



Laser induced fluorescence and machine learning: a novel approach to microplastic identification

Nikolaos Merlemis¹ · Eleni Drakaki^{1,2} · Evangelini Zekou¹ · Georgios Ninos³ · Anastasios L. Kesidis¹

Received: 27 February 2024 / Accepted: 20 August 2024 / Published online: 30 August 2024
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2024

Abstract

Identifying the types of materials such as plastics, microplastics, and oil pollutants is essential for understanding their effects on marine life. We propose a new methodology for the real-time detection and identification of microplastics in aquatic environments. Our experiments are based on a compact Laser Induced Fluorescence (LIF) device, with machine learning techniques applied to classify the materials. A 405 nm CW laser excitation source effectively induces fluorescence spectra in the visible spectrum from material samples that are either floating or submerged in water. We examine known plastic pollutants in seawater, including polyethylene (PE), polypropylene (PP), polystyrene (PS) and polyethylene terephthalate (PET), as well as maritime fuels, lubricating oils, and other organic substances that are abundant in the marine environment. Our two-step identification process first employs machine learning algorithms to distinguish microplastics from other organic materials with a high degree of accuracy (97.6%). Subsequently, the type of plastic is determined with an accuracy of 88.3% in a second application of machine learning techniques.

1 Introduction

Anthropogenic waste has been recognized as a significant ocean pollution environmental problem, with the expanding presence of plastic and oil pollutants into the oceans around the world arising big concerns about their effects in the environment and marine life and consequently to human health. Oil contamination, which is caused by wastewater discharged from industrial facilities or offshore oil leakage, has been recently accurately detected using fluorescence spectroscopy and pattern recognition algorithms [1]. Human negligence has resulted in the accumulation of plastic materials in the oceans, where they typically remain near the surface. Over time, these plastics undergo degradation through various physical processes and fragment into microplastics (< 5 mm particles). Polymers like polyethylene (PE),

polypropylene (PP) and polystyrene (PS) have been found to be a major source of plastic and microplastic pollutants within the western Mediterranean Sea [2]. Microplastics can cause severe problems to ecosystems [3], and their existence into the food chain, sometimes ending with human consumption, can also pose significant concerns for human health. This issue has motivated researchers to examine accurate and compact methods for the detection and identification of plastics, usually present in the form of microplastics [4–7].

Manual counting of microplastics by optical microscopy, scanning electron microscopy (SEM), transmission electron microscopy (TEM) and atomic force microscopy (AFM) have been extensively applied to detect microplastics, combined with spectroscopic techniques in order to characterize the chemical composition of the materials [8–10]. Spectroscopic methods, such as Fourier transform infrared spectroscopy (FTIR), Raman spectroscopy, and hyperspectral imaging have been recently applied for the identification of the material type [9]. Other analytical techniques that are accurate in identifying microplastics, such as pyrolysis/gas chromatography and mass spectrometry, are either invasive or costly and time-consuming [10]. However, most of the characterization methods need expensive apparatus, expert operators, time-consuming preprocessing processes, and sophisticated data analysis. In addition, plastics can be

✉ Nikolaos Merlemis
merlemis@uniwa.gr

¹ Department of Surveying and Geoinformatics Engineering, University of West Attica, Athens 12243, Greece

² Department of Naval Architecture, University of West Attica, Athens 12243, Greece

³ Department of Biomedical Sciences, University of West Attica, Athens 12243, Greece

easily degraded in the environment and become more hazardous, thus making imperative the development of a rapid and real-time screening method for the timely identification of plastic pollutants in the aquatic environment.

Spectroscopy, as an analytical technique, relies on the interaction between matter and light, such as absorption, emission and scattering. Specifically, fluorescence spectroscopy arises from the emission of radiation by a target following its excitation due to light absorption. The molecular target emits photons with lower energy than those absorbed, attributable to energy loss through non-radiative or radiative decay. Excitation can be achieved using a lamp, a Light-Emitting Diode (LED) or a laser source. When excitation is performed with a laser beam, the resulting fluorescence is termed Laser Induced Fluorescence (LIF). LIF radiation can be collected at various angles relative to the incident laser beam, because it is incoherent and is emitted in all directions. The LIF spectra reveal information about the transitions from the excited state to various lower energy levels of the target molecules, and the delay between excitation and detection signals further elucidate the physical processes involved (time-resolved LIF).

Photoluminescence spectroscopy and LIF have been proposed as cost effective alternatives for monitoring microplastics, due to the simplicity of their experimental setups, which are feasible for integration on unmanned aerial vehicles (UAVs), unmanned surface vehicles (USVs), or other vehicles to assess water pollutants [11–13]. A comprehensive review of various techniques and optical methods for the in situ detection of microplastics in water is detailed in the work by Asamoah et al. [14]. Recently, principal component analysis (PCA) and LIF have been proved capable of identifying pure microplastic samples [15]. Furthermore, the study of the component ratio in mixed seawater samples revealed a linear dependency between the first two principal components [15]. However, their microplastic samples, whether pure or mixed, were meticulously prepared in glass containers devoid of any impurities or common marine substances such as oils, organic matter, seaweed, etc., ensuring that the laser beam fully encompassed the sample surface. Consequently, this method is only effective after the separation of microplastics from other marine substances and is not suitable for real-time measurements.

While traditional spectroscopic techniques have enabled precise material identification, the increasing complexity of the data sets that are being analyzed necessitates the development of more accurate, faster and automated analysis techniques. This has resulted in the incorporation of supervised machine learning (ML) algorithms into material classification tasks. Supervised learning methods utilize a set of labeled data for training and include algorithms such as decision trees, support vector machines (SVM), k-nearest

neighbors (KNN), random forests, and neural networks. These algorithms categorize spectral data into predefined classes, identifying materials based on their spectral signatures. KNN identify the data points most similar to new inputs, while decision trees and random forests create models predicting target values by learning decision rules. SVM finds optimal hyperplanes separating different classes, while neural networks, particularly deep learning models, automatically extract pertinent features and process intricate data relationships. The efficacy of these techniques, such as SVM, is contingent upon the nature of the data. SVM is particularly effective for two-class classification problems, as it identifies the optimal hyperplane that separates the classes with the maximum margin. In multiclass classification cases, where multiple materials need to be identified, random forests and neural networks are typically the most effective due to their capacity to handle complex relationships and interactions within the data. PCA is often used in conjunction with these algorithms to reduce data dimensionality, preserving essential features while facilitating more efficient analysis and enhancing the performance of the classification models. The integration of these supervised algorithms enhances accuracy, speed, and the ability to manage large datasets.

In the present study, we assess the capability of the compact LIF apparatus proposed in [11] and [12] for the real-time and in situ detection of plastic pollutants and microplastics in water. The experimental setup employs a focused laser beam. This will be advantageous in a future application of the methodology in real-time measurements in the sea, as it allows for the adequate detection of small particles from different materials in mixed samples by their spatial separation through the use of a small focal point. Oil contaminants from the maritime industry such as fuel and lubricating oils and organic substances prevalent in marine environments are also examined. A two-step methodology is proposed for the evaluation of the apparatus' suitability for real-time measurements. Initially, we employ PCA and ML algorithms to differentiate between plastic and other organic materials. In a second step, we apply the PCA technique, as proposed in [15], in combination with ML analysis, solely to the verified plastic samples, in order to evaluate the success rate for the correct characterization of the microplastic type. The ultimate objective is to develop an LIF system for the in situ identification of marine pollutants from an unmanned surface vehicle (USV) in real-time.

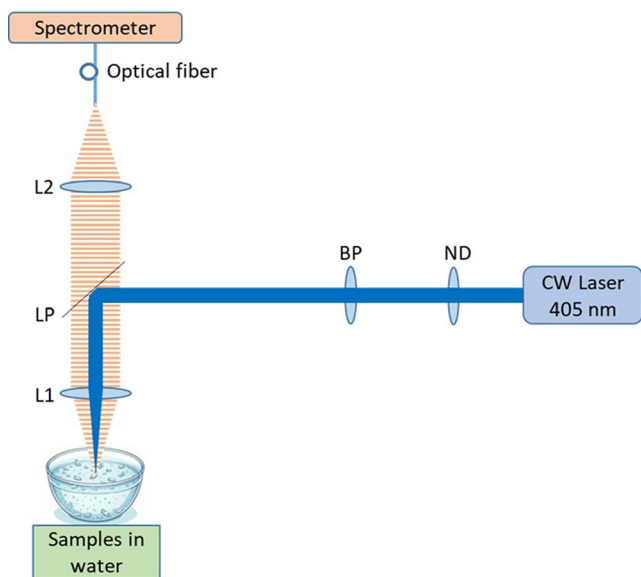


Fig. 1 The LIF experimental apparatus. ND stands for neutral density filters of various densities; BP is a bandpass filter (center wavelength 405, Edmund Optics); LP is a long-pass filter with a cut-on wavelength of 420 nm (Edmund Optics). L1 and L2 are lenses with 40 mm focal length (25 mm diameter)

2 Methods and experimental setup

Figure 1 illustrates the simplified experimental setup, which is similar to the apparatus described by Drakaki et al. [12]. We used a low-cost continuous wave (CW) laser diode system emitting at 405 nm, as our excitation laser source, with a maximum output power of 100 mW. To avoid intensity saturation of the recorded spectra, we adjusted the power output with neutral density (ND) filters as needed. A bandpass filter at 405 nm was utilized to filter the laser beam from possible amplified spontaneous emissions. The filtered laser light was reflected to a lens (L1) by a long-pass filter with a cut-on wavelength of 420 nm. The asymmetric diode laser beam profile was estimated using beam-scanning techniques [16]. The full angle beam divergence was approximately 1.5 mrad/3 mrad (horizontal/vertical), resulting in a focal spot diameter of approximately $60 \mu\text{m} \times 120 \mu\text{m}$. The water surface of the cuvettes enclosing the floating or submerged microplastics or other materials was positioned on or close to the focal plane of the lens L1. The generated LIF signal from the samples was collimated by passing from the same lens L1. The long-pass filter also provided filtering of any laser radiation reflected by the samples. The filtered signal was coupled into a multimode optical fiber by a second 40 mm lens (L2) and a compact spectrometer recorded the spectra (Ocean Optics S2000 from Ocean Insights, ~ 2 nm resolution).

The samples (water/contaminants mixtures) were generated in the laboratory by placing tiny amounts (~ 0.3 g)

Table 1 The natural materials, maritime pollutants (lubricants-oils, fuels and paints) and plastics included in this work. Most plastic materials used were uncolored, with the exception of the Bakelite and PVC (black)

Material	Origin	Material	Origin
Poly(methyl methacrylate) (PMMA)	Industry	Naval oil SAE 40	Shipping industry
Polyethylene (PE)	Industry & retail	Fuel oil VLSFO	Shipping industry
Polypropylene (PP)	Industry	Fuel oil HSFO	Shipping industry
Polystyrene (PS)	Industry & retail	Paint SKF9005	Shipping industry
Polyethylene terephthalate (PET)	retail	Bakelite	retail
Polyurethane (PUR)	retail	Wood (Palm tree)	Natural
Polyvinyl chloride (PVC)	retail	Olive Oil	Natural
Polycarbonate (PC)	retail	Posidonia Oceanica	Natural

of the contaminants in 40 ml of water within an open top glass cuvette. Industrial grade plastics and retail product plastics were both used to form the samples. Uncolored plastics were mainly used for the calibration measurements (training and evaluation of the machine learning models). Microplastics were produced by breaking down the plastic materials into particles of different geometries and sizes of less than 2 mm. To simulate actual marine conditions, the samples were stirred before each measurement to slightly homogenize the composition and elevate materials that are submerged at the bottom of the cuvettes, although perfect homogenization inside the cuvettes and at the surface of the water was not desired. In the majority of the cases, the concentration of pollutants is higher near the water surface or at the bottom of the cuvettes, depending on the material density. Sixty (60) measurements were recorded for every sample in order to examine different conditions of particles concentration and size, as well as different laser intensities exciting the particles (since particles are not always placed at the surface of the water where laser is focused and the laser intensity is maximum). In addition, cuvettes could be shifted in all three directions to get measurements from various locations inside the sample. Every measurement recorded is the average of ten spectra taken over a 50 ms integration period of the spectrometer.

Substances that may be present in the marine environment were examined in this study, as shown in Table 1. A comparative analysis was conducted on a range of natural materials, including wood, olive oil, and seaweed (*Posidonia Oceanica*), as well as oils, fuels and paints derived from the maritime industry. The research includes olive oil as an indicative example of natural oils, as well as due to the possibility that the coastal zone may be impacted by the discharge of olive oil mills wastewater into waterways.

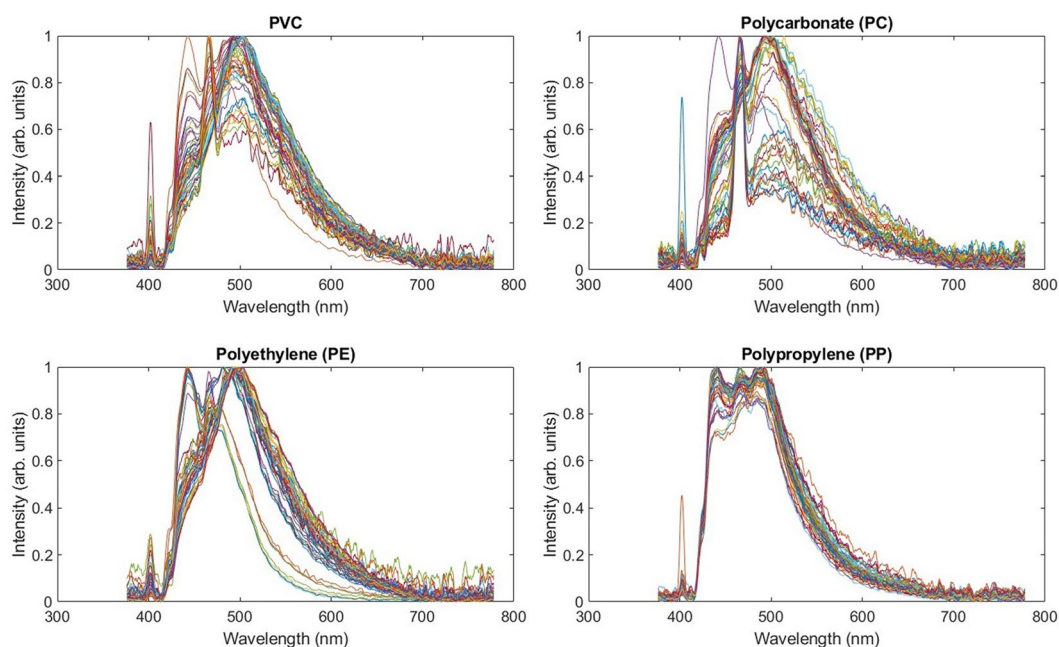


Fig. 2 Normalized and smoothed fluorescence spectra of PVC, PC, PE and PP. Multiple (60) spectra are shown for each material

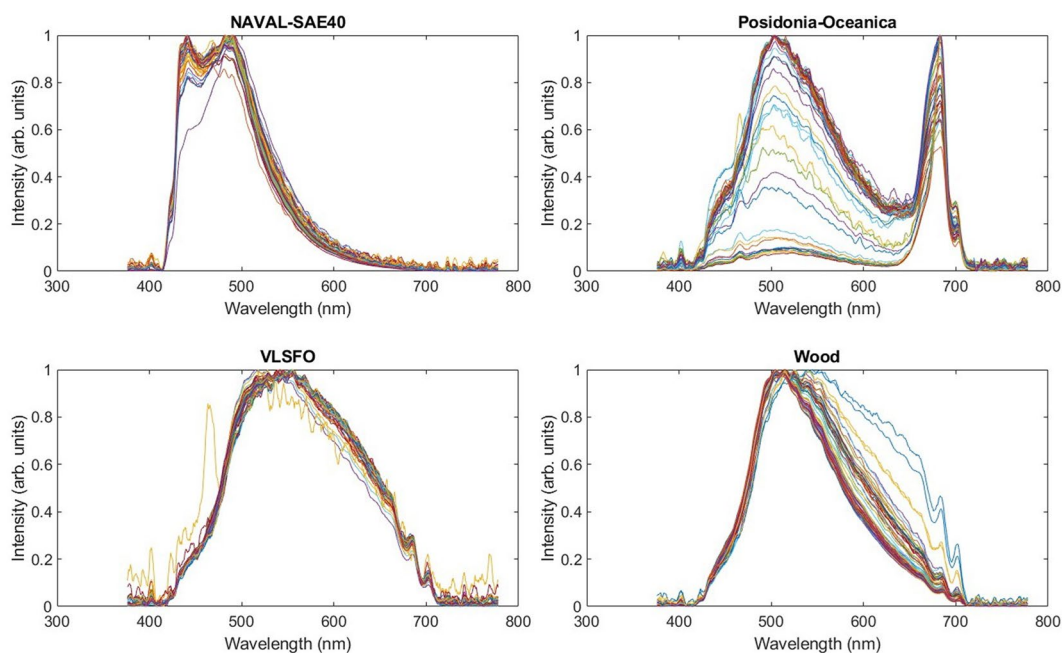


Fig. 3 Normalized and smoothed fluorescence spectra of naval oil (SAE40), naval fuel (VLSFO), wood (Palm Tree) and seaweed (*Posidonia-Oceanica*). Multiple (60) spectra are shown for each material

Furthermore, chlorophyll fluorescence present in olive oil represents a significant emission that we have considered in our analysis, given that it is also anticipated in other marine organic materials (e.g., phytoplankton).

3 Experimental results

LIF spectra were recorded after background emission, with the laser deactivated, was subtracted from the fluorescence signal. Figure 2 depicts representative LIF spectra of plastic materials and Fig. 3 shows similar spectra of some maritime oils, fuels and natural materials. All spectra have been

smoothed with a Savitzky–Golay filter function of order 2 with a frame length of 27 in order to minimize the relatively high noise for materials with low fluorescence signal. Spectra with signal-to-noise ratio less than 5 were discarded. Multiple measurements (60) were taken for each sample, for different laser intensities and material densities, since, in most cases, the studied materials and microplastics are outside the exact focal point of the laser beam.

Our methodology emphasizes in the high variability of the parameters during the measurements for each sample, in order to simulate in the laboratory field measurements in the sea (in situ), where material density, laser intensity and focal spot at different depths are highly fluctuating parameters. Therefore, the fluctuating LIF spectra are normalized, as shown in Figs. 2 and 3. The small intensity peak at 405 nm seen in some of the spectra is due to some residual laser intensity after the dichroic mirror (LP), which is visible only in the case of low signal spectra.

A fluorescence peak is observed around 500 nm for all plastics, as well as for other materials, such as wood, seaweed and marine oils. In addition, Raman emissions are also observed at 427 nm and 462 nm in the case of plastics in Fig. 2, comparable to the spectra recorded in [11]. These emissions are generated by nonlinear Raman mechanisms, and they have different excitation intensity dependence from the rest of the emission spectra, as shown in Fig. 2 by the highly variable shape between measurements. For example, in the case of PE, several spectra appear to be less broad than the rest, due to the fact that the excitation intensity was higher, making Raman emissions dominant.

To add to the complexity of the observed spectra, *Posidonia Oceanica* (shown in Fig. 3) and olive oil also show a strong peak near 680 nm. The complexity of the spectra observed from natural materials is expected due to the complex structure of natural organic materials consisting of different chromophores. For the intensities used in our study no spectral peaks are observed in the red or infrared region of the spectrum for the uncolored plastic samples used. However, some dyed plastic samples (which are unfortunately abundant in the environment) may produce more complex spectra, which will be discussed in the next section.

Although there may be some discrepancies between our spectra and those reported in [11, 12] and [13], mainly due to the lack of established standards for both the measurement of plastic samples and the construction of a fluorescence setup [13], it is evident that each material's spectrum can yield unique and distinct spectral characteristics. The ability to discriminate between plastic and non-plastic materials was proposed in [11] using two classifiers, namely the ratio of the intensity at 427 nm and 462 nm to the intensity observed at 550 nm (named parameters P1 and P2). Peaks P1 and P2 are not sufficient to identify different types of plastic

materials under different experimental conditions (excitation intensity), thus requiring the use of machine learning techniques with a broader set of classifiers for proper characterization [13]. However, the results presented in [13] are not suitable for the real-time classification of the pollutants, since some classification parameters were included in their research in a non-automatic way (for example the color of the samples). In addition, maritime oils, which may be present in real seawater measurements and add complexity to the identification process, were not considered in [11] and [13].

In accordance with the proposed indicators presented in [12], a thorough inspection of the spectra can reveal a variety of spectral characteristics. An important characteristic of LIF emissions from plastics is the shape of the spectral region above 500 nm. As observed in [12], uncolored plastic materials exhibited an exponential decrease in intensity variation for wavelengths λ after the peak at 500 nm, which was different to that of other materials studied. However, the observed differences among plastic materials appeared to be negligible. To improve the analytical method outlined in [12], the number of collected spectra was augmented from 20 to 60. This characteristic of exponential decline is also discernible in the 60 spectra shown in Fig. 2. Nonetheless, it is not distinctly observable in other materials, such as VLSFO oil and wood in Fig. 3. The exponential fit coefficient b and the coefficient of determination R^2 are estimated by the exponential fit of the spectral intensity versus wavelength λ :

$$I(\lambda) = I_{max} \cdot e^{b \cdot \lambda} \quad (1)$$

To evaluate whether b and R^2 are useful indicators for plastics materials, box plots are shown in Figs. 4 and 5, respectively. As illustrated in Fig. 5, all the examined plastics exhibit an excellent exponential fit (median of R^2 very close to 1). Naval oil SAE40 also follows an exponential fit well enough, but other materials display more spectral features, leading to median R^2 values as low as 0.6 and 0.1 for olive oil and maritime fuels, respectively. Consequently, the median value of the parameter b seems to vary between plastics and other materials. However, this parameter, when used alone, is not efficient to discriminate between different types of plastics.

In addition, the 3D plot of the parameter b versus R^2 and P2 (Raman peak at ~ 462 nm) is shown in Fig. 6. It is clear that non-plastic materials are already easily classified, although differentiating between plastic types is difficult.

Summarizing the characteristics of the recorded spectra, we observe the following: (a) a broad fluorescence emission around 500 nm, with peak wavelengths that vary slightly depending on the type of material (such as plastics,

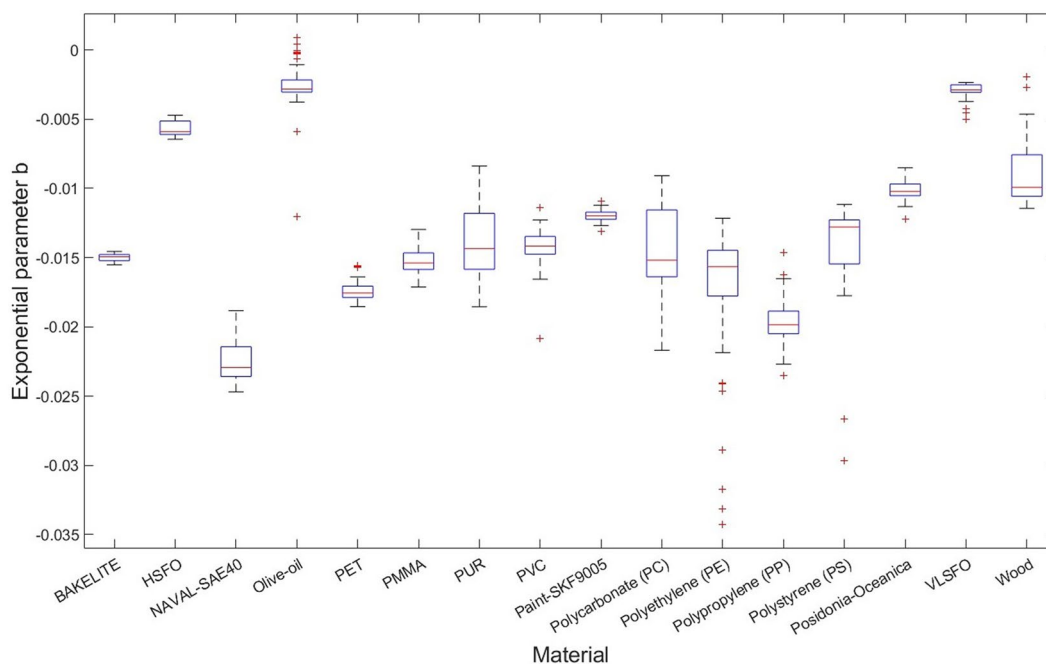


Fig. 4 Boxplot of the exponential fitting parameter b versus the material type. The central red mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively

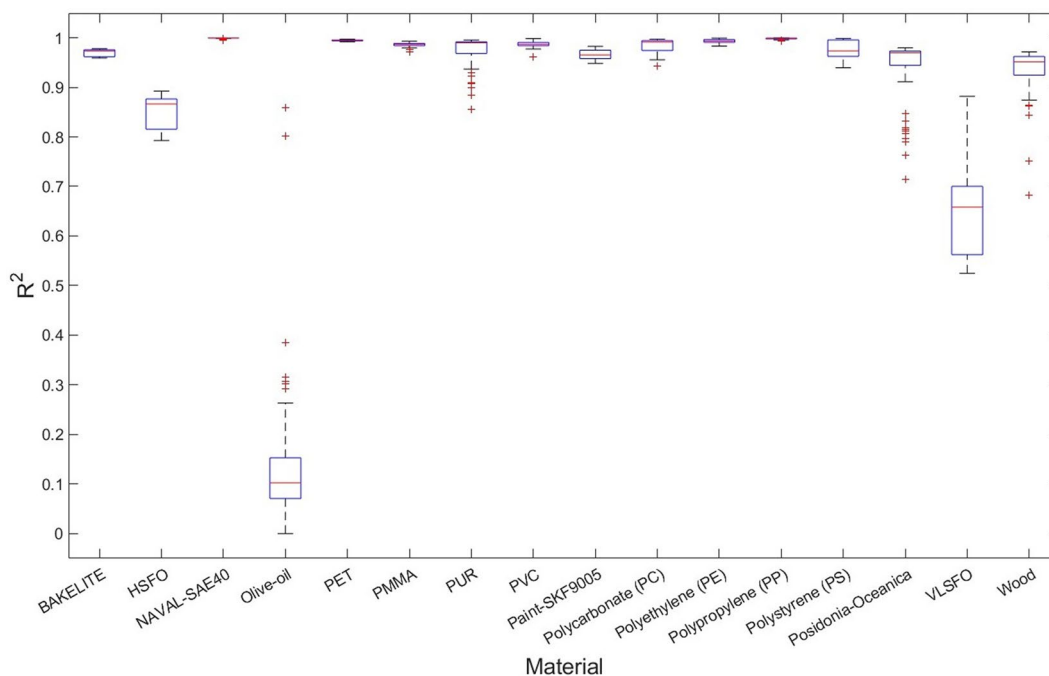


Fig. 5 Boxplot of the coefficient of determination R^2 versus the material type. The central red mark indicates the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively

maritime lubricants-oils, and natural organic materials). (b) pronounced nonlinear emissions (e.g. Raman scattering) that become stronger at higher intensities; (c) a secondary broad emission region around 680 nm that is present in natural organic materials spectra (explained by the chlorophyll emission), as has also been recorded in field measurements

of chlorophyll-a and chromophoric dissolved organic matter (CDOM) under 405 nm excitation [17]. The emission around 500 nm is a prominent characteristic of LIF under 405 nm excitation in organic materials, e.g. in the case of CDOM in water [18]. This emission is recorded in every material used in our study. For the laser intensities used in

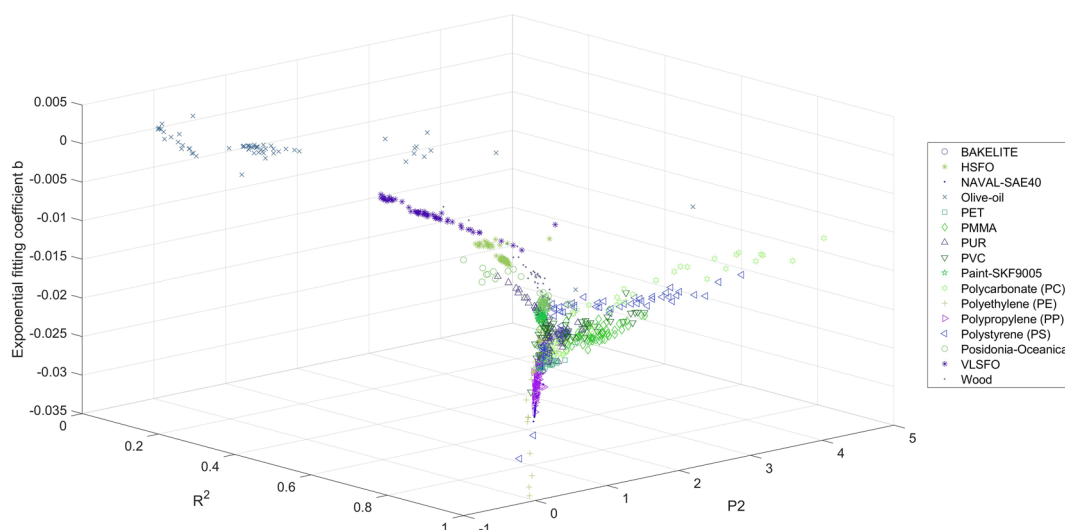


Fig. 6 3D plot of the exponential parameter b versus R^2 and versus P2

our experiments, emissions at wavelengths above 716 nm and near infrared are extremely low or absent and are not taken into account in this study.

It is also clear that there are no other distinctive spectral peaks observed that would facilitate the classification of the materials. Typically, the broadband nature of plastic fluorescence makes the analysis of the obtained spectra difficult since the spectra of distinct chromophores may significantly overlap. However, the shape of the recorded wideband spectra varies slightly for each material, which means that the excitation energy is distributed differently in each spectral region, depending on the material type. In the next section, we propose a two-step identification process: initially, we employ the most promising ML models to distinguish plastic samples from other materials, and subsequently, we develop distinct ML models to characterize the types of plastics in the identified microplastics.

4 Machine learning classification method evaluation and discussion

Machine learning for classification tasks plays a crucial role in the realm of material identification, where the goal is to automatically categorize substances based on their unique properties and characteristics. In material science, the vast array of compounds and materials requires sophisticated methods to discern and classify them accurately. Machine learning models, ranging from traditional algorithms to advanced deep learning approaches, are employed to analyze various data sources such as spectroscopic data, chemical compositions, and structural information. This enables the automatic identification of materials, revolutionizing processes like quality control, forensic analysis, and the

discovery of new materials with specific properties. The integration of machine learning in material identification not only enhances efficiency but also opens avenues for breakthroughs in fields like chemistry, physics, and engineering, where precise categorization of materials is paramount for advancements and innovations.

In principle, it is easy to implement machine learning classification methods directly on the raw spectral data (after smoothing), since it contains all available information. However, the training process is time consuming and the testing is also complicated due to the large number of input parameters. In addition, it is evident from the recorded spectra that many of the spectral features are in fact dependent. For this reason, a preprocessing of the raw spectral data using PCA is chosen as an effective way to reduce the dimensionality of the problem and improve the training process of the machine learning techniques evaluated. This preprocessing methodology will be also helpful for future works and implementations of the proposed methodology, where the number of materials studied may be much higher than the 16 materials in this work.

4.1 Identification of microplastics from other non-plastic organic materials

We posit that the differentiation of plastics from other organic materials is pivotal as an initial step. This is not only vital for environmental surveys, particularly in real-time measurements, but it also simplifies subsequent plastic type characterization. Our study encompassed 16 distinct materials, of which nine were plastics and seven were representative natural or synthetic materials (refer to Table 1). Our dataset consists of 540 microplastic samples and 420 other organic material samples, totaling 960 spectra. Each

Table 2 Classification results of machine learning models tested to identify plastics from non-plastics, shown with descending test accuracy

Model type	Validation accuracy	Test accuracy
SVM (<i>Cubic</i>)	99.1%	97.6%
kNN ($k=1$)	98.7%	97.2%
Neural Networks (<i>two fully connected layers</i>)	98.4%	96.9%
Ensemble (<i>bagged trees</i>)	98.1%	96.2%
Logistic regression (<i>kernel</i>)	96.9%	94.4%
Naïve Bayes (<i>kernel</i>)	92.4%	92.3%
Decision tree	92.9%	89.9%
Discriminant (<i>linear</i>)	86.0%	85.8%

spectrum includes 878 spectral channels (smoothed and normalized spectra from 420 nm to 716 nm), serving as predictors, despite a degree of interdependence among channels. PCA was applied to these predictors, yielding five components that collectively account for at least 99% of the data variance. Specifically, the first five components explained 51.6%, 37.7%, 5.7%, 2.9% and 1.1% of the variance, respectively. The classification task then utilized these five PCA components as predictors in various machine learning models. We set aside as a test dataset the 30% of the measurements (288 out of 960 samples). In addition, we implemented cross-validation (cross-validation folds was set to 5). Specifically, we divided the data into five subsets, selected at random (5 folds), each approximately the same size. The model is then trained on four of these subsets, while the fifth serves as the validation set. This method is carried out five times, rotating the validation set so that each of the five subsets is used for validation exactly once. Table 2 details the most precise models for distinguishing plastics from non-plastics.

Test results are similar to validation results, verifying the prediction accuracy of the methodology. Moreover, due to the PCA preprocessing, the prediction speed is

approximately 2000 obs/s, which is very important for the application of the method in real-time measurements. The training time is also adequately low, which will be essential for applying the method to a larger number of materials.

Figure 7 depicts the validation (Fig. 7a) and test (Fig. 7b) confusion matrices for the most efficient model (SVM with cubic kernel function), which exhibits a validation accuracy of 99.1% and a test accuracy of 97.6%. The diagonal cells indicate materials that have been correctly classified, whereas the off-diagonal cells denote those that have been incorrectly classified. It is evident that plastics are accurately distinguished from non-plastic organic materials with a very high efficiency.

4.2 Identification of different types of microplastics

The second step of our identification process is aimed at characterizing the chemical types of the microplastics identified in the first step. The previous analysis has shown accurate differentiation of plastic materials from non-plastics, prompting the use of the full dataset of microplastic materials studied, comprising 540 samples. PCA was also employed as a preprocessing method for the 878 predictors, resulting in 6 components that explain 99% of the data variance, with the first six components accounting for 64.5%, 26.9%, 4.4%, 2.1%, 0.8% and 0.4% of the variance, respectively. For the classification task, these six PCA components were used as predictors in various machine learning models. Similar to the first step, we conducted cross-validation (5-fold) and reserved 30% of the measurements (162 out of 540 samples) as a test dataset. Table 3 presents a summary of the most accurate models for predicting one of the nine different types of microplastic materials utilized in this study.

Figures 8 and 9 display the validation and test confusion matrices for the most efficient model, which is the SVM

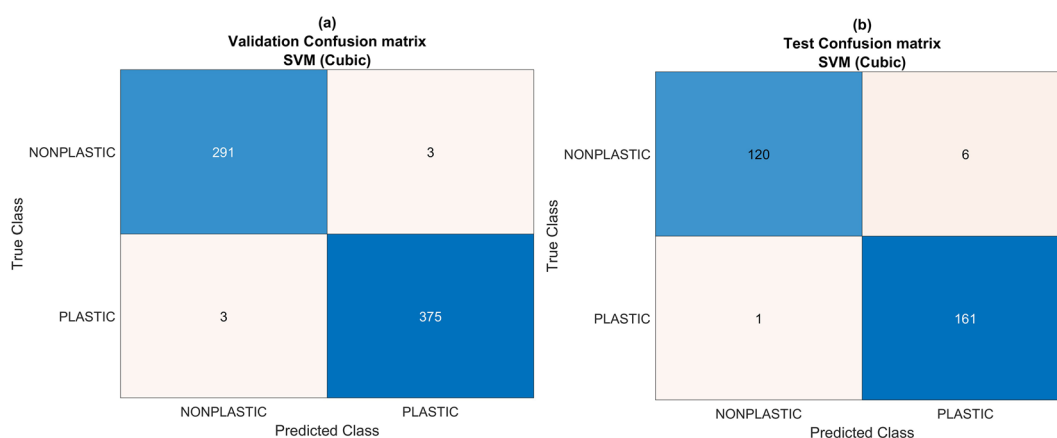


Fig. 7 Neural Network model: (a) validation confusion matrix, (b) test confusion matrix. The rows correspond to the true materials and the columns signify the predicted ones

Table 3 Classification results of the machine learning models tested for the prediction of the type of microplastic, shown in descending order of test accuracy

Model type	Validation accuracy	Test accuracy
SVM (<i>Cubic</i>)	91.8%	88.3%
Discriminant (Quadratic discriminant)	87.3%	87.0%
kNN ($k = 1$)	88.6%	85.8%
Neural Networks (<i>one fully connected layer</i>)	89.2%	84.6%
Ensemble (<i>bagged trees</i>)	86.2%	80.9%
Naïve Bayes (<i>Gaussian kernel</i>)	76.7%	76.4%
Decision tree	78.3%	75.9%
Efficient Logistic Regression	72.2%	69.1%

with cubic kernel function. This model achieves a validation accuracy of 91.8% and a test accuracy of 88.3%. The results clearly indicate that ML models can also adequately characterize different types of microplastics, as demonstrated by the analysis of the nine plastic materials included in this study.

This analysis should be considered a proof of concept, especially since the plastics used were predominantly uncolored. Future studies should investigate a broader range of materials, including non-plastic natural substances found in seawater, such as chitin. Nevertheless, it is anticipated that the presence of such materials will not significantly affect the efficiency of the machine learning models. This

expectation is supported by the findings in reference [11], which demonstrate that the spectral signatures of these organic materials (specifically from cuttlefish bone, sea snail shells, sea urchin skeletons, and black mussel shells) are distinct from those of plastics.

In addition, the fluorescence of pigments in phytoplankton (e.g., cyanobacteria and microalgae), which may form biofilms on the surface of microplastics, presents a significant challenge for microplastics real-time identification in situ. Biofilms are expected to exhibit strong fluorescence peaks at longer wavelengths, for example LIF using either 405–532 nm excitation show biofilm emission peaks at 680 nm [19, 20]. As a result, their spectra may resemble those of organic materials as seaweed and olive oil (see Fig. 3), which are included in our study and exhibit similar peaks due to chlorophyll fluorescence. These materials will be correctly identified as such if they form biofilms on microplastics. This effect complicates the correct identification of the plastic material and poses a significant problem for the real-time application of the method. However, this problem may be mitigated by conducting multiple successive LIF measurements. The photoablation induced by the focused laser beam on the biofilm could eventually expose the underlying plastic material, enabling its correct identification as plastic in subsequent measurements.

In the case of colored plastics, pigments are utilized not only to enhance aesthetic appeal but also to fulfill the intended purpose of the product. Pigments can also operate as light-shielding

Fig. 8 Cubic SVM model validation confusion matrix of the microplastics type. The rows correspond to the true materials and the columns signify the predicted ones. TPR is the True Positive Rate and FNR is the False Negative Rate respectively

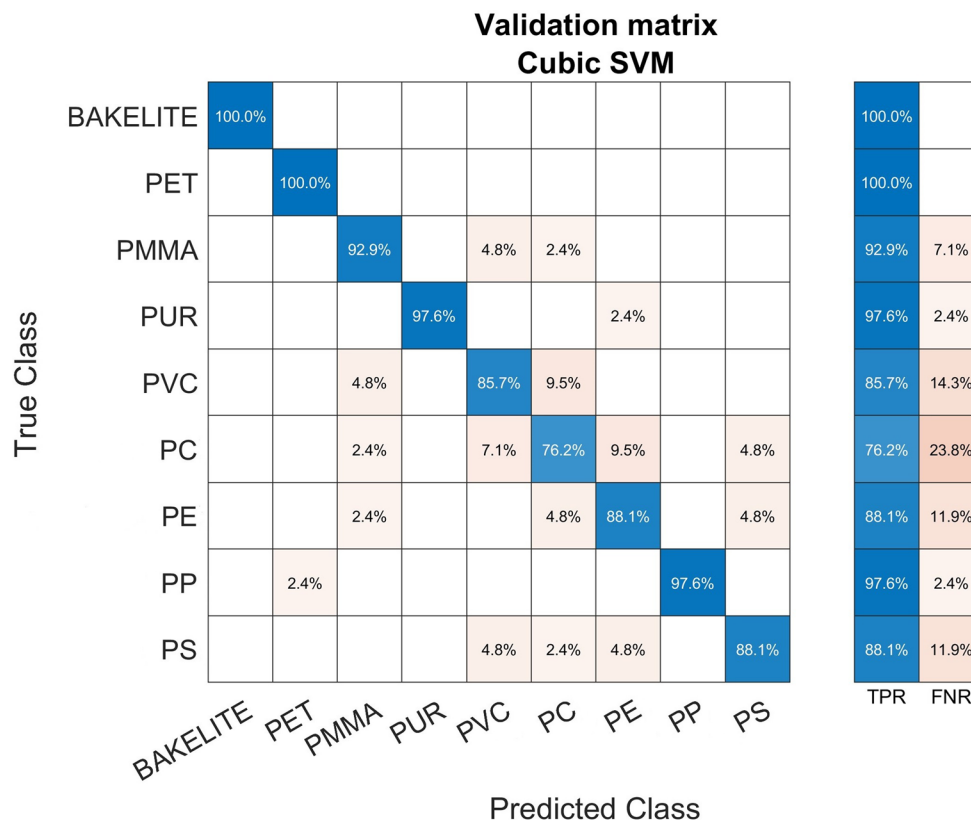
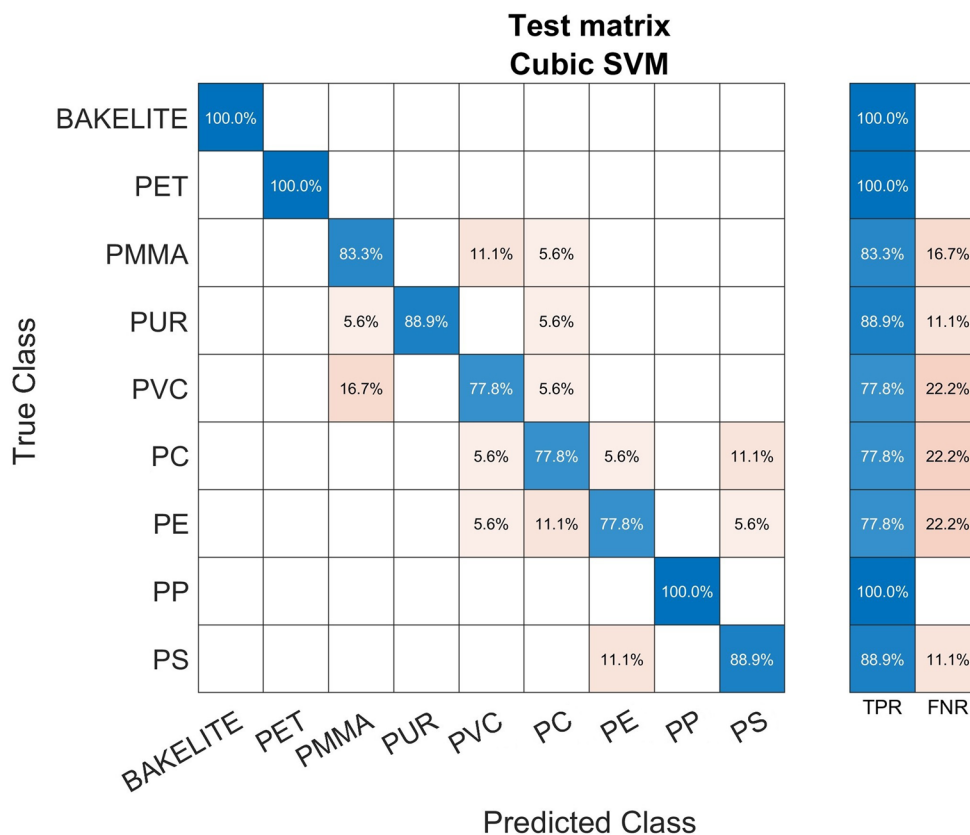


Fig. 9 SVM (with cubic kernel) model test confusion matrix of the microplastics type. The rows correspond to the true materials and the columns signify the predicted ones. TPR is the True Positive Rate and FNR is the False Negative Rate respectively



agents, absorbing a portion of ultraviolet radiation, which helps prevent or delay photodegradation, thereby extending the operational lifespan of plastic items. In general, the most abundant colors are white and transparent/translucent (47%), similar to the plastics used in our study, with yellow/brown and blue plastics comprising the 26% and 9% correspondingly [21]. Certain black colored polymer types, such as PVC in our study, which might be undetected and underestimated [22], can be adequately characterized with our method, as is shown in Figs. 8 and 9. To assess the broader applicability of our proposed methodology, we conducted additional measurements with colored plastic samples (blue, green and red) taken from the environment. In the case of blue and green samples excited by the 405 nm laser, the recorded spectra are not significantly different to those of the uncolored material samples in most cases. Nonetheless, certain colored samples, particularly red ones, exhibited additional spectral emissions, complicating the accurate characterization of the material. Consequently, the characterization accuracies presented in Tables 2 and 3 were lower for these samples. Despite this, the method effectively distinguished plastics from non-plastics in over 80% of the colored materials tested.

5 Conclusions

This study demonstrates that a comprehensive analysis of LIF spectra, generated by 405 nm CW laser excitation, can yield a select set of predictors. These predictors are derived from Principal Component Analysis and are effective for training machine learning models using a two-step approach. The potential of these predictors to differentiate microplastics from other organic materials, such as maritime oils, fuels, paints, and natural organic substances found in water bodies, is significant (accuracy 97.6%). In a second step, the accuracy of identifying different types of the microplastics studied can reach up to 88.3%. Looking forward, the application of the proposed methodology in future research involving a larger variety of materials could establish the foundation for developing an affordable pollution monitoring device. We suggest this methodology, which employs the principles of LIF spectroscopy and machine learning, for the implementation of sensors capable of real-time measurement and detection of microplastics in seawater by an Unmanned Surface Vehicle (USV). This could significantly contribute to environmental conservation efforts.

Acknowledgements E. Drakaki and E. Zekou gratefully acknowledge the financial support from the University of West Attica.

Author contributions Conceptualization and supervision: N. M.; methodology and experiments: N. M., A. L. K., E. D.; validation: N.

M., A. L. K., E. D., E. Z., G. N.; writing and editing, N. M., A. L. K., E. D., E. Z., G. N. All authors reviewed the manuscript.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

- P. Cheng, Y. Zhu, C. Cui, F. Wang, J. Pan, *Meas. Control.* **55**, 1078 (2022). <https://doi.org/10.1177/00202940221114902>
- W.P. de Haan, A. Sanchez-Vidal, M. Canals, *Mar. Pollut. Bull.* **140**, 523 (2019). <https://doi.org/10.1016/j.marpolbul.2019.01.053>
- E. Agathokleous, I. Iavicoli, D. Barceló, E.J. Calabrese, *J. Hazard. Mater.* **417**, 126035 (2021). <https://doi.org/10.1016/j.jhazmat.2021.126035>
- G. Xu, H. Cheng, R. Jones, Y. Feng, K. Gong, K. Li, X. Fang, M.A. Tahir, V.K. Valev, L. Zhang, *Environ. Sci. Technol.* **54**(24), 15594–15603 (2020). <https://doi.org/10.1021/acs.est.9b01517>
- X.F. Wei, T. Rindzevicius, K. Wu, M. Bohlen, M. Hedenqvist, A. Boisen, A. Hakonen, *Chemical Engineering Journal* **442**, Part 2, 136117 (2022). <https://doi.org/10.1016/j.cej.2022.136117>
- W.J. Shim, S.H. Hong, S.E. Eo, *Anal. Methods.* **9**, 1384 (2017). <https://doi.org/10.1039/C6AY02558G>
- J.C. Prata, J.P. da Costa, A.C. Duarte, T. Rocha-Santos, *TRAC Trends Anal. Chem.* **110**, 150 (2019). <https://doi.org/10.1016/j.trac.2018.10.029>
- S. Liu, E. Shang, J. Liu, Y. Wang, N. Bolan, M.B. Kirkham, Y. Li, *Front. Environ. Sci. Eng.* **16**, 8 (2022). <https://doi.org/10.1007/s11783-023-1700-6>
- W. Fu, J. Min, W. Jiang, Y. Li, W. Zhang, *Sci. Total Environ.* **721**, 137561 (2020). <https://doi.org/10.1016/j.scitotenv.2020.137561>
- F. Zhou, X. Wang, G. Wang, Y. Zuo, *Toxics.* **10**, 118 (2022). <https://doi.org/10.3390/toxics10030118>
- J. Ornik, S. Sommer, S. Gies, M. Weber, C. Lott, J.C. Balzer, M. Koch, *Appl. Phys. B* **126** (2019). <https://doi.org/10.1007/s00340-019-7360-3>
- E. Drakaki, E. Zekou, M. Serris, N. Merlemis, *Environ. Sci. Proc.*, **26**, 12 (2023). <https://doi.org/10.3390/envirosciproc2023026012>
- B. Lotter, S. Konde, J. Nguyen, M. Grau, M. Koch, P. Lenz, *Sci. Rep.* **12** (2022). <https://doi.org/10.1038/s41598-022-23414-3>
- B.O. Asamoah, E. Uurasjärvi, J. Rätty, A. Koistinen, M. Roussey, K.-E. Peiponen, *Polymers.* **13** (2021). <https://doi.org/10.3390/polym13050730>
- X. Meng, S. Chen, D. Li, Y. Song, L. Sun, *J. Hazard. Mater.* **465** (2024). <https://doi.org/10.1016/j.jhazmat.2023.133352>
- N. Merlemis, A.L. Kesidis, I. Sianoudis, *Eur. J. Phys.* **42**, 015304 (2020). <https://doi.org/10.1088/1361-6404/abba01>
- A. Chekalyuk, M. Hafez, *Opt. Express.* **21**, 29255–29268 (2013). <https://doi.org/10.1364/OE.21.029255>
- P. Chen, D. Pan, Z. Mao, *Opt. Laser Technol.* **64**, 213–219 (2014). <https://doi.org/10.1016/j.optlastec.2014.05.021>
- S. Khayatian, A. Bavali, S. Moradi, M. Farhadi, S.B. Jameie, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **284**, 121791 (2023). <https://doi.org/10.1016/j.saa.2022.121791>
- J. Marques da, Silva, A. Borissovitch, Utkin, *Processes.* **6**, 227 (2018). <https://doi.org/10.3390/pr6110227>
- E. Martí, C. Martín, M. Galli, F. Echevarría, C.M. Duarte, A. Cózar, *Environ. Sci. Technol.* **54**(11), 6594–6601 (2020). <https://doi.org/10.1021/acs.est.9b06400>
- Y. Huang, E.G. Xu, *Water, Emerg. Contam. Nanoplastics.* **14**(1) (2022). <https://doi.org/10.20517/wecn.2022.10>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.