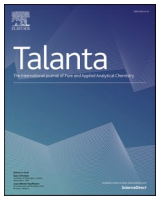
Talanta 298 (2026) 128834



Contents lists available at ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta



Inter-instrument definition of valid criteria for the automatic identification of microplastics by micro-Raman spectroscopy

Rafaela Fernandes [a], Paul-Tiberiu Miclea [b], Marta Fadda [c], Mara Putzu [c], A. Sacco [c], Andrea M. Rossi [c], Andrea M. Giovannozzi [c], Marta Barbaresi [d], Matteo Masino [d], Monica Mattarozzi [d], Maria Careri [d], Carla Palma [e], José Almeida [e], Claudia Drago [f], Olivier Pellegrino [g], Raquel Quendera [g], Ulrike Braun [f], Ricardo J.N. Bettencourt da Silva [a,*]

[a] *Centro de Química Estrutural, Institute of Molecular Sciences, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal*
[b] *Fraunhofer Center for Silicon Photovoltaics (CSP), Germany*
[c] *Istituto Nazionale di Ricerca Metrologica (INRiM), Italy*
[d] *Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy*
[e] *Instituto Hidrográfico, Portugal*
[f] *Umweltbundesamt, German Environment Agency, Germany*
[g] *Instituto Português da Qualidade, Portugal*

ARTICLE INFO

Handling editor: A Campiglia

Keywords:
Microplastics
Automatic identification
μ-Raman spectroscopy
Match threshold
Method validation
True results rate

ABSTRACT

The assessment of the impact of microplastic contamination on the environment and human health requires a reliable identification of the polymer type of these particles. μ-Raman spectroscopy is a popular technique for identifying microplastics by comparing the reference spectra with those of the particles analysed. Automatic identification of microplastics requires defining an algorithm for the match between these spectra and setting a minimum match above which identification is performed with adequately high true and low false results rates. Ideally, the algorithm and match threshold should apply to different spectrometers and spectra collection parameters. This research presents a methodology to determine the best match algorithm for polymer identification using μ-Raman spectroscopy data collected on different instruments and laboratories, associated with a true positive rate (*TP*) of 95 % and a false positive rate (*FP*) lower than 5 %. Determining the match threshold (*P*5»P) by the bootstrap method does not require assumptions regarding match distribution. The normal distribution of the match between the reference and a particle's spectra from a different material allows *FP* determination. Identifying PET microparticles was optimal using Pearson's correlation coefficient (*P*5»P = 0.6244, *TP* = 95 %, $FP = 4.9 \times 10^{-7}$ %). Identification quality was tested based on three unweighted and three weighted correlation coefficients. Spectra with signal-to-noise ratios lower than 10 were forwarded to manual identifications. The MS Excel files used in the research are available as supporting information. The developed methodology for setting up identification criteria of microplastics by spectroscopy proved to be adequate for μ-Raman assessments and robust to different spectrometers and spectra collection conditions.

1. Introduction

The characteristics of plastic materials that drive their commercial success, namely low cost and high chemical and physical resistance, combined with the inadequate disposal of these inexpensive materials, contribute to the ubiquitous presence of long-lasting plastic particles in the environment. The type, size, and shape of plastic materials influence their environmental impact, with smaller particles even able to travel inside living organisms. While significant research has evaluated the abundance of plastic waste [1,2], its impact on ecosystems and human health remains largely unknown due to the diversity and unpredictable consequences of this type of pollution. Plastic particles can be classified according to their size as macroplastics (>25 mm), mesoplastics (from 5 mm to 25 mm), microplastics (from 1 μm to 5 mm) and nanoplastics (<1

* Corresponding author.
E-mail address: rjsilva@ciencias.ulisboa.pt (R.J.N. Bettencourt da Silva).

https://doi.org/10.1016/j.talanta.2025.128834
Received 3 July 2025; Received in revised form 4 September 2025; Accepted 6 September 2025
Available online 9 September 2025
0039-9140/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

μm) [3].

Parallel to determining the toxicological impact of these particles, it is necessary to determine their abundance, respective trends, and the most relevant sources of contamination by these materials. This information should support the establishment and monitoring of policies to mitigate the environmental and health impacts of plastic pollution.

The harmonisation of procedures for monitoring plastic contamination in food products and environmental matrices is still lacking. Additionally, further developments are needed to assess the performance and evaluate the uncertainty associated with these monitoring methods [4,5].

When analysing food or environmental samples, it is necessary to identify which particles are indeed plastic before counting them. μ-FTIR (Fourier-transform infrared spectroscopy) and μ-Raman are the most popular tools for plastic particle identification before characterising their size and shape [6].

Compared to μ-FTIR spectroscopy, μ-Raman allows for the identification of smaller particles — down to the micrometer and even sub-micrometer scale — that are beyond the capabilities of FTIR [7]. Raman spectroscopy is also largely unaffected by the strong water absorption that constrains FTIR, making it particularly well suited for the analysis of aquatic samples [8]. In addition, it produces sharp and polymer-specific vibrational spectra, which facilitates better differentiation of chemically similar polymers than FTIR. Despite these advantages, Raman spectroscopy can be affected by fluorescence interference arising from dyes, additives, or biofilms, which may mask characteristic signals and hinder accurate polymer identification [9,10].

The spectroscopic identification of microplastics involves comparing the unknown particle spectrum with reference polymer spectra. Reference and particle spectra can be compared manually or automatically. While manual identifications are time-consuming and must be performed by qualified analysts, automatic identifications are faster and require less analytical expertise. Given the large number of plastic particles observed in some samples, automatic identifications are thus recommended. The most common method for quantifying the similarity between particle and reference spectra is the determination of their correlation using a correlation coefficient, such as the Pearson or Spearman correlation coefficient. The popularity of this method stems from its computational efficiency and the ease of interpreting the coefficients. However, the decrease in correlation values due to fluorescence interferences, baseline variations, signal noise, and particle weathering must be considered in data interpretation. Weighted correlation coefficients can be used to highlight relevant spectra features, improving identification reliability.

Spectral similarity can also be assessed using more complex chemometric tools, such as Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA). More recently, machine learning (ML), through techniques such as Neural Networks (NN), has been used to establish flexible and accurate methods for identifying microplastics even in the presence of noise and spectral overlap. Nevertheless, ML approaches require substantial computational resources, large training datasets, careful validation, and computational expertise, and they are often less transparent than traditional correlation methods. Therefore, the overcome of correlation coefficient fragilities in microplastics identification is particularly welcome.

A study by Jin et al. [11] used Raman spectra combined with PCA-LDA (LDA - Linear Discriminant analysis) followed by Support Vector Machine (SVM) classification to differentiate seven common polymers, i.e. polypropylene (PP), polyethylene terephthalate (PET), polyvinyl chloride (PVC), polycarbonate (PC), polyamide (PA), high density polyethylene (HDPE), and low density polyethylene (LDPE), reporting fitting accuracies over 98 % (ratio between true and total performed identifications) for most polymers and around 70 % for HDPE and LDPE. A study published by Xie et al. [12] extended these methods to nanoplastics, using Raman data with Random Forest (RF) classification to achieve an average accuracy of 98.8 % in identifying individual nanoplastic particles.

A more recent work by Zhang et al. [13] introduced a one-dimensional Convolutional Neural Network (1DCNN) trained on Raman spectra from ten polymer types, achieving a classification accuracy of approximately 96.4 %. Additionally, a CNN model applied to Surface-Enhanced Raman Scattering (SERS) spectra of six common microplastics demonstrated a mean identification accuracy of 99.5 %, outperforming classical classifiers such as SVM, PCA-LDA, PLS-DA, RF, and K-Nearest Neighbours [14].

Studies using hyperspectral Raman imaging paired with multivariate curve resolution have enabled visual mapping of microplastic particles in complex matrices, though classification metrics are less frequently reported [15]. Recent comprehensive reviews confirm that combining Raman spectroscopy with ML techniques — especially PCA, RF, SVM, and deep learning architectures — yields robust identification across varying environmental matrices [16].

Regarding performance statistics, supervised chemometric models consistently achieve high true positive rates (*TP*) (≥98 %) and low false positive rates (*FP*) (≤1 %), while deep learning approaches maintain strong *TP* (around 96 % to 99 %) with comparably low *FP*. Techniques based on SERS combined with CNNs have shown exceptional accuracy (~99.5 %) even with unpreprocessed data.

However, the listed chemometric tools require extensive data preprocessing, which is performed by complex software that consumes relevant computational resources. On the other hand, correlation coefficients are straightforward and computationally efficient, making them the preferred tool for rapid comparisons. Correlation determination is also easier to disseminate in harmonised analysis protocols.

Some guidelines and standards designate the quantified spectral similarity as a Match or Hit Quality Index (HQI) with a maximum value of 100 %. Unfortunately, HQIs are frequently presented and discussed without specifying the algorithm or computational tool used. This omission makes it impossible to compare values determined by different authors and to discern which spectral characteristics contribute most to the similarity value. Some documents even define thresholds for reliable identifications, such as HQI >60 % [17] or HQI >80 % [18], without specifying how HQI should be determined and what data and statistics support the threshold. The harmonisation of protocols is only effective if supported by adequately described and performing protocols applicable to data collected in various laboratories using different instruments. To protect match determinations from noisy signals, a signal-to-noise ratio (S/N) threshold can be defined, below which spectra should be assessed manually or using an alternative tool.

Morgado et al. [4,5] developed a strategy for the automatic identification of microparticles isolated from river and marine sediments based on μ-FTIR, where reference and particle spectra were collected using the same equipment and spectral parameters. Microparticles were manually identified and designated as positive or negative cases based on whether they were from the specific polymer type being studied or not, respectively. For instance, when identifying PET particles, all particles from this polymer were considered positive cases, while all non-PET particles were considered negative cases. After selecting a specific match algorithm and reference spectrum, match values between the reference and positive or negative case spectra were determined. The 5^{th} percentile of match values for positive cases, determined by the bootstrap method [19], was used as the minimum match for polymer identification, with a true positive rate (*TP*) of 95 %. When the particle and reference are from the same polymer, in 95 % of cases, the match exceeds the threshold. This threshold was subsequently tested for the chance of a particle from a material different from the reference polymer producing a spectrum with a match with the reference spectra above the threshold, i.e. the false positive rate (*FP*). Assuming the normal distribution of the match with negative cases, this probability was estimated using regular parametric statistics. An identification method applicable to microplastic identification in environmental samples is considered adequate if associated with a $TP = 95$ % and $FP \leq 5$ % [20]. Various

algorithms for match determination were tested to identify signal transformations and correlation formulas that drive lower *FP*. Since spectral comparison is affected by the low intensity of FTIR bands and the absorbance band from non-oxidised biofilm, a minimum value for the most intense spectral band and a maximum intensity for biofilm bands were defined to exclude spectra from automatic identification. Such spectra should undergo manual identification due to observed signal fragilities.

This paper advances the state-of-the-art in microplastic identification by establishing identification criteria applicable to μ-Raman spectra collected using different instruments and experimental conditions. The μ-Raman spectra improve identification due to their high selectivity and narrow peaks, compared with the broader spectral bands observed in FTIR. An algorithm for determining the S/N was developed to identify spectra that require manual identification by experienced analysts. In this study, PET was chosen as a representative polymer to assess the viability of the approach to determine match thresholds and evaluate match methods, as a result of its extensive use in food packaging and its limited stability under hard oxidative treatments, high temperatures or strong alkaline conditions. Furthermore, the reference material provides a realistic representation of the morphological variability of microplastics, including different shapes and size distribution, commonly found in environmental and food samples. Spectra collected from three different laboratories and in two matrices, ultrapure water and milk, were studied to assess the robustness of the methodology for these factors. The assessment of the ability of the developed method to identify other polymers and weathered particles requires collecting additional spectra and applying a method development and validation equivalent to that performed in this research.

2. Experimental

2.1. Particles and spectra collection conditions

From the 175 particles analysed, 93 were PET and 82 were non-PET. PET particles were provided by the German Federal Institute for Materials Research and Testing (BAM) within the framework of the EU-funded PlasticTrace project. These reference materials were in the form of soluble tablets containing secondary PET microparticles, with irregular shapes and sizes ranging from 10 μm to 100 μm.

Non-PET plastic polymers were analysed in different physical forms (e.g., pellets, powders, films) with a size distribution from 0.05 mm to 5 mm.

While most PET tablets analysed were dissolved in ultrapure water, some were dissolved in powdered milk (infant formula) after reconstitution in ultrapure water. As a result, 4 out of the 175 Raman spectra were acquired from particles in the milk-derived matrix.

The Raman spectrometers and their instrumental conditions used by the three laboratories (Lab. 1 to Lab. 3) are listed in Table 1. The table specifies the wavenumber interval and number of spectral points (Detector pixels) of the original spectra.

Although the Diffraction Grating expresses the dispersion of wavenumber per length unit, since detector pixels can have different lengths, this parameter is not directly related to the spectral resolution. Instead, the ratio between the spectral range (cm^{-1}) and the number of detector pixels, the Spectral Sampling Interval, $\Delta\nu$, is a better way to express spectral resolution. In this work, the spectra were collected with a $\Delta\nu$ between 1.19 cm^{-1} and 5.22 cm^{-1}, subsequently converted to $\Delta\nu =$ 6.51 cm^{-1} in the interval 856 cm^{-1} and 1787 cm^{-1}, considering 143 equidistant points/pixels to allow for spectrum comparison. An Excel spreadsheet was used for spectra harmonisation. No obvious resolution loss was observed when comparing spectra sampled at $\Delta\nu = 1.19$ cm^{-1} and 6.51 cm^{-1}.

Since the Raman shift is referenced to the excitation wavelength (in this work, 514 nm, 532 nm, or 633 nm), although different lasers can produce the same Raman shift peaks, their relative intensities can differ due to resonance effects, fluorescence background, or absorption. Therefore, the different laser wavelengths of the used spectrometers impact the comparability of the spectra.

Though the Raman signal is proportional to the laser output power, excessively high power may excite fluorescence that can overwhelm the Raman peaks. Too much power can also cause photodegradation of the sample. Therefore, this parameter should be optimized to achieve an acceptable S/N without damaging the sample.

The microscope objective magnification (*MOM*) affects the collection of spectra from particles smaller than the scanned area, but it does not influence the Raman shift positions, the relative intensities of the peaks, or the spectral resolution. Depending on the *MOM*, the numerical aperture determines how tightly the laser is focused and how efficiently scattered light is collected, and thus requires optimization to achieve a better S/N ratio.

Longer acquisition times and signal accumulation reduce the signal's noise. Very short acquisition times can hinder peak detectability, while excessively long acquisition times may cause detector saturation, nonlinearity, or sample damage, which can compromise polymer identification [21].

2.2. Method validation strategy

The validation of the method for identifying PET microparticles by μ-Raman spectroscopy involved collecting spectra from PET and non-PET particles using the spectrometers and instrumental conditions listed in Section 2.1. The particles were manually identified, considering at least the three most intense characteristic Raman spectra peaks of PET occurring at 1615 cm^{-1}, 1120 cm^{-1}, and 1000 cm^{-1}. The Raman shift range for spectra comparison was limited between 856 cm^{-1} and 1787 cm^{-1}. All collected spectra are made available as Supplementary Material 01 being identified with the following code: "A B (C)", where A identifies the polymer type ("PET", "non-PET" or specified polymer acronym such as "PE" [22]), B is a sequential number, and (C) specifies the laboratory that produced the spectra (F – Fraunhofer CSP, I – Istituto Nazionale di Ricerca Metrologica (INRiM) and UP – University of Parma).

After selecting a match algorithm identified by the code presented in

Table 1
Raman spectrometers and relevant instrumental conditions used by the three laboratories. Lab.1 – Fraunhofer CSP (two equipment used, Sp. 1 and Sp. 2), Lab.2 – INRiM and Lab.3 – University of Parma.

Lab.	Spectrometer	*DF*/ (gr. nb./mm)	*LW*/nm| *OP*/mW	*MOM* |*NA*	Detector	*DP*	*WI*/cm^{-1}	*AT*/s| *AN*
1 (Sp. 1)	Triple 557 TriVista	1500	514|100	50 × |0.75	LN2| Si-CCD	450	850 to 3200	1 to 10| 5 to 10[a]
1 (Sp. 2)	HORIBA LabRAM HR Evolution	1800	532|100	50 × |0.75	LN2| Si-CCD	450	850 to 3200	1 to 10|5 to 10[a]
2	HORIBA LabRAM Odyssey	600	633|15	10 × |0.25	LN2| Si-CCD	1024 or 512	664 to 1786	1| 1
3	HORIBA LabRAM HR Evolution	600	532|25	50 × |0.50	LN2| Si-CCD	1024	415 to 2060	1 to 5| 3 to 20[a]

DF – Diffraction grating in groove number by mm; *LW* – Laser wavelength in nm; *OP* – Output power in mW; *MOM* – Microscope objective magnification; *NA* – Numerical aperture; *DP* – Number of detector pixels; *WI* – Wavenumber interval in cm^{-1}; *AT* – Acquisition time in s; *AN* – Accumulation number; LN2\| Si-CCD - silicon-based charge-coupled device cooled with liquid nitrogen.

[a] Optimized according to the specific polymer type and its physical characteristics.

Table 2
Match algorithm identification code, where *CC* is the studied correlation coefficient [4]. "*S*" specifies the type of signal considered (*I*, 1/*I* and (1-I_N) for original, inverse and complementary intensity after signal normalisation – numbers 1, 2 or 3 from code "*S*|#|*CC*|#|#"). The "*d*" specifies the use of the original signal, *S*, or its first or second derivative (f' (*S*) and f'' (*S*)) (*S*|*d*|*CC*|#|#) (*d* = 1, 2 or 3). For weighted *CC*, the signal was weighed (*SW* equal to *y* for yes – Code "*S*|*d*|*CC*|1|#"), and the wavenumber of the Raman shift was weighted directly, $\tilde{\nu}$, ("1″) or inversely, $1/\tilde{\nu}$ ("2")) (*S*|*d*|*CC*|*SW*|*RW*).

Match	*S*	*d*	*SW*	*RW*	Match	*S*	*d*	*SW*	*RW*
1|1|*CC*|1|1	*I*	*S*	*y*	$\tilde{\nu}$	1|1|*CC*|1|2	*I*	*S*	*y*	$1/\tilde{\nu}$
2|1|*CC*|1|1	(1/*I*)	*S*	*y*	$\tilde{\nu}$	2|1|*CC*|1|2	(1/*I*)	*S*	*y*	$1/\tilde{\nu}$
3|1|*CC*|1|1	(1 - I_N)	*S*	*y*	$\tilde{\nu}$	3|1|*CC*|1|2	(1- I_N)	*S*	*y*	$1/\tilde{\nu}$
1|2|*CC*|1|1	*I*	$f'(S)$	*y*	$\tilde{\nu}$	1|2|*CC*|1|2	*I*	$f'(S)$	*y*	$1/\tilde{\nu}$
2|2|*CC*|1|1	(1/*I*)	$f'(S)$	*y*	$\tilde{\nu}$	2|2|*CC*|1|2	(1/*I*)	$f'(S)$	*y*	$1/\tilde{\nu}$
3|2|*CC*|1|1	(1 - I_N)	$f'(S)$	*y*	$\tilde{\nu}$	3|2|*CC*|1|2	(1 - I_N)	$f'(S)$	*y*	$1/\tilde{\nu}$
1|3|*CC*|1|1	*I*	$f''(S)$	*y*	$\tilde{\nu}$	1|3|*CC*|1|2	*I*	$f''(S)$	*y*	$1/\tilde{\nu}$
2|3|*CC*|1|1	(1/*I*)	$f''(S)$	*y*	$\tilde{\nu}$	2|3|*CC*|1|2	(1/*I*)	$f''(S)$	*y*	$1/\tilde{\nu}$
3|3|*CC*|1|1	(1 - I_N)	$f''(S)$	*y*	$\tilde{\nu}$	3|3|*CC*|1|2	(1 - I_N)	$f''(S)$	*y*	$1/\tilde{\nu}$

Table 2, including the considered correlation coefficient (*CC*), the match between the reference spectra (PET 01 (F)) and the spectra of all PET or non-PET particles was determined. The studied *CC* are *r*, *ρ*, and C_{PE} for unweighted Pearson, Spearman, and alternative correlation coefficients, and r_w, ρ_w, and C_{PEw} are the weighted versions of the first [4]. The total number of studied matches is 81 (9 × 3 matches involving unweighted *CC* and 18 × 3 matches involving weighted *CC*).

For the unweighted *r* and *ρ*, the identification performance is the same for signal (*I*) or (1 - I_N). For the unweighted C_{PE}, and the weighted correlation coefficients, both signal types ((*I*) or (1 - I_N)) can produce different identification performance due to algorithm features.

Positive cases correspond to PET particles, and negative cases refer to non-PET particles. The 5th percentile of the match values for positive cases was determined using the bootstrap resampling method, which simulates its distribution [4,5]. The *P*5»P, the 5th percentile of the 5th percentile distribution of positive cases, is the minimum value of the specific match algorithm above which a particle can be defined as PET with a true positive rate of not less than 95 %, meaning that, in studied PET spectra, there is at least a 95 % probability that a PET microparticle spectrum has a match with "PET 01 (F)" spectrum above *P*5»P. Afterwards the mean, *M*, standard deviation, *s*, and total number, *n*, of match values between "PET 01 (F)" and negative cases is calculated being the false positive rate, *FP*, the complementary of the cumulative t-distribution of (*P*5»P – *M*)/*s* for *n* - 1 degrees of freedom. The *FP* is the estimated chance of a non-PET particle, from the studied non-PET population, producing a match with the "PET 01 (F)" spectra above *P*5»P, therefore being wrongly identified as PET. All match algorithms (Table 1) were tested and considered adequate for PET identification if, together with the *TP* set at 95 %, the *FP* is not larger than 5 %. In this case, the identification performance quantified as a likelihood ratio *LR* (*LR* = *TP*/*FP*) or accuracy, *A* [*A* = (*TP* + *TN*)/(*TP* + *TN* + *FP* + *FN*); where *TN* and *FN* are true and false negative rates complementary to *FP* and *TP* respectively], should not be lower than 19 [4,5,20] or 95 %, respectively.

An *FP* of 5 % is considered adequate for the analysis of environmental and food samples with one microplastic, given the number of required identifications and the health impact of microplastic contamination [4,5,20].

Since the reliability of particle identification is affected by the low intensity of peaks compared to the signal's noise, an algorithm was developed to determine the S/N. This algorithm, described in Section 2.2.1, is applicable regardless of the scale of absolute peak intensities, which vary depending on the spectrometers and spectral collection conditions used. Spectra with S/N lower than 10 should be forwarded to manual identification.

2.2.1. *Determination of signal-to-noise ratio*

To characterise Raman spectra regarding the spectral information clarity, the highest S/N was determined as the ratio between the most intense peak of the spectrum and the background noise. The intensity of the most intense peak was determined as the difference between the maximum signal observed in the spectrum and the minimum signal predicted at the wavenumber of the maximum signal by a regression line. This regression line, referred to as "baseline", is defined from minimum signals of 20 wavenumber segments with equal wavenumber ranges. The absolute difference between the average signal and the minimum signal within each segment is calculated and then averaged across all segments, yielding a value that quantifies the spectrum noise. Fig. 1 presents examples of spectra collected by three laboratories with S/N equal to 15.

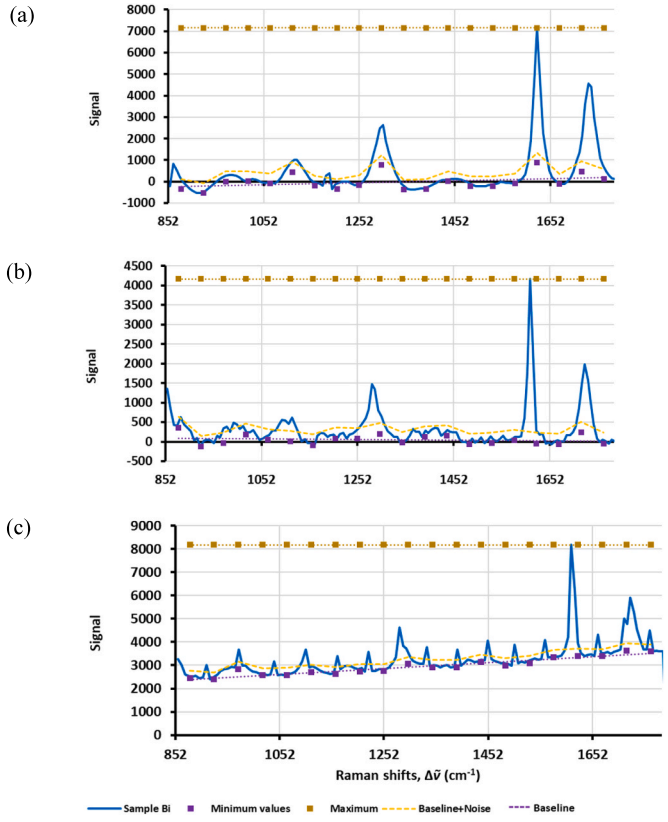


Fig. 1. Spectra (a), (b) and (c) present spectra with S/N of 15 obtained from Fraunhofer CSP, INRiM and the University of Parma, respectively. The mean, minimum and base regression lines of 20 wavenumber ranges are also presented. The developed algorithm was implemented in the MS Excel file used to quantify the match between the studied spectra (Supplementary Material 02).

3. Results and discussion

Table 3 presents the best-performing match algorithms, i.e. those associated with lower *FP*, considering all collected spectra and spectra with a S/N greater than 10. For *r* and ρ, only *I* and 1/*I* signals were considered because *I* and (1 - I_N) produce the same results. The *P*5»P estimated by the resampling bootstrap method slightly varies with the simulation run, with an impact on estimated *FP* and *LR*. It is reported the *P*5»P, *FP*, *LR* and the number of positive, n_P, and negative, n_N, cases considered. For the unweighted correlation coefficients, *SW* and *RW* are not specified.

Supplementary Material 03 presents the S/N of all spectra, match values determined for positive and negative cases for the 81 studied match methods and the *FP* and *LR* for the match methods.

As expected, identification performance improves and *P*5»P raises when considering a S/N threshold. Several match algorithms enable identification with an *LR* larger than 19 and an accuracy larger than 95 %, where the unweighted Pearson's correlation of the original signals distinguishes between positive and negative cases more effectively when a S/N threshold of 10 is considered.

Several weighted correlation coefficients can support valid PET identifications. The weighted alternative correlation coefficient, C_{PE_w}, for complementary normalised signals (1 - I_N) and the two studied types of weighing are the most successful weighted algorithms (3|1| C_{PE_w} |1|1 and 3|1| C_{PE_w} |1|2).

The *TP* of 95 % and *FP* of 4.90×10^{-7} % of identifications based on Person's correlation coefficient and S/N > 10 are converted into an identification accuracy of 97.5 %, comparable to the observed accuracy using advanced chemometrics tools that require significantly more computational resources.

The identification criteria can be implemented for new spectra using the Excel file available as Supplementary Material 04. Detailed instructions of how to use this spreadsheet are available in video file Supplementary Material 05.

Supplementary data related to this article can be found online at https://doi.org/10.1016/j.talanta.2025.128834

Fig. 2 presents a graphical representation of Match values for positive and negative cases and the *P*5»P for the Match methods listed in Table 3 considering (a) all spectra or (b) only signals with a S/N > 10. The Match of positive and negative cases are rather distinct being worth highlighting the very low dispersion of Match values of both positive and negative cases for the Match method 3|1| C_{PE} valid for S/N > 10.

4. Conclusions

The proposed methodology to develop a valid procedure for the automatic identification of microplastics using μ-Raman spectroscopy applicable to spectra collected from different spectrometers and spectra collection conditions was successfully applied to the identification of PET microplastics. Different match algorithms were tested, considering the use of original or inverse signals, before or after derivatisation, and based on unweighted Pearson's, Spearman's, an alternative algorithm, or weighted versions of these correlation coefficients. For the more complex weighted coefficients, signal and two types of wavenumber weightings were considered. The minimum match value (*P*5»P) for identifying a PET particle corresponds to the 5th percentile of the match between spectra of the PET reference and particle, and it is determined by the resampling bootstrap method. The bootstrap method does not require match distribution normality. The match values between the PET reference spectrum and particle spectra proved not to be PET allowed estimating the probability of a non-PET particle producing a match value with the PET reference above *P*5»P. This probability is the false positive rate. A methodology for quantifying the signal-to-noise ratio was developed, and spectra with an S/N lower than 10 were

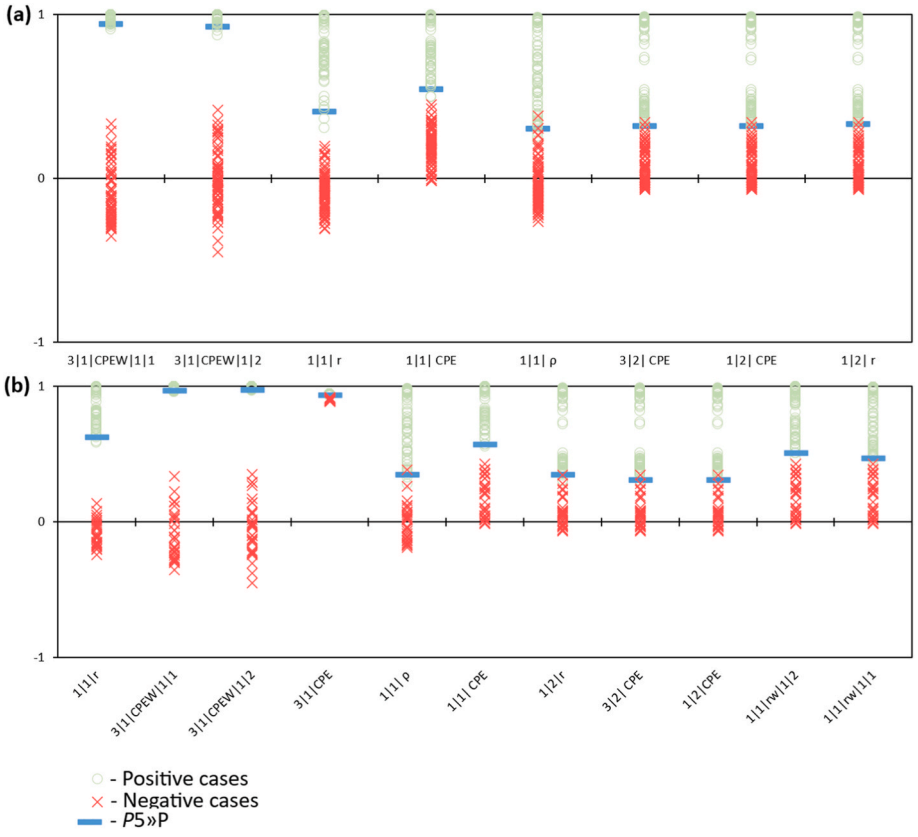


Fig. 2. Graphical representation of Match values for positive and negative cases and the 5th percentile for positive cases, *P*5»P, for the Match methods presenting *LR* larger than 19. (a) All spectra considered; (b) Only spectra with S/N > 10 are considered.

Table 3
Best-performing match algorithms, considering all collected spectra and spectra with S/N larger than 10. The *P5»P* is the match threshold, *FP* the false positive rate, *LR* the likelihood ratio (*LR* = *TP/FP*; *TP* is the true positive rate), and n_P and n_N the number of considered PET and non-PET cases.

All spectra (n_P = 92 and n_N = 82)				Spectra with S/N > 10 (n_P = 80 and n_N = 33)			
Match	*P5»P*	*FP*/%	*LR*	Match	*P5»P*	*FP*/%	*LR*
3|1| C_{PE_w} |1|1	0.9386	4.46×10^{-7}	2.13×10^{8}	1|1|r	0.6244	4.90×10^{-7}	1.94×10^{8}
3|1| C_{PE_w} |1|2	0.9212	5.15×10^{-5}	1.85×10^{6}	3|1| C_{PE_w} |1|1	0.9664	8.09×10^{-5}	1.17×10^{6}
1|1|*r*	0.4049	0.00660	1.44×10^{4}	3|1| C_{PE_w} |1|2	0.9734	0.000537	1.77×10^{5}
1|1| C_{PE}	0.5418	0.190	501	3|1| C_{PE}	0.9325	0.282	336
1|1|ρ	0.3000	0.985	96.5	1|1|ρ	0.3471	0.507	188
3|2| C_{PE}	0.3167	1.30	73.2	1|1| C_{PE}	0.5703	0.620	153
1|2| C_{PE}	0.3167	1.30	73.2	1|2|*r*	0.3438	0.627	151
1|2|*r*	0.3166	1.32	71.9	3|2| C_{PE}	0.3089	1.33	71.2
				1|2| C_{PE}	0.3051	1.45	65.6
				1|1| r_w |1|2	0.5072	1.68	56.5
				1|1| r_w |1|1	0.4692	2.93	32.4

excluded from the automatic identification. The unweighted Pearson's correlation coefficient of the original signal proved to be the most adequate match algorithm for PET particle identification for a match threshold, *P5»P*, of 0.6244. The identification with this match is associated with a true positive rate of 95 % and a very low false positive rate of 4.90×10^{-7} %, fit for microplastic identifications in environmental and food samples. This performance, which guarantees an accuracy of 97.5 %, is equivalent to that observed using complex chemometric tools that require extensive data processing of a large number of spectra.

The user-friendly spreadsheets used to set the *P5»P* and for the subsequent identification of microparticles are made available as supplementary material.

Although promising, the developed methodology should be evaluated for its ability to produce a reliable method for the identification of other polymer types, aged polymers, and particles contaminated with biofilm. Ideally, the identification approach should also be tested considering spectra collected by additional spectrometers.

CRediT authorship contribution statement

Rafaela Fernandes: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Paul-Tiberiu Miclea:** Writing – review & editing, Writing – original draft, Resources, Investigation, Funding acquisition, Data curation, Conceptualization. **Marta Fadda:** Writing – review & editing, Resources, Investigation, Data curation. **Mara Putzu:** Writing – review & editing, Resources, Investigation, Data curation. **A. Sacco:** Writing – review & editing, Resources, Investigation, Data curation. **Andrea M. Rossi:** Writing – review & editing. **Andrea M. Giovannozzi:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Marta Barbaresi:** Writing – review & editing, Writing – original draft, Investigation, Data curation. **Matteo Masino:** Writing – review & editing, Investigation, Data curation. **Monica Mattarozzi:** Writing – review & editing, Supervision, Investigation, Data curation, Conceptualization. **Maria Careri:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Carla Palma:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **José Almeida:** Writing – review & editing. **Claudia Drago:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Olivier Pellegrino:** Writing – review & editing, Funding acquisition, Conceptualization. **Raquel Quendera:** Writing – review & editing. **Ulrike Braun:** Writing – review & editing, Resources, Funding acquisition. **Ricardo J. N. Bettencourt da Silva:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

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.

Acknowledgment

The authors thank the European Commission for funding the PlasticTrace project (21GRD07) from the European Partnership on Metrology (Funder ID: 10.13039/100019599), co-financed from the European Union's Horizon Europe Research and Innovation Program and by the Participating States. This work was supported by Fundação para a Ciência e a Tecnologia (FCT) (Projects UIDB/00100/2020 and UIDP/00100/2020), and Institute of Molecular Sciences (Project LA/P/0056/2020).

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.talanta.2025.128834.

Data availability

Data available as supplementary material and deposited in Zenodo with DOI 10.5281/zenodo.17079479.

References

[1] V. Redko, L. Wolska, M. Potrykus, E. Olkowska, M. Cieszyńska-Semenowicz, M. Tankiewicz, Environmental impacts of 5-Year plastic waste deposition on municipal waste landfills: a Follow-up study, Sci. Total Environ. 906 (2024) 167710, https://doi.org/10.1016/J.SCITOTENV.2023.167710.
[2] Y. Yang, M. Jalalah, S.A. Alsareii, F.A. Harraz, N. Thakur, Y. Zheng, M. Koutb, Y. Yoon, E.S. Salama, Plastic wastes (PWs) and microplastics (MPs) formation: management, migration, and environmental impact, J. Environ. Chem. Eng. 12 (3) (2024), https://doi.org/10.1016/J.JECE.2024.112926.
[3] N.B. Hartmann, T. Hüffer, R.C. Thompson, M. Hassellöv, A. Verschoor, A. E. Daugaard, S. Rist, T. Karlsson, N. Brennholt, M. Cole, M.P. Herrling, M.C. Hess, N.P. Ivleva, A.L. Lusher, M. Wagner, Are we speaking the same language? Recommendations for a definition and categorization framework for plastic debris, Environ. Sci. Technol. 53 (3) (2019) 1039–1047, https://doi.org/10.1021/acs.est.8b05297.
[4] V. Morgado, L. Gomes, R.J.N. Bettencourt da Silva, C. Palma, Validated spreadsheet for the identification of PE, PET, PP and PS microplastics by Micro-ATR-FTIR spectra with known uncertainty, Talanta 234 (April) (2021) 122624, https://doi.org/10.1016/j.talanta.2021.122624.
[5] V. Morgado, C. Palma, R.J.N. Bettencourt da Silva, Microplastics identification by infrared spectroscopy – evaluation of identification criteria and uncertainty by the bootstrap method, Talanta 224 (October 2020) (2021) 121814, https://doi.org/10.1016/j.talanta.2020.121814.
[6] J.L. Xu, K.V. Thomas, Z. Luo, A.A. Gowen, FTIR and raman imaging for microplastics analysis: state of the art, challenges and prospects, TrAC, Trends Anal. Chem. 119 (2019), https://doi.org/10.1016/j.trac.2019.115629.
[7] C.F. Araujo, M.M. Nolasco, A.M.P. Ribeiro, P.J.A. Ribeiro-Claro, Identification of microplastics using raman spectroscopy: latest developments and future prospects,

Water Res. 142 (2018) 426–440, https://doi.org/10.1016/J.WATRES.2018.05.060.
[8] L. Cabernard, L. Roscher, C. Lorenz, G. Gerdts, S. Primpke, Comparison of Raman and fourier transform infrared spectroscopy for the quantification of microplastics in the aquatic environment, Environ. Sci. Technol. 52 (22) (2018) 13279–13288, https://doi.org/10.1021/ACS.EST.8B03438/SUPPL_FILE/ES8B03438_SI_002.XLSX.
[9] A. Käppler, D. Fischer, S. Oberbeckmann, G. Schernewski, M. Labrenz, K. J. Eichhorn, B. Voit, Analysis of environmental microplastics by vibrational microspectroscopy: FTIR, raman or both? Anal. Bioanal. Chem. 408 (29) (2016) 8377–8391, https://doi.org/10.1007/S00216-016-9956-3/TABLES/3.
[10] N.P. Ivleva, A.C. Wiesheu, R. Niessner, Microplastic in aquatic ecosystems, Angew. Chem. Int. Ed. 56 (7) (2017) 1720–1739, https://doi.org/10.1002/ANIE.201606957.
[11] N. Jin, Y. Song, R. Ma, J. Li, G. Li, D. Zhang, Characterization and identification of microplastics using raman spectroscopy coupled with multivariate analysis, Anal. Chim. Acta 1197 (2022) 339519, https://doi.org/10.1016/j.aca.2022.339519.
[12] L. Xie, S. Luo, Y. Liu, X. Ruan, K. Gong, Q. Ge, K. Li, V.K. Valev, G. Liu, L. Zhang, Automatic identification of individual nanoplastics by Raman spectroscopy based on machine learning, Environ. Sci. Technol. 57 (46) (2023) 18203–18214, https://doi.org/10.1021/ACS.EST.3C03210/ASSET/IMAGES/LARGE/ES3C03210_0005.JPEG.
[13] W. Zhang, W. Feng, Z. Cai, H. Wang, Q. Yan, Q. Wang, A deep one-dimensional convolutional neural network for microplastics classification using raman spectroscopy, Vib. Spectrosc. 124 (2023) 103487, https://doi.org/10.1016/J.VIBSPEC.2022.103487.
[14] Y. Luo, W. Su, D. Xu, Z. Wang, H. Wu, B. Chen, J. Wu, Component identification for the SERS spectra of microplastics mixture with convolutional neural network, Sci. Total Environ. 895 (2023) 165138, https://doi.org/10.1016/J.SCITOTENV.2023.165138.
[15] K. Liu, X. Pang, H. Chen, L. Jiang, Visual detection of microplastics using raman spectroscopic imaging, Analyst 149 (1) (2023) 161–168, https://doi.org/10.1039/D3AN01270K.
[16] M. Sunil, N.N. Pallikkavaliyaveetil, I. Mi, A. Gopinath, S. Chidangil, S. Kumar, J. Lukose, Machine learning assisted raman spectroscopy: a viable approach for the detection of microplastics, J. Water Proc. Eng. 60 (2024) 105150, https://doi.org/10.1016/J.JWPE.2024.105150.
[17] European Commission, MSDF guidance on monitoring marine litter. https://doi.org/10.2788/99475, 2013.
[18] ISO/FDIS 16094-2 - water quality — analysis of microplastic in water — part 2: vibrational spectroscopy methods for waters with low content of suspended solids including drinking water. https://www.iso.org/standard/84460.html. (Accessed 20 June 2025).
[19] B. Efron, R.J. Tibshirani, An introduction to the bootstrap method, in: An Introduction to the Bootstrap Method, Chapman & Hall/CRC, 1993.
[20] R. Bettencourt da Silva, S.L.R. Elison, EURACHEM/CITAC Guide Assessment of Performance and Uncertainty in Qualitative Chemical Analysis, 2021.
[21] R.L. McCreery, Raman spectroscopy for chemical analysis, Raman Spectroscopy for Chemical Analysis (2005) 1–420, https://doi.org/10.1002/0471721646.
[22] R.G. Jones, J. Kahovec, R. Stepto, E.S. Wilks Michael Hess, T. Kitayama, W. Val Metanomski, A. Jenkins, P. Kratochvíl, Compendium of polymer terminology and nomenclature IUPAC recommendations 2008 with advice from, Compendium of Polymer Terminology and Nomenclature IUPAC Recommendations (2008).