



Microplastic detection and identification by Nile red staining: Towards a semi-automated, cost- and time-effective technique

Nelle Meyers^{a,b,*}, Ana I. Catarino^{a,1}, Annelies M. Declercq^{a,c,1}, Aisling Brenan^a, Lisa Devriese^a, Michiel Vandegehuchte^a, Bavo De Witte^b, Colin Janssen^d, Gert Everaert^a

^a Flanders Marine Institute (VLIZ), InnovOcean Site, Wandelaarkaai 7, 8400 Ostend, Belgium

^b Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Animal Sciences Unit - Aquatic Environment and Quality, Ankerstraat 1, 8400 Ostend, Belgium

^c Department of Animal Sciences and Aquatic Ecology, Laboratory of Aquaculture & Artemia Reference Center, Ghent University, Coupure Links 653, 9000 Gent, Belgium

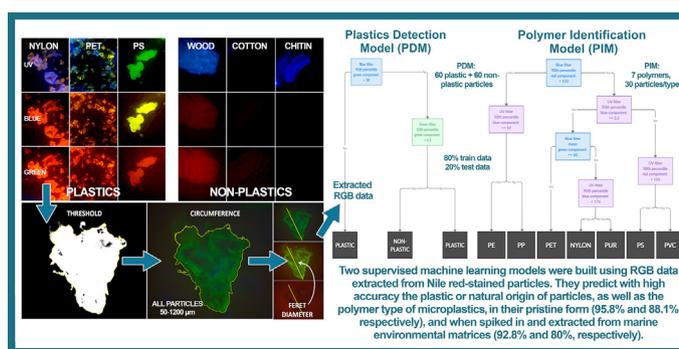
^d Department of Animal Sciences and Aquatic Ecology, GhEnToxLab, Ghent University, Coupure Links 653, 9000 Gent, Belgium



HIGHLIGHTS

- We developed a cost- and time-effective, automated method to identify microplastics.
- We present two machine learning models that use RGB color quantification.
- Method combines image analysis of fluorescent particles with classification models.
- Our developed models showed high classification accuracies of 95.8% and 88.1%.
- Method can detect and identify particles spiked in marine environmental matrices.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 27 October 2021

Received in revised form 17 January 2022

Accepted 22 January 2022

Available online 3 February 2022

Editor: Jay Gan

Keywords:

Microplastic classification

Automation

Machine learning

Nile red fluorescence

Image processing

RGB

ABSTRACT

Microplastic pollution is an issue of concern due to the accumulation rates in the marine environment combined with the limited knowledge about their abundance, distribution and associated environmental impacts. However, surveying and monitoring microplastics in the environment can be time consuming and costly. The development of cost- and time-effective methods is imperative to overcome some of the current critical bottlenecks in microplastic detection and identification, and to advance microplastics research. Here, an innovative approach for microplastic analysis is presented that combines the advantages of high-throughput screening with those of automation. The proposed approach used Red Green Blue (RGB) data extracted from photos of Nile red-fluorescently stained microplastics (50–1200 μm) to train and validate a 'Plastic Detection Model' (PDM) and a 'Polymer Identification Model' (PIM). These two supervised machine learning models predicted with high accuracy the plastic or natural origin of particles (95.8%), and the polymer types of the microplastics (88.1%). The applicability of the PDM and the PIM was demonstrated by successfully using the models to detect (92.7%) and identify (80%) plastic particles in spiked environmental samples that underwent laboratorial processing. The classification models represent a semi-automated, high-throughput and reproducible method to characterize microplastics in a straightforward, cost- and time-effective yet reliable way.

1. Introduction

Increasing global plastic production, together with poor waste management (Kershaw et al., 2019), has led to major concerns about the widespread of plastic pollution in marine ecosystems. This includes

Abbreviations: PDM, Plastics Detection Model; PIM, Polymer Identification Model.

* Corresponding author at: Flanders Marine Institute (VLIZ), InnovOceanSite, Wandelaarkaai 7, 8400 Ostend, Belgium.

E-mail address: nelle.meyers@vliz.be (N. Meyers).

¹ These authors contributed equally to this work and share first authorship.

<http://dx.doi.org/10.1016/j.scitotenv.2022.153441>

0048-9697/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

microplastics (MPs), defined as plastic particles smaller than 5 mm (Hartmann et al., 2019), which enter the environment either through direct, intentional sources (primary MPs), or indirectly through abrasion and degradation of larger plastic debris (secondary MPs). Because of their small dimensions they can be readily taken up by, amongst others, aquatic organisms (Wang et al., 2020; Avio et al., 2015). In consequence of this and due to their omnipresence in the marine environment, they pose a potential risk to populations and ecosystems (Everaert et al., 2020). Exposure to MPs assessed in laboratory settings indicates potential physical harm on species of various taxa, e.g. through inhibition of digestive processes as well as through abrasion and lesions in the gut, and by generating a misleading sensation of satiation, resulting in toxic anorexia, ultimately leading to malnutrition and starvation (Scholten, 2005), in a similar way as macroplastics do (Ahrendt et al., 2020; Bucci et al., 2020). Although still under debate (Koelmans et al., 2021), after being ingested, MPs could transfer chemicals adsorbed or added to their matrix to organisms, via desorption of co-contaminants or via leaching of chemical additives (Avio et al., 2015; Luo et al., 2019), where the direction of the chemical flux often depends on the trophic level of the organism considered (Diepens and Koelmans, 2018). Related impacts to MPs exposure include delayed ovulation and reproductive failure, as well as reduced growth rate and oxidative stress (Galloway et al., 2018).

Assessors require data on environmental concentrations of MPs to define exposure of organisms, as these in-situ concentrations are used to quantify the risk of MPs (and inform stakeholders), by comparing with safe concentrations derived from effect studies (Everaert et al., 2018). A wide range of MP detection and identification methods have been developed and applied to characterize the environmental exposure of organisms to MPs. However, to date, no routine, high-throughput, and cost-effective method is available to quantify environmental concentrations of MPs. This research has been recognized as a priority action by various international organizations (ICES, 2021). In current identification and characterization methods, a combination of physical (e.g., microscopy) and chemical (e.g. spectroscopy and thermal analysis) analyses is often used following the extraction steps for MP characterization. Each of these conventional methods has its advantages and limitations, but many are still prone to human error and/or are time-consuming and laborious, with associated high costs (Cowger et al., 2020). Moreover, the lack of method uniformity adds complexity to cross-study comparability, since harmonization of methods is imperative to provide reliable and meaningful data that allows to effectively observe both temporal and spatial trends of MP pollution. Recently, considerable efforts are being undertaken towards the automation of MP detection and identification, to improve cost- and time-effectiveness of laboratorial methodologies. Automated methods are less sensitive to human subjectivity, less labor intensive, can reduce analysis time, and are consequently often cheaper than conventional methods (Kedzierski et al., 2019; Bianco et al., 2020). Even though recent advances on automation have enabled substantial developments in the field of MPs research, most of these methodologies still require the use of expensive equipment (e.g. μ -FTIR analysis or μ -Raman spectroscopy), and often require a long computing time (Primpke et al., 2017).

New approaches are being developed using methods that combine the advantage of high-throughput screening options with the advantage of automation. These approaches aim to overcome the above mentioned bottlenecks in MP sample and data analysis, and in this way advance MPs detection research. One of the suggested methodologies for automated MP detection and identification is through predictive modeling, e.g. by using supervised machine learning techniques in combination with microscopy (Cowger et al., 2020). In this context, the use of fluorescence staining methodologies can enable the development of high-throughput, cost-effective screening methods for MP detection and identification as an alternative to presently-used MP identification methodologies (Shruti et al., 2022). The fluorescent staining of MPs using the hydrophobic dye Nile red (NR), frequently used in histology, enables the rapid screening of environmental samples for the presence of MPs using fluorescence microscopy (Maes et al., 2017; Bakir et al., 2020; Erni-Cassola et al., 2017;

de Witte et al., 2022). Based on the solvatochromic nature of NR, whose emission spectrum shifts depending on the polarity of its environment, MPs can be classified into “polar” and “hydrophobic”, based on their polymer characteristics (Maes et al., 2017). The interaction of NR with different polymers varies according to the chemical characteristics of the plastic, which has led to the successful use of NR staining of particles for polymer identification based on fluorescence (Sancataldo et al., 2020; Nel et al., 2021). However, despite this potential, so far an automated polymer classification process based on NR fluorescence has not been established.

The goal of this research was to develop a straightforward, reliable methodology to detect and identify MPs, based on a semi-automated, cost- and time-effective approach using MPs representative of the most abundantly produced plastic polymers worldwide (Geyer et al., 2017; Suaria et al., 2020), as well as natural materials with high prevalence in the marine environment. A procedure that combines Red Green Blue (RGB)-color quantification based on image analysis of NR stained particles with a supervised machine learning classification tree model was established, developed in an open source environment. To do so, a two-step model was developed: first, to predict with high accuracy whether particles are of plastic origin or of non-plastic origin, and second, to identify plastic polymer types. As proof of principle, the applicability of the developed models was demonstrated by successfully using the models to identify plastic and non-plastic particles in spiked environmental matrices (seawater and biota).

2. Materials and methods

2.1. Reference materials and staining procedure

Seven amongst the most abundantly produced plastic polymers worldwide that are commonly observed in marine environmental samples were selected (Geyer et al., 2017; Suaria et al., 2020). Besides their omnipresence and prevalence in the environment that was used as a first criterion, it was ensured a broad range of densities was covered. The selection of reference materials was complemented with ten different non-plastic, nature-based source materials (Tables A1 and A2 in Supplementary materials). To create the datasets (dataset 1 (Meyers et al., 2021a) and dataset 2 (Meyers et al., 2021b)), cryomilled (Spex Sample Prep 6875D-Freezer/Mill Dual-Chamber Cryogenic Grinder) particles of nylon, polyethylene (PE), polyethylene terephthalate (PET), polypropylene (PP), polystyrene (PS), polyurethane (PUR) and polyvinyl chloride (PVC), chitin, cotton, flax (raw and bleached), hemp (raw and bleached), silk, wood and wool (alpaca and sheep) were used. All particles were sized between 50 μ m to 1200 μ m. The plastic polymer types were confirmed using Micro-Fourier Transform Infrared (μ -FTIR) Spectroscopy (PerkinElmer FTIR spectrometer Frontier with μ -FTIR microscope Spotlight 200i, Zaventem, Belgium). All particles were analyzed with a magnification of 10 x and the detector type was liquid nitrogen-cooled mercury cadmium telluride. Beam splitter OptKBr and mid-infrared (MIR) source set the infrared spectral range of 4000–600 cm^{-1} (the PTFE (Polytetrafluoroethylene) filters that supported MP samples had high absorption in the range of 1250–1150 cm^{-1} , so this range was excluded from the analysis). All spectra were recorded in transmittance mode with a resolution of 4 cm^{-1} and an average of 64 scans per particle. Generated spectra were compared with commercial spectra libraries (Perkin Elmer) and the polymer composition of each particle was accepted for matches above 70%.

To stain particles with NR, a portion (<0.5 mg) of each reference material was selected and transferred to a labelled scintillation glass vial, containing 20 mL Milli-Q water. Samples were sonicated for 10 min (level 9, VWR Ultrasonic Cleaner USC-THD) to ensure particle dispersion. Next, the content of each vial was vacuum filtered on PTFE filters (47 mm diameter, 10 μ m pore size, Millipore Ltd.) using a Millipore manifold system (3 + 3 workstations) (Merck Millipore). Particles were subsequently stained with 1 mL NR solution (10 $\mu\text{g mL}^{-1}$ in acetone), dosed with a glass pipette, while filters still laid on the filter head (for choice of solvent, please see Fig. A2 in Supplementary materials). After 15 min, filters were

thoroughly rinsed with Milli-Q water and vacuum filtered to discard any accumulated liquid. Finally, filters were carefully transferred onto labelled glass slides, stored inside covered petri dishes and left to dry to the air in the dark. A total of six filters per reference material were made. To avoid background contamination throughout the process, all work was performed in a laminar flow cabinet, only cotton clothes were worn, no plastic materials were used in the filtration set-up and all materials were always first rinsed with Milli-Q water and air dried on a metal rack.

2.2. Selection of photo acquisition settings

To select the photo acquisition parameters to acquire images for the decision tree analyses, a dataset of images of plastic and non-plastic particles was first obtained using a fluorescence microscope (LEICA DM 1000). The microscope was equipped with a camera and a LED for fluorescence, and the software LAS Core software (Leica Application Suite version 4.13.0). Default microscope settings used were: gain: 1.0 ×; saturation: 1.50; and gamma: 0.60. Based on previous studies (Erni-Cassola et al., 2017; Prata et al., 2019; Iannilli et al., 2019), three types of fluorescence filters (Leica) were used here to obtain images of each particle type: ultraviolet (UV) (Filter System A S, Band Pass (BP) 340–380 nm), blue (Filter System I3 S, BP 450–490 nm) and green (Filter system N2.1 S, BP 515–560 nm). Images were acquired for particles exposed to each of these filters at the following exposure times: 5.22 ms, 26 ms, 80 ms and 150.8 ms for the UV filter; 2.91 ms, 4.96 ms, 12.5 ms and 38.5 ms for the blue filter, and 1.01 ms, 1.48 ms and 9.83 ms for the green filter. To avoid color distortion in acquired photos, the work was performed in a dark environment and the white balance was first set, in bright field mode, in a filter position with no particles. Then, the hue and saturation settings were set to 70 and 100 for the UV filter, 0 and 100 for the blue filter, and automatic settings were used for the green filter. Exposure times were chosen so that the lowest number of images possible would be overexposed, to avoid loss of information. To determine the best selection of parameters, the overexposure of 165 images (8 bit, grayscale) of plastic particles were analyzed using the free software ImageJ (Schindelin et al., 2012). Images of each particle acquired under each fluorescence filter at increasing exposure times (see Fig. A4 in Supplementary materials) were stack imported to ImageJ. Then, overexposure of each acquired image was identified using the histogram analysis for information loss (Glatthorn and Beckschäfer, 2014). Subsequently, the best performing exposure time per fluorescence filter was selected for further analyses.

2.3. Photo acquisition of particles using fluorescence microscopy

The selected exposure times were 26 ms for the UV filter, 2.91 ms for the blue filter, and 1.01 ms for the green filter, at a magnification of 10 × 10. To construct and validate the Plastics Detection Model (PDM), images of 60 plastic particles belonging to seven polymer types were acquired (Figs. A5 and A6 in Supplementary materials), as well as images of 60 particles belonging to ten types of non-plastic materials, and this for each of the fluorescence filters (blue, green, UV). Similarly, for the Plastics Identification Model (PIM), the same seven polymer types were used, but here 30 stained particles per polymer type were selected for image analysis: five particles were chosen randomly on each of the six PTFE-filters made per polymer. The number of particles considered ensured that both analyses had enough statistical power for a medium effect size (Cohen's $f = 0.257$, $\alpha = 0.05$, power = 0.8). The average fluorescence intensity of random particles ($n = 10$) per reference material was calculated based on the determined 'mean gray value' feature in ImageJ, as well as the average fluorescence intensity of the plastics ($n = 70$) and of the non-plastics ($n = 70$). To obtain the fluorescence intensity of each particle, particle images taken under the blue filter at an exposure time of 2.91 ms were imported into ImageJ and converted from RGB to grayscale (8 bit) images for processing. Next, a region of interest (ROI) that included at least 50% of the photographed particle was selected, and the mean gray value of all pixels situated within the selected area was calculated. The mean gray

value of the background was calculated following the same criteria. Particle fluorescence intensity was subsequently determined by subtracting the mean gray value of the background from the mean gray value of the selected particle ROIs in each photo.

2.4. Decision tree models for plastics prediction and identification

Two separate decision tree models were created: one decision tree for the automatic classification of the plastic particles versus non-plastic particles (PDM), and one decision tree for the classification of the seven plastic polymers (PIM). The created models can be used to predict the class to which a targeted particle of unknown origin belongs to: hence, whether it is of plastic or non-plastic origin, and once classified as plastic, classify particles according to their polymer type. Two datasets were needed: one for training and validating the PDM (dataset 1), and one for the PIM (dataset 2). Dataset 1, used to train and validate the PDM, contained RGB statistics calculated through image analysis of 60 plastic and 60 non-plastic particles. Dataset 2, used to train and validate the PIM, contained RGB statistics from 30 particles for each of the seven plastic polymers.

To obtain the RGB statistics of each dataset (dataset 1 and 2), the respective series of three images per particle were first analyzed, i.e., one under each of the three fluorescence filters (see Section 2.3), in an automated way using a macro in ImageJ (Meyers, 2021). First, each series of images was imported in stack into ImageJ and converted to 8-bit grayscale. Next, a color threshold was set manually to segment the grayscale image into the particle of interest and the background, after which a selection of the particle was created. Then, for each particle above a specified min. maximum Feret diameter, the RGB (Red-Green-Blue) values, which define color intensity as an integer between 0 and 255, were extracted from the pixels laying along the maximum Feret diameter of that particle. Following this, the RGB statistics were calculated in R (R Core Team, 2020) as the 10th, 50th and 90th percentile as well as the mean value, for the red, green, and blue color component of each particle, and were automatically exported as a dataset. In this way, the two separate datasets were built, each compiling all percentiles and means per color component of all relevant particles, for images acquired under each of the three fluorescence filters (Figs. A5 and A6 in Supplementary materials).

For the PDM, to infer the decision rules from the training data and to discriminate plastic particles from non-plastic particles, four fifth of dataset 1 (i.e. 96 particles) was randomly selected to serve as training data (training dataset 1) and the remaining one fifth (i.e. 24 particles) was kept to be used as independent validation data (validation dataset 1), based on the Pareto principle ratio (Dunford et al., 2014). Accordingly, for the PIM, dataset 2 was split into training dataset 2 (i.e. 168 particles) and validation dataset 2 (i.e. 42 particles). Training dataset 1, used to construct the PDM, consisted for each of the 96 training particles (48 plastic and 48 non-plastic particles) of the μ -FTIR confirmed particle type (plastic/non-plastic), and the RGB statistics along the maximum Feret diameter collected under the three different fluorescence filters. Training dataset 2, used to construct the PIM, consisted of the same information (polymer type for μ -FTIR analysis), but extracted from each of the 168 plastic particles (24 particles / polymer * 7 polymer types). Both classification models were built through recursive binary splitting of the respective training datasets, using simple decision rules inferred from the emission spectra features of the NR-stained particles.

The CART (Classification and Regression Tree) supervised machine learning algorithm was used to generate both decision trees (Meyers, 2021) using the Gini coefficient as a computational metric to establish the best rule to split the spectral data each time into more refined categories, until mutually exclusive categories were formed at the terminal nodes. While developing the PDM, to avoid overfitting of the data, the minimum amount of records that must exist in a node in order for a split to be attempted was set to 8, the minimum amount of records that end up in an end node was set to 3 and any splits that did not improve the model fit by 0.02 were discarded. For the PIM, the minimum amount of records that must exist in a node in order for a split to be attempted was set to 39, the

minimum amount of records that end up in an end node was set at 13, and any splits that did not improve the model fit by 0.02 were discarded.

Both models were independently validated by their respective validation datasets through appropriate procedures (Witten et al., 2011). Model fit of the PDM was assessed based on the associated confusion matrix, constructed using validation dataset 1, in which known plastic and non-plastic particle identification was compared with the predicted plastic and non-plastic particle identification. From that matrix, the percentage of correctly classified instances (CCI %) was calculated as well as a Cohen's kappa statistic (κ) (see Table A3 in Supplementary materials for the terminology of the generated confusion matrices). In ecology, typical thresholds to accept a decision tree model is a CCI % of minimum 70% and a κ of at least 0.4 (de Troyer et al., 2020; Goethals et al., 2007), and κ -values of 0.80–0.90 generally indicate 'strong' and >0.90 'almost perfect' levels of agreement (McHugh, 2012). To assess model fit of the PIM, a confusion matrix where known polymers identification was compared with the predicted polymer identification was constructed using validation dataset 2, and associated evaluation metrics CCI % and κ were calculated as was done for the PDM. Both models were constructed in R using the 'rpart' package (Therneau et al., 2015), and validated using the 'caret' package (Kuhn, 2008).

To determine significant differences between groups of classified particles in both decision trees models, an analysis of similarity (ANOSIM) with 9999 Monte Carlo permutations using the Vegan (Dixon, 2003) and ggplot2 packages (Wickham, 2007) was performed in R. Here, the R-value represents the ratio of dissimilarities between particles within a reference material type (plastic/non-plastic) to the dissimilarities between particles of different reference material types for the PDM, and the ratio of dissimilarities between particles within a plastic polymer type to the dissimilarities between particles of different plastic polymer types for the PIM. The ANOSIM is a robust test as it does not assume equal group variances. Subsequently, to visualize classification performance of the models, a non-metric multidimensional scaling (NMDS) using the Bray-Curtis similarity index was performed for each of the two training datasets. To test whether certain polymer types show a greater variability in fluorescent coloration than others, a permutation test was also performed. Lastly, the varImp function of the caret package in R was used to assess the contribution of each fluorescence filter in the classification procedure.

2.5. Applicability of the models to identify particles from spiked environmental samples

After validating the models, the ability of the PMD and the PIM to detect and identify a variety of MP types and non-plastics that 1) had undergone laboratory processing, and that 2) were spiked in and extracted from biota and seawater samples, was tested.

2.5.1. H₂O₂-treated MPs

Reference MPs were submerged in H₂O₂ (33%, 10 mL) for 24 h, in a glass beaker covered with aluminum foil. Next, the solution was filtered and MP were NR-stained as described under 'Section 2.1. Reference materials and staining procedure'. MP particles used were PE, PET, PP, PS and PVC (commercial origin). Background measures were taken and a quality control was performed (see Method details A1 in Supplementary materials).

2.5.2. Mussel matrix

Commercial mussels, bought in a Belgian storehouse on 11 October 2021 and originating from Zeeland (The Netherlands), as well as non-commercial mussels (collected at a breakwater in Oostende, Belgium; N 51°0.241; E 2°0.930) were used as matrix to spike reference MPs into (20 g wet weight mussel tissue/sample). Four different polymer types were used for spiking: PET (Carat), PP (Carat), PS (Carat), and PVC (two types: commercial and Carat). The particle size ranged from 250 to 1000 μ m. A spatula tip of each MP type was added to three samples of both commercial and non-commercial mussels.

2.5.3. Seawater matrix

Seawater was collected from the harbor in Oostende (N 51°0.236, E 2°0.982) to prepare three seawater samples of 200 mL each in glass beakers, covered with aluminum foil. Next, each sample was spiked with MPs of various types: nylon, PE, PET, PS, PP, and PVC (commercial), and with various types of natural material: chitin, cotton, flax raw, silk and sheep wool (Tables A1 and A2 in Supplementary materials).

2.5.4. Sample analysis

After the H₂O₂-treatment of pristine MPs, and a two-step digestion in case of the spiked mussel and seawater samples (see Method details A1 and Method details A2 in Supplementary materials), the previously described procedure for particle staining was followed, and an additional μ -FTIR analysis for polymer identification validation was performed. Image series (blue – green – UV filter) of the extracted particles were acquired using the same fluorescence microscope settings as the ones used to build the models, under the exact same conditions (see Section 2.3). Next, an RGB-statistics dataset derived from the analyzed particles was built (as under 2.4., without subdivision into training and test datasets) and used to test both models.

3. Results

A two-step semi-automated classification method that allows to distinguish plastic from non-plastic particles (50–1200 μ m) and further identifies and discriminates between seven plastic polymers was developed (process diagram Fig. A1 in Supplementary materials). The method relies on image analysis of NR-stained particles combined with a modeling methodology based on a decision tree classification to distinguish plastic from non-plastic particles (Plastic Detection Model, PDM), and subsequently for plastic polymer identification (Polymer Identification Model, PIM).

3.1. Plastic Detection Model: identification of polymer-based material versus natural material

Using the PDM, plastics could be distinguished from non-plastics with an accuracy of 95.8% (here expressed as the percentage of correctly classified instances (CCI %); $n = 24$), a Cohen's kappa statistic (κ) of 0.92 was obtained (Table 1). Only two decision rules were generated to distinguish plastics from non-plastics (Fig. 1). The first decision rule of the PDM dealt with the 90th percentile of the green component (G) of the RGB spectrum acquired from particles photographed under the blue filter. This first decision rule indicates that particles with values smaller than 58 are to be classified as plastics. The second decision rule of the PDM dealt with the 50th percentile of the green component (G) of the RGB spectrum of particles photographed under the green filter. In this second decision rule, particles with values lower than 6.5 were classified as non-plastics while higher values were classified as plastics. Based on the model's variable importance evaluation (Table A4 in Supplementary materials), the blue filter was more important than the green filter for distinguishing plastic from non-plastic particles, and both were more important than the UV filter.

The CCI % calculated from the confusion matrix (Table 2) obtained using validation dataset 1 independently evaluated the accuracy of the PDM classification by comparing the actual target materials with the materials predicted by the model. For the PDM based on training dataset 1, all 12

Table 1

Evaluation metrics Plastic Detection Model (PDM) and Polymer Identification Model (PIM). Evaluation metrics of the PDM and the PIM based on training datasets 1 and 2, calculated from the associated confusion matrices using validation datasets 1 and 2, respectively.

	PDM	PIM
Total number of tested particles	24	42
Number of correctly classified particles (CCI %)	23 (95.8%)	37 (88.1%)
Number of wrongly classified particles	1 (4.2%)	5 (11.9%)
Kappa statistic (κ)	0.92	0.86

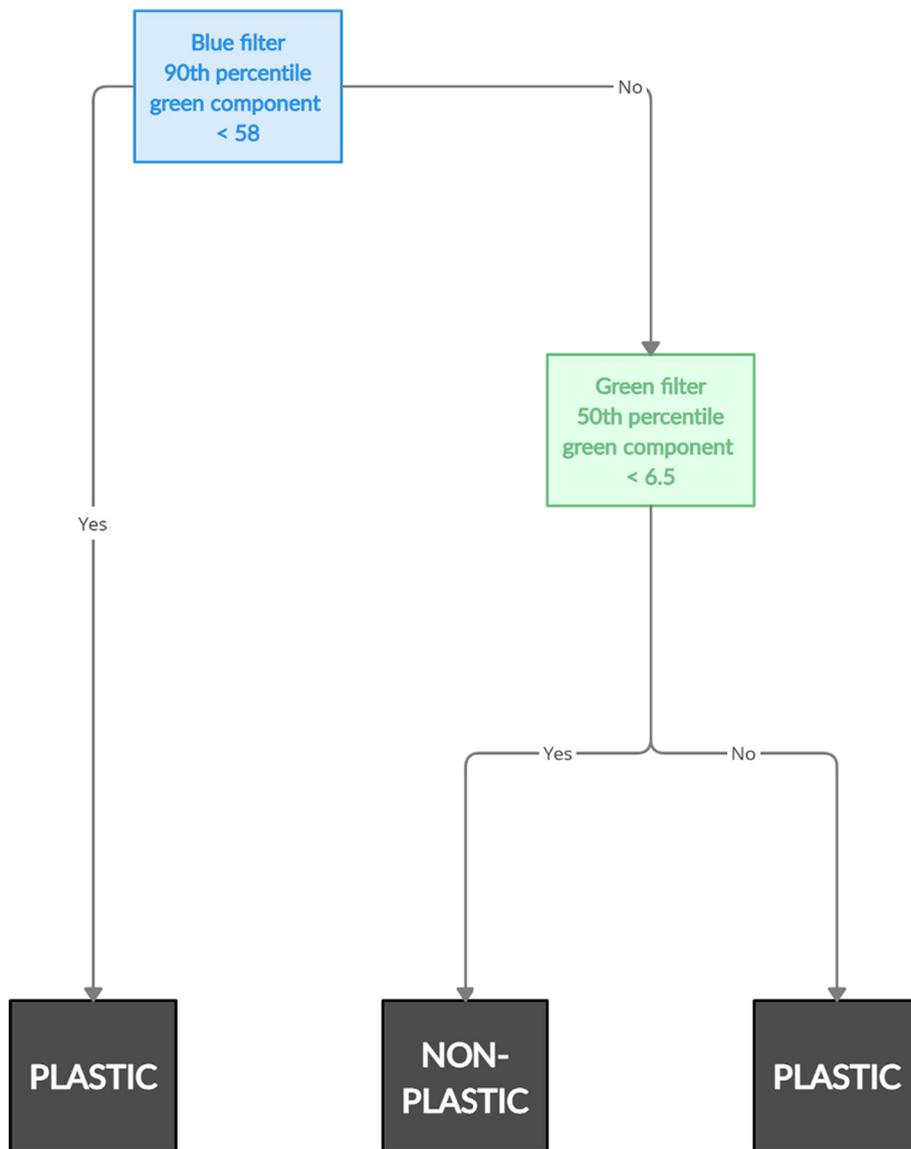


Fig. 1. Plastic Detection Model (PDM). The PDM, a decision tree model based on RGB statistics derived from RGB color values extracted from frequently observed particles in the environment: NR-stained plastic and non-plastic material particles (training dataset 1).

particles from validation dataset 1 belonging to ten different types of non-plastic reference materials were correctly identified by the algorithm as non-plastics. Eleven out of 12 plastic particles were true positives and thus correctly categorized within the plastics class. One false negative was detected because a PP particle was wrongly classified and predicted as being of non-plastic origin. Fluorescence intensity under the blue filter was lower for the PP polymer ($16.1 \pm 4.4, n = 10$) compared to other, more polar polymers such as nylon ($54.2 \pm 11.4, n = 10$), PET ($65.4 \pm$

Table 2

Plastic Detection Model (PDM) confusion matrix. Confusion matrix to assess classification performance of the PDM based on training dataset 1, generated using validation dataset 1. Each column represents the particles in an actual material type class, while each row represents the particles in a predicted material type class. All correct model predictions of whether a particle is of plastic or non-plastic origin are observed on the diagonal of the matrix, while prediction errors are represented outside the diagonal.

Material	Non-plastic	Plastic
Non-plastic	12	1
Plastic	0	11

$12.6, n = 10$), PUR ($40.2 \pm 9.5, n = 10$) and PVC ($114.6 \pm 14.2, n = 10$), which can be linked to a darker particle coloration and consequently distinct RGB statistics. Fluorescence intensity of chitin, wood and silk was moderate to low ($5.8 \pm 8.1, n = 6$; $10.8 \pm 5.9, n = 6$; and $7.3 \pm 8.9, n = 6$ respectively), while very weak fluorescence or an absence of fluorescence was observed for the organic fibers alpaca wool ($-5.4 \pm 1.5, n = 6$), sheep wool ($-1.4 \pm 1.8, n = 6$), cotton ($-4.0 \pm 0.6, n = 6$), flax bleached ($0.6 \pm 4.9, n = 6$), flax raw ($-3.0 \pm 2.5, n = 6$), hemp bleached ($-0.8 \pm 1.5, n = 6$), and hemp raw ($2.9 \pm 2.3, n = 6$). Overall, non-plastic particles were observed to be less fluorescent than plastic particles (respectively $1.3 \pm 6.6, n = 60$ and $58.7 \pm 8.6, n = 60, p < 0.01$). Polyethylene particles show similar fluorescence intensities ($19.7 \pm 4.5, n = 10$) compared to PP particles ($16.1 \pm 4.4, n = 10$), and are similarly fluorescent as can be observed in the confusion matrix and on the NMDS plot of the PIM (see Section 3.2). The PDM recognizes the PE particles as being of plastic origin, while a single PP particle was misclassified as plastic. The visualization of the classification performance of the PDM, performed by means of a non-metric multidimensional scaling (NMDS) using the Bray-Curtis similarity index on the untransformed training dataset 1, showed a clear segregation of plastic versus non-plastic materials (Fig. 2) (stress-value NMDS ordination = 0.072). The significant

difference between groups ($R = 0.59, P < 0.01$) was confirmed by means of an analysis of similarity (ANOSIM).

3.2. Polymer Identification Model: identification of plastic polymers

The CART (Classification and Regression Tree) algorithm allowed to categorize the polymers into seven polymer classes commonly observed in marine litter (nylon, PE, PET, PP, PS, PUR and PVC). With the PIM, the MP's polymer type was assessed with an accuracy of 88.1% (here expressed as CCI %), obtaining a κ of 0.86 ($P < 0.01$) (Table 1). To do so, the CART algorithm used six decision rules applied to spectral data extracted from the UV and blue filter images, as is visualized in the PIM (Fig. 3). The root node dealt with the 90th percentile of the red color (R) component of the RGB spectrum of particles photographed under the blue filter. This allowed for the separation of plastic polymers for which these percentile values were lower than 103, which represent the PE and PP plastics. All other considered plastic polymers had values higher than 103. Polypropylene was subsequently separated from PE at the second level of the PIM based on a decision rule established using the 50th percentile value of the blue component (B) of the RGB spectrum of particles photographed under the UV filter, with values lower than 19. In the following step, also at the second level of the PIM, PVC and PS plastics were separated from nylon, PET and PUR plastics with 90th percentile values of the blue component (B) of the RGB spectrum of particles photographed under the UV filter lower than 3.3. At the third PIM level, PS plastics were distinguished from PVC plastics with 50th percentile values of the red component (R) of the RGB spectrum of particles photographed under UV filter lower than 155. In addition, PET plastics were isolated from nylon and PUR plastics with mean values of the green component (G) of the RGB spectrum of particles photographed under the blue filter higher than or equal to 60. In the last level of the PIM, nylon plastics were separated from the similarly fluorescent polyurethane plastics with 90th percentile values of the blue component (B) of the RGB spectrum of particles photographed under the UV filter lower than 176. Based on the model's variable importance evaluation

(Table A5 in Supplementary materials), the UV filter contributed more to the plastic polymer classification than the blue filter, and both filters were more important than the green filter.

The obtained CCI and κ indicate a close agreement between the predicted identity of the polymers and their known identity (Table 3). Of the 42 particles used to validate the model, only 11.9% of the particles was misclassified. Despite the high overall accuracy, the identification performance was not the same for all polymers. For example, PET, PS and PVC plastics all showed a CCI % of 100.0%, Nylon, PE and PUR plastics showed a slightly lower CCI % of 83.3% and PP obtained a CCI % of 66.7%. Confusion/misclassification mostly occurred between PP and PE, between PUR and nylon, and between PE and nylon, with misclassifications of 7.1%, 2.4%, and 2.4%, respectively. The two-dimensional NMDS plot (Fig. 4) for training dataset 2 confirms the fluorescence resemblance within the plastic polymer classes (ANOSIM, $R = 0.70, P < 0.01$), where similarly fluorescing particles are grouped together while particles with low similarity are distributed further from each other (stress-value = 0.048). Based on fluorescent coloration, six groups can be distinguished (Fig. 4): PVC, PS, PET, Nylon, PUR, and PE and PP. A clear separation between PVC and PS particles is visible. Polystyrene and PET also differ from each other, as well as PET and PUR, and PET and Nylon. Nylon and PUR overlap slightly, while PP and PE show a somewhat larger overlap. PVC particles show a low within-polymer fluorescence variability compared to nylon, PET, PS, PE and PP ($P < 0.01$), indicating the particles of this polymer type fluoresce in a very similar way.

3.3. Applicability of the models to identify particles from spiked environmental samples

Overall, 65 out of 70 particles (92.8%) were correctly identified by the PDM (Table 4), where 88.6% of all plastics and 97.1% of all non-plastics were correctly identified. The PIM could identify the polymer type of 28 out of 35 plastics (80%) accurately. When considering the H_2O_2 -treated particles only, 90% of all particles was correctly classified as being plastic or

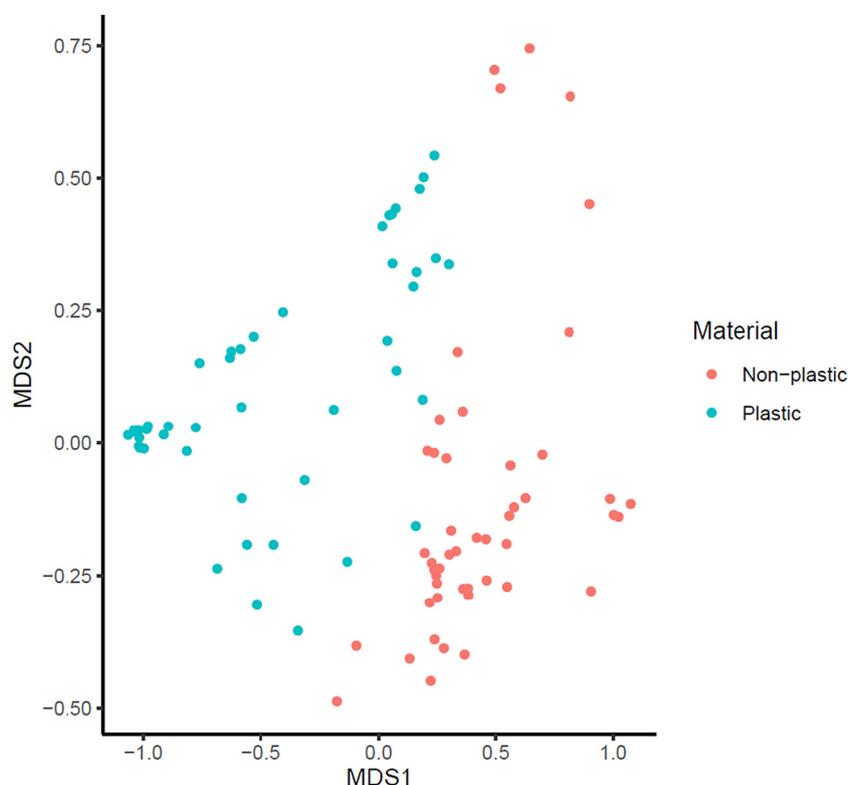


Fig. 2. NMDS plot of the Plastic Detection Model (PDM). A non-metric multidimensional scaling (NMDS) plot of training dataset 1, based on the Bray-Curtis similarity index, to visualize the classification performance of the PDM.

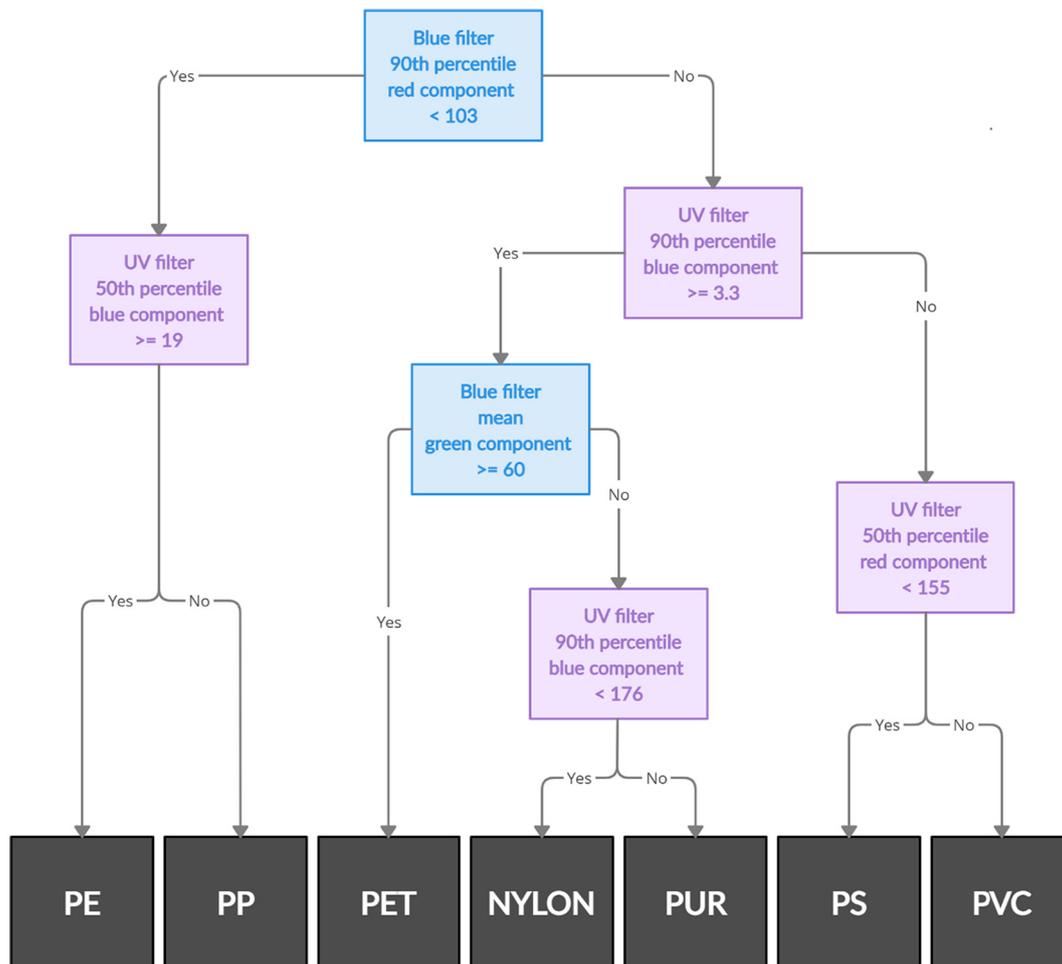


Fig. 3. The Polymer Identification Model (PIM). The PIM, a decision tree model based on RGB statistics derived from RGB color values extracted from NR-stained microplastics of frequently observed polymers in the environment (training dataset 2).

non-plastic. More specifically, 80% of the plastic particles were identified as being plastic, and a 100% accuracy was obtained for detecting non-plastics. Indeed, one PP particle was misidentified as being of non-plastic origin (Table A7 in Supplementary materials). The PIM obtained an accuracy of 80% for all tested particles: one PET particle was misclassified as nylon. Of all particles spiked in mussel tissue matrices, 96.7% was identified correctly by the PDM. Here, all tested plastics were identified correctly as well as 93.3% of all tested non-plastics, where one mussel matrix particle was mistaken for plastic. The PIM was able to identify the polymer type of 86.7% of all plastics, two PP particles were inaccurately identified as PE

Table 3

Polymer Identification Model (PIM) confusion matrix. Confusion matrix to assess classification performance of the PIM based on training dataset 2, generated using validation dataset 2. Each column represents the plastic particles in an actual polymer type class, while each row represents the plastic particles in a predicted material polymer class. All correct model predictions of a particle's polymer origin are visible on the diagonal of the matrix, while prediction errors are represented outside the diagonal.

Polymer	Nylon	PE	PET	PP	PS	PUR	PVC
Nylon	5	0	0	0	0	1	0
PE	1	5	0	2	0	0	0
PET	0	0	6	0	0	0	0
PP	0	1	0	4	0	0	0
PS	0	0	0	0	6	0	0
PUR	0	0	0	0	0	5	0
PVC	0	0	0	0	0	0	6

particles. Similarly, of all particles spiked in seawater samples, 92.9% was classified correctly: 88.6% and 100% of all plastics and non-plastics, respectively. Two PP particles were not classified as plastics. The PIM obtained an accuracy of 73.3%, where two PE particles were mistakenly identified as PP, a nylon particle as PUR and a PET particle as nylon (Table 4 & Table A7 in Supplementary materials).

4. Discussion

4.1. Microplastic identification using Nile red fluorescence imaging and classification models

An innovative, semi-automated methodology to detect and identify MP particles (50–1200 μm) was developed based on quantification of fluorescent coloration after NR staining. The approach involves a predictive modeling tool based on the extraction of RGB-derived statistics from images of fluorescent NR-dyed particles acquired under exposure to microscope UV, blue and green fluorescence filters. Data extracted from image analysis was used to build two easily interpretable classification models through recursive binary splitting using simple decision rules, inferred from emission spectra features. The PDM was built to distinguish plastic from non-plastic particles, while the PIM identified the polymer type of a given plastic particle. The obtained κ-values for the PDM and PIM of respectively 0.92 and 0.86, combined with the obtained CCI % values of respectively 95.8% and 88.1%, indicate a strong (PIM) to excellent (PDM) classification performance of the constructed models, demonstrating the reliability of the models (Everaert et al., 2011).

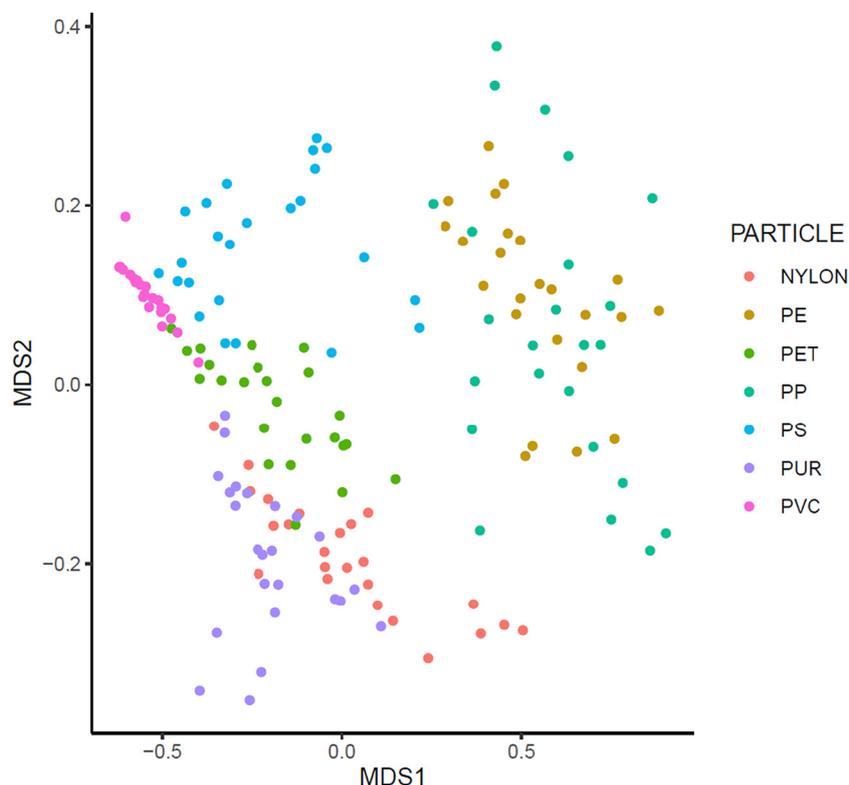


Fig. 4. NMDS plot of the Polymer Identification Model (PIM). A non-metric multidimensional scaling (NMDS) plot of training dataset 2, based on the Bray-Curtis similarity index, to visualize the classification performance of the PIM.

The choice of solvent affects the fluorescence of NR stained particles (Maes et al., 2017). In our work, we have selected acetone as the stain solvent, as in the preparatory phase of the qualitative analysis (see Figs. A2 and A3 in Supplementary materials), the use of methanol was observed not to add any particular advantages to the use of acetone, the latter being a much less hazardous solvent. Earlier research had demonstrated that ethanol (Prata et al., 2019) or methanol when used as alternative solvents to acetone (NR solution), also allow to distinguish MPs from non-plastic particles. The use of DMSO was also further tested, as well as exposing plastics to warmer solutions for improved dye incorporation due to plastics thermal expansion (Lv et al., 2019). However, in our work no improved fluorescence was observed when compared to NR dissolved in acetone or to plastics exposed at room temperature (Fig. A2 in Supplementary materials). As such, acetone was considered the optimal NR solvent for the development of the classification models.

Table 4

Applicability of the models to identify particles from spiked environmental samples. The ability of the PDM and the PIM was tested to detect and identify a variety of particles: particles that had undergone a H₂O₂-treatment (10 particles), and particles that were spiked in and extracted from mussel (30 particles) and seawater samples (30 particles) (each time of both plastic and non-plastic origin). The overall accuracy of the PDM as well as the detection accuracy of plastic and non-plastics particles are presented for all tested particles (70 particles in total). The accuracy of the PIM is presented as well in the case of plastic particles (35 particles).

	Accuracy [%]			PIM
	PDM		PIM	
	Overall	ID plastics		
H ₂ O ₂ -treatment	90.0	80.0	100.0	80.0
Mussel matrix	96.7	100.0	93.3	86.7
Seawater matrix	90.0	80.0	100.0	73.3
Overall	92.8	88.6	97.1	80.0

The conducted classification analyses were successfully applied to particles with a size range of 50–1200 μm. Currently, most MPs monitoring and survey programs are constrained by the sampling methodology in terms of particle size (e.g. manta net mesh size often >300 μm). However, available data on MPs quantification from field samples (>300 μm) contrasts with size ranges (Shim et al., 2016; Beiras and Schönemann, 2020) used in most ecotoxicological assessments (<100 μm) (Beiras and Schönemann, 2020), which is also predicted to be the larger proportion of particles in the field (Kooi et al., 2021). This method thus addresses the current data gap on the distribution and abundance of smaller size ranged MPs, in agreement with previous reports that have demonstrated the potential to use NR fluorescence analysis for particles down to a few μm or even <1 μm (Maes et al., 2017; Erni-Cassola et al., 2017; Prata et al., 2019; Iannilli et al., 2019).

The workflow developed in this work (Fig. A1 in Supplementary materials) can be reproduced with any available equipment for fluorescence observation of NR-stained particles, which should consider fixed settings for photo acquisition and analysis. The classification performance of both models for particles photographed with a different fluorescence microscope remains to be tested, but the process to build a training dataset and validate the decision tree is semi-automated and applicable to any models adjusted for microscope-specific observations. When building training and validation datasets, image acquisition settings for the different fluorescence filters may need to be adapted for other fluorescence microscopes, but settings should be manually kept constant to avoid variation between images and improve comparability. Furthermore, the NR concentration (Shim et al., 2016) as well as NR adsorption time (Lv et al., 2019) should be kept as constant as possible to avoid large variations in fluorescence intensity and coloration of the particles, and consequently to ensure good model performance.

4.1.1. Plastic Detection Model: particle classification into plastic and non-plastic

The PDM incorporated decision rules based on color components of the blue filter and the green filter. While the blue and green filter were of great

importance, the UV filter had a lower discriminative power and was not integrated in the final model. This confirms visual observations which demonstrated that plastic and non-plastic materials appeared similarly fluorescent under the UV filter. Other studies support the exclusion of the UV filter in the PDM, as natural materials stained with NR dissolved in acetone have already been reported to fluoresce under UV light (Tamminga, 2017). Contrarily, one study which used ethanol instead of acetone for the NR solution, indicated that UV light can distinguish some plastics from natural organic matter, while the organic matter was fluorescent under blue light using an orange filter (Prata et al., 2019). Nile red staining has proven efficient at distinguishing MPs from non-plastic materials such as amphipod carapaces, algae, seaweeds, wood, feathers, mollusk shells, chalk and sand particles using only blue light microscopy filters, or in combination with orange filters (Maes et al., 2017; Shim et al., 2016; Sturm et al., 2021). This enables the identification of MPs extracted from environmental samples as these, even after digestion and further processing, often contain residual biological material and sediment, which is known to interfere in other MPs identification and characterization methods (Ruggero et al., 2020). Even though previous assessments indicate that NR has low affinity for natural materials frequently observed in the marine environment (Maes et al., 2017), other studies indicated that debris such as chitin-based and natural fibers can still be stained and show fluorescence, leading to a potential misidentification of the particles present (Stanton et al., 2019). To overcome overestimation of the number of MPs identified using fluorescence analysis, authors have suggested co-staining techniques using NR combined with a dye with high affinity for biological samples (Stanton et al., 2019; Maxwell et al., 2020), but this approach would imply further complexity in the sample processing and analysis. Here, the distinction of plastics from non-plastic materials was achieved by using one single dye, facilitating the laboratory procedures, and by using a simple machine learning based model that classifies materials based on their fluorescent coloration.

4.1.2. Polymer Identification Model: classification of plastic particles into polymer groups

The seven polymers considered in the models are amongst the most commonly observed MPs in field samples (Geyer et al., 2017; Suaria et al., 2020). The fluorescence exhibited by NR-stained plastic particles is polymer dependent, but some polymers can show a similar irradiation behavior, even under multiple fluorescence microscopy filters. The model performance of the PIM demonstrated for instance that during the classification procedure, mismatches occurred between PP and PE particles. For a defined excitation wavelength, the Stokes shift in the emission wavelength is related to the size and chemical characteristics of the polymer's chromophores (Brahme, 2014). As PE and PP are both polyolefins, a class of non-polar hydrocarbons, they have a similar chemical structure. These polymers cluster together in the results (Fig. 4), indicating a similarity with respect to fluorescence as well. Polystyrene is another non-polar hydrocarbon polymer, but this polymer carries aromatic rings, which may explain the difference in fluorescence from other analyzed polyolefins (PE, PP). Nylon, PET, and PUR are relatively more polar polymers due to their carbonyl groups. In the results, particles of these plastics showed similarities in fluorescence classification and grouping (Figs. 3 and 4) with a shift to longer emissions wavelengths compared to PE, PP and PS. Similarly, the polymer PVC has a higher polar nature, but its polarity is caused by CCl bonds instead of CO bonds, which is translated in a difference in fluorescence compared to the aforementioned polar polymers.

The PIM with the best fit used six decision rules based on color components of both the UV and blue fluorescence filters. The RGB data under the UV filter had the largest discriminative power for identifying polymer types. The green filter was not incorporated in the model as it was of less importance for the polymer classification compared to the other two fluorescence filters. Out of the six nodes in the plastic polymers identification model, four had decision rules based on the UV filter. Opposite to what was observed in the PDM, the UV filter is highly important for the classification performance of the PIM. Under a UV filter, NR-stained particles

exhibit a wide diversity of colors according to their polymeric identity (Fig. A5 in Supplementary materials) compared to the colors exhibited under the blue or green filters (visual observations). Visual examinations done by experienced observers can potentially be used to identify the most commonly observed MPs. However, identification based on observations with the human eye is prone to human errors, and not accessible to all users (e.g., due to color blindness). The method here uses RGB color information acquired through image analysis using standardized photo acquisition parameters, which enables obtaining unbiased data and information. Fluorescence microscopy images of others studies confirm the findings that polymers fluoresce in a variety of colors when irradiated with UV light (Prata et al., 2019).

4.1.3. Applicability of the models to identify particles from spiked environmental samples

The model performance of the PIM and PDM was assessed for an additional 70 particles in laboratory processed and spiked samples, to emulate environmental samples. The results of these preliminary tests indicate a good model performance of both the PDM and PIM for most MPs (accuracy of 92.8% and 80%, respectively). A few inaccurate classifications were noted, primarily for PP particles that were not recognized as plastics (3 out of 6 particles), and PP particles that were identified as PE particles, or the other way around (4 out of 9 particles) (see Table A7 in Supplementary materials). The results indicate that overall the fluorescent coloration is not prone to matrix effects and is not affected by digestive reagents, and that the models are robust enough to handle variability in fluorescent coloration. Based on these exploratory results, the method appears promising for the implementation of large scale routine analyses of field samples. More into-depth research is however advised to support these preliminary results. For example, the application of these models for naturally weathered MPs from real marine environmental samples needs to be further investigated. However, the flexibility enabled by the use of the classification methodology allows for possible necessary adjustments and expansion of the RGB training datasets to maintain good model performance. For example, the number of polymers used to develop the PIM should not limit future users in adding new polymer types or other materials.

4.2. Advancements in automated microplastic detection and identification

The models developed and evaluated in the current work are a reliable methodology for MPs identification, with a high time efficiency. The performance of the models for MP detection and identification had a similar accuracy to μ -FTIR polymer identification reported accuracy, i.e. ranged between 80 and 95.8% between all models when compared to $96 \pm 7\%$ for μ -FTIR (Therneau et al., 2015). The use of FTIR for MPs identification is a reliable and well-established method based on the analysis of the chemical structure of a polymer molecule and its associated IR spectrum. However, the working time per particle is quite high and increases quickly with particle number (often over 20 h per filter) (Veerasingam et al., 2021). Next to an increased analysis time for μ -FTIR-based analyses, the accuracy of this methodology decreases with decreasing MP size, due to the inclusion of surrounding spectral signals, and with increased exposure towards UV induced photodegradation, thermal degradation, and biodegradation due to alteration of the pristine polymer composition (Maes et al., 2017). Furthermore, due to the need for visual presorting, the chance of sample contamination or MP loss increases. The use of μ -Raman spectroscopy, a powerful but rather expensive and labor-intensive technique for MP analysis, may also introduce visual bias when analyzing small MPs (Erni-Cassola et al., 2017). Three factors determine the speed of processing a sample using the developed models presented here: 1) particle load per PTFE-filter, 2) the number of image series (blue – green – UV filter) that needs to be taken to cover all particles on a filter, and 3) the computational time needed to apply the knowledge rules of both models. Based on our experience in the lab, the most important factor is the number of image series needed, which is determined by the particle load and the homogeneity of their distribution on the filter. Typically, it takes 2.5 min to scan a filter

for particles and to acquire one image series. Predicting the origin of the photographed particles using the models takes around 5 min per image series, with a negligible increase in processing time with an increase in particle number (Table A6 in Supplementary materials). For samples potentially containing a larger number of plastics, such as samples taken by Manta nets where large volumes of water are filtered, multiple pictures are required per filter, which increases the processing time. In the case of more heavily loaded filters, if 30 image series can cover all MPs present on the filter, the overall analysis is estimated to take 80 min ($30 \times 2.5 \text{ min} + 5 \text{ min}$). Still, the automated classification analysis of a sample containing 20 particles is over three times faster compared to μ -FTIR analysis when each particle needs to be photographed separately (Table A6 in Supplementary materials, 55 min vs. 170 min), and around six times faster when the particles can be captured in 10 image series (30 min vs. 170 min). This drastic decrease in processing time implies a reduction in costs invested in MPs analysis, such as staff time and computer power. When using a fluorescence stereomicroscope, analysis time can be reduced even more by acquiring whole filter images instead of separate images per particle. Here, analysis time can be brought down to around 13 min per sample of 20 MPs. In this way, the proposed models tackle the above-mentioned limitations of μ -FTIR-based MP analysis. To obtain optimal model performance, particles should however not greatly overlap during image acquisition. Therefore, for samples with an expected high content of MPs (>150 MPs), it is advisable to dilute over multiple PTFE-filters.

Automated approaches for MPs identification and characterization are a rapidly evolving field, which aims to contribute to and enhance environmental observations and monitoring programs. In general, automation methods can strongly reduce the time of sample analysis, decrease human bias, and allow larger datasets to be analyzed in a standardized and cost-effective manner. Previously suggested automated methods for MP assessments combined spectroscopic techniques with 1) SVR (support vector machine regression) and 2) PLS-DA (partial least squares discriminant analysis) (Paul et al., 2019; da Silva et al., 2020), 3) random decision forests (Weisser et al., 2021; Hufnagl et al., 2022), and 4) KNN algorithms (k-nearest neighbors) (Kedzierski et al., 2019). However, although promising and quite efficient in identifying MPs, the referred methodologies are still dependent on obtaining outputs based on infrared spectroscopy, which are themselves highly dependent on the access to expensive equipment. Benchtop FTIRs are available from US\$25 k onwards (Primpke et al., 2020), but for FTIR microscope systems with options crucial for MP analysis, costs increase quickly. Particle finder systems range from US\$100–250 k, not yet considering the liquid nitrogen supply needed for certain detector types. Raman spectrometers are available starting at US\$50 k (Primpke et al., 2020). For the method developed here, equipment costs include a fluorescence microscope equipped with camera and fluorescence filters (<US\$10 k), while staining costs per sample are <US\$5 per filter (1 PTFE filter, 10 μg NR and 1 mL acetone). The supervised machine learning models were developed using open source software, as opposed to the use of commercial spectra libraries (e.g. for FTIR polymer identification). Without the need for expensive tools, this method can be a way forward on a global scale as it enables research based in emerging economy countries, where the access to sophisticated equipment can be limited.

The automation of inexpensive fluorescence staining methodologies as an alternative to often used, spectroscopy-based methodologies has already been demonstrated to be promising for the development and advancement of high-throughput, cost-effective screening methods for MPs analysis (Shruti et al., 2022). Most of these approaches however focused on MP detection only (Primpke et al., 2017; Maes et al., 2017; Prata et al., 2019), or were limited to a MP classification into “polar” and ‘hydrophobic’ polymers (Maes et al., 2017). A fundamental issue previously reported in most NR-based studies was the co-staining of organic residues such as residual fat, which can interfere with MP analysis (Prata et al., 2021). An effective matrix removal prior to MP analysis is critical, while for FTIR-based methods this is of less concern. Nonetheless, the NR method was recently trialed successfully for biota samples (de Witte et al., 2022). Here, a method based on NR-staining developed for MPs in fish GIT matrices achieved an overall

85% of correct predictions (for plastics and non-plastic particles), confirmed using μ -FTIR analysis. Nevertheless, there is no established method that ensures the complete removal of all types of organic contaminants in environmental samples nowadays (Shruti et al., 2022), and consequently, the chance of inaccurately accounting for MPs in environmental samples remains. However, the proposed ML models proposed in the current work take the NR method a step further by looking not just at the fluorescence intensity, but also at the RGB-coloration of photographed particles. This way, the models enable the researchers to tackle the fundamental problem of co-staining of natural materials. Because of the unique combination of RGB-statistics derived from each particle's emitted fluorescent coloration under three different microscope fluorescence filters, plastics particles can be distinguished from non-plastics, and plastic polymers can be distinguished from each other.

Classification models, such as the classification trees used in the current work, have been broadly and successfully used in the field of ecological modeling, e.g., to determine soil moisture of an area, which is of high importance in agriculture (Pekel, 2020), to construct groundwater spring potential maps (Chen et al., 2020), or to allow for plant identification and determination of informative traits to distinguish different plant taxa (Almeida et al., 2020). Once classification trees have been constructed, these models can analyze large datasets of spectral statistics from unidentified particles, using only low computer power. Neither complex nor large datasets based on acquired RGB data are needed for their development (Fig. A1 in Supplementary materials) (Everaert et al., 2011). Moreover, the used algorithm eliminates the subjectivity of human sorting and is therefore an unbiased method. A drawback of automated research techniques is often their difficulty to be grasped by non-experts, as they often require expertise in the field of artificial intelligence (AI) and machine learning. Here, this challenge was overcome, as an alternative to complex models is introduced, with a methodology for MPs detection and polymer identification that is high-performance, simplified and reproducible. The classification models in this work and their branching methodology are easily interpretable while being capable of achieving high accuracy, provided that extrapolation outside the spectral ranges used in the training dataset is avoided. Contrary to black-box models used in machine learning (e.g. neural networks), white-box models, such as the decision tree models used here, have a transparent inner structure and represent information in a visual and clear way. These models can give insight in complex, unbalanced, non-linear data where commonly used exploratory and statistical modeling techniques often fail to find meaningful patterns (De'ath and Fabricius, 2000). Through the use of patterns, rule-based classifiers like the models obtained here can provide a clear view of the interrelationships between the different variables considered (i.e. the RGB statistics), can obtain comparable results, and these results can be explained in a way that can be grasped by the human mind (Loyola-Gonzalez, 2019). A lower structural complexity of decision tree models, which can be measured through parameters such as the number of terminal nodes, question depth, and branching factor (Piltaver et al., 2016) usually leads to a higher interpretability, and therefore a higher usability (Witten et al., 2011; Zhou et al., 2018). The obtained classification models PDM and the PIM had, respectively, three and seven terminal nodes, a question depth of three and four, and both had a branching factor of two. This illustrates their simplicity and high comprehensibility (Witten et al., 2011).

The integration of this innovative methodology based on RGB color-codes quantification holds potential for marine environmental MPs survey and monitoring programs, as it enables the distinction between plastic and natural-based materials, and as it can further categorize plastics according to polymer type. Microplastics extracted from field samples show varying degrees of environmental degradation (photodegradation, thermal degradation and biodegradation) and biofouling, processes which may affect the plastics' integrity and properties, and consequently also their fluorescence after NR staining (Shim et al., 2016; Tu et al., 2020). Preliminary results indicated that spiked MPs extracted from environmental matrices can be identified by the developed models. The models demonstrated to be robust to potential matrix effects, to effects of reagents used in sample

processing, to variability in fluorescent coloration, and to polymeric matrices with (commercial plastics) and without additional additives (Carat). More into-depth research is required to support these exploratory results, and the application of both classification models for naturally weathered MPs needs to be further trialed. The models can be adjusted for any specifications required by the user, in the same way as described in the workflow (Fig. A1 in Supplementary materials), though the current method is expected to be robust in the detection and identification of MPs collected during large survey and/or monitoring campaigns. If deemed necessary, RGB data of weathered MPs can be used to expand the already existing training dataset based on pristine MPs, or to build new models based on weathered plastics only.

5. Conclusion

Both models presented in this work were independently validated for particles prepared in three ways. A first independent validation set of plastics were pristine particles. A second independent validation set consisted of particles that had undergone digestive treatment. For the third validation set, particles were spiked in environmental samples and underwent laboratory processing (i.e. digestion and filtration). Each of the outcomes of the validation exercises demonstrated the general applicability of the PDM and the PIM for analyzing MPs in marine samples. In conclusion, the classification models represent a new, semi-automated, high-throughput, reproducible method that enables detection and identification of commonly found MPs (50–1200 μm) in the marine environment in a reliable way, without needing expertise in the field of AI, through the quantification of fluorescence spectra emitted by NR-stained particles. This classification approach lays the foundation for a new, promising, cost- and time-effective method for environmental MP characterization, contributing to the harmonization of MPs identification and quantification methods for the accurate assessment of MP pollution in the marine environment.

CRedit authorship contribution statement

Nelle Meyers: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Ana I. Catarino:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Annelies M. Declercq:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Aisling Brennan:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Lisa Devriese:** Methodology, Supervision, Writing – review & editing. **Michiel Vandegheuchte:** Funding acquisition, Supervision, Writing – review & editing. **Bavo De Witte:** Funding acquisition, Methodology, Supervision, Writing – review & editing. **Colin Janssen:** Funding acquisition, Supervision, Writing – review & editing. **Gert Everaert:** Conceptualization, Funding acquisition, Methodology, Project administration, Software, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We want to thank Jonas Mortelmans, Hans Pirlet, Julie Muyle, Mattias Bossaer, Zhiyue Niu and Maarten De Rijcke from VLIZ for the technical support and the content feedback on the early draft of the manuscript, as well as the lab technicians of the Marine Analytical Laboratory of ILVO. The assistance provided was greatly appreciated and helped with finalizing the project. Furthermore, we also wish to show our appreciation for Henk Vrielinck and Samira Khelifi from Ghent University for the assistance

provided for the initial work with the $\mu\text{-FTIR}$. We also thank Centexbel for the natural reference materials and the cryomilled reference plastics which we needed to build and test our classification models. Lastly, we thank the project partners of the Andromeda project (JPI Oceans) who supported this work, and BELSPO [contract no B2/20E/P1/Andromeda] who funded this work.

Funding

This work was supported by the Andromeda project (JPI Oceans) with funding of the Belgian partners by BELSPO [contract no B2/20E/P1/Andromeda].

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2022.153441>.

References

- Ahrendt, C., Perez-Venegas, D.J., Urbina, M., Gonzalez, C., Echeveste, P., Aldana, M., Pulgar, J., Galbán-Malagón, C., 2020. Microplastic ingestion cause intestinal lesions in the intertidal fish *Girella laevis*. *Mar. Pollut. Bull.* 151. <https://doi.org/10.1016/j.marpolbul.2019.110795>.
- Almeida, B.K., Garg, M., Kubat, M., Afkhami, M.E., 2020. Not that kind of tree: assessing the potential for decision tree-based plant identification using trait databases. *Appl. Plant Sci.* 8. <https://doi.org/10.1002/aps3.11379>.
- Avio, C.G., Gorb, S., Milan, M., Benedetti, M., Fattorini, D., D'Errico, G., Paoletto, M., Bargelloni, L., Regoli, F., 2015. Pollutants bioavailability and toxicological risk from microplastics to marine mussels. *Environ. Pollut.* 198, 211–222. <https://doi.org/10.1016/j.envpol.2014.12.021>.
- Bakir, A., Desender, M., Wilkinson, T., van Hoytema, N., Amos, R., Airahui, S., Graham, J., Maes, T., 2020. Occurrence and abundance of meso and microplastics in sediment, surface waters, and marine biota from the South Pacific region. *Mar. Pollut. Bull.* 160. <https://doi.org/10.1016/j.marpolbul.2020.111572>.
- Beiras, R., Schönemann, A.M., 2020. Currently monitored microplastics pose negligible ecological risk to the global ocean. *Sci. Rep.* 10. <https://doi.org/10.1038/s41598-020-79304-z>.
- Bianco, V., Memmolo, P., Carcagnì, P., Merola, F., Paturzo, M., Distanti, C., Ferraro, P., 2020. Microplastic identification via holographic imaging and machine learning. *Adv. Intell. Syst.* 2, 1900153. <https://doi.org/10.1002/aisy.201900153>.
- Brahme, A., 2014. *Comprehensive Biomedical Physics*. Newnes, Stockholm.
- Bucci, K., Tulio, M., Rochman, C.M., 2020. What is known and unknown about the effects of plastic pollution: a meta-analysis and systematic review. *Ecol. Appl.* 30. <https://doi.org/10.1002/eap.20444>.
- Chen, W., Li, Y., Tsangaratos, P., Shahabi, H., Ilia, I., Xue, W., Bian, H., 2020. Groundwater spring potential mapping using artificial intelligence approach based on kernel logistic regression, random forest, and alternating decision tree models. *Appl. Sci.* 10. <https://doi.org/10.3390/app10020425>.
- Cowger, W., Gray, A., Christiansen, S.H., DeFronzo, H., Deshpande, A.D., Hemabessiere, L., Lee, E., Mill, L., Munno, K., Ossmann, B.E., Pittroff, M., Rochman, C., Sarau, G., Tarby, S., Pimpke, S., 2020. Critical review of processing and classification techniques for images and spectra in microplastic research. *Appl. Spectrosc.* 74, 989–1010. <https://doi.org/10.1177/0003702820929064>.
- da Silva, V.H., Murphy, F., Amigo, J.M., Stedmon, C., Strand, J., 2020. Classification and quantification of microplastics (<100 μm) using a focal plane array-Fourier transform infrared imaging system and machine learning. *Anal. Chem.* 92. <https://doi.org/10.1021/acs.analchem.0c01324>.
- de Troyer, N., Eurie Forio, M.A., Roels, K., de Meester, L., Lemmens, P., Declercq, S.A.J., Martens, K., Goethals, P., 2020. Key management rules for agricultural alpine newt breeding ponds based on habitat suitability models. *Glob. Ecol. Conserv.* 23. <https://doi.org/10.1016/j.gecco.2020.e01086>.
- de Witte, B., Catarino, A.I., Vandecasteele, L., Dekimpe, M., Meyers, N., Deloof, D., Pint, S., Hostens, K., Everaert, G., Torreale, E., 2022. Feasibility study on biomonitoring of microplastics in fish gastrointestinal tracts. *Front. Mar. Sci.* 8. <https://doi.org/10.3389/fmars.2021.794636>.
- De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology* 81, 3178. [https://doi.org/10.1890/0012-9658\(2000\)081\[3178:CARTAP\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2).
- Diepens, N.J., Koelmans, A.A., 2018. Accumulation of plastic debris and associated contaminants in aquatic food webs. *Environ. Sci. Technol.* 52, 8510–8520. <https://doi.org/10.1021/acs.est.8b02515>.
- Dixon, P., 2003. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14 (6), 927–930.
- Dunford, R., Su, Q., Tamang, E., 2014. *The Pareto principle*. *The Plymouth Student Scientist* 7, pp. 140–148.
- Erni-Cassola, G., Gibson, M.I., Thompson, R.C., Christie-Oleza, J.A., 2017. Lost, but found with Nile red: a novel method for detecting and quantifying small microplastics (1 mm to 20 μm) in environmental samples. *Environ. Sci. Technol.* 51, 13641–13648. <https://doi.org/10.1021/acs.est.7b04512>.

- Everaert, G., Boets, P., Lock, K., Džeroski, S., Goethals, P.L.M., 2011. Using classification trees to analyze the impact of exotic species on the ecological assessment of polder lakes in Flanders, Belgium. *Ecol. Model.* 222. <https://doi.org/10.1016/j.ecolmodel.2010.08.013>.
- Everaert, G., van Cauwenbergh, L., de Rijkke, M., Koelmans, A.A., Mees, J., Vandegehuchte, M., Janssen, C.R., 2018. Risk assessment of microplastics in the ocean: modeling approach and first conclusions. *Environ. Pollut.* 242, 1930–1938. <https://doi.org/10.1016/j.envpol.2018.07.069>.
- Everaert, G., de Rijkke, M., Lonnaville, B., Janssen, C.R., Backhaus, T., Mees, J., van Sebille, E., Koelmans, A.A., Catarino, A.I., Vandegehuchte, M.B., 2020. Risks of floating microplastic in the global ocean. *Environ. Pollut.* 267. <https://doi.org/10.1016/j.envpol.2020.115499>.
- Galloway, T.S., Cole, M., Lewis, C., 2018. ORE open research exeter a note on versions interactions of microplastic debris throughout the marine ecosystem 1 2. available at <http://hdl.handle.net/10871/34727> doi:10.1038/s41559-017-0116.
- Geyer, R., Jambeck, J.R., Law, K.L., 2017. Production, use, and fate of all plastics ever made. available at <http://advances.sciencemag.org/> doi:10.1016/j.ecolmodel.2010.08.013.
- Glatthorn, J., Beckschäfer, P., 2014. Standardizing the protocol for hemispherical photographs: accuracy assessment of binarization algorithms. *PLoS ONE* 9. <https://doi.org/10.1371/journal.pone.0111924>.
- Goethals, P.L.M., Dedecker, A.P., Gabriels, W., Lek, S., de Pauw, N., 2007. Applications of artificial neural networks predicting macroinvertebrates in freshwaters. *Aquat. Ecol.* 41, 491–508. <https://doi.org/10.1007/s10452-007-9093-3>.
- Hartmann, N.B., Hüffer, T., Thompson, R.C., Hassellöv, M., Verschoor, A., Daugaard, A.E., Rist, S., Karlsson, T., Brennholt, N., Cole, M., Herrling, M.P., Hess, M.C., Ivleva, N.P., Lusher, A.L., Wagner, M., 2019. Are we speaking the same language? Recommendations for a definition and categorization framework for plastic debris. *Environ. Sci. Technol.* 53, 1039–1047. <https://doi.org/10.1021/acs.est.8b05297>.
- Hufnagl, B., Stibi, M., Martirosyan, H., Wilczek, U., Möller, J.N., Löder, M.G.J., Laforsch, C., Lohninger, H., 2022. Computer-assisted analysis of microplastics in environmental samples based on μ FTIR imaging in combination with machine learning. *Environ. Sci. Technol. Lett.* 9, 90–95. <https://doi.org/10.1021/acs.estlett.1c00851>.
- Iannilli, V., Pasquali, V., Setini, A., Corami, F., 2019. First evidence of microplastics ingestion in benthic amphipods from Svalbard. *Environ. Res.* 179. <https://doi.org/10.1016/j.envres.2019.108811>.
- ICES, 2021. Working Group on Marine Litter (WGML; Outputs From 2020 Meeting), ICES Sci. Reports 3 (51). <https://doi.org/10.17895/ices.pub.8185>.
- Kedzierski, M., Falcou-Préfol, M., Kerros, M., Emmanuelle, Henry, M., Pedrotti, L., Bruzard, S., 2019. A machine learning algorithm for high throughput identification of FTIR spectra: 1 application on microplastics collected in the Mediterranean Sea 2. available at <https://www.elsevier.com/open-access/userlicense/1.0/> doi: 10.1016/j.chemosphere.2019.05.113.
- Kershaw, P., Turra, A., Galgani, F., 2019. Guidelines for the Monitoring and Assessment of Plastic Litter in the Ocean. United Nations Environment Programme (UNEP), Nairobi <https://doi.org/10.25607/OBP-435>.
- Koelmans, A.A., Diepens, N.J., Hazimah, N., Nor, M., Koelmans, A.A., Diepens, N.J., Nor, N.H.M., 2021. Weight of Evidence for the Microplastic Vector Effect in the Context of Chemical Risk Assessment. Springer https://doi.org/10.1007/978-3-030-78627-4_6.
- Kooi, M., Primpke, S., Mintenig, S.M., Lorenz, C., Gerdt, G., Koelmans, A.A., 2021. Characterizing the multidimensionality of microplastics across environmental compartments. *Water Res.* 202, 117429. <https://doi.org/10.1016/j.watres.2021.117429>.
- Kuhn, M., 2008. Journal of statistical software building predictive models in R using the caret package. available at <http://www.jstatsoft.org/>.
- Loyola-Gonzalez, O., 2019. Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view. *IEEE Access* 7, 154096–154113. <https://doi.org/10.1109/ACCESS.2019.2949286>.
- Luo, H., Xiang, Y., He, D., Li, Y., Zhao, Y., Wang, S., Pan, X., 2019. Leaching behavior of fluorescent additives from microplastics and the toxicity of leachate to *Chlorella vulgaris*. *Sci. Total Environ.* 678, 1–9. <https://doi.org/10.1016/j.scitotenv.2019.04.401>.
- Lv, L., Qu, J., Yu, Z., Chen, D., Zhou, C., Hong, P., Sun, S., Li, C., 2019. A simple method for detecting and quantifying microplastics utilizing fluorescent dyes - safranin T, fluorescein isophosphate, Nile red based on thermal expansion and contraction property. *Environ. Pollut.* 255. <https://doi.org/10.1016/j.envpol.2019.113283>.
- Maes, T., Jessop, R., Wellner, N., Haupt, K., Mayes, A.G., 2017. A rapid-screening approach to detect and quantify microplastics based on fluorescent tagging with Nile red. *Sci. Rep.* 7. <https://doi.org/10.1038/srep44501>.
- Maxwell, H.S., Melinda, K.F., Matthew, G., 2020. Counterstaining to separate Nile red-stained microplastic particles from terrestrial invertebrate biomass. *Environ. Sci. Technol.* 54, 5580–5588. <https://doi.org/10.1021/acs.est.0c00711>.
- McHugh, M.L., 2012. Interrater reliability: the kappa statistic. *Biochem. Med.* <https://doi.org/10.11613/BM.2012.031>.
- Meyers, N., 2021. NelleMeyers/RGB models: updated release RGB models for microplastic analysis. available at Zenodo. <https://zenodo.org/record/5596531>.
- Meyers, N., Catarino, A.I., Declercq, A., Brennan, A., Devriese, L., Vandegehuchte, M., de Witte, B., Janssen, C., Everaert, G., 2021. RGB-statistics derived from Nile red-stained reference plastics for the construction of the PDM (Plastics Detection Model). *Mar. Data Arch.* <https://doi.org/10.14284/512>.
- Meyers, N., Catarino, A.I., Declercq, A., Brennan, A., Devriese, L., Vandegehuchte, M., de Witte, B., Janssen, C., Everaert, G., 2021. RGB-statistics derived from Nile red-stained reference plastics for the construction of the PIM (Polymer identification Model). *Mar. Data Arch.* <https://doi.org/10.14284/511>.
- Nel, H.A., Chetwynd, A.J., Kelleher, L., Lynch, I., Mansfield, I., Margenat, H., Onoja, S., Goldberg Oppenheimer, P., Sambrook Smith, G.H., Krause, S., 2021. Detection limits are central to improve reporting standards when using Nile red for microplastic quantification. *Chemosphere* 263. <https://doi.org/10.1016/j.chemosphere.2020.127953>.
- Paul, A., Wander, L., Becker, R., Goedecke, C., Braun, U., 2019. High-throughput NIR spectroscopic (NIRS) detection of microplastics in soil. *Environ. Sci. Pollut. Res.* 26. <https://doi.org/10.1007/s11356-018-2180-2>.
- Pekel, E., 2020. Estimation of soil moisture using decision tree regression. *Theor. Appl. Climatol.* 139, 1111–1119. <https://doi.org/10.1007/s00704-019-03048-8>.
- Piltaver, R., Luštrek, M., Gams, M., Martinčić-Ipsić, S., 2016. What makes classification trees comprehensible? *Expert Syst. Appl.* 62, 333–346. <https://doi.org/10.1016/j.eswa.2016.06.009>.
- Prata, J.C., Reis, V., Matos, J.T.V., da Costa, J.P., Duarte, A.C., Rocha-Santos, T., 2019. A new approach for routine quantification of microplastics using Nile red and automated software (MP-VAT). *Sci. Total Environ.* 690, 1277–1283. <https://doi.org/10.1016/j.scitotenv.2019.07.060>.
- Prata, J.C., Sequeira, I.F., Monteiro, S.S., Silva, A.L.P., da Costa, J.P., Dias-Pereira, P., Fernandes, A.J.S., da Costa, F.M., Duarte, A.C., Rocha-Santos, T., 2021. Preparation of biological samples for microplastic identification by Nile Red. *Sci. Total Environ.* 783, 147065. <https://doi.org/10.1016/j.scitotenv.2021.147065>.
- Primpke, S., Lorenz, C., Rascher-Friesenhausen, R., Gerdt, G., 2017. An automated approach for microplastics analysis using focal plane array (FPA) FTIR microscopy and image analysis. *Anal. Methods* 9, 1499–1511. <https://doi.org/10.1039/C6AY02476A>.
- Primpke, S., Christiansen, S.H., Cowger, W., de Frond, H., Deshpande, A., Fischer, M., Holland, E.B., Meyns, M., O'Donnell, B.A., Ossmann, B.E., Pittroff, M., Sarau, G., Scholz-Böttcher, B.M., Wiggan, K.J., 2020. Critical assessment of analytical methods for the harmonized and cost-efficient analysis of microplastics. *Appl. Spectrosc.* 74. <https://doi.org/10.1177/0003702820921465>.
- R Core Team, 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria Available online: <https://www.r-project.org/>.
- Ruggero, F., Gori, R., Lubello, C., 2020. Methodologies for microplastics recovery and identification in heterogeneous solid matrices: a review. *J. Polym. Environ.* 28, 739–748. <https://doi.org/10.1007/s10924-019-01644-3>.
- Sancataldo, G., Avellone, G., Vetri, V., 2020. Nile red lifetime reveals microplastic identity. *Environ. Sci. Process Impacts* 22, 2266–2275. <https://doi.org/10.1039/D0EM00348D>.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.Y., White, D.J., Hartenstein, V., Eliceiri, K., Tomancak, P., Cardona, A., 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682.
- Scholten, M.C.Th., 2005. *Eutrophication Management and Ecotoxicology*. Springer.
- Shim, W.J., Song, Y.K., Hong, S.H., Jang, M., 2016. Identification and quantification of microplastics using Nile red staining. *Mar. Pollut. Bull.* 113, 469–476. <https://doi.org/10.1016/j.marpolbul.2016.10.049>.
- Shruti, V.C., Pérez-Guevara, F., Roy, P.D., Kutralam-Muniasamy, G., 2022. Analyzing microplastics with Nile red: emerging trends, challenges, and prospects. *J. Hazard. Mater.* 423, 127171. <https://doi.org/10.1016/j.jhazmat.2021.127171>.
- Stanton, T., Johnson, M., Nathanael, P., Gomes, R.L., Needham, T., Burson, A., 2019. Exploring the efficacy of Nile red in microplastic quantification: a costaining approach. *Environ. Sci. Technol. Lett.* 6. <https://doi.org/10.1021/acs.estlett.9b00499>.
- Sturm, M.T., Horn, H., Schuhen, K., 2021. The potential of fluorescent dyes—comparative study of Nile red and three derivatives for the detection of microplastics. *Anal. Bioanal. Chem.* 413. <https://doi.org/10.1007/s00216-020-03066-w>.
- Suaría, G., Achtypi, A., Perold, V., Lee, J.R., Pierucci, A., Bornman, T.G., Aliani, S., Ryan, P.G., 2020. OCEANOGRAPHY Microfibers in oceanic surface waters: a global characterization. <http://advances.sciencemag.org/>.
- Tamminga, M., 2017. Nile red staining as a subsidiary method for microplastic quantification: a comparison of three solvents and factors influencing application reliability. *SDRP J. Earth Sci. Environ. Stud.* 2. <https://doi.org/10.15436/JESES.2.2.1>.
- Therneau, T., Atkinson, B., Ripley, B., 2015. Package 'rpart' (Available online: cran.ma.ic.ac.uk/web/packages/rpart/rpart).
- Tu, C., Chen, T., Zhou, Q., Liu, Y., Wei, J., Wanik, J.J., Luo, Y., 2020. Biofilm formation and its influences on the properties of microplastics as affected by exposure time and depth in the seawater. *Sci. Total Environ.* 734. <https://doi.org/10.1016/j.scitotenv.2020.139237>.
- Veerasingam, S., Ranjani, M., Venkatachalapathy, R., Bagaev, A., Mukhanov, V., Litvinyuk, D., Mugilarasan, M., Gurumoorthi, K., Gaganathan, L., Aboobacker, V.M., Vethamony, P., 2021. Contributions of Fourier transform infrared spectroscopy in microplastic pollution research: a review. *Crit. Rev. Environ. Sci. Technol.* 51, 2681–2743. <https://doi.org/10.1080/10643389.2020.1807450>.
- Wang, W., Ge, J., Yu, X., 2020. Bioavailability and toxicity of microplastics to fish species: a review. *Ecotoxicol. Environ. Saf.* 189. <https://doi.org/10.1016/j.ecoenv.2019.109913>.
- Weisser, J., Beer, I., Hufnagl, B., Hofmann, T., Lohninger, H., Ivleva, N.P., Glas, K., 2021. From the well to the bottle: identifying sources of microplastics in mineral water. *Water (Switzerland)* 13. <https://doi.org/10.3390/w13060841>.
- Wickham, M.H., 2007. The ggplot package. available at <http://had.co.nz/ggplot/>.
- Witten, I.H., Frank, E., Hall, M.A., 2011. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier.
- Zhou, Q., Liao, F., Mou, C., Wang, P., 2018. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 11154 LNAL. Springer Verlag, pp. 295–308.