Supplementary Information -- Metadata

What you net depends on if you grab: A meta-analysis of sampling method's impact on measured aquatic microplastic concentration

Watkins L[1], Sullivan PJ[2], Walter MT[1]

[1]Department of Biological and Environmental Engineering, Cornell University

[2]Department of Natural Resources, Cornell University

Metadata associated with the following .csv data files:

A. Literature Review, Full

B. Literature Review, Paired

Disclaimer: These tables are a compilation of information from available datasets (no authors-sourced data is included). They were gathered according to the methods detailed in and utilized for the analysis included in the associated manuscript. We share them as an act of transparency and so that they may serve as a reference of the many important works that have made our own analysis possible. We caution, however, the use of these tables as raw data for future analysis. To avoid the propagation of any errors in transcription or interpretation that may occur in these tables, we encourage future work to refer back to the original dataset to confirm values and details.

A. Dataset: Literature Review, Full

For data collection information, refer to the methods section of the associated manuscript.

Number of columns/variables

Number of columns: 35

Number of rows

Number of rows: 140

Variable list

- **Author_date**: First author & publication date of study
- **doi**: DOI link to the row's study
- **earliestyear_sampled**: For studies that provide dates for when samples are collected, this column includes the year when first samples of the study were collected. For studies where sampling dates could not be found, this lack of data is designated by "nd".
- **latestyear_sampled**: For studies that provide dates for when samples are collected, this column includes the year when last samples were collected. For studies where sampling dates could not be found, this column includes date of paper submission.
- **method**: the method of microplastic collection. Samples that were collected by scooping water and then pouring it through a handnet are marked as grab samples here.
    - Options: Grab, net, pump
- **method_specific**: defining the exact method used, in an attempt to preserve some of the more specific language from the paper describing exact sampling apparatus used.

    - Options for grab method rows: bucket+handnet, bulk
    - Options for net method rows: conical, manta, rectangular
    - Options for pump method rows: centrifugal pump, flow sampler, in situ pump, jet pump, large flow sampler, pump, PVC pump, submersible pump, Teflon pump
- **volume_measure**: the method or apparatus used to measure or calculate volume sampled, according to the study
    - Options: vessel (for grab method samples only), GPS distance (for net method samples only), flowmeter, mechanical flowmeter, digital flowmeter, water sampler (unknown method preserved from study language), current meter, river velocity (for net method samples only), time (for pump method samples only), time x speed (for towed net method samples only), not mentioned.
- **river_ocean**: the kind of waterbody sampled
    - Options: marine (ocean/sea), wetland, lake, river, estuary, lagoon, reservoir, pond
- **lowestsize_mm**: The smallest filtration size used to separate the sample contents from discarded particles, in units of millimeters. Depending on methods, this may be the size of filter paper openings, a sieve size, or a net mesh.
- **lowest_mesh**: Description to inform lowestsize_mm column. The type of filtration used to enforce the lowest size.
    - Options: mesh (plastic), sieve (metal), or filter (paper or plastic).
- **vol_L**: Sample volume, in units of liters. For studies with inconsistent volumes across multiple samples, this is the averaged volume.
- **vol_source**: How vol_L value was determined for this table.

    - Options: given (found in the text); averaged (several volumes in the text were averaged to include in table); calculated, distance (net area multiplied by tow distance); calculated, speed (net area multiplied times tow time multiplied times average tow speed)
- **uppersize_mm**: Microplastics are typically defined as plastics < 5mm. This column notes the upper size limit included in the concentrations and counts of the given study. For studies where largest included particle size was not found, this lack of data is designated by "nd".
- **Cavg.L**: Average reported concentration, in units of particles per liter.
- **Csd.L**: Standard deviation of reported concentration, in units of particles per liter. For studies where a standard deviation on average concentration could not be found or easily calculated, "nd" is used.
- **num_samples**: The number of samples collected. This value would indicate the number of values included in average concentration (Cavg.L) and average volume (vol_L). For some studies, the number of samples for different methods or locations is difficult to discern, but we have done our best to enter the best guess. "nd" is used for studies where number of samples was not found.
- **total_particles_collected**: The number of particles counted overall in the analyzed samples. For some studies, the number of particles for different methods or locations is difficult to discern, but we have done our best to enter the best guess. "nd" is used for studies where number of samples was not found.

- **secondary_id**: The second (or most advanced) method or apparatus used to confirm a particle is plastic and should be included in the sample count, in addition to or in place of a visual inspection.
- **top_shape**: The dominant particle type detected in the sample. "nd" indicates mention of most prominant particle type was not found in the paper.
    - Options: fiber, fragment, bead (includes pellet-type), film, foam.
- **excluded_particles**: Notes whether a certain particle type was omitted from all counts in the study.
- **topsize_bottom_mm**: The lower bound of the most common size range of particles, in units of mm.
- **topsize_upper_mm**: The upper bound of the most common size range of particles, in units of mm.
- **top_polymer**: The most common polymer type identified.
- **second_polymer:** The second most common polymer type identified.
- **top_color**: The most common particle color identified.
- **second_color**: The second most common particle color identified.
- **location**: General area where sample was collected.
- **lat**: The latitude of the general area where samples were collected. This should not be used for replicating sampling. It is merely intended to be used, on a global scale, to show the general region where samples were collected.
- **long**: The longitude of the general area where samples were collected. This should not be used for replicating sampling. It is merely intended to be used, on a global scale, to show the general region where samples were collected.

- **waterblanks**: The number of procedural blanks run for a given sample. 0 indicates no mention of blanks were found in the paper.
- **airblanks**: The number of blanks meant to measure potential contamination from lab air by leaving samples exposed. 0 indicates no mention of blanks were found in the paper.
- **blanks_subtracted**: An indication in the methods or results of whether reported concentrations are corrected by subtracting any blank sample measurements from the sample counts.
    - Options: yes (concentrations corrected by blanks), no (concentrations not corrected by blanks), unknown (no mention of a correction), some (imprecise explanation of whether correction has been or should be made)
- **comment**: Column with notes about assumptions made, method specifics, where values were found or how they were calculated. These are a commentary from Lisa Watkins as this table was being compiled.

Missing data codes

When no information could be found in the publication, supplementary materials, or linked data (though possibly just overlooked!), "no data" is designated in the tables by "nd".

A. Dataset: Literature Review, Paired

For data collection information, refer to the methods section of the associated manuscript.

Number of columns/variables

Number of columns: 12

Number of rows

Number of rows: 447

Variable list

- **study**: author and publication date of the dataset from which the row's values come
- **doi**: DOI link associated with study
- **location**: the kind of waterbody sampled
    - Options: ocean (marine, estuarine samples), river, lake
- **study_type:** A category of paired sample comparisons
    - Options: mesh_compare (pairs of samples from same time, location and method but with different mesh sizes), grab-net (pairs of samples from same time and location, but one collected with a grab and the other with a net—not necessarily the same mesh size), net-pump (pairs of samples from the same time and location, but one collected with a pump and the other with a net—not necessarily the same mesh size).
- **source**: the location within the study's manuscript where the information used to fill in the row's volume and concentration values were found

    - Options: varies by the study manuscript's numbering. Note: Those of the style "S#" are from a figure or table within the study's supplementary materials.
- **sample_num**: Indicates which sample's data is included in a given row.
    - Options: Varies. Sample labels and format vary by study.
- **method**: The microplastic sample collection method used to collect the data included in the given row
    - Options: grab, pump, net
- **mesh_mm**: The smallest filtration size used to separate the sample contents from discarded particles, in units of millimeters
- **vol_L**: Sample volume, in units of liters, for the sample noted in row's sample_num column.
- **Cavg.L**: Reported concentration, in units of particles per liter, of the sample noted in row's sample_num column.
- **contam_numpersample_avg**: The average number of particles of contamination measured in a blank sample collected by the given study. For those that did not include blank sample counts, "nd" denotes "no data available".
- **comment**: notes from Lisa Watkins as this data was compiled to keep track of methodological differences, calculations or assumptions made, and other information about the study or the row's values.