Efficient Prediction of Microplastic Counts from Mass Measurements

Shuyao Tan,[†] Joshua Taylor,[‡] and Elodie Passeport[*,†,¶]

†*Department of Chemical Engineering and Applied Chemistry, University of Toronto, 200 College Street, Toronto, ON M5S 3E5, Canada*

‡*The Edward S. Rogers Sr. Department of Electrical & Computer Engineering, University of Toronto, 10 King's College Road, Toronto, ON M5S 3G4, Canada*

¶*Department of Civil and Mineral Engineering, University of Toronto, 35 St. George Street, Toronto, ON M5S 1A4, Canada*

E-mail: elodie.passeport@utoronto.ca

Abstract

Microplastics must be characterized and quantified to assess their impact. Current quantification procedures are time-consuming and rely on expensive equipment. This study evaluates the use of machine learning to estimate the number of microplastic particles based on aggregate particle weight measurements. Synthetic datasets are used to test the performance of linear regression, kernel ridge regression and decision trees. Kernel ridge regression achieves the strongest performance, and it is also tested with experimental datasets. The numerical results show that the algorithm is better at predicting the counts of larger and more homogenous samples, and that contamination by organics does not significantly increase error. In mixed samples, prediction error is lower for heavier particles, with an error rate comparable to or better than that of manual counting. Overall, the proposed method is faster, cheaper and easier than current approaches.

Synopsis: Using generated and real datasets, this study demonstrates that Kernel Ridge Regression can estimate microplastic counts from weight measurements as accurately as traditional visual sorting techniques.

Introduction

Plastic particles less than 5 mm in size are referred to as microplastics.[1] They are widespread in the environment, both in densely populated and remote areas.[2–5] Much remains unknown about their impact on organisms; while negative effects have been observed in several studies, a clear, causal picture has not yet emerged.[6] This is in part due to the diversity of microplastics, i.e., shape, size, polymer identity, chemical mixture, color, and count, which are not accounted for in most studies on their ecological effects.[6,7]

Research on microplastics can have several different objectives, e.g., source identification, fate assessment, and (eco)toxicological impact evaluation, which all require the characterization and quantification of microplastics. Depending on the subject, the quantity of microplastics can be referred to as count, concentration, abundance, or dose, expressed as particles per sampling area, per volume or per mass of sample.[8–10] Here, we will use the term count. The count of microplastics, though not sufficient alone to address all research questions, remains an important metric. Indeed, higher concentrations of microplastics are found near more populated areas, thus helping estimate proximity to sources. Microplastic counts are also used to evaluate the performance of water treatment systems.[11,12] Finally, the count of microplastics is also one of the drivers of observed effects on organisms.[6]

While currently there are no formal standards for microplastic characterization and quantification, the most common method involves the following steps: organic digestion, in which chemicals or enzymes are used to remove organics; density separation, in which a solution of known density is used to separate lighter particles from heavier sediments; sieving, in which the particles are passed through sieves of different mesh sizes and grouped into different size

ranges; visual identification, in which particles are manually sorted under a microscope for counting, sometimes particle size measurement, and classification for colors and shape categories.[8–10,13–15] The counting step is time-consuming and subject to human error. Usually a technique such as Raman or Fourier-transform infrared spectroscopy (FTIR) is also used to verify that the particles are indeed plastic, further increasing cost and time. As a result, characterization and quantification have become the rate-limiting steps in the analysis of microplastic contamination.

Recent efforts have attempted to make this process cheaper and faster with machine learning. Various machine learning algorithms have been coupled with spectroscopic techniques to auto-identify microplastic particles.[16] For example, a convolutional neural network was trained to classify microbeads based on microscopic images, and support vector machines, adaptive boosting and random forests have been trained to identify particles from scanned images.[17,18] These methods are highly effective, but their applicability is limited for the following reasons: (i) they rely on expensive high-resolution scanners or cameras attached to microscopes, which require expertise to use, (ii) cannot reliably distinguish plastics from organic particles, and (iii) require the particles to be loosely distributed over a surface. Our objective is to develop an alternative that is cheap, fast, and simple enough that anyone can use it to reliably quantify microplastic contamination.

The objective of this study is to use machine learning to estimate the number of microplastic particles in a sample from aggregate weight measurements. Microplastic samples will be sieved into different size ranges, and the total particle weight of each size range will be measured. From the resulting set of weight measurements, we estimate the count of each type of microplastic particle. Koelmans *et al.* have proposed empirical relationships to calculate a microplastic mass-volume-number conversion factor based on data averaged over many microplastic studies.[19] In this study, we aim to develop a method for quantifying microplastic particles without assuming that all samples share similar compositions. We use three machine learning algorithms: linear regression, kernel ridge regression, and decision

trees. We have chosen these techniques for their simplicity and low computational cost. While neural networks could potentially also work well, we prefer these techniques for several reasons. First, the effectiveness of modern, deep neural networks comes from massive training datasets, which are not available in this application. Second, neural networks are more computationally intensive and often entail extensive parameter tuning, thus increasing the complexity of the procedure. In our numerical results, we will compare the performance of the algorithms to manual counting and evaluate the impact of mass measurement error on count prediction.

Methods

Experimental and Synthetic Datasets

Two sets of experimental data are taken from the literature. Isobe *et al.*[20] made samples by adding microplastic particles to sea water, and sent them to different laboratories around the world. These samples were sieved into multiple size ranges, weighed, and counted, resulting in 10 pairs of particle weight and count measurements. McIlwraith *et al.*[21] washed blankets in a washing machine, collected fibers from the effluent water, and reported 18 pairs of total weight and count. These datasets will be referred to as *Interlab* and *Washing Machine*, respectively.

Due to the limited quantity and variation of published experimental data, a computational procedure was developed to generate synthetic data. These synthetic datasets were used to evaluate the performance of different algorithms under various circumstances. The procedure is summarized in Table S1. A particle's density depends on its type, polymer identity and the presence of biofilm. The shape of a particle is defined by its largest, medium, and smallest dimensions. The number of each type of particle in each of the samples described below was randomly drawn from the ranges given in Table S1 in Supporting Information. On average, there can be around 10 rubbers, 40 fragments, 50 films, 50 beads and 300 fibers

in a sample. These values are approximates to analysed samples collected from freshwater systems that are subject to human activities.[3,22–24] The probability density plots of the number of each type of particle can be found in Section S1 in Supporting Information. The dimensions of each particle were set as follows.

1. Rubber particles and fragments had similar dimensions but different densities. Their largest dimension was uniformly distributed between 153 μm and 1500 μm. The medium dimension was 70 to 100% of the largest dimension. The smallest dimension was 40 to 100% of the medium dimension. Rubber particles were lighter than fragments.

2. Films had similar densities to fragments, but their smallest dimensions were smaller. The largest and medium dimensions of films were drawn from the same range as those of fragments. The smallest dimension for films was 5 to 10% of the medium dimension.

3. Beads were assumed to be perfect spheres. Each bead's diameter was a uniformly drawn random number between 107 μm and 1500 μm.

4. Fibers were randomly divided into long and short fibers. Their diameters were between 10 and 40 μm. The length of each long fiber was 30 to 50 times of its diameter, and the length of each short fiber was 10 to 30 times of its diameter.

5. The largest dimension of organic particles was between 214 and 2000 μm. The medium dimension was 50 to 100% of the largest dimension. The smallest dimension was between 40 and 150 μm.

The sieving process for non-fiber particles was simulated by comparing the size of each particle with the mesh size. If both the largest and medium dimensions were larger than the mesh size, the particle remained on the sieve; if not, the particle passed through the sieve to the next one. In this study, the sieve mesh sizes were: 1000 μm, 500 μm, 300 μm and 106 μm. Although fibers were thin enough to pass through every sieve, they could still become entangled with the mesh or with other particles. To approximate this randomness, the probability for long fibers to stay on each sieve was set to be 0.3, 0.4, 0.2, 0.1, and for short fibers, the distribution was 0.15, 0.35, 0.3, 0.2. These probabilities were estimated

from storm water sample data collected from a parking lot using the same four sieve sizes specified above.[22]

After sieving, the total mass of all particles on each sieve was calculated by summing up all individual particle masses. Individual masses were obtained by multiplying the volume and density of each particle. For cuboid-shaped particles, volume was calculated by multiplying the three dimensions. Fiber volume was the product of its cross-sectional area and length. Spherical particles were treated as perfect spheres.

In this study, 10 datasets with different particle combinations were generated, each containing 100 samples (Section S2). These 10 combinations were:

1. Film and fiber.
2. Bead and fiber.
3. Rubber, fragment and fiber.
4. Rubber, fragment, fiber and half as many organic particles as fragments.
5. Rubber, fragment, fiber and as many organic particles as fragments.
6. Rubber, fragment, fiber and film.
7. Rubber, fragment, fiber, film and as many organic particles as fragments.
8. Rubber, fragment, fiber and bead.
9. Rubber, fragment, fiber, film and bead.
10. Rubber, fragment, fiber, film, bead and as many organic particles as fragments.

Fibers appeared in all samples because they are the most common type of particle found in water, sediment and air.[22,25,26] Fragments appeared in most samples because they are the second most common particle in water and sediment.[25] The above synthetic datasets form the base case for the numerical experiments in later sections. To test the dependence of algorithm performance on sample size, two more cases were generated with composition 9, but with more particles, as described in Table S2.

Machine Learning Algorithms

Let $x_i = [x_{i1}, x_{i2}, ..., x_{ik}]$ be a row vector of the masses in each of k sieves for the ith sample. The matrix $X = [x_1; x_2; x_3; ...; x_n]$ contains the mass measurements for all n samples. Let $y_i = [y_{i1}, y_{i2}, ..., y_{im}]$ be a row vector of the number of each type of particle in each sieve for the ith sample, and $Y = [y_1; y_2; y_3; ...; y_n]$ be the matrix of all counts. Let $\hat{Y} = f(X)$ be an estimate of Y based on some function of X. In each of the following learning frameworks, we assumed some simple structure for f, and then used training data consisting of known (X, Y) pairs to solve for the parameters of f that minimize prediction error.

Linear Regression (LR)

Linear regression assumes a relationship of the form

$$\underbrace{\begin{bmatrix} \hat{y}_{11} & \hat{y}_{12} & \dots & \hat{y}_{1m} \\ \hat{y}_{21} & \hat{y}_{22} & \dots & \hat{y}_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{y}_{n1} & \hat{y}_{n2} & \dots & \hat{y}_{nm} \end{bmatrix}}_{\hat{Y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} b_1 & b_2 & \dots & b_m \\ a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k1} & a_{k2} & \dots & a_{km} \end{bmatrix}}_{A} \tag{1}$$

Given training data (X, Y), the matrix A that minimizes the mean squared error is given by

$$A = (X^T X)^{-1} X^T Y. \tag{2}$$

Kernelized Ridge Regression (KRR)

Ridge regression is a modification of linear regression.[27] Linear regression assumes that there is no relationship between the independent variables. If there is some relationship among the independent variables, ridge regression often gives more reliable results than linear regression.[28] Ridge regression also has the form $\hat{Y} = XA$, but the matrix A is given

by

$$A = (\lambda I + X^T X)^{-1} X^T Y, \tag{3}$$

where $\lambda \geq 0$ is a complexity penalty and I is the identity matrix. The complexity penalty discourages large magnitude entries in A, hence reducing overfitting. This equation can also be written as

$$A = X^T (X X^T + \lambda I)^{-1} Y. \tag{4}$$

In KRR, the term XX^T is replaced by the Gaussian kernel function, given by

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right), \tag{5}$$

where σ^2 is referred to as the bandwidth of the kernel. The Gaussian kernel measures the similarity between any two points. Given a new input x, KRR identifies the portions of the training data with the highest similarity, which therefore have larger influences on the resulting estimate. Let K be an n by n matrix in which the ij^{th} entry is $K(x_i, x_j)$, and define

$$r = (K + \lambda I)^{-1} Y. \tag{6}$$

The parameter matrix A is given by

$$A = X^T r. \tag{7}$$

Given new input data x, the predicted output is given by

$$\hat{Y} = A^T x. \tag{8}$$

Note that this is equivalently written as

$$\hat{Y} = \sum_{i=1}^{n} r_i x_i^T x = \sum_{i=1}^{n} r_i K(x_i, x). \tag{9}$$

In this study, the Gaussian kernel parameters were found using grid search, as described in Section S2.

Decision Tree (DT)

Decision trees recursively partition a given dataset to minimize information entropy.[27] Information entropy measures the disorder of a dataset, and is given by

$$H = -\sum_{i=1}^{n} p(x_i) \log_2 p(x_i), \tag{10}$$

where p stands for probability. After each partition, the total entropy of the resulting subsets should be smaller than before. Training a decision tree only requires that one specifies its depth, that is, the level of partition. For example, a tree of depth two splits the original dataset into four subsets. Each final subset is called an end node. A prediction is made by passing a new data point through the decision tree, and outputting the average value of the training data at the resulting end node.

Training and Evaluation Methods

Training is the process of learning model parameters from given (X, Y) pairs. In this case, the (X, Y) pairs that make up the training data could be obtained in several ways, for example, from traditional counting-based quantification, or from an existing dataset. Evaluation of a model is done by inputting data X, then comparing the estimated $\hat{Y}$ with the known Y. In this study, the learning algorithms were trained and tested through leave-one-out cross-validation. In each dataset of 100 samples, 99 were used for training and the left-out sample was used to test the trained algorithm's performance. This process was repeated for every sample in each dataset, hence yielding 100 predictions. The accuracy of each algorithm was quantified by mean absolute error (MAE), mean percentage error (MPE) and normalized

root mean square error (NRMSE), which are defined as

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{11}$$

$$\text{MPE} = \frac{1}{n}\sum_{i=1}^{n}|\frac{\hat{y}_i - y_i}{y_i}| \tag{12}$$

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{y_{\max} - y_{\min}}, \tag{13}$$

where y and $\hat{y}$ are respectively the true and predicted values. There were occasional samples in which the number of a certain type of particle was zero, and these cases were not included in MPE calculation. Another normalization factor, $y_{max} - y_{min}$, was used to calculate NRMSE to avoid zero denominators. We conducted our numerical experiments as described above, and also with measurement error, which reflects non-idealities such as balance roundoff. We simulated these errors by adding random numbers uniformly distributed between -30% and 30% of the original value to the two finer mesh sieve weights. Random numbers uniformly distributed between -10% and 10% were added to the two coarser mesh sieve weights.

The 95% confidence intervals of the synthetic datasets were calculated using 1000 bootstrapped samples, in order to achieve a normal distribution. The statistical difference between two sets of predictions were evaluated using the Wilcoxon Signed-Rank Test. A significance level $\alpha = 0.05$ was used.

To test the effect of training dataset size on prediction accuracy, the following number of training samples were also tested: 90, 80, 70, 60, 50, 40, 30, 20, 10, 8, 6, 4, and 2. The corresponding testing dataset sizes were 10, 20, 30 and so on. For each synthetic sample composition, the training-and-testing procedure was repeated 20 times, and each time the training samples were randomly selected. Final result was reported as MPE for each type of particle against the number of training samples.

Results and Discussion

Training with Synthetic Data

Table 1: Mean Absolute Prediction Error

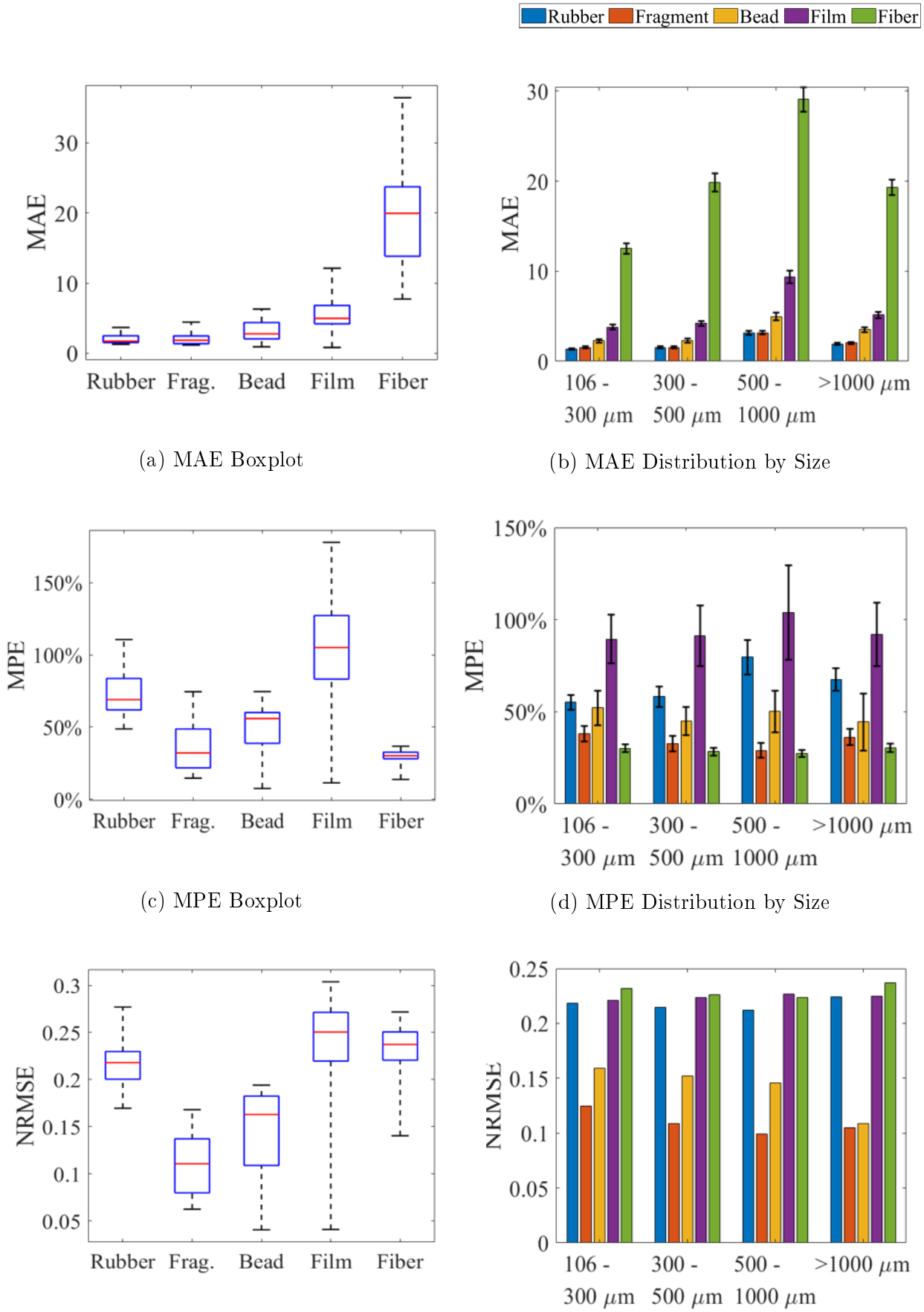
	KRR[a]	LR[b]	DT[c]	KRR-E[d]	LR-E[d]	DT-E[d]
Rubber	1.98 ± 0.07	2.14 ± 0.07	2.37 ± 0.10	2.02 ± 0.08	2.14 ± 0.07	2.34 ± 0.09
Fragment	2.06 ± 0.08	2.10 ± 0.07	2.99 ± 0.10	2.19 ± 0.08	2.23 ± 0.08	3.04 ± 0.10
Bead	3.25 ± 0.18	3.41 ± 0.18	4.17 ± 0.20	3.40 ± 0.18	3.60 ± 0.18	4.14 ± 0.20
Film	5.62 ± 0.25	5.79 ± 0.23	6.21 ± 0.25	5.72 ± 0.23	5.88 ± 0.25	6.50 ± 0.28
Fiber	20.2 ± 0.53	20.7 ± 0.53	22.3 ± 0.59	20.7 ± 0.49	21.1 ± 0.53	22.7 ± 0.62
Non-Fiber[e]	3.67 ± 0.12	3.74 ± 0.14	5.07 ± 0.18	3.89 ± 0.14	3.97 ± 0.14	5.10 ± 0.17

[a] Kernel Ridge Regression; [b] Linear Regression; [c] Decision Tree, implemented with Scikit-Learn[29]; [d] With weight measurement error; [e] Total number of microplastic particles excluding fibers

The performances of the three algorithms were evaluated by comparing the MAE with 95% confidence interval for the synthetic samples. The results are reported in Table 1. The KRR prediction results were statistically different from LR and DT (p-values $< 10^{-8}$). The results of LR and DT were similar (p-value = 0.57). Among the three algorithms, KRR always resulted in the lowest average MAE, both with and without incorporating weight measurement error. After adding weight measurement error, the performance of each algorithm did not significantly decrease (p-values were 0.71, 0.13 and 0.97 for KRR, LR and DT respectively). This suggests that all algorithms are somewhat robust to weight measurement error. We remark that DT is often poor at predicting continuous values because its output comes from a discrete set of values, the size of which is the number of end nodes.

KRR can be expected to outperform LR because, whereas LR assumes a linear relationship between the data, KRR can accommodate linear and some non-linear relationships. While highly scalable, KRR is more complicated than LR in that one must tune its model parameters (λ and σ). When simplicity is more important than accuracy, LR may be a superior choice.

We hereon focus on KRR because it achieved the best performance. Its error for the



(a) MAE Boxplot

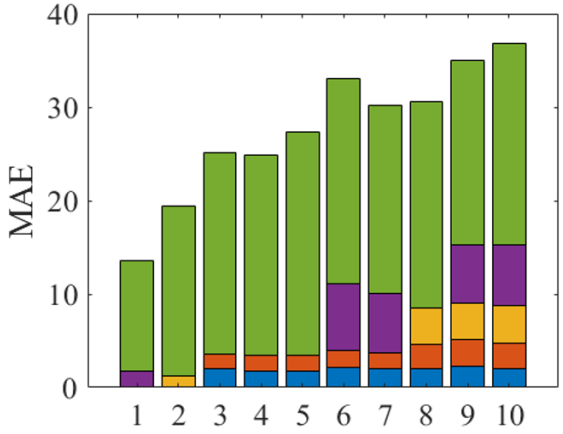
(b) MAE Distribution by Size

(c) MPE Boxplot

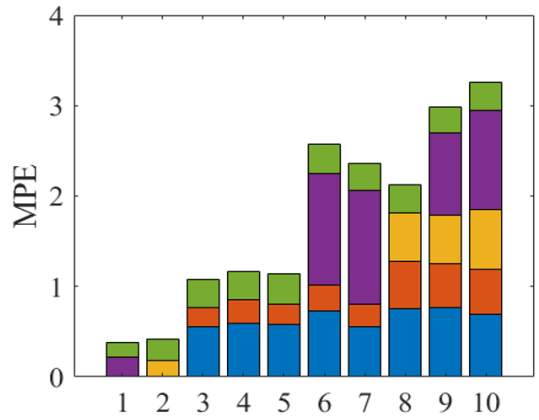
(d) MPE Distribution by Size

(e) NRMSE Boxplot

(f) NRMSE Distribution by Size



(g) Stacked MAE for each type of particle in each dataset.



(h) Stacked MPE for each type of particle in each dataset.

Figure 1: (Previous page.) Error of KRR on synthetic data. In the boxplots, the upper and lower bars represent the upper and lower bounds of the data. The box represents the 25 and 75 percentile, and the middle line represents the median value. Error bars on the right hand side are 95% bootstrapped confidence interval. Frag. represents fragment. Sample complexity increases from composition 1 to 10.

synthetic datasets is summarized in Figure 1. On rubber, KRR achieves low MAE, high MPE and high NRMSE because rubber is usually present in smaller numbers than other particle types. As a result, even an absolute error of one or two particles can result in a percentage error that is more than 50%. On the contrary, fibers are the most numerous. This leads to a large MAE, which, when divided by the large number of fiber particles, leads to a small MPE. For other particles, despite the fact that they are present in similar numbers, films have larger median and average errors than beads, and beads have larger errors than fragments. MAE differs for each size fraction because each size fraction has different amounts of particles. The $500-1000$ μm size range has the most particles and thus the largest MAE. NRMSE for rubber, fragment, film and fiber is almost constant for all size ranges. Altogether, these results show that the larger the number of microplastics in a given category, the higher its MAE and the lower its MPE.

Figure 1g shows the stacked MAE for each of the ten synthetic datasets. The smallest

prediction errors (MAE, MPE and NRMSE) for films and beads occur in their respectively purest samples, e.g., dataset 1 is composed entirely of films and fibers, and dataset 2 of beads and fibers. These values are even lower than that of fragments, because fragments always appear with rubbers in the synthetic datasets. MAE for beads and films increases dramatically with the addition of fragments and rubbers, while MAE for fragments only increases notably with the addition of beads. Also note that the smallest MAE for fibers occurs in dataset 1, made of only films and fibers. This is most likely because films are very light, so that fibers have a higher weight percentage in this dataset than in others. All these observations suggest that, when aggregate weight is used as the sole feature, heavier particle types have consistently lower prediction error because of their higher weight percentage in each sample. We can also see that MAE does not change significantly after adding various amounts of organics, most likely because the total weight of organics is smaller than that of the plastics, and the addition of organics does not change the order of weight percentages of other particles. For samples severely contaminated by organic particles, e.g., a lake with large amount of algae or plant debris sampled along with microplastics, an organic digestion step will be necessary prior to drying and measuring the sample weight, in order to limit the contribution of organics to the total sample weight. To summarize, the procedure is accurate for particles that account for a significant fraction of a sample's weight. This is why it is relatively difficult to predict fiber counts in mixed samples, which are numerous but might account for only a few percent of the overall weight. On the other hand, it is relatively easy to predict fragment counts in mixed samples because they account for a moderate fraction of the overall weight.

The total error associated with microplastic count prediction by KRR remains comparable to or lower than that of manual counting, even for the more heterogenous samples. Isobe *et al.*[20] used artificial samples to estimate the relative uncertainty of microplastic counts performed by humans. In their study, samples with $n_0 \approx 90$ fragments less than 2 mm were sent to 12 different laboratories. The relative error of a given laboratory's count, n, was

calculated as $(n - n_0)/n_0$. The average error of all laboratories was around $\pm 50\%$, which we interpret as the error of a typical human count. In other visual sorting microplastic studies, researchers have also reported a false positive ratio ranging from 20% to 70%, meaning that, among the particles visually sorted as microplastics, 20% to 70% of them were determined via spectroscopic analysis to be non-plastic.[8,22,30,31]

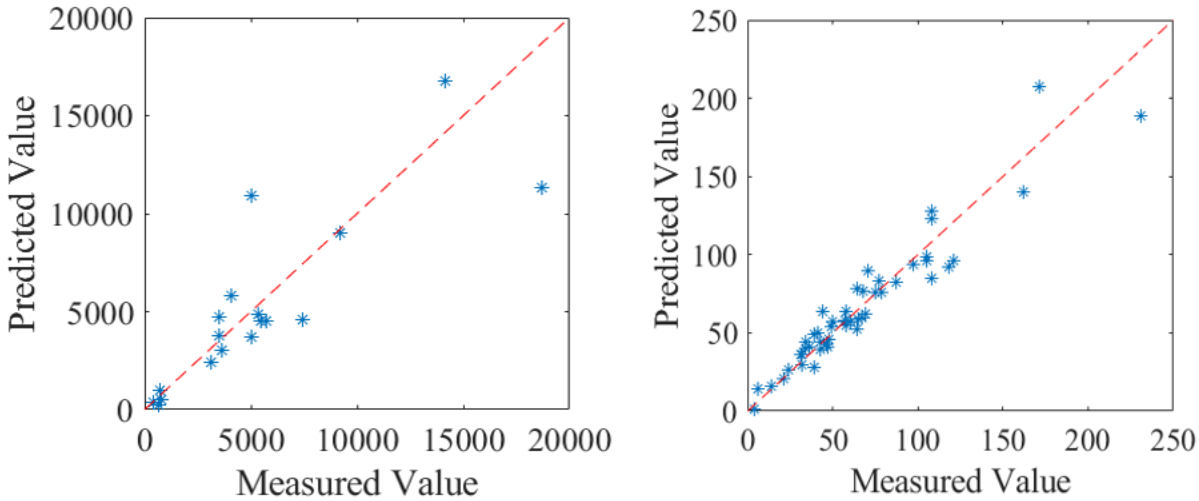
In our synthetic datasets, the average number of fragments, beads and films is around 50 for each, whenever present. Therefore, based on the similarity of *Interlab* samples to our synthetic data, it is reasonable to compare our results with theirs, with respect to each particle type. As discussed earlier, when these particles are present in their respective purest samples, their MPE can be less than 10%. In mixed samples, however, MPE for beads can go beyond 60% and films can go beyond 100%. Figure 1h shows MPE for each type of particle in the ten synthetic datasets. In mixed samples, MPE for films is always above 90%, due to their low contribution to the total weight, and that for rubbers is always above 50%, due to their small amount in each sample. MPE for fragments and fibers is below 35% except for datasets 8, 9, and 10, where the presence of beads increases the prediction error for fragments. When all particles are present, e.g., datasets 9 and 10, the prediction accuracy is better for fibers, comparable for beads and fragments, and worse for films and rubbers, compared to the accuracy of visual sorting. However, MPE decreases with increasing sample size, as indicated in Table S3. As the number of each type of particle increases, MPE for fragment, bead and film gradually decreases. MPE for fibers does not change significantly, but MPE for rubber increases notably. This is because the number of rubber particles is a fraction of the number of fragments, as stated in Table S3. Then, there are several cases where the number of rubber particles is small, and the prediction is large, resulting in a large average MPE.

The above leave-one-out experiments correspond to training data with 99 samples, which may be impractically large in some cases. The effect of the number of training samples on KRR prediction accuracy can be found in Section S4. In general, MPE bottoms out when

the number of training samples exceeds 20, regardless of sample composition and particle type. Therefore, in this case, there is little further accuracy to be gained from having more than 20 samples in the training data, and the leave-one-out results can be seen as the typical KRR performance.

These findings suggest that KRR, if given a moderate amount of accurate training data, can estimate counts more accurately than visual sorting on samples with homogeneous composition, on heavier particles in mixed samples, and on larger samples, i.e., samples with larger particle number and total weight.

Training with Experimental Data



(a) Prediction of *Washing Machine* Data

(b) Prediction of *Interlab* Data

Figure 2: KRR Performance Summary on Real Data

Figure 2 shows the performance of KRR on two datasets obtained from experiments: *Interlab* and *Washing Machine*. Training and testing was done through leave-one-out cross-validation. Figure 2a and 2b show the predicted values plotted against the measured values. In most cases, the outcomes scatter closely around the diagonal line, which represents perfect prediction. The MPE for both datasets are mainly below 50%, with median values at 25% for *Washing Maching* and 13% for *Interlab*, shown by Figure S16. These results demonstrate that

predictions from a well-trained KRR can be more accurate than the results of traditional counting. However, there are a few large prediction errors, which could be the result of patterns undetected by KRR, or errors in the measured counts.

If the original count error is Gaussian with zero mean and constant variance, it will not affect the expected value of the model parameters.[32] In the case of microplastic quantification, it is possible for human counts to be biased, violating the zero-mean requirement, e.g., by always over-counting cotton fibers from air contamination or ignoring a type of particle thought to be non-plastic. To reduce the effect of biased measurement errors, training datasets should be as accurate as possible, for example, verified by FTIR or Raman spectroscopy and corrected to report plastic-only particles. Most studies that used such methods to confirm particle identities (plastic vs. non-plastic), showed that between 20% to 70%[8,22,30,31] of the total number of microparticles are not plastics, and for sediment samples this rate can be as high as 98%.[33] As a consequence, accurate training data is necessary for making accurate predictions.

To optimize prediction quality, a minimum training dataset size is required. The recommended training dataset size increases with the number of predicted variables. When the number of predicted quantities is small, e.g., only fiber counts were predicted for the *Washing Machine* data, a training dataset with 8 or more samples can be adequate, as suggested by Jenkins *et al.*[34] and shown in Figure S15. In cases like synthetic dataset 9, which has 4 size fractions and 5 types of particle, a training dataset of 20 samples is recommended, as demonstrated by our numerical experiments. The training samples should cover as much the data range as possible, meaning that, the samples should be representative of the full range of inputs and outcomes.

Implementation

Assessing the extent and impact of microplastic contamination requires the collection of samples at high temporal and spatial resolutions. This section briefly compares our approach

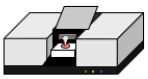
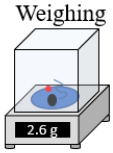
	Current Method			Proposed Method			
	Counting	Spectroscopic Analysis 	Total	Drying	Weighing 	Predicting	Total
Time /Sample	10-40 hr	12-15 hr	**20-50 hr**	4 hr - overnight	1 min	1 min	**4 hr - overnight**
Cost (USD) /Equipment	1k	150k	**150k**	3k	1-3k	1k	**5-7k**
Level of Expertise	Low	High	**High**	Low	Low	Low	**Low**
Performance	MPE is around 50%			Median MPE for fragment, fiber and bead are below or around 50%. MPE for rubber and film are above 70%.			

Figure 3: Comparison between the current and the proposed method. k stands for thousand. The time needed to manually count microplastic particles using a microscope was estimated from laboratory experiences, at around 10 to 40 hours per sample, with each sample containing around 50 to 100 particles. The time needed to analyse 100 particles using FTIR was estimated as 12 to 15 hours.[35] Predicting counts from sample weights consists of three parts: 1) drying, which ranges from 4 hours to overnight depending on the temperature used, 2) weighing, estimated as 1 minute, and 3) prediction, estimated as 1 minute for an already trained algorithm.[21,36] To this, one can add the time needed to process the required number of samples through the traditional manual counting method to generate a training dataset, i.e., an extra 10 to 40 hours per training sample. The equipment cost for the traditional method is dominated by the spectroscopic equipment, estimated as 150,000 USD. The cost for the prediction method is estimated as the total cost of the oven ($\sim 3,000$ USD), balance ($\sim 1,000$ to $3,000$ USD) and computer ($\sim 1,000$ USD). The cost for the balance can be lower for heavier samples where two decimals are enough; whereas, for lighter samples a four-decimal balance will be preferred.

to standard quantification procedures and demonstrates its power to quickly collect large amounts of data. Figure 3 shows the two quantification procedures after organic digestion, density separation, and sieving. This comparison mainly focuses on time, cost and level of expertise required. The equipment needed for traditional manual counting method costs approximately 150,000 USD. This is 30 times more than the cost of prediction based on sieved mass, which is approximately 5,000 USD to 7,000 USD in equipment for sample mass determination. Commercial laboratories charge between 1,000 to 5,000 USD for microplastic sample analysis, and around 50 to 250 times less for total suspended solid concentration measurement ($\sim$ 20 USD), which is comparable to the proposed method on determining the weight of microplastics. The level of expertise required was estimated based on the fact that the use of a spectroscopic technique such as FTIR or Raman spectroscopy requires a specialized degree often offered in graduate school; whereas, drying, weighing, and running a trained algorithm are all relatively straightforward tasks. Our proposed method can achieve better MPE than manual counting results reported by Isobe *et al.* for heavier particles in a complicated sample, and our MPE for lighter particles can be reduced to below 30% when the sample size increases.[20] In conclusion, by using machine learning to predict microplastic counts, we can reduce the experimental labour by 50% to 80%, and avoid using expensive equipment that requires expertise to operate.

We now discuss three hypothetical use cases. Graphical representations can be found in Section S6. In general, depending on the nature of different microplastic studies, extra features can be added if they are suspected to have some influence on microplastic counts. In the case of assessing environmental contamination by microplastics, as chosen here as an example, these features can include time of the year, season, longitude, latitude, distance to the nearest urban or industrial center, and so on.

In the first hypothetical use case, microplastic contamination is monitored at a single site, e.g., a lake, over the course of a year (Figure S17). If one collects several samples each month, one sample from each collection period can be counted using the traditional method

to form a training dataset. This training data would be used to train KRR. The rest of the samples would be sieved and weighed, and the microplastic counts would be estimated using the trained KRR. Here, the time of the year could be added as an extra feature.

In the second hypothetical use case, one wishes to monitor a large geographical area (Figure S18). One could perform traditional counting on samples from a small number of locations, use this data to train KRR, and then use the algorithm to quickly estimate counts at a larger number of remaining locations. In this case, location might be included as an additional feature in the training data.

The third hypothetical use case involves samples with homogeneous composition, e.g., atmospheric or washing machine effluent samples composed of mainly fibers (Table S4). As demonstrated before, KRR would perform better if the sample contains only one type of particle. Ideally, one would weigh all collected samples and choose from them several samples that depict the range of the measurements, count them using traditional manual method and use them to train KRR. The rest of the samples can be predicted using the trained model.

Environmental Significance

This study demonstrated the possibility of using regression methods for microplastic quantification based on sieved weight measurements. Because weight is the only feature considered, this method works better for pure samples made of one type of particles, or for heavier particles in mixed samples. This method is cheaper, faster and easier than current visual quantification methods. As shown in Figure 3, the new method takes less than half of the time needed for manual counting, and it only involves fundamental laboratory equipment. We therefore believe that it could be widely useful to both experts and non-experts. The accuracy of predictions from KRR are better than that attained by human counting, suggesting that this approach has the potential to accelerate experimental research on this topic

without sacrificing the quality of microplastic quantification.

Acknowledgements

This research was supported by the University of Toronto and the Natural Sciences and Engineering Research Council of Canada, Canada Research Chairs, grant no. 950-230892.

References

(1) Moore, C. J. Synthetic polymers in the marine environment: A rapidly increasing, long-term threat. *Environmental Research* **2008**, *108*, 131 – 139, The Plastic World.

(2) Wright, S. L.; Ulke, J.; Font, A.; Chan, K. L.; Kelly, F. J. Atmospheric microplastic deposition in an urban environment and an evaluation of transport. *Environment International* **2020**, *136*, 105411.

(3) Liu, F.; Olesen, K. B.; Borregaard, A. R.; Vollertsen, J. Microplastics in urban and highway stormwater retention ponds. *Science of the Total Environment* **2019**, *671*, 992–1000.

(4) Allen, S.; Allen, D.; Phoenix, V. R.; Le Roux, G.; Durántez Jiménez, P.; Simonneau, A.; Binet, S.; Galop, D. Atmospheric transport and deposition of microplastics in a remote mountain catchment. *Nature Geoscience* **2019**, *12*, 339–344.

(5) González-Pleiter, M.; Velázquez, D.; Edo, C.; Carretero, O.; Gago, J.; Barón-Sola, Á.; Hernández, L. E.; Yousef, I.; Quesada, A.; Leganés, F.; Rosal, R.; Fernández-Piñas, F. Fibers spreading worldwide: Microplastics and other anthropogenic litter in an Arctic freshwater lake. *Science of the Total Environment* **2020**, *722*, 137904.

(6) Bucci, K.; Tulio, M.; Rochman, C. M. What is known and unknown about the effects

of plastic pollution: A meta-analysis and systematic review. *Ecological Applications* **2020**, *30*.

(7) Rochman, C. M. et al. Rethinking microplastics as a diverse contaminant suite. *Environmental Toxicology and Chemistry* **2019**, *38*, 703–711.

(8) 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. *Environmental Science & Technology* **2012**, *46*, 3060–3075, PMID: 22321064.

(9) Li, J.; Liu, H.; Paul Chen, J. Microplastics in freshwater systems: A review on occurrence, environmental effects, and methods for microplastics detection. *Water Research* **2018**, *137*, 362–374.

(10) Methods for sampling and detection of microplastics in water and sediment: A critical review. *TrAC - Trends in Analytical Chemistry* **2019**, *110*, 150–159.

(11) Eerkes-Medrano, D.; Thompson, R. In *Microplastic Contamination in Aquatic Environments*; Zeng, E. Y., Ed.; Elsevier, 2018; pp 95 – 132.

(12) Sun, J.; Dai, X.; Wang, Q.; van Loosdrecht, M. C.; Ni, B.-J. Microplastics in wastewater treatment plants: Detection, occurrence and removal. *Water Research* **2019**, *152*, 21 – 37.

(13) Silva, A. B.; Bastos, A. S.; Justino, C. I.; da Costa, J. P.; Duarte, A. C.; Rocha-Santos, T. A. Microplastics in the environment: Challenges in analytical chemistry - A review. *Analytica Chimica Acta* **2018**, *1017*, 1–19.

(14) Andrady, A. L. Microplastics in the marine environment. *Marine Pollution Bulletin* **2011**, *62*, 1596 – 1605.

(15) Mai, L.; Bao, L.-J.; Shi, L.; Wong, C. S.; Zeng, E. Y. A review of methods for measuring

microplastics in aquatic environments. *Environmental Science and Pollution Research* **2018**, *25*, 11319–11332.

(16) Michel, A. P. M.; Morrison, A. E.; Preston, V. L.; Marx, C. T.; Colson, B. C.; White, H. K. Rapid Identification of Marine Plastic Debris via Spectroscopic Techniques and Machine Learning Classifiers. *Environmental Science & Technology* **2020**, *54*, 10630–10637, PMID: 32697577.

(17) Yurtsever, M.; Yurtsever, U. Use of a convolutional neural network for the classification of microbeads in urban wastewater. *Chemosphere* **2019**, *216*, 271 – 280.

(18) Lorenzo-Navarro, J.; Castrillón Santana, M.; Gómez, M.; Herrera, A.; Marín-Reyes, P. Automatic Counting and Classification of Microplastic Particles. 2018.

(19) 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. *Environmental Science & Technology* **2020**,

(20) Isobe, A. et al. An interlaboratory comparison exercise for the determination of microplastics in standard sample bottles. *Marine Pollution Bulletin* **2019**, *146*, 831 – 837.

(21) McIlwraith, H. K.; Lin, J.; Erdle, L. M.; Mallos, N.; Diamond, M. L.; Rochman, C. M. Capturing microfibers – marketed technologies reduce microfiber emissions from washing machines. *Marine Pollution Bulletin* **2019**, *139*, 40 – 45.

(22) Smyth, K.; Drake, J.; Li, Y.; Rochman, C.; Van Seters, T.; Passeport, E. Bioretention cells remove microplastics from urban stormwater. Submitted to Water Research.

(23) Olesen, K. B.; Stephansen, D. A.; van Alst, N.; Vollertsen, J. Microplastics in a stormwater pond. *Water (Switzerland)* **2019**, *11*.

(24) Lima, A.; Costa, M.; Barletta, M. Distribution patterns of microplastics within the plankton of a tropical estuary. *Environmental Research* **2014**, *132*, 146 – 155.

(25) Burns, E. E.; Boxall, A. B. Microplastics in the aquatic environment: Evidence for or against adverse impacts and major knowledge gaps. *Environmental Toxicology and Chemistry* **2018**, *37*, 2776–2796.

(26) Zhang, Y.; Kang, S.; Allen, S.; Allen, D.; Gao, T.; Sillanpää, M. Atmospheric microplastics: A review on current status and perspectives. *Earth-Science Reviews* **2020**, *203*, 103118.

(27) Murphy, K. P. *Machine Learning: A Probabilistic Perspective*; The MIT Press, 2012.

(28) El-Dereny, M.; Rashwan, N. Solving multicollinearity problem using ridge regression models. *International Journal of Contemporary Mathematical Sciences* **2011**, *6*.

(29) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V., et al. Scikit-learn: Machine learning in Python. *Journal of machine learning research* **2011**, *12*, 2825–2830.

(30) Eriksen, M.; Mason, S.; Wilson, S.; Box, C.; Zellers, A.; Edwards, W.; Farley, H.; Amato, S. Microplastic pollution in the surface waters of the Laurentian Great Lakes. *Marine Pollution Bulletin* **2013**, *77*, 177 – 182.

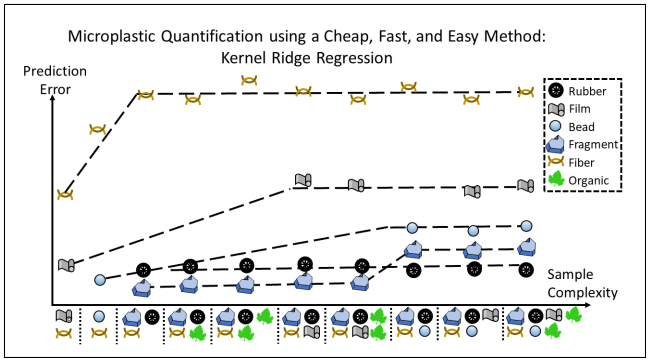
(31) Lenz, R.; Enders, K.; Stedmon, C. A.; Mackenzie, D. M.; Nielsen, T. G. A critical assessment of visual identification of marine microplastic using Raman spectroscopy for analysis improvement. *Marine Pollution Bulletin* **2015**, *100*, 82 – 91.

(32) Cawley, G.; Talbot, N.; Chapelle, O. Estimating Predictive Variances with Kernel Ridge Regression. 2005; pp 56–77.

(33) Löder, M. G. J.; Gerdts, G. In *Marine Anthropogenic Litter*; Bergmann, M., Gutow, L., Klages, M., Eds.; Springer International Publishing: Cham, 2015; pp 201–227.

(34) Jenkins, D.; Quintana-Ascencio, P. A solution to minimum sample size for regressions. *PLOS ONE* **2020**, *15*, e0229345.

(35) Primpke, S.; Lorenz, C.; Rascher-Friesenhausen, R.; Gerdts, G. An automated approach for microplastics analysis using focal plane array (FPA) FTIR microscopy and image analysis. *Anal. Methods* **2017**, *9*, 1499–1511.

(36) Rodrigues, S.; R. Almeida, C. M.; Ramos, S. Adaptation of a laboratory protocol to quantity microplastics contamination in estuarine waters. *MethodsX* **2019**, *6*, 740 – 749.

Graphical TOC Entry