle morphology. These parameters were not reported with sufficient details. To avoid introducing compounding assumptions and uncertainties, mass-based concentration was not converted to particle counts. That record was excluded from standardized concentration analyses, though original values were retained in the dataset for future mass-based syntheses. Measurements reported only as "below detection limit" were also excluded from quantitative analyses due to absence of specific values, though their occurrence was documented as it provides information about the detection frequency.

After the implementation of the standardization procedures, eight records were excluded as one study reported only mass-based concentrations (mg/g) without particle counts, preventing conversion to particles per liter without substantial assumptions about particle size and density; the rest had MP concentration reported as below detection limit. These eight excluded records were retained in the dataset with documented limitations as they contain valuable metadata and may be suitable for future mass-based or qualitative analyses. Consequently, the standardization process achieved a 95.6% success rate (172 out of the original 180 records for concentration-based analyses). All subsequent statistical analyses were based on the 172 records with successfully standardized particle concentrations, unless otherwise specified.

Particle size standardization. Particle size information was reported inconsistently across studies with formats, including single values, ranges with various delimiters, descriptive text, and mixed units. A two-stage standardization process was developed: first, extracting numeric size values in standardized units (micrometers); second, assigning particles to standardized size bins for categorical analysis. For each size field, all numeric values were extracted along with their unit indicators (micrometers, millimeters, "mm", or "µm"). When millimeters were detected, values were converted to micrometers by multiplying by one thousand.

For size ranges (e.g., "100–500 µm", "1 to 5 mm", or "20–45 µm"), both boundary values were extracted. The midpoint (arithmetic mean of minimum and maximum) was calculated and used as the representative numeric size for bin assignment as this approach best represents the central tendency of particles detected within that range. For single size values without ranges, the reported value was used directly. Entries with ambiguous formatting (e.g., "median 50 µm ± 20") were parsed to extract the central value (50 µm in this example), with standard deviations documented but not used for bin assignment. Studies that did not report any size range of the identified particles could not be reliably converted to numeric sizes and were assigned to a "Not reported" category, documented separately from the quantitative size data.

Standardized size bins were established to capture ecologically and analytically relevant size fractions: 1 to 50 µm, 50 to 100 µm, 100 to 500 µm, 500 to 1000 µm, and greater than 1000 µm. These bins align with common detection limit transitions (100 µm separating visual from microscopic techniques, 1000 µm separating MPs from larger debris) and enable comparisons across studies with different lower detection limits.

Particles were assigned to bins based on their calculated numeric size. For particles reported with ranges spanning multiple bins (e.g., "50 to 300 µm" spanning three bins), assignment was based on the midpoint (175 µm, assigned to 100 to 500 µm bin). When studies explicitly reported the dominant size range rather than full distributions, the dominant range was used for bin assignment to best represent the modal particle population.

Correction for size range detection limits. Systematic differences in analytical detection limits across studies create artifactual concentration variability because MP size distributions follow a power-law relationship where smaller particles occur at exponentially higher abundances[32,33]. To enable quantitative comparisons, we applied the rescaling method[32] to correct all concentration measurements to an equivalent 1–5000 µm detection range using power-law correction factors (CFs) with exponent $\alpha = 1.6$. Detection limits were extracted from reported methodological details in source publications; when only a single size class was reported, this range was used as the detection window. Studies detecting smaller particles (e.g., ≥1 µm via Raman spectroscopy) required minimal correction (CF ≈ 1), whereas studies limited to coarser detection (e.g., ≥100 µm via visual microscopy) required substantial upward adjustment (CF ≈ 5–10) to account for undetected sub-100 µm particles. All statistical analyses and concentration comparisons use these rescaled, detection-limit-corrected concentrations unless explicitly noted otherwise.

Morphological and polymer characterization. Morphological classifications required standardization due to overlapping terminology and varying levels of specificity across studies. Shape categories were consolidated into five primary types: fibers, fragments, films, granules, and tire wear particles (specifically identified particles from tire abrasion, often reported separately due to their distinct source and composition).

When studies reported multiple morphologies for individual samples (e.g., "predominantly fibers with some fragments"), each shape was recorded separately to capture the full morphological diversity. This provided a comprehensive inventory of particle types present in urban runoff, with the understanding that individual samples typically contain mixed morphologies. Shape distributions were calculated by counting all shape occurrences across records, with multiple shapes per sample counted separately when reported.

Polymer composition standardization involved consolidating various polymer names and abbreviations to standard nomenclature. Common polymers were identified through keyword matching: polyethylene (PE) captured "polyethylene," "PE," "LDPE," "HDPE"; polypropylene (PP) included "polypropylene" and "PP"; polyethylene terephthalate (PET) encompassed "PET," "polyethylene terephthalate," and "polyester." Specialized polymers like styrene-butadiene rubber (SBR) and ethylene propylene diene monomer (EPDM) were identified through exact matches. When studies reported only "plastic" or "synthetic polymer" without specific identification, these were categorized as "unidentified polymer" rather than being assigned to specific polymer classes.

Similar to the morphological analyses, when multiple polymers were identified in individual samples, each polymer type was recorded separately. This approach acknowledged that urban runoff typically contains diverse polymer compositions, and enabled assessment of polymer prevalence across the full dataset rather than forcing

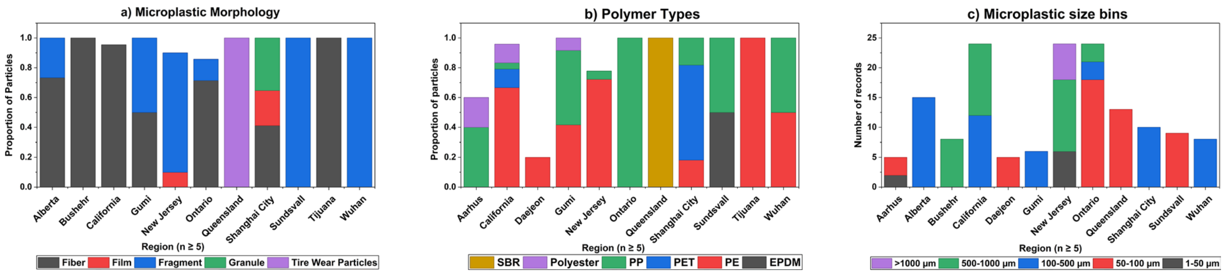

Fig. 4 Distribution of (**a**) microplastic morphology, (**b**) polymer type, and (**c**) size bin distribution across Dataset of Urban RUnoff Microplastics (DURUM).

assignment of single dominant types. Figure 4 shows the morphological, polymer type, and size bin distribution after applying the standardization methods included in DURUM.

Environmental metadata. Environmental and meteorological variables were extracted as reported in source publications to enable assessments of how weather conditions influence MP transport. The documented variables included antecedent dry days (number of days since last precipitation event), total rainfall depth during the sampling events, air temperature, and average wind speed. However, reporting consistency varied substantially across these variables. Antecedent dry days were reported in 40 of the 172 standardized concentration records (23%). Definitions varied across studies, with some employing rainfall thresholds (e.g., days with a total depth below 1 millimeter), others using discharge-based criteria (days without measurable runoff), and many not specifying their operational definition. Spatial scale of meteorological data also varied, ranging from site-specific rain gauges installed at or near sampling locations to regional weather station data located at varying distances from sampling sites. This spatial heterogeneity can introduce uncertainty in the representativeness of reported meteorological conditions for actual sampling locations.

Total rainfall depth was reported for a similar subset of sampling events, while air temperature and wind speed data were documented even less frequently. The inconsistent reporting of environmental metadata poses a limitation for temporal variability analyses and mechanistic transport modeling as these variables are theoretically important controls on MP accumulation and mobilization processes.

Data Records

DURUM is stored as a single Excel file (Urban Runoff MP Dataset.xlsx) on the Open Science Framework (OSF) repository (https://osf.io/9cgu4/)[34] under a CC-By Attribution 4.0 International license. The file contains 180 records representing individual sampling efforts from 28 peer-reviewed publications spanning 2018 to 2024, with 42 variables documenting MP concentrations, sampling methodologies, environmental conditions, and particle characteristics. Each record corresponds to a unique sampling event extracted from the published literature on MP in urban stormwater.

Repository content and structure. The Open Science Framework (OSF) repository provides three primary components to ensure data accessibility and reproducibility. The main dataset file contains all records with complete variable definitions and standardized values. A comprehensive data dictionary accompanies the dataset, defining each variable's meaning, acceptable values, units, and format specifications. Jupyter notebook script written in python documents the standardization workflow, including unit conversions, power-law size corrections, and categorical harmonization procedures. This repository structure enables users to understand both the final dataset and the methodological steps taken to create it from heterogeneous source literature.

Dataset architecture. The dataset organizes 42 variables into six thematic groups that reflect the data compilation workflow and intended use cases. The identification group captures bibliographic metadata linking each record to its source publication through author names, publication year, and Digital Object Identifier (DOI). The geographic group documents sampling location through multiple complementary fields including standardized regions, coordinate pairs in both text and decimal formats, and site-specific descriptions. The methodological group preserves both original and standardized classifications for sampling approaches and analytical quantification techniques, enabling users to assess methodological bias or select subsets based on analytical rigor. The environmental conditions group contains meteorological variables such as rainfall, air temperature, antecedent dry days, and wind speed, though data availability varies substantially across these fields. The MP characteristics group provides comprehensive information on particle concentrations, size distributions, polymer compositions, and morphological types. The standardization and correction group contains derived variables essential for cross-study comparison, including size detection bounds, and harmonized concentration values adjusted for methodological differences in detection limits.

Variable definitions. Table 1 provides complete definitions for all 42 variables in the dataset, organized by their column position in the Excel file. Variables are presented with their exact column names as they appear in the dataset to facilitate direct lookup and use.

Column	Variable Name	Definition
1	ID	Unique study identifier assigned in the DURUM dataset
2	Author(s) and publication year	Primary authors and year of publication of the study
3	DOI	Digital Object Identifier for the referenced publication
4	Sampling location	Site description where samples were collected (e.g., bioretention cell, highway, residential catchment)
5	Land use tags	Standardized land-use categories assigned for analysis (10 categories: Urban Infrastructure, Green infrastructure, Residential, Urban catchment, Transportation, Engineered infrastructure, Mixed, Industrial, Commercial, Agricultural)
6	Country	Country where the sampling study was conducted (field not populated; use Region instead)
7	Location	City/town or general geographic locality of the sampling site
8	Region	Larger geographic region or state/province (primary geographic classifier)
9	Latitude	Reported latitude coordinate of the sampling site in original format
10	Longitude	Reported longitude coordinate of the sampling site in original format
11	Area	Watershed or drainage area as reported in the literature
12	Sampling method	Microplastic sampling approach used in the study as described by authors
13	Sampling method (standardized)	Harmonized sampling method category (5 categories: Composite, Grab, Automated, Engineered Flow, Other)
14	Quantification method	Analytical technique used to identify and count microplastics as described by authors
15	Quantification method (standardized)	Harmonized microplastic quantification category (7 categories: FTIR, Multi-method, FTIR + Raman, Raman, Hot Needle, Microscopy, GC-MS)
16	Date	Sampling date or date range reported in the study
17	Total rainfall	Total rainfall depth during the event or sampling period
18	Temperature	Air temperature conditions during sampling
19	Antecedent dry days	Number of days without rainfall before the sampling event
20	Average wind speed	Mean wind speed during the sampling period
21	Measured MPs	Reported microplastic count from the study before standardization
22	Unit	Unit originally used by the authors to report microplastic counts
23	Sample volume	Volume of water sampled as reported by the authors in original units
24	Sample volume (L)	Sample volume converted to liters for standardization
25	Sample volume bin	Standardized category for sample volume range (6 categories: Very Small, Small, Medium, Large, Very Large, Not reported)
26	MP concentration	Reported microplastic concentration from the literature in original units
27	Unit.1	Unit associated with MP concentration as reported in source publication
28	Standardized Concentration (particles/L)	MP concentration converted to particles per liter before size correction
29	Size range	Size range of microplastics analyzed in the original study
30	Controls	Description of procedural, field, or laboratory blanks used in the study
31	Identified polymer (s)	Complete list of polymer types identified in the study
32	Dominant polymer (s)	Most frequently occurring polymer type or types in the sample
33	Dominant size range	Most abundant microplastic size range reported in the study
34	Numeric size (µm)	Midpoint of the dominant size range expressed in µm
35	Standardized size bin	Assigned size category based on dominant size (6 categories: 1–50 µm, 50–100 µm, 100–500 µm, 500–1000 µm, >1000 µm, Unknown)
36	Dominant shape	Most frequently observed microplastic morphology (e.g., fiber, fragment, tire wear particle)
37	Latitude	Latitude converted to decimal degrees for computational use
38	Longitude	Longitude converted to decimal degrees for computational use
39	x1M_um	Lower size detection limit (µm) for microplastics in the study
40	x2M_um	Upper size detection limit (µm) for microplastics in the study
41	CF_1_5000µm	Correction factor used to normalize concentration to the 1–5000 µm size domain
42	Conc_1_5000µm (particles/L)	Size-corrected MP concentration normalized to the 1–5000 µm detection range

Table 1. Complete variable definitions for the Dataset of Urban RUnoff Microplastics (DURUM).

Technical Validation

Source literature quality assessment. All data in DURUM were extracted exclusively from peer-reviewed publications indexed in the Web of Science Core Collection. The 28 source studies underwent rigorous peer review prior to publication in well reputed journals. This publication-level quality control ensures that original sampling protocols, analytical methods, and reported concentrations met scientific standards prior to inclusion in DURUM.

Data extraction verification. Data extraction accuracy was ensured through a systematic verification protocol. Each record was extracted by the primary author and cross-checked against source publications for numerical accuracy. Ambiguous values or formatting inconsistencies in source publications were resolved through examination of supplementary materials. When original reporting units required conversion (e.g., particles/m^3 to particles/L), conversion calculations were documented in the dataset and are reproducible from the recorded original values. The standardization process achieved a 95.6% success rate (172 of 180 records), with eight records excluded due to incompatible units (mass-based concentrations) or values reported as below detection limit.

Internal consistency checks. Four main internal consistency checks were applied to validate the data quality. (1) Geographic coordinate verification: Decimal latitude and longitude values were verified against reported location names using mapping tools to confirm spatial accuracy. (2) Unit consistency: All concentration values were verified to ensure proper unit conversion and standardization to particles per liter. (3) Size range plausibility: Reported particle size ranges were checked against stated analytical detection limits to ensure physical consistency (e.g., studies using visual microscopy could not report particles below ~100 μm detection thresholds). (4) Methodological coherence: Quantification methods were cross-referenced with reported size detection limits to verify analytical plausibility (e.g., Raman spectroscopy studies should report finer size fractions than visual microscopy studies).

Power-law correction validation. The size-range correction methodology applied to harmonize concentrations across different detection limits follows an established approach in current literature[32], which has been validated across multiple aquatic systems. The power-law exponent ($\alpha = 1.6$) represents a conservative estimate within the range reported in MP literature (1.4–2.0). Application of this correction reduced the interquartile ratio from 90.8 to 32.1, indicating that approximately two-thirds of apparent concentration variability in the original dataset arose from methodological differences in detection limits rather than genuine environmental heterogeneity. This reduction in artifactual variability supports the validity of the correction approach for cross-study comparisons.

Comparison with published literature. Concentration ranges reported in the DURUM dataset are broadly consistent with those documented in published reviews[6] of MP contamination in urban runoff systems. Measured, standardized concentrations in DURUM span 0.16 to 1.9×10^4 particles/L, which aligns with the order-of-magnitude ranges reported in global synthesis studies of urban runoff MP. When concentrations are extrapolated to a common 1–5000 μm size range using size CFs, concentration values extend up to 2.1×10^6 particles/L, reflecting differences in analytical size cutoffs and emphasizing the strong influence of particle size normalization on reported concentrations. Regional concentration patterns in DURUM are consistent with findings documented in literature. Elevated concentrations are reported in Mediterranean and East Asian urban centers, including studies from Venice (Italy), Shanghai (China), and Daejeon (South Korea). In contrast, North American regions exhibit substantially lower and more variable concentrations, with reported values from Alberta and Ontario (Canada) and California and New Jersey (United States) approximately two orders of magnitude below peak Mediterranean values. The lowest concentrations are observed in Scandinavian regions, exemplified by studies from Aarhus, Denmark, consistent with stringent environmental regulations and advanced stormwater management practices. These patterns provide external validation that DURUM captures genuine environmental variability rather than systematic extraction errors.

Usage Notes

DURUM represents measurements from diverse sampling protocols, analytical methods, and reporting formats that have undergone systematic standardization to enable comparative analyses. The standardization process converted all concentrations to particles per liter and applied power-law corrections (exponent $\alpha = 1.6$) to account for differences in size detection limits across studies. Missing or excluded data occurred when source publications reported incompatible units, provided insufficient metadata for standardization, or documented interrupted monitoring without final concentration values.

The corrected concentrations reflect harmonization across size detection limits ranging from 1 to 1000 μm for lower bounds and 50 to 5000 μm for upper bounds. The correction methodology assumes a power-law size distribution validated across multiple aquatic systems but may introduce uncertainty for samples with substantially different distributions. Studies employing coarse detection limits receive proportionally larger CFs that amplify measurement uncertainty. Polymer identification reliability varies substantially between spectroscopic methods (FTIR, Raman) and visual or thermal methods (microscopy, hot needle tests), with 88% of records employing spectroscopic approaches capable of definitive identification.

Antecedent weather conditions significantly influence MP concentrations through accumulation and mobilization dynamics. The dataset demonstrates positive correlations between antecedent dry days and observed concentrations, consistent with models where longer dry periods allow greater surface accumulation before washoff. However, meteorological metadata is limited, with antecedent dry days available for 51% of records, rainfall for 55%, temperature for 3%, and wind speed for 6%. Users exploring environmental relationships should recognize that analyses based on available meteorological data may not represent the complete dataset. Sampling timing also influences concentrations as first-flush protocols typically yield higher values than whole-event or baseflow sampling.

For modeling applications, users should incorporate both corrected concentration data and associated metadata, including land use classification, sampling methods, and environmental conditions. CF may serve as a useful covariate to account for residual methodological effects not fully captured by size correction alone. Including these contextual variables will improve model performance and generalizability across different urban environments.

Users should recognize geographic limitations in the dataset. Ninety-five percent of records originate from high-income temperate regions in North America, Europe, and East Asia, with minimal representation from tropical climates, lower middle-income, low-income nations, and the Global South. Climate affects rainfall patterns and hydrologic regimes, while economic development influences plastic consumption and waste management infrastructure. Extrapolation to underrepresented geographic contexts requires caution and highlights the need for expanded sampling in currently understudied regions.

This dataset focuses on MPs in the 1–5000 μm size range and does not capture nanoplastics (<1 μm) or many macroplastic items. This size window matches the analytical capabilities of most literature studies but systematically underestimates total plastic particle burden. Nanoplastics may dominate particle number in some environments, while larger items contribute disproportionately to mass. Users developing mass budgets or fate models should recognize these size-based constraints when interpreting DURUM concentration data.

Finally, users should recognize inherent limitations in compiled datasets. Methodological heterogeneity across source studies introduces systematic uncertainty that cannot be fully eliminated through standardization. Geographic representation is skewed toward high-income and upper middle-income temperate regions (95% of records from North America, Europe, and East Asia), limiting extrapolation to underrepresented tropical, lower middle-income and low-income nation contexts. These limitations are documented in the dataset through metadata fields that enable users to filter records based on methodological characteristics and assess uncertainty for specific applications.

Data availability

DURUM is publicly available on the Open Science Framework (OSF) repository (https://osf.io/9cgu4/)[34] under a CC-By Attribution 4.0 International license. The repository contains the complete standardized dataset in the data directory, along with a comprehensive data dictionary describing all variables, standardization protocols, and controlled vocabularies. Researchers can download the dataset directly, contribute additional data via pull requests, or submit issues for dataset improvements.

Code availability

All analytical code used to process, standardize, and visualize the DURUM is available at https://github.com/abdulmobinibnahafiz/DURUM-Urban-Runoff-Microplastics under the MIT license. The notebooks directory contains Jupyter notebooks that reproduce all statistical analyses, figures, and summary outputs presented in this study. The repository includes documentation for replicating the standardization protocols and extending the dataset with new measurements.

Received: 25 November 2025; Accepted: 26 March 2026;
Published online: 02 April 2026

References

1. Österlund, H. *et al*. Microplastics in urban catchments: Review of sources, pathways, and entry into stormwater. *Sci. Total Environ.* **858**, 159781 (2023).
2. Ross, M. S. *et al*. Estimated Discharge of Microplastics via Urban Stormwater During Individual Rain Events. *Front. Environ. Sci.* https://doi.org/10.3389/fenvs.2023.1090267 (2023).
3. Salehi, M., Hafiz, A. M. I., Ahmadisharaf, E., & Jazaei, F. Release of Microplastics by Terrestrial Sources into Freshwater Environments. *In Occurrence, Detection, and Fate of Microplastics in Freshwater Ecosystems* (pp. 41-76). Singapore: SpringerNature Singapore (2025).
4. Dris, R., Gaspéri, J., Saad, M., Mirande, C. & Tassin, B. Synthetic Fibers in Atmospheric Fallout: A Source of Microplastics in the Environment? *Mar. Pollut. Bull.* https://doi.org/10.1016/j.marpolbul.2016.01.006 (2016).
5. Werbowski, L. M. *et al*. Urban Stormwater Runoff: A Major Pathway for Anthropogenic Particles, Black Rubbery Fragments, and Other Types of Microplastics to Urban Receiving Waters. *Acs Est Water* https://doi.org/10.1021/acsestwater.1c00017 (2021).
6. Wang, C. *et al*. Microplastics in urban runoff: Global occurrence and fate. *Water Res.* **225**, 119129 (2022).
7. Alimi, O. S. *et al*. Weathering pathways and protocols for environmentally relevant microplastics and nanoplastics: What are we missing? *J. Hazard. Mater.* **423**, 126955 (2022).
8. Birch, Q. T., Potter, P. M., Pinto, P. X., Dionysiou, D. D. & Al-Abed, S. R. Sources, transport, measurement and impact of nano and microplastics in urban watersheds. *Rev. Environ. Sci. Biotechnol.* **19**, 275–336 (2020).
9. Horton, A. A., Svendsen, C., Williams, R. J., Spurgeon, D. J. & Lahive, E. Large Microplastic Particles in Sediments of Tributaries of the River Thames, UK – Abundance, Sources and Methods for Effective Quantification. *Mar. Pollut. Bull.* https://doi.org/10.1016/j.marpolbul.2016.09.004 (2017).
10. Do, T. *et al*. Effect of the first-flush phenomenon on the quantification of microplastics in rainwater. *Mar. Pollut. Bull.* **187**, 114559 (2023).
11. García-Haba, E., Hernández-Crespo, C., Martín, M. & Andrés-Doménech, I. The role of different sustainable urban drainage systems in removing microplastics from urban runoff: A review. *J. Clean. Prod.* **411**, 137197 (2023).
12. Chen, H. *et al*. The occurrence of microplastics in water bodies in urban agglomerations: Impacts of drainage system overflow in wet weather, catchment land-uses, and environmental management practices. *Water Res.* **183**, 116073 (2020).
13. Ibna Hafiz, A. M. *et al*. A review of processes and models for the export of microplastics from terrestrial to aquatic systems. *Wiley Interdiscip. Rev. Water* **12**, e70004 (2025).
14. Nyadjro, E. S. *et al*. The NOAA NCEI marine microplastics database. *Sci. Data* **10**, 726 (2023).
15. Čerkasova, N. *et al*. A Public Database for Microplastics in the Environment. *Microplastics* **2**, 132–146 (2023).
16. Wilkens, J. L., Calomeni-Eck, A. J., Boyda, J., Kennedy, A. & McQueen, A. D. Microplastic in Dredged Sediments: From Databases to Strategic Responses. *Bull. Environ. Contam. Toxicol.* **112** (2024).
17. Chen, C. *et al*. Significant regional disparities in riverine microplastics. *J. Hazard. Mater.* **472**, 134571 (2024).
18. Cho, Y. *et al*. Microplastic emission characteristics of stormwater runoff in an urban area: Intra-event variability and influencing factors. *Sci. Total Environ.* **866**, 161318 (2023).
19. Shruti, V. C., Pérez-Guevara, F., Elizalde-Martínez, I. & Kutralam-Muniasamy, G. Current trends and analytical methods for evaluation of microplastics in stormwater. *Trends Environ. Anal. Chem.* **30**, e00123 (2021).
20. Wang, W. & Wang, J. Investigation of microplastics in aquatic environments: An overview of the methods used, from field sampling to laboratory analysis. *TrAC Trends Anal. Chem.* **108**, 195–202 (2018).

21. Gwinnett, C. & Miller, R. Z. Are we contaminating our samples? A preliminary study to investigate procedural contamination during field sampling and processing for microplastic and anthropogenic microparticles. *Mar. Pollut. Bull.* **173**, 113095 (2021).
22. Ahmed, W., Hamilton, K., Toze, S., Cook, S. & Page, D. A review on microbial contaminants in stormwater runoff and outfalls: Potential health risks and mitigation strategies. *Sci. Total Environ.* **692**, 1304–1321 (2019).
23. Eriksen, M. *et al.* Microplastic pollution in the surface waters of the Laurentian Great Lakes. *Mar. Pollut. Bull.* **77**, 177–182 (2013).
24. Hidalgo-Ruz, V., Gutow, L., Thompson, R. C. & Thiel, M. Microplastics in the Marine Environment: A Review of the Methods Used for Identification and Quantification. *Environ. Sci. Technol.* **46**, 3060–3075 (2012).
25. Lin, J., Liu, H. & Zhang, J. Recent advances in the application of machine learning methods to improve identification of the microplastics in environment. *Chemosphere* **307**, 136092 (2022).
26. Löder, M. G. J. & Gerdts, G. Methodology Used for the Detection and Identification of Microplastics—A Critical Appraisal. in *Marine Anthropogenic Litter* (eds Bergmann, M., Gutow, L. & Klages, M.) 201–227. https://doi.org/10.1007/978-3-319-16510-3_8 (Springer International Publishing, Cham, 2015).
27. Constant, M., Billon, G., Breton, N. & Alary, C. Extraction of microplastics from sediment matrices: Experimental comparative analysis. *J. Hazard. Mater.* **420**, 126571 (2021).
28. Watkins, L., Sullivan, P. J. & Walter, M. T. What You Net Depends on if You Grab: A Meta-analysis of Sampling Method's Impact on Measured Aquatic Microplastic Concentration. *Environ. Sci. Technol.* **55**, 12930–12942 (2021).
29. Gimiliani, G. T. & Izar, G. Difficulties in Comparison Among Different Microplastic Studies: The Inconsistency of Results and Lack of Guide Values. *Environ. Toxicol. Chem.* **41**, 820–821 (2022).
30. Weerasekara, P. G. Y. W. *et al.* Sampling and Analysis of Microplastics in the Coastal Environments of Sri Lanka: Estuaries of the Kelani River to Mahaoya. *Water* **16**, 1932 (2024).
31. Braga Moruzzi, R. *et al.* Stormwater Detention Reservoirs: An Opportunity for Monitoring and a Potential Site to Prevent the Spread of Urban Microplastics. *Water* **12**, 1994 (2020).
32. Koelmans, A. A., Redondo-Hasselerharm, P. E., Mohamed Nor, N. H. & Kooi, M. Solving the Nonalignment of Methods and Approaches Used in Microplastic Research to Consistently Characterize Risk. *Environ. Sci. Technol.* **54**, 12307–12315 (2020).
33. Schreyers, L. *et al.* Suitability of river plastic monitoring methods for citizen science. *Camb. Prisms Plast.* **3**, e31 (2025).
34. Ibna Hafiz, A. M. *et al. DURUM: Database of Urban Runoff Microplastics.* https://doi.org/10.17605/OSF.IO/9CGU4 (2026).

Acknowledgements
Abdul Mobin Ibna Hafiz, Ebrahim Ahmadisharaf, and Maryam Salehi were financially supported by United States Department of Agriculture (USDA) award number NR234209XXXXG004 and National Science Foundation (NSF) award numbers 2347006, 2446126, and 2305189. The content of this paper should not be considered as an endorsement by the funding agencies.

Author contributions
Abdul Mobin Ibna Hafiz: Conceptualization, Methodology, Formal analysis, Investigation, Validation, Visualization, Data curation, Writing—original draft, Writing—reviewing and editing. Ebrahim Ahmadisharaf: Conceptualization, Methodology, Writing—reviewing and editing, Project administration, Resources, Funding acquisition, Supervision. Maryam Salehi: Conceptualization, Methodology, Writing—reviewing and editing, Project administration, Resources, Funding acquisition, Supervision. Tim H. M. van Emmerik: Conceptualization, Methodology, Writing—reviewing and editing. Tamara Galloway: Conceptualization, Writing—reviewing and editing. Jeffrey Farner: Writing—reviewing and editing. Jason C. White: Writing—reviewing and editing. Naipeng Han: Writing—reviewing and editing. Wen Zongguo: Writing—reviewing and editing. Yanxu Zhang: Writing—reviewing and editing.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to E.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026