HAL
open science

A machine learning algorithm for high throughput identification of FTIR spectra: Application on microplastics collected in the Mediterranean Sea

Mikael Kedzierski, Mathilde Falcou-Préfol, Marie Emmanuelle Kerros, Maryvonne Henry, Maria Luiza Pedrotti, Stephane Bruzaud

To cite this version:

Mikael Kedzierski, Mathilde Falcou-Préfol, Marie Emmanuelle Kerros, Maryvonne Henry, Maria Luiza Pedrotti, et al.. A machine learning algorithm for high throughput identification of FTIR spectra: Application on microplastics collected in the Mediterranean Sea. Chemosphere, 2019, 234, pp.242-251. 10.1016/j.chemosphere.2019.05.113 . hal-02407542

HAL Id: hal-02407542

https://hal.science/hal-02407542v1

Submitted on 18 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Chemosphere
November 2019, Volume 234, Pages 242-251
https://doi.org/10.1016/j.chemosphere.2019.05.113
https://archimer.ifremer.fr/doc/00501/61247/

Archimer
https://archimer.ifremer.fr

A machine learning algorithm for high throughput identification of FTIR spectra: Application on microplastics collected in the Mediterranean Sea

Kedzierski Mikaël [1, *], Falcou-Préfol Mathilde [1], Kerros Marie Emmanuelle [2], Henry Maryvonne [3], Pedrotti Maria Luiza [2], Bruzaud Stéphane [1]

[1] Université Bretagne Sud, UMR CNRS 6027, IRDL, F-56100, Lorient, France
[2] Sorbonne Universités, UMR CNRS 7093, LOV, F-06230, Villefranche sur mer, France
[3] IFREMER, LER/PAC, F-83500, La Seine-sur-Mer, France

* Corresponding author : Mikaël Kedzierski, email address : mikael.kedzierski@univ-ubs.fr

Abstract :

The development of methods to automatically determine the chemical nature of microplastics by FTIR-ATR spectra is an important challenge. A machine learning method, named k-nearest neighbors classification, has been applied on spectra of microplastics collected during Tara Expedition in the Mediterranean Sea (2014). To realize these tests, a learning database composed of 969 microplastic spectra has been created. Results show that the machine learning process is very efficient to identify spectra of classical polymers such as poly(ethylene), but also that the learning database must be enhanced with less common microplastic spectra. Finally, this method has been applied on more than 4000 spectra of unidentified microplastics. The verification protocol showed less than 10% difference in the results between the proposed automated method and a human expertise, 75% of which can be very easily corrected.

Highlights

► A machine learning algorithm was developed to determine the chemical nature of microplastics. ► This method allows a fast and reliable automated identification even when several thousand of FTIR spectra have to be studied.► This method is the first part of a software dedicated to the study of microplastics: POSEIDON.

Keywords : Microplastic, Tara Mediterranean campaign, FTIR spectra, Machine learning, k-nearest neighbor classification

Please note that this is an author-produced PDF of an article accepted for publication following peer review. The definitive publisher-authenticated version is available on the publisher Web site.

1. INTRODUCTION

Pollution of the marine environment by plastics, and especially by microplastics (Thompson et al., 2004), is the subject of an increasing number of studies (Barboza and Gimenez, 2015). Many expeditions have been carried out at sea (Brach et al., 2018; Dussud et al., 2018; Kanhai et al., 2017; Pedrotti et al., 2018, 2016), on lakes (Eriksen et al., 2013; Free et al., 2014; Imhof et al., 2013) or in the ice pack (Obbard et al., 2014; Peeken et al., 2018) to determine the state of contamination of the environment. Among the parameters studied on these particles, the chemical nature of microplastics is increasingly determined (Gewert et al., 2017; Imhof et al., 2017; Löder and Gerdts, 2015; Qiu et al., 2016; Wang et al., 2017). As the number of particles to be analyzed can range from several hundred to several thousand (Falcou-Préfol et al., 2018; Pedrotti et al., 2018, 2016), it is necessary to develop methods for efficient microplastics analysis (Kedzierski et al., 2019). When determining the chemical nature of samples, methods based on thermal degradation as pyrolysis-gas chromatography in combination with mass spectrometry (GC/MS) are fast and efficient (Dümichen et al., 2017, 2015; Gimeno et al., 2015). Nevertheless, this type of method results in the destruction of the analyzed sample (Primpke et al., 2018) which is considered as a limitation in the study of microplastics (Rocha-Santos and Duarte, 2015). Indeed, with this type of analysis, important parameters such as the size distribution, particle mass or the plastisphere are lost (Amaral-Zettler et al., 2015; Dussud et al., 2018; Frère et al., 2018; Primpke et al., 2018; Zettler et al., 2013). Furthermore, in the case of large sampling campaigns, the samples, taken at a given location and time, are unique and must be

ACCEPTED MANUSCRIPT

preserved over the long term in library. Thus, destructive analyses must be kept to a strict minimum. Other methods, based on Raman (Frère et al., 2016; Imhof et al., 2017, 2012; Lenz et al., 2015; Zhao et al., 2017) or Infrared (IR) spectroscopy (Ioakeimidis et al., 2016; Löder and Gerdts, 2015; Primpke et al., 2018, 2017), have shown their efficiency for the identification of the chemical nature of microplastics. However, the analysis of the spectra is time-consuming and can be a source of misinterpretation (Elert et al., 2017; Käppler et al., 2016; Löder et al., 2015; Primpke et al., 2018). It is indeed often necessary to compare one by one the different spectra obtained with reference spectra. Moreover, a significant part of misinterpretations is due to the fact that, during the aging of the plastics, additional bands appear on spectra (Albertsson et al., 1995; Andrady, 2017; Fotopoulou and Karapanagioti, 2012; Küpper et al., 2004). This phenomenon also become noticeable because of the fouling present on particles (Howe et al., 2002; Maquelin et al., 2002; Sudhakar et al., 2007). However, the reference spectra are made on new and clean plastics. Thus, an aged plastic is sometimes matched to the spectrum of a reference plastic due to the fact that this reference plastic has characteristic bands similar to those which appear during plastic aging. So, the greater the weight of the bands associated with aging and fouling, the more automated interpretation based on new and clean reference plastics may diverge from the actual nature of the polymer.

The development of methods to automatically determine the spectra of microplastics is therefore an important challenge of the coming years. Very recent publications began to take an interest in the issue and proposed answers. In 2017, an automated method based on the identification of the most relevant bands was tested (Renner et al., 2019, 2017). In 2018, an automated identification method based on hierarchical cluster analysis was published (Primpke et al., 2018). More recently a new method based on Random Decision Forest was proposed (Hufnagl et al., 2019).

These three methods have shown a high degree of spectrum identification. However, other approaches are possible and have not been yet tested for the recognition of microplastic spectra. This is the case, for example, of methods belonging to the machine learning family which are quite

commonly used in the field of spectrometry and are known to give good results (Wehrens, 2011). Recently, a machine learning process has been successfully used on the determination of the chemical nature of microplastics (Hufnagl et al., 2019). Among these methods, one of them, called K-Nearest Neighbors (KNN), is known for its simplicity and efficiency in automated spectrum recognition (Ripley, 1996; Venables et al., 2002; Wehrens, 2011). The originality of this work is the test of this machine learning method in the context of the study of microplastics. This project is part of the Tara Expedition in the Mediterranean Sea (2014). Tara Foundation is a French non-profit organization acting for the environment since 2003. In 2014, Tara Expedition conducted sampling for 7 months across the Mediterranean Sea. The objective of this expedition was to characterize microplastic contamination in order to better describe the effects of plastic litter on marine ecosystems. In particular, the Tara Mediterranean consortium is trying to evaluate the spatial distribution of floating debris in the Mediterranean Sea, to chemically characterize the different types of plastics and to characterize bacteria communities (Dussud et al., 2018) as well as the micro- and macro-organisms fixed on these plastic particles (Pedrotti et al., 2018). The scientific knowledge provided by this project will allow in the long run to better target the actions to be taken to reduce contamination by plastic waste in the Mediterranean Sea. In this context, the development of an automated spectrum identification tool would facilitate the analysis of microplastics collected during the expedition. This study therefore attempted to answer the following questions: can k-nearest neighbors classification method be used for the automated identification of the chemical nature of the FTIR spectra of microplastics? What is the origin of the misinterpretations and how could they be reduced? Finally, this method was applied for the analysis of more than 4,000 microplastic spectra.

2. MATERIALS AND METHODS

2.1. Sample collection

Microplastic samples were collected from Mediterranean Sea waters during the Tara Expedition which was conducted between May and November 2014. Sampling was conducted using a 4.4 m long

manta net (mesh size: 333 µm; net opening: 16 x 60 cm), from 120 sites which were selected based on ocean color satellite images supplied by ACRI-ST and analyzed with the Mercator circulation model (Fig. 1). Geographical coordinates and dates of sampling analyzed are available at Pangea Data Publisher http://www.pangaea. At each site, the manta net was towed on the sea surface for ca. 60 min behind the boat at an average speed of 2.5 knots, enabling thus the filtration of about 507 m^3 of seawater.

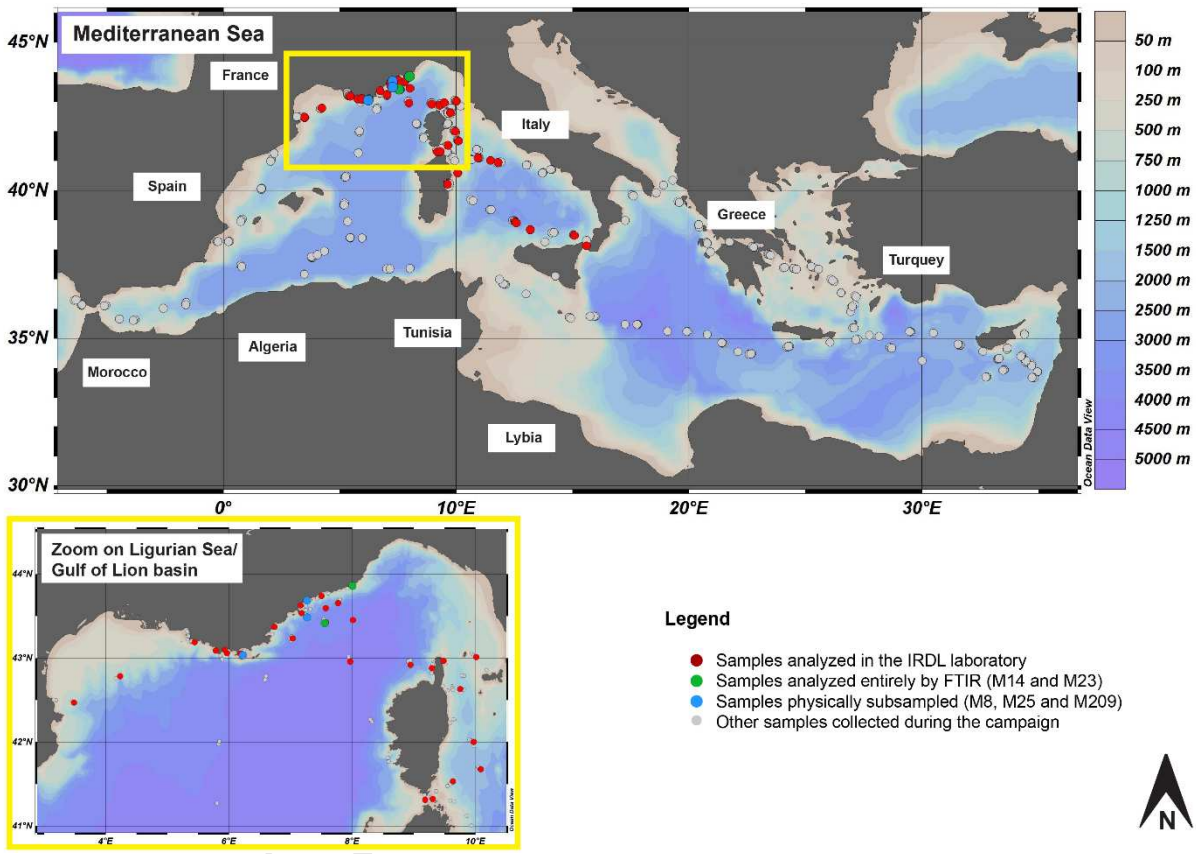

Fig. 1. Sampling effort during the Tara campaign in the Mediterranean Sea. Ocean Data View (Schlitzer, 2015).

2.2. Laboratory preparations and subsampling

Under dissecting microscope, preserved organic material was removed from microplastics. Particles were then counted and measured using the ZooScan image analysis method at the Laboratoire d'Océanographie de Villefranche-sur-Mer (LOV, Villefranche-sur-Mer, France) (Pedrotti et al., 2016).

A total of 13,374 particles, larger than 315 µm from 42 mantas trawls, were then transferred to 96-well microplates and named with a unique identifier at the Institut de Recherche Dupuy de Lôme (IRDL, Lorient, France) (Kedzierski et al., 2019). To perform the tests of the machine learning method, 3,090 particles were randomly selected. For samples M14 (767 particles) and M23 (207 particles), the particles were entirely analyzed during a previous study (Falcou-Préfol et al., 2018; Kedzierski et al., 2019). Therefore, a total of 4,064 spectra, performed on the samples of Tara Expedition, were used in the study. The particles chosen for analysis came mainly from M14 (762 particles; 18.9%), M8 (289 particles; 7.1%), M209 (272 particles; 6.7%), M25 (212 particles; 5.2%) and M23 (207 particles; 5.1%), all sampled in the Ligurian sea (Fig. 2.A).

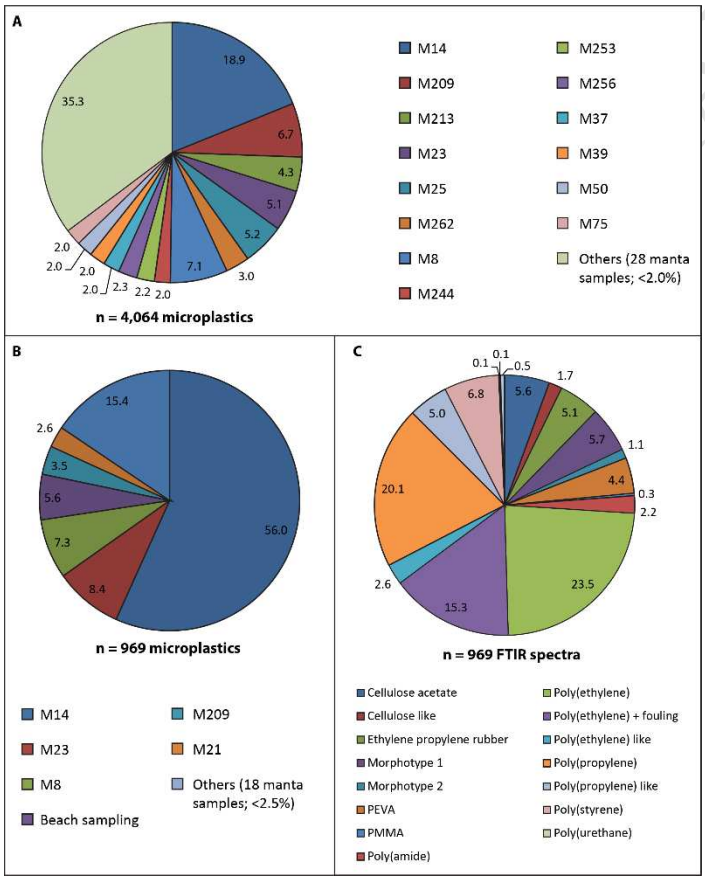

Fig. 2. Characteristics of the data used. A) Origin of the spectra available for the study (%). B) Origin of the spectra used for the machine learning database (%). C) Proportions (%) of the different classes of the machine learning database.

The preparation step was performed in an area dedicated to the treatment of Tara Mediterranean Sea samples in controlled conditions (chemical laboratory). Contamination risks were avoided during the sample preparation stage by cleaning the different parts of the apparatus, especially glassware, with distilled water, ethanol and/or acetone. The use of plastic apparatus was avoided as far as possible. If this was impossible, the Fourier-transform infrared spectroscopy (FTIR) spectra from these materials were obtained to check whether potential contamination of the samples had occurred.

2.3. Fourier-transform infrared spectroscopy (FTIR)

The spectra of the particles were acquired using a non-destructive Attenuated Total Reflection Fourier Transform Infrared spectrometer (ATR-FTIR Vertex70v, Bruker). All spectra were recorded in absorbance mode in the 4,000-600 cm^{-1} region with 4 cm^{-1} resolution and 16 scans. Each particle was placed onto the germanium diamond cell (ATR Golden Gate) used to obtain a better quality spectrum. After each analysis by ATR-FTIR, the sample holder was cleaned with ethanol or acetone. The sample chamber was also cleaned out with a vacuum cleaner after every sixty analyses. A total of 4,064 particles was analyzed.

2.4. Machine learning database

The learning database consists of 969 spectra, with 64.4% of them came from the M14 and M23 sites already analysed during a previous study (Falcou-Préfol et al., 2018; Kedzierski et al., 2019) (Fig. 2.B). 5.6% of the database's spectra were acquired from cigarette filters, based on cellulose acetate, collected on the beach of Toulhars (Larmor-Plage, France). The latter data were not part of the Tara Mediterranean campaign, but were used to complete and to test the learning database. The remaining 30.0% came from 22 different manta samples (Fig. 2.B). The spectra of the learning database were selected to represent the diversity present in the initial sample. In order to facilitate automated learning, "rare" spectra were systematically kept for the machine learning database while

more "common" spectra (mostly poly(ethylene) and poly(propylene)) were often discarded so as not unnecessarily slow down the learning process. As a result, "rare" spectra are over-represented in the machine learning database compared to more "common" spectra. The diversity of the database is currently representative of the diversity of plastic pieces found on the surface of the Mediterranean Sea. The database must be gradually enriched with new spectra such as PVC, PET, paints.

Each of the 969 spectra were then compared with the OMNIC specta 2.0 software (Thermo Scientific). The reference database was the Hummel Polymer Sample Library. This library, developed by Professor Dieter Hummel of the Institute of Physics and Chemistry at the University of Cologne (Köln), included about 2,000 spectra of polymers, polymer impurities and additives. The software interpretation was then systematically validated or rejected and reinterpreted by a polymer expert.

Thus, 17 classes were created (Fig. 2.C). The best represented class was poly(ethylene) (23.5%). It was closely associated with two other classes: "poly(ethylene) + fouling" (15.3%) and "poly(ethylene) like" (2.6%). The "poly(ethylene) + fouling" class was characterized by bands more or less intense between about 900 and 1,100 cm^{-1}, as well as around 1,641 cm^{-1}, which are not normally present on poly(ethylene) spectra. These bands can be associated with the presence of fouling on the plastic. The "poly(ethylene) like" class was characterized by the presence of one or more additional bands compared to a poly(ethylene) spectrum. The presence of these additional bands casted some doubts on the precise identification of the spectrum. It was therefore decided not to associate them directly with the "poly(ethylene)" class. The “poly(propylene)” class was the second well represented class with 20.1% of microplastics. This class was very similar to that of "poly(propylene) like" since it included spectra with the same bands as those of “poly(propylene)”, but with supernumerary bands. The classes of "poly(styrene)", "cellulose acetate", "ethylene-propylene rubber", "poly(ethylene-vinyl acetate)", "poly(methyl methacrylate)", "poly(amide)" and "poly(urethane)" were composed with spectra of the corresponding polymers. It should be noted here that ethylene-propylene rubber is a copolymer which has a spectrum sharing similar characteristics to those of the poly(ethylene) and

176 poly(propylene). If the ethylene/propylene ratio is low or very high, the distinction between
177 poly(ethylene) or poly(propylene) and this copolymer can be confusing. Two classes were named
178 morphotype 1 and 2. A “morphotype” was defined in this context as a set of spectra, numbering ten
179 or more, with strong similarities, but for which the identification of the precise chemical nature
180 remained unclear. This choice was made to describe and name new spectra, despite a chemical
181 nature that remains not clearly identified at the time of the study. The last class, called "cellulose
182 like" grouped all spectra showing strong similarities with cellulose. These spectra could therefore
183 correspond to poorly sorted plant debris or plastics with a largely colonized surface.

184 *2.5. Spectra pre-processing and machine learning process*

185 All the programming was done using the software R i386 3.1.2 which is a free and open source
186 software (The R Core Team, 2019). The pre-processing procedure of the infrared spectra was classic
187 (Renner et al., 2019; Wehrens, 2011). Firstly, the baseline of the spectra was corrected using function
188 “baseline()” (library “baseline”) and more particularly with the 4S Peak Filling baseline estimation
189 procedure (Liland, 2015). This procedure, based on an iterative spectrum suppression consisting of a
190 moving window minimum replacement, was chosen due to the fact that it is particularly adapted for
191 non-linear baseline. The suggested starting values of the four parameters necessary for the function
192 were respected (Liland, 2015). The second derivative penalty for smoothing (*lambda*) was equal to 4
193 (centered in noise band). The number of buckets for subsampling (*int*) was equal to 176 (about 1/10
194 of the number of wavelengths). The initial half width of windows used for suppression (*hwi*) was
195 equal to 25. The number of iteration for suppression (*it*) was equal to 10. Finally, after the baseline
196 correction, spectra were normalized.

197 To test the process, the machine learning database was randomly divided: two third for the learning
198 process and one third to test the process. The sampling was performed using the package “base”
199 version 3.1.2 and the “sample()” function (Becker et al., 1988; Ripley and Wiley, 1987). The machine
200 learning process was performed using *k*-nearest neighbor classification (Ripley, 1996; Venables et al.,

2002) of the “knn()” function (library “class”). This technique is based on the calculation of the distances between objects (Wehrens, 2011). The number of neighbor (*k*=10) was determined by recurrence. If the entire *k*-nearest neighbor belongs to the same class, the spectrum is directly identified; else a majority of vote is performed. If there are less than 3 votes, the spectrum is classified in the "unknown" category. Then, the model inspection was made with a confusion matrix (“confusionMatrix()” function; library “caret). The number of spectra well categorized (true positive; *TP*) or placed by the model in the wrong class (false negative; *FN*) were counted, as well as the number of spectra misclassified (false positive; *FP*) and well-classified (true negative; *TN*) in the predicted category. The sensitivity and the specificity were then calculated (Fawcett, 2006):

$$Sensitivity = \frac{TP}{TP+FN} \quad \text{(eq. 1)}$$

$$Specificity = \frac{FP}{FP+TN} \quad \text{(eq. 2)}$$

In order to test the reliability of the proposed methodology, the automated learning was carried out 1,000 times. Each time, a new set of training data was randomly drawn, each time leading to a slightly different training and therefore to results that may differ. The results presented are the averages with standard deviations calculated from these thousand tests.

In order to highlight the origin of classification error, a representation of the confusion matrix was made using the “alluvial()” function (library “alluvial”) and a synthetic version was designed using Adobe Illustrator CS5.

2.6. Identification of FTIR spectra with k-nearest neighbors classification and verification of the results

At this stage, 3,095 FTIR spectra still needed to be analysed. For this purpose, the KNN method and the learning database were used.

In order to test the accuracy of the final classification, a verification step was performed. This step was carried out for each of the identified classes in two sub-steps. First, a hierarchical cluster analysis was performed on all spectra of a same class (“hclust” function, centroid method). The number of

225 sub-clusters was estimated visually. A principal component analysis (PCA; “prcomp” function) was
226 used to check whether or not the estimated number of sub-clusters was relevant. The objective here
227 was to identify possible outliers. In a second step, the average spectrum of each subcluster
228 generated was then calculated and verified. If the average spectra of a subcluster did not match the
229 correct cluster, the spectrum or the spectra group was manually identified and possibly reallocated
230 to another class.

231 *2.7. POSEIDON*

232 The computer programs developed on R software in this publication are integrated into a more
233 global software called POSEIDON (Plastic pOllutionS ExtractIon, DetectiOn and aNalysis). POSEIDON is
234 an under development (Beta version, 2018-2022) free and open source software (Supplementary
235 material). It aims to integrate in a single tool functionalities developed on R software to help in the
236 study of pollution by plastics.

237 **3. RESULTS AND DISCUSSION**

238 *3.1. Test of the machine learning process*

239 The machine learning process test showed efficient learning (Fig. 3). Thus, based on the classification
240 proposed by (Renner et al., 2019), nine classes were placed in the "working well" category (sensitivity
241 greater than 75%). These included “poly(amide)”, “poly(propylene)”, “poly(styrene)” and “cellulose
242 acetate” classes for which sensitivity and specificity were higher than 95%. For morphotypes 1 and 2,
243 as well as for the “poly(ethylene)+fouling” class, the results were also very good with a sensitivity
244 ranging between 87 and 90%. The results for the "ethylene-propylene rubber" class were slightly
245 lower but the sensitivity remained on average above 75%. Four classes fell into the "working"
246 category: “PEVA”, “poly(ethylene) like”, “cellulose like” and “poly(propylene) like”. Finally, the
247 machine learning process did not work actually for four categories: “PMMA”, “poly(urethane)”,
248 “poly(vinylchloride)” and “unknown”.

These results for “poly(ethylene)”, “poly(propylene)” and “polystyrene” were very similar to those obtained by Renner et al. (2017), with sensitivity values close to 100%. However, the KNN method showed higher results (100%) than Renner et al. (2017; 87.5%) and Primpke et al. (2018; 91%) for “polyamide” class. Nonetheless, the results of sensitivity for Renner et al. (2017; 50%), and especially for Primpke et al. (2018; 99%) were better for “poly(urethane)” class.

These differences are mainly due to the difference in methodology and database. For example, in the particular case of "poly(urethane)", it is clearly a lack of spectra in the learning database that explain these poor results. The addition of poly(urethane) spectra collected at sea will quickly improve the results. However, this type of microplastic was relatively rare in the samples collected by Tara Expedition in the studied campaign. The contribution of other teams, that have already acquired this type of FTIR spectra, could be a solution to quickly improve the quality of the database.

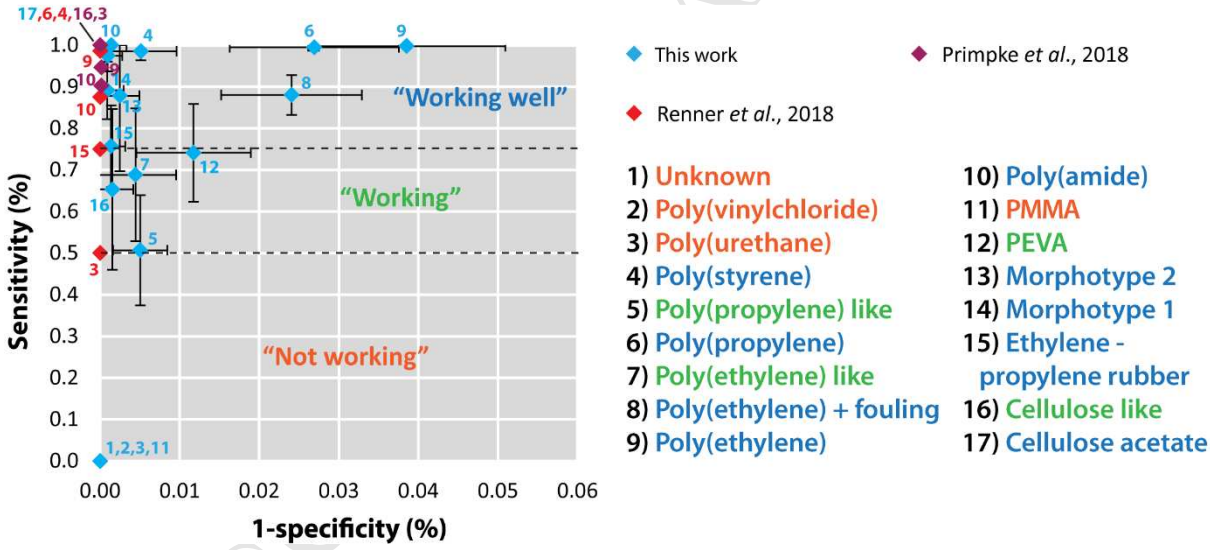

Fig. 3. Sensitivity (see eq. 1) as a function of 1-specificity (see eq. 2).

3.2. Sources of errors

The false negative (spectra placed in the wrong category) of the classes “poly(ethylene)+fouling”, “ethylene-propylene rubber”, and “poly(ethylene vinyl acetate)” were essentially related to spectra placed by the model in the category "poly(ethylene)" (Fig. 4). This type of error was also observed between the “poly(ethyelene vinyl acetate)” and “poly(ethylene)+fouling classes”, as well as

267 between the “poly(ethylene) like” and “poly(ethyelene vinyl acetate)” classes. Thus, it was possible
268 to observe that most of these errors corresponded to misclassified spectra of classes whose chemical
269 nature is very marked by ethylene. The same observation could be made for propylene and vinyl.
270 Thus, spectra of the “ethylene-propylene rubber” and “poly(propylene) like” classes tended to be
271 placed by the model in the “poly(propylene)” class. Most of the spectra of the “poly(vinyl chloride)”
272 class were classified by the model in the “poly(ethylene vinyl acetate)” class. Two reasons can explain
273 this type of errors. First, it was possible that the spectra had too many similarities and so the model
274 failed to distinguish them correctly. This was probably the case, for example, between the
275 “poly(ethylene)+fouling” and “poly(ethylene)” classes, where the differences are only related to the
276 intensity of the band at 1,015 cm^{-1}. It was also the case for the classes of “poly(ethylene)”,
277 “poly(propylene)” and “ethylene-propylene rubber”, for which there are no clear limits. The co-
278 polymer ethylene propylene rubber (EPR) is made from two different monomers (ethylene and
279 propylene), which are also used in the manufacture of poly(ethylene) (PE) and poly(propylene) (PP).
280 Therefore, EPR being a mixture of monomers of ethylene and propylene, it has a transition spectrum
281 between its two "pure" chemical poles, PE and PP. Depending on the desired properties, the
282 proportions between ethylene and propylene are adjusted during manufacture. There are therefore
283 EPRs containing high levels of propylene and others of low levels. These levels will give spectra
284 similar to those of a poly(propylene) if the propylene content in the EPR is high. On the contrary, if it
285 is low, the spectrum is close to that of a PE. When the mixture is balanced, the EPR spectrum has
286 clearly identifiable characteristic bands of ethylene, as well as propylene. It is thus possible to
287 gradually move from the "pure" PE pole to the "pure" PP pole by a set of transition spectra
288 corresponding to different mixtures of PE and PP and where only the intensity of certain bands
289 varied.

290 In some cases, spectra could belong to both classes at the same time. However, it was still necessary
291 to place it, when creating the learning database, in a class rather than in the two. If the model in the
292 machine learning test makes a different choice, it will automatically be counted as an error, but from

the expert's point of view it is not. Thus, the sensitivity of the classes “poly(propylene) like” (51%), “poly(ethylene) like” (69%) or “ethylene-propylene rubber” (76%) was probably reduced by 20 to 30% because of transfers of these transition spectra. Thus, 35.4% of the spectra of the “poly(propylene) like” class was transferred to the “poly(propylene) class”.

A FTIR spectra of a weathered microplastic is composed by three types of bands. The first one is the bands characteristic of the polymer, the second one is linked to the aging of the polymer, and the last one to the presence of fouling. Thus, microplastic spectra may have the same bands related to the presence of organisms on their surface or related to relatively similar aging process. This is probably the case, for example, of certain bands (between 1,650 and 1,700 cm^{-1}) observed in some poly(ethylene) and which could be linked to the presence of proteins on their surface (Maquelin et al., 2002). The impact of biofouling on the spectra of plastics should be better studied. On the one hand, it may be possible to obtain information about living organisms from the surface of microplastics and on the other hand, it is important to better control the risks in terms of misinterpretation that these additional bands can generate. The bands between 1,700 and 1,760 cm^{-1} are another example. These bands, linked to the formation of carbonyl bonds, appear classically when the polymer ages (Andrady, 2011). The aging of the polymer may vary according to its environment and it is possible that the spectra may not be exactly the same as a function of the history of the microplastic. These small potential variations would be worth studying on large databases.

Another source of error was the number of spectra in the database. For some classes, the database did not have enough spectra to allow an effective learning by the model. This was the case for the classes “poly(methyl methacrylate)”, “poly(urethane)”, “poly(vinyl chloride)” and “unknown”, which do not currently have enough spectra to allow learning by the model. In these cases, the progressive enrichment of the database should make it possible to achieve learning outcomes close to those of “poly(ethylene)” or “poly(amide)”.

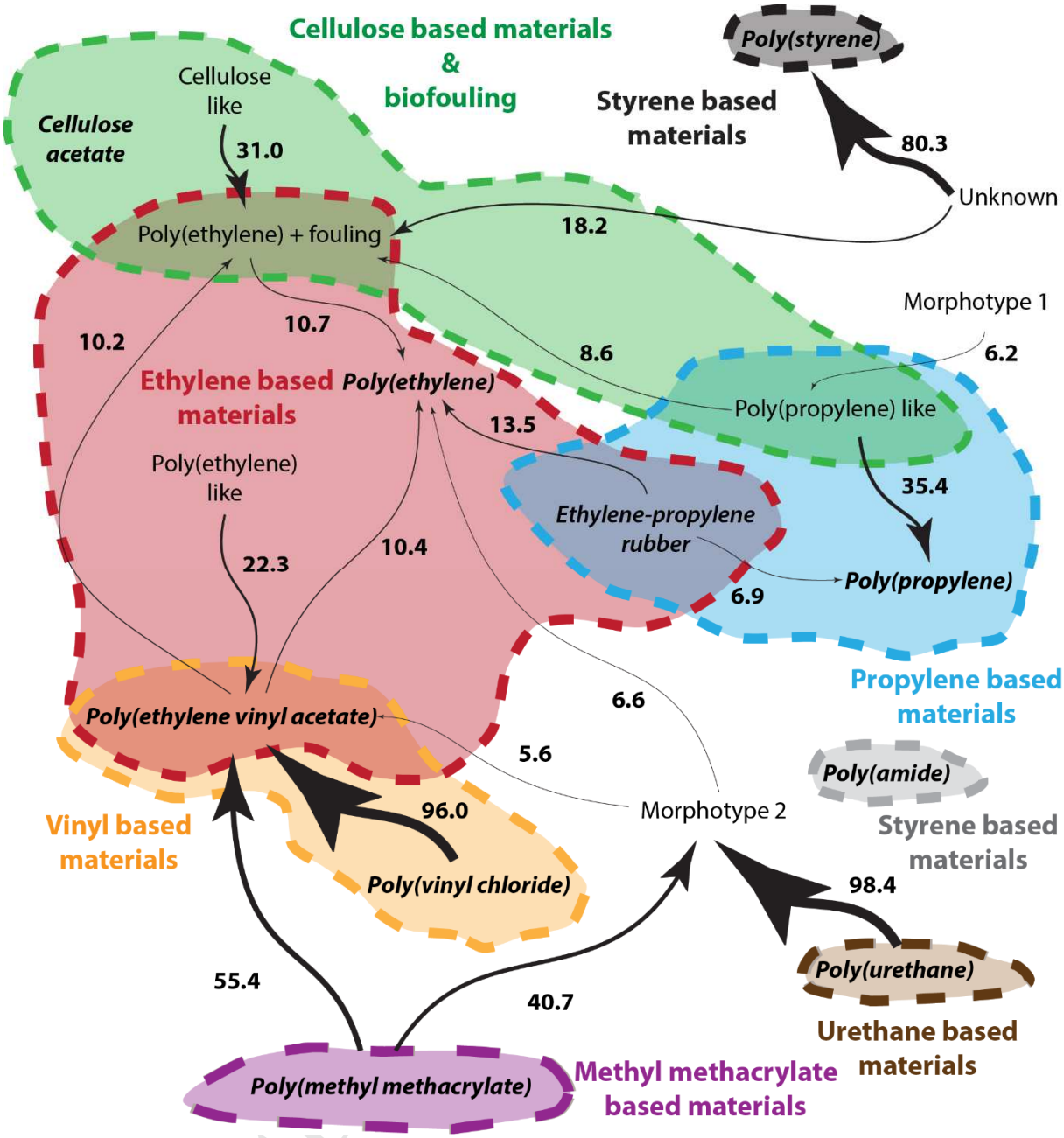

Fig. 4. Percentage of false negatives (greater than 5%) of the different classes in the learning database. The arrows indicate in which class the spectra have been placed by the model.

3.3. Analysis of the full spectra database and verification

After validation of the KNN method, this method was then applied to the remaining 3,095 spectra to be identified. Spectra were classified into 13 classes, 171 sub-clusters whose average spectra, verified during the verification step, were then created. Thus, 9.5% of the average spectra differ in terms of interpretation between the model and the human expert. These differences were explained

by the presence of a significant number of poly(ethylene) spectra in “morphotype 1” class. This error, which was very easily identified by the expert, corresponded to 45% of the errors made by the model. The addition of spectra in the “morphotype 1” class, in future versions of the database, should enable to reduce this kind of error. The second main error done by the model was the attribution to the “poly(propylene)” class of many spectra belonging rather to the “poly(propylene) like” class. This error was more complex to determine by the expert, because it involved spectra that varied subtly to few wavelengths. It represented about 30% of the errors made by the model. The addition of spectra in the “poly(propylene) like” class could be a small improvement in the learning. However, it can be assumed that this class will always be at the origin of more or less significant variation in interpretation between human expertise and the KNN model. In fact, the presence in this class of transition spectra and the significant disparity in the variations observed in the spectra of this class make the machine learning more difficult. This will probably also be the case for ethylene based materials classes. Errors made, related to spectra exchange between these different classes, represented about 11% of the differences in interpretation between the model and the expert. Finally, all other errors represented about 15% of the differences in interpretation.

With a success rate of 90.5%, the automated learning method proposed in this study is slightly inferior to the method proposed by Renner et al. (2017; 96.1%). However, our method has been tested on 3,000 microplastic spectra compared to 300 microplastics previously and by adding a step of rapid visualization and validation of the results proposed by the model. A user can then quickly correct the most obvious model errors and increase the effects results rate to nearly 97%.

After verification and validation of the data, the 4,064 microplastics collected and analyzed by IR spectrophotometry consisted mainly of poly(ethylene) (65.5%), 10.6% of which showed traces of significant fouling and poly(propylene) (21.2%) (Fig. 5). The low represented classes where polystryrene (3%), ethylene-propylene rubber (2.3%) and PEVA (1.5%). Therefore, these results do not stand as the real microplastics type distribution observed in the Mediterranean Sea. The

application of the statistical method previously developed (Falcou-Préfol et al., 2018; Kedzierski et al., 2019) to these data will enable to refine the obtained results and determine the percentages of the various polymers in the North-West Mediterranean basin, as well as manta by manta.

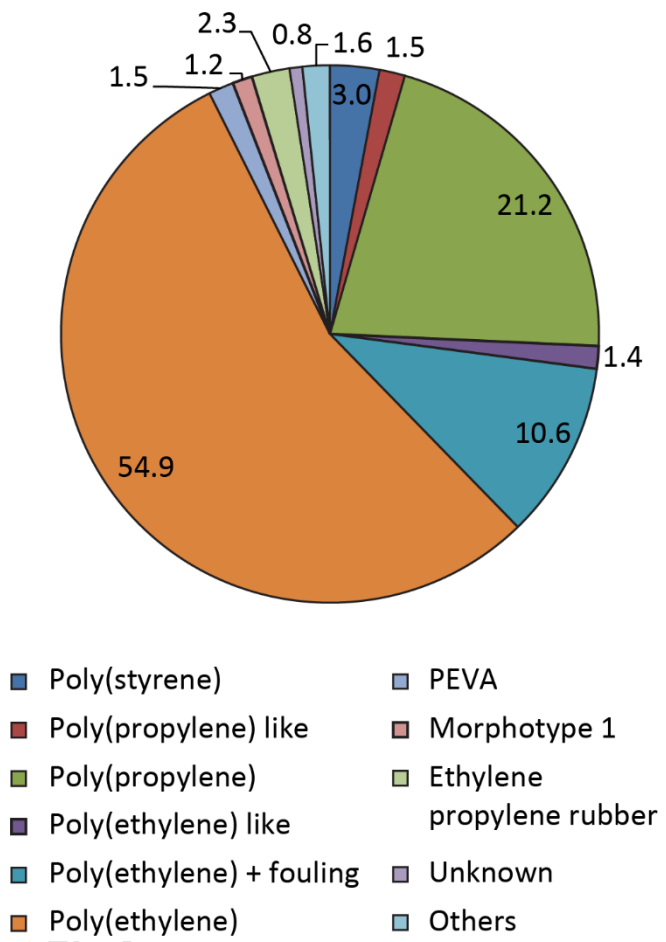

Fig. 5. Distribution of the different classes constituting the database of 4,064 spectra (learning database and undetermined spectra).

3.4. Strengths and areas for improvement

POSEIDON has several advantages over the previously developed tools. First of all, it is a collaborative tool written with R software. The code can therefore be consulted and modified. New R programs facilitating the analysis of microplastic can be integrated into POSEIDON. POSEIDON is therefore a tool designed to evolve over the long term.

In the particular case of modules developed for the automated determination of microplastic spectra, several strengths appear. First of all, the tool allows to process and determine a large number of spectra at once (up to 50,000 currently) which is a real advantage compared to the spectrum-by spectrum determination offered by many specialized software. The work carried out in the study is based on the analysis of 4,064 different particles, which is one of the largest databases on FTIR spectra of microplastics available to date. The learning process is carried out on microplastic spectra collected at sea, which is more realistic than methods based on spectra of virgin sample. Working from the spectra of plastics aged at sea makes it possible to better take into account the fouling and aging of plastics. Finally, through the process proposed in this work, which combines hierarchical clustering and PCA, the user can easily identify if, in the interpretation, there are outliers. This identification is quick and allows the user to judge the interpretations made by the machine. The KNN method applied to IR spectra could also be applied to data obtained using automated methods such as focal plane array (FPA) detector combined with BaF2 window or, with another learning database, to data from Raman spectrometry (Cabernard et al., 2018; Hufnagl et al., 2019; Primpke et al., 2019). Raman spectrometry is indeed often used for the analysis of microplastics and is likely to provide additional information (Frère et al., 2017; Hahn et al., 1997; Hiejima et al., 2018; Zhao et al., 2017). In addition, it has recently been shown that Raman spectrometry can identify microplastics slightly more efficiently than with FTIR analysis (Cabernard et al., 2018).

However, there is some evolution that could improve the program over the time. First of all, the program cannot identify plastics that it does not know or does not know well. This is the case here, for example, of poly(urethane) or PVC. However, it will be possible to identify in the interpretation results an outlier corresponding to a new spectrum and therefore misinterpreted by POSEIDON. To improve learning, it will therefore be necessary to improve the learning database with new spectra. The user can also do this himself if he has additional spectra by adding them in the file "D4_4_publication.csv" (Poseidon_files_V0.1.1.1\Data\IR_References). However, the characteristics of the new spectra must be the same as those of the reference spectra. Another solution may be to

send them to authors with associated metadata, so that they can be added to other reference spectra and distributed to the community. The improvement of the learning database is one of the main research approaches to improve the results. Nevertheless, the addition of new spectra will slow down the learning process as the KNN method, although very efficient, is very computationally intensive. The identification and removal of the least informative spectrum's areas could be a way to compensate in the future the slowdown associated with the addition of new spectra. This research approach has already been explored and seems to yield promising results (Hufnagl et al., 2019; Renner et al., 2019, 2017). Another limitation is the use of a programming language that is not necessarily mastered by the user. This first version of POSEIDON attempts to facilitate its use by limiting the use of computer code by the user. A fully interfaced version of POSEIDON could be developed. It would make it easier to process data and share results in connection with the study of pollution by microplastics without any particular knowledge of a programming language. To this end, adequate resources must be made available to develop and maintain this future version of POSEIDON.

4. CONCLUSION

Sampling microplastics at sea sometimes involves collecting a very large number of particles that are then analysed with spectrophotometry methods. In the case of the Tara Mediterranean Sea campaign, more than 4,000 FTIR spectra have been made. The analysis of these spectra is time consuming and skill dependent. To solve this problem, an automated learning method based on the KNN method has been implemented. The method test showed a high efficiency of this method on certain polymers (e.g. poly(ethylene), poly(propylene), poly(amide)). For others, however, such as poly(vinyl chloride), the learning database lacked spectra for the machine learning to be effective. Additional spectra need to be added to the learning database in the near future to strengthen the results. In view of the FTIR analyses carried out on microplastics by other studies, other FTIR spectrum databases exist. Sharing them would make it possible to quickly and significantly increase

the current database. Nevertheless, the application of this method on more than 4,000 spectra confirmed its high efficiency. The synthetic verification step, based on different clustering methods, further improves the good results ratio.

Another research approach would be to integrate the various programs developed during this study, as well as those previously developed (Kedzierski et al., 2019, 2018, 2017), into the POSEIDON software. Thus, POSEIDON would allow people working on microplastics to use a set of digital tools to facilitate their work.

Acknowledgments

We thank the commitment of the following institutions, persons and sponsors: CNRS, UPMC, LOV, Genoscope/CEA, the Tara Expeditions Foundation and its founders: agnès b.®, Etienne Bourgois, Romain Troublé, the Veolia Environment Foundation, Lorient Agglomeration, Serge Ferrari, the Foundation Prince Albert II de Monaco, IDEC, the "Tara" schooner and teams. We thank MERCATOR-CORIOLIS and ACRI-ST for providing daily satellite data during the expedition. We are also grateful to the French Ministry of Foreign Affairs for supporting the expedition and to the countries that graciously granted sampling permission. This article is contribution number 4 of Tara Mediterranean.

References

Albertsson, A.C., Barenstedt, C., Karlsson, S., Lindberg, T., 1995. Degradation product pattern and morphology changes as means to differentiate abiotically and biotically aged degradable polyethylene. Polymer (Guildf). 36, 3075–3083. https://doi.org/10.1016/0032-3861(95)97868-G

Amaral-Zettler, L.A., Zettler, E.R., Slikas, B., Boyd, G.D., Melvin, D.W., Morrall, C.E., Proskurowski, G., Mincer, T.J., 2015. The biogeography of the Plastisphere: implications for policy. Front. Ecol. Environ. 13, 541–546. https://doi.org/10.1890/150017

Andrady, A.L., 2017. The plastic in microplastics: A review. Mar. Pollut. Bull. 119, 12–22.

https://doi.org/https://doi.org/10.1016/j.marpolbul.2017.01.082

Andrady, A.L., 2011. Microplastics in the marine environment. Mar. Pollut. Bull. 62, 1596–1605. https://doi.org/10.1016/j.marpolbul.2011.05.030

Barboza, L.G., Gimenez, B.C., 2015. Microplastics in the marine environment: Current trends and future perspectives. Mar. Pollut. Bull. 97, 5–12. https://doi.org/10.1016/j.marpolbul.2015.06.008

Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The New S Language: A Programming Environment for Data Analysis and Graphics, Wadsworth & Brooks/Cole computer science series. Chapman & Hall.

Brach, L., Deixonne, P., Bernard, M.-F., Durand, E., Desjean, M.-C., Perez, E., van Sebille, E., ter Halle, A., 2018. Anticyclonic eddies increase accumulation of microplastic in the North Atlantic subtropical gyre. Mar. Pollut. Bull. 126, 191–196. https://doi.org/https://doi.org/10.1016/j.marpolbul.2017.10.077

Cabernard, L., Roscher, L., Lorenz, C., Gerdts, G., Primpke, S., 2018. Comparison of Raman and Fourier Transform Infrared Spectroscopy for the Quantification of Microplastics in the Aquatic Environment. Environ. Sci. Technol. 52, 13279–13288. https://doi.org/10.1021/acs.est.8b03438

Dümichen, E., Barthel, A.K., Braun, U., Bannick, C.G., Brand, K., Jekel, M., Senz, R., 2015. Analysis of polyethylene microplastics in environmental samples, using a thermal decomposition method. Water Res. 85, 451–457. https://doi.org/10.1016/j.watres.2015.09.002

Dümichen, E., Eisentraut, P., Bannick, C.G., Barthel, A.K., Senz, R., Braun, U., 2017. Fast identification of microplastics in complex environmental samples by a thermal degradation method. Chemosphere 174, 572–584. https://doi.org/10.1016/j.chemosphere.2017.02.010

Dussud, C., Meistertzheim, A.L., Conan, P., Pujo-Pay, M., George, M., Fabre, P., Coudane, J., Higgs, P., Elineau, A., Pedrotti, M.L., Gorsky, G., Ghiglione, J.F., 2018. Evidence of niche partitioning

among bacteria living on plastics, organic particles and surrounding seawaters. Environ. Pollut. 236, 807–816. https://doi.org/https://doi.org/10.1016/j.envpol.2017.12.027

Elert, A.M., Becker, R., Duemichen, E., Eisentraut, P., Falkenhagen, J., Sturm, H., Braun, U., 2017. Comparison of different methods for MP detection: What can we learn from them, and why asking the right question before measurements matters? Environ. Pollut. 231, 1256–1264. https://doi.org/https://doi.org/10.1016/j.envpol.2017.08.074

Eriksen, M., Mason, S., Wilson, S., Box, C., Zellers, A., Edwards, W., Farley, H., Amato, S., 2013. Microplastic pollution in the surface waters of the Laurentian Great Lakes. Mar. Pollut. Bull. 77, 177–182. https://doi.org/10.1016/j.marpolbul.2013.10.007

Falcou-Préfol, M., Kedzierski, M., Villain, J., Kerros, M.E., Elineau, A., Pedrotti, M.L., Bruzaud, S., Cocca, M., Di Pace, E., Errico, M.E., Gentile, G., Montarsolo, A., Mossotti, R., 2018. Statistical Methodology for Identifying Microplastic Samples Collected During TARA Mediterranean Campaign. Springer International Publishing, Cham, pp. 31–35.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27, 861–874. https://doi.org/https://doi.org/10.1016/j.patrec.2005.10.010

Fotopoulou, K.N., Karapanagioti, H.K., 2012. Surface properties of beached plastic pellets. Mar. Environ. Res. 81, 70–77. https://doi.org/https://doi.org/10.1016/j.marenvres.2012.08.010

Free, C.M., Jensen, O.P., Mason, S.A., Eriksen, M., Williamson, N.J., Boldgiv, B., 2014. High-levels of microplastic pollution in a large, remote, mountain lake. Mar. Pollut. Bull. 85, 156–163. https://doi.org/https://doi.org/10.1016/j.marpolbul.2014.06.001

Frère, L., Maignien, L., Chalopin, M., Huvet, A., Rinnert, E., Morrison, H., Kerninon, S., Cassone, A.-L., Lambert, C., Reveillaud, J., Paul-Pont, I., 2018. Microplastic bacterial communities in the Bay of Brest: Influence of polymer type and size. Environ. Pollut. 242, 614–625. https://doi.org/https://doi.org/10.1016/j.envpol.2018.07.023

Frère, L., Paul-Pont, I., Moreau, J., Soudant, P., Lambert, C., Huvet, A., Rinnert, E., 2016. A semi-automated Raman micro-spectroscopy method for morphological and chemical characterizations of microplastic litter. Mar. Pollut. Bull. 113, 461–468. https://doi.org/10.1016/j.marpolbul.2016.10.051

Gewert, B., Ogonowski, M., Barth, A., MacLeod, M., 2017. Abundance and composition of near surface microplastics and plastic debris in the Stockholm Archipelago, Baltic Sea. Mar. Pollut. Bull. In Press. https://doi.org/https://doi.org/10.1016/j.marpolbul.2017.04.062

Gimeno, P., Spinau, C., Lassu, N., Maggio, A.F., Brenier, C., Lempereur, L., 2015. Identification and quantification of bisphenol A and bisphenol B in polyvinylchloride and polycarbonate medical devices by gas chromatography with mass spectrometry. J Sep Sci 38, 3727–3734. https://doi.org/10.1002/jssc.201500552

Howe, K.J., Ishida, K.P., Clark, M.M., 2002. Use of ATR/FTIR spectrometry to study fouling of microfiltration membranes by natural waters. Desalination 147, 251–255. https://doi.org/https://doi.org/10.1016/S0011-9164(02)00545-3

Hufnagl, B., Steiner, D., Renner, E., Löder, M.G.J., Laforsch, C., Lohninger, H., 2019. A methodology for the fast identification and monitoring of microplastics in environmental samples using random decision forest classifiers. Anal. Methods. https://doi.org/10.1039/C9AY00252A

Imhof, H.K., Ivleva, N.P., Schmid, J., Niessner, R., Laforsch, C., 2013. Contamination of beach sediments of a subalpine lake with microplastic particles. Curr. Biol. 23, R867–R868.

Imhof, H.K., Schmid, J., Niessner, R., Ivleva, N.P., Laforsch, C., 2012. A novel, highly efficient method for the separation and quantification of plastic particles in sediments of aquatic environments. Limnol. Oceanogr. Methods 10, 524–537.

Imhof, H.K., Sigl, R., Brauer, E., Feyl, S., Giesemann, P., Klink, S., Leupolz, K., Loder, M.G., Loschel, L.A., Missun, J., Muszynski, S., Ramsperger, A.F., Schrank, I., Speck, S., Steibl, S., Trotter, B., Winter, I.,

Laforsch, C., 2017. Spatial and temporal variation of macro-, meso- and microplastic abundance on a remote coral island of the Maldives, Indian Ocean. Mar. Pollut. Bull. 116, 340–347. https://doi.org/10.1016/j.marpolbul.2017.01.010

Ioakeimidis, C., Fotopoulou, K.N., Karapanagioti, H.K., Geraga, M., Zeri, C., Papathanassiou, E., Galgani, F., Papatheodorou, G., 2016. The degradation potential of PET bottles in the marine environment: An ATR-FTIR based approach. Sci. Rep. 6, 1–8. https://doi.org/10.1038/srep23501

Kanhai, L.D.K., Officer, R., Lyashevska, O., Thompson, R.C., O'Connor, I., 2017. Microplastic abundance, distribution and composition along a latitudinal gradient in the Atlantic Ocean. Mar. Pollut. Bull. 115, 307–314. https://doi.org/https://doi.org/10.1016/j.marpolbul.2016.12.025

Käppler, A., Fischer, D., Oberbeckmann, S., Schernewski, G., Labrenz, M., Eichhorn, K.-J., Voit, B., 2016. Analysis of environmental microplastics by vibrational microspectroscopy: FTIR, Raman or both? Anal. Bioanal. Chem. 408, 8377–8391. https://doi.org/10.1007/s00216-016-9956-3

Kedzierski, M., D'Almeida, M., Magueresse, A., Le Grand, A., Duval, H., Cesar, G., Sire, O., Bruzaud, S., Le Tilly, V., 2018. Threat of plastic ageing in marine environment. Adsorption/desorption of micropollutants. Mar. Pollut. Bull. 127, 684–694. https://doi.org/10.1016/j.marpolbul.2017.12.059

Kedzierski, M., Le Tilly, V., Bourseau, P., Bellegou, H., Cesar, G., Sire, O., Bruzaud, S., 2017. Microplastics elutriation system. Part A: Numerical modeling. Mar. Pollut. Bull. 119, 151–161. https://doi.org/10.1016/j.marpolbul.2017.04.060

Kedzierski, M., Villain, J., Falcou-Préfol, M., Kerros, M.E., Henry, M., Pedrotti, M.L., Bruzaud, S., 2019. Microplastics in Mediterranean Sea: A protocol to robustly assess contamination characteristics. PLoS One 14, e0212088. https://doi.org/10.1371/journal.pone.0212088

Küpper, L., Gulmine, J. V., Janissek, P.R., Heise, H.M., 2004. Attenuated total reflection infrared spectroscopy for micro-domain analysis of polyethylene samples after accelerated ageing

within weathering chambers. Vib. Spectrosc. 34, 63–72. https://doi.org/10.1016/j.vibspec.2003.05.002

Lenz, R., Enders, K., Stedmon, C.A., Mackenzie, D.M., Nielsen, T.G., 2015. A critical assessment of visual identification of marine microplastic using Raman spectroscopy for analysis improvement. Mar. Pollut. Bull. 100, 82–91. https://doi.org/10.1016/j.marpolbul.2015.09.026

Liland, K.H., 2015. 4S Peak Filling – baseline estimation by iterative mean suppression. MethodsX 2, 135–140. https://doi.org/10.1016/j.mex.2015.02.009

Löder, M.G.J., Gerdts, G., 2015. Methodology Used for the Detection and Identification of Microplastics—A Critical Appraisal, in: Bergmann, M., Gutow, L., Klages, M. (Eds.), Marine Anthropogenic Litter. Springer International Publishing, Cham, pp. 201–227. https://doi.org/10.1007/978-3-319-16510-3_8

Löder, M.G.J., Kuczera, M., Mintenig, S., Lorenz, C., Gerdts, G., 2015. Focal plane array detector-based micro-Fourier-transform infrared imaging for the analysis of microplastics in environmental samples. Environ. Chem. 12, 563–581. https://doi.org/https://doi.org/10.1071/EN14205

Maquelin, K., Kirschner, C., Choo-Smith, L.P., van den Braak, N., Endtz, H.P., Naumann, D., Puppels, G.J., 2002. Identification of medically relevant microorganisms by vibrational spectroscopy. J. Microbiol. Methods 51, 255–271. https://doi.org/https://doi.org/10.1016/S0167-7012(02)00127-6

Obbard, R.W., Sadri, S., Wong, Y.Q., Khitun, A.A., Baker, I., Thompson, R.C., 2014. Global warming releases microplastic legacy frozen in Arctic Sea ice. Earth's Futur. 2, 315–320. https://doi.org/doi:10.1002/2014EF000240

Pedrotti, M.L., Mazzocchi, M.G., Lombard, F., Galgani, F., Kerros, M.E., Henry, M., Elineau, A., Petit, S., Fernandez-de-Puelles, M.L., Gasparini, S., Tirelli, V., Jamet, J.-L., Gorsky, G., 2018. TARA

Mediterranean Expedition: Assessing the Impact of Microplastics on Mediterranean Ecosystem, in: Cocca, M., Di Pace, E., Errico, M.E., Gentile, G., Montarsolo, A., Mossotti, R. (Eds.), Proceedings of the International Conference on Microplastic Pollution in the Mediterranean Sea. Springer International Publishing, Cham, pp. 25–29.

Pedrotti, M.L., Petit, S., Elineau, A., Bruzaud, S., Crebassa, J.C., Dumontet, B., Marti, E., Gorsky, G., Cozar, A., 2016. Changes in the Floating Plastic Pollution of the Mediterranean Sea in Relation to the Distance to Land. PLoS One 11, e0161581. https://doi.org/10.1371/journal.pone.0161581

Peeken, I., Primpke, S., Beyer, B., Gütermann, J., Katlein, C., Krumpen, T., Bergmann, M., Hehemann, L., Gerdts, G., 2018. Arctic sea ice is an important temporal sink and means of transport for microplastic. Nat. Commun. 9, 1505. https://doi.org/10.1038/s41467-018-03825-5

Primpke, S., A. Dias, P., Gerdts, G., 2019. Automated identification and quantification of microfibres and microplastics. Anal. Methods 11, 2138–2147. https://doi.org/10.1039/C9AY00126C

Primpke, S., Lorenz, C., Rascher-Friesenhausen, R., Gerdts, 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., Wirth, M., Lorenz, C., Gerdts, G., 2018. Reference database design for the automated analysis of microplastic samples based on Fourier transform infrared (FTIR) spectroscopy. Anal. Bioanal. Chem. 410, 5131–5141. https://doi.org/10.1007/s00216-018-1156-x

Qiu, Q., Tan, Z., Wang, J., Peng, J., Li, M., Zhan, Z., 2016. Extraction, enumeration and identification methods for monitoring microplastics in the environment. Estuar. Coast. Shelf Sci. 176, 102–109. https://doi.org/10.1016/j.ecss.2016.04.012

Renner, G., Nellessen, A., Schwiers, A., Wenzel, M., Schmidt, T.C., Schram, J., 2019. Data preprocessing & evaluation used in the microplastics identification process: A critical review &

practical guide. TrAC - Trends Anal. Chem. 111, 229–238. https://doi.org/10.1016/j.trac.2018.12.004

Renner, G., Schmidt, T.C., Schram, J., 2017. A New Chemometric Approach for Automatic Identification of Microplastics from Environmental Compartments Based on FT-IR Spectroscopy. Anal. Chem. 89, 12045–12053. https://doi.org/10.1021/acs.analchem.7b02472

Ripley, B.D., 1996. Pattern recognition and neural networks. Cambridge University Press, Cambridge ; New York.

Ripley, B.D., Wiley, J., 1987. Stochastic Simulation. Wiley.

Rocha-Santos, T., Duarte, A.C., 2015. A critical overview of the analytical approaches to the occurrence, the fate and the behavior of microplastics in the environment. TrAC Trends Anal. Chem. 65, 47–53. https://doi.org/https://doi.org/10.1016/j.trac.2014.10.011

Schlitzer, R., 2015. Data Analysis and Visualization with Ocean Data View. Can. Meteorol. Oceanogr. Soc. 43, 9–13.

Sudhakar, M., Trishul, A., Doble, M., Suresh Kumar, K., Syed Jahan, S., Inbakandan, D., Viduthalai, R.R., Umadevi, V.R., Sriyutha Murthy, P., Venkatesan, R., 2007. Biofouling and biodegradation of polyolefins in ocean waters. Polym. Degrad. Stab. 92, 1743–1752. https://doi.org/https://doi.org/10.1016/j.polymdegradstab.2007.03.029

The R Core Team, 2019. R: A Language and Environment forStatistical Computing.

Thompson, R.C., Olsen, Y., Mitchell, R.P., Davis, A., Rowland, S.J., John, A.W.G., McGonigle, D., Russell, A.E., 2004. Lost at sea: where is all the plastic? Science (80-.). 304, 838.

Venables, W.N., Ripley, B.D., Venables, W.N., 2002. Modern applied statistics with S, 4th ed, Statistics and computing. Springer, New York.

Wang, J., Peng, J., Tan, Z., Gao, Y., Zhan, Z., Chen, Q., Cai, L., 2017. Microplastics in the surface

sediments from the Beijiang River littoral zone: Composition, abundance, surface textures and interaction with heavy metals. Chemosphere 171, 248–258. https://doi.org/10.1016/j.chemosphere.2016.12.074

Wehrens, R., 2011. Chemometrics with R Multivariate Data Analysis in the Natural Sciences and Life Sciences. Use R.

Zettler, E.R., Mincer, T.J., Amaral-Zettler, L.A., 2013. Life in the “plastisphere”: Microbial communities on plastic marine debris. Environ. Sci. Technol. 47, 7137–7146. https://doi.org/10.1021/es401288x

Zhao, S., Danley, M., Ward, J.E., Li, D., Mincer, T.J., 2017. An approach for extraction, characterization and quantitation of microplastic in natural marine snow using Raman microscopy. Anal. Methods 9, 1470–1478. https://doi.org/10.1039/c6ay02302a

ACCEPTED MANUSCRIPT