Classification of diverse plastic samples by LIBS and Raman data fusion

Daniel Holub [a,b,*], Dávid Jenő Palásti [c], Krisztian Fintor [d], Pavel Pořízka [a,b,e], Gábor Galbács [c], Jozef Kaiser [a,b,e]

[a] Faculty of Mechanical Engineering, Brno University of Technology, Technická 2896, 616 69, Brno, Czech Republic
[b] CEITEC Brno University of Technology, Purkyňova 123, 612 00, Brno, Czech Republic
[c] Department of Molecular and Analytical Chemistry, University of Szeged, Dóm Square 7-8, 6720, Szeged, Hungary
[d] Department of Mineralogy, Geochemistry and Petrology, University of Szeged, Egyetem Street 2, 6722, Szeged, Hungary
[e] Lightigo s.r.o., Renneská třída 329/13, 639 00, Brno, Štýřice, Czech Republic

ARTICLE INFO

ABSTRACT

The plastic production and usage in the world is steadily increasing. This leads to increased amounts of plastic waste. Most of the waste could be potentially recycled, but only 14 % of plastic waste is recycled. In order to increase the share of recycling in plastic waste management, the recycling process should be completely automated. The problematic part of sorting is being solved by either manual (labor-intensive) or spectroscopy-based (still in development) methods. In this work, we propose the data fusion of Laser-Induced Breakdown Spectroscopy (LIBS) and Raman spectroscopy as a fast, robust, and reliable way to sort/classify any potential polymer material. The sample set of this work consists of several types of polymers in clear, colored, and even mixture versions. So far, no LIBS/Raman classification works involved all these categories in one experiment. Additionally, the low and medium level of data fusion is discussed, and the performance is compared. By using LIBS and Raman data fusion method and both linear and nonlinear chemometric techniques, increased accuracy reaching more than 98 % in the classification of investigated plastic samples was achieved, which was a significant improvement when compared with singular methods classification accuracy.

1. Introduction

There has been a steady rise in the production and use of objects made from plastics in the world for several decades. This results in an increasing volume of plastic waste which needs to be dealt with. Around 75 % of all produced plastic objects end up as waste. Out of these 75 % only 14 % is being recycled. The rest of the plastic waste is either incinerated (14 %), put in a landfill (40 %), or released into the environment without any control (32 %) [1]. However, the majority of the plastic waste could be recycled and thus, there is a big attention to increase the share of recycling of plastic waste. The need for an increased share of recycling can also be seen in the pollution of the most remote ecosystems [2] or the human bodies [3].

If the post-consumer plastic object enters the recycling process, there is a need to correctly specify the type of the polymer. Firstly, the plastic debris is cleared of any material of other nature such as glass or metal. Afterward, plastic waste is put on a conveyor belt and either manually or automatically sorted by the polymer type. The automatic sorting can be done by a spectroscopic method in which case a version of infrared spectroscopy (near, or short-wave infrared spectroscopy – NIR, SWIR) is used. For more challenging samples like high-density and low-density polyethylene (HDPE/LDPE) or heavily colored samples some other method for sorting is needed as the NIR/SWIR method is not able to sort these samples [4]. The full automatization of the process could bring additional capacity to recycling plants while also decreasing the cost of recycling. We believe that a combination of Laser-Induced Breakdown Spectroscopy (LIBS) and Raman spectroscopy is a method versatile and quick enough to be the perfect candidate for reliable plastic sorting.

Laser-Induced Breakdown Spectroscopy (LIBS) is an analytical technique which utilizes a high power pulsed laser. The laser pulse which usually lasts several ns is focused in a typically 10–100 μm wide spot on the surface of a sample, where it evaporates the material. The ablated material is further atomized by the laser beam and a microplasma is formed. The analytical information is gained by recording the plasma emission which carries information mainly about the elemental composition of the sample including light elements like H, Li, etc. In

some cases even molecular information can be observed, especially when organic material is investigated [5,6]. The result is usually qualitative, but it can also give quantitative results [7]. After the boom of the LIBS method at the turn of this century, the method has been accepted as a modern spectroscopic method across various fields. Its versatility enables the use of LIBS even in the most remote of places [8,9]. The needed instruments are reasonably simple and fairly robust, easy to operate with low installation and maintenance costs, which makes these instruments desirable for industrial applications. The method also started to be used in industrial applications for sorting [10,11] or detection of trace elements [12,13]. The ability of LIBS to be used in the polymer industry has been shown in several works already [14–21].

In the work by J. Anzano et al. [14] a simple LIBS system was used to classify different polymers by extracting ratios of different spectral lines or molecular bands. The most important ratios for the classification were H/C_1 and C_2/C_1. Other older polymer related LIBS papers deal with classification [15,16] or double-pulse LIBS optimization [17].

Some of the more thorough polymer classification papers are by Unnikrishnan et al. [18] and Boueri et al. [19]. The work by Unnikrishnan and his team applied Principal Component Analysis (PCA) to a CN band area of four different polymer samples and reached more than 80 % classification accuracy. The work by Boueri includes 8 different polymer species and their classification is done via artificial neural networks (ANN). The classification accuracy of unknown PVC samples reaches above 95 %. Both works highlight especially the speed of the LIBS approach to the polymer sorting, but they encounter problems either in the similarity of polymer spectral responses or with variance in the spectral signal of the same polymer species when an additive is present.

The work by Chamradová et al. [20] explores the performance of LIBS in the discrimination of five polymer samples under various experiment parameters. The most noteworthy are the changes in the experiment surrounding gas conditions and the discussion from the perspective of the monomer composition. It has been found out, that helium atmosphere is not suitable for LIBS analysis of polymer samples, while argon and air atmospheres lead to satisfactory results. This work does not deal with polymers with additives and no classification algorithm was trained. Similar work with an increased number of polymers showed better results in Argon atmosphere [21]. These results are contradictory and this work aims to use conditions relevant to polymer recycling plants, so air atmosphere is chosen.

Raman spectroscopy is an already established spectroscopy method used in various fields. It is widely used because of the simplicity of the instrumentation and minimal sample preparation requirements [22]. It can be used for a wide range of applications and the limits of the method are being sought [23–26]. The Raman spectroscopy method is very powerful in polymer discrimination because of their different Raman responses [27]. The drawback of Raman spectroscopy is the danger of fluorescence of the sample. The fluorescence is usually much more intense than the Raman signal, thus the Raman signal is lost and cannot be analyzed. To tackle this problem the employment of different laser wavelengths is needed, or more elaborate data evaluation has to be used [28,29].

In the main part of this work, we are using the fusion of LIBS and Raman spectroscopy data. This fusion is beneficial due to the complementary nature of these two techniques. LIBS yields elemental information about the sample in the form of spectral lines corresponding to elemental and molecular species in the sample. Raman spectroscopy on the other hand provides information about molecular vibrational and rotational responses to the inducing laser signal. This atomic and molecular information together has already been proven to offer complementary information confirming statements or leading to new results [30,31]. Some of these works also use plastic samples, but as the information is only discussed side-by-side, the presented results do not improve noticeably [32–34].

One of the work employing both LIBS and Raman spectroscopy on

plastic samples is a work by M. Shameem et al. [33]. By using a custom system able to gather both LIBS and Raman spectra, the work then uses PCA with different metrics to classify 4 polymer types and additionally several unknown PVC samples. Both data sets in all phases of the work are analyzed separately. On the other hand, the results do correspond to each other, and the side-by-side comparison serves as a good validation of obtained results by one technique or the other. However, no fusion of the data sets is done and thus does not use the full potential of the combined system.

A work which connects more than two spectroscopic techniques to detect and analyze plastic samples is the work by P. Pořízka et al. [34] In this work, laser-based spectroscopic methods (LIBS, Raman spectroscopy and Laser ablation inductively coupled plasma mass spectrometry – LA-ICP-MS) is used to detect and analyze microplastics aged in various aquatic environments. The work shows that all three methods are able to detect the polymer signal on the sample no matter the aging setting. Additionally, Raman spectroscopy and LA-ICP-MS are also able to classify the polymer type of the aged microplastic. On the other hand, the three methods are analyzed and discussed separately, so no data fusion is applied.

The fusion of the data sets can further enhance the complementary information as the information from both data sets is taken into account simultaneously. The methodology of this data fusion on LIBS and Raman spectroscopy was first introduced by M. Hoehse [35] and used for the classification of inks and pigments. The benefits of the LIBS and Raman fusion were also confirmed by group based in Brno, Czech Republic in a work about the classification of different bacteria [36].

Other works employing LIBS + Raman data fusion are the works by K. Rammelkamp et al. [37] on Mars-relevant salts, E. Gibbons et al. [38] on soils or Lihui Ren et al. [39] on fish samples. All these works contain some discussion on the LIBS + Raman data fusion methodology, but in some cases, the LIBS and Raman data fusion benefits are not clear (the sample set is not challenging enough), or a discussion of the data fusion requirements and influence is missing. Data fusion can be applied to other spectroscopic techniques as well. In one of the more thorough works LIBS data is fused with FTIR-ATR (attenuated total reflectance Fourier-transform mid-infrared spectroscopy) to increase the accuracy of soil organic matter (SOM) prediction. The work provides a thorough discussion of the effects of data fusion on its results and clearly states the benefits of data fusion [40].

Up to our knowledge, no previous work on LIBS and Raman spectroscopy data fusion was presented on plastic samples. Moreover, none of the LIBS + Raman data fusion works discuss the influence of the preprocessing options on the results. While the number of polymer types is not high, the number of different colors and the presence of polymer mixtures in the data set cover the vast majority of plastic debris that can be encountered in plastic sorting lines. With a nonstandard approach to data fusion of LIBS and Raman spectroscopy, this work highlights the benefits of such methodology and presents its performance on a highly challenging data set.

2. Experimental

2.1. Samples

Samples used in this project are selected to contain the most common polymer types that can be encountered in post-consumer plastic waste – it included polymers with minimal modifications, mixtures of polymers, and colored polymers. To have variability in the sample set, samples consisted of four basic polymer types – polystyrene (PS - General purpose PS Synthos 154, Synthos, POL), polycarbonate (PC – Polycarbonate standard, Polymer Institute Brno, CZE), polypropylene (PP - Mosten GB 218, Unipetrol Orlen, CZE) and polyethylene (PE - General purpose LDPE Bralen FA 03–01, Slovnaft, SVK). Out of these types, PS, PP, and PE could also be colored by extrusion with PE-based color additives (99 % bulk polymer, 1 % PE-based color additive, VIBATAN pigments, VIBA

group, ITA) to five different colors – white, yellow, orange, red, and dark blue. The sample set also contains two sets of polymer mixture samples. Each set consists of four samples with the following ratios of a pair of polymer types – 20/80, 40/60, 60/40, and 80/20. These two polymer-type pairs are PC/PS and PP/PE. Also counting the color additives which are separately available for analysis, there are 32 available samples altogether (see Table 1).

2.2. Laser induced breakdown spectroscopy

LIBS experiments were performed on a J-200 Tandem LA/LIBS instrument (Applied Spectra, USA) equipped with a 266 nm, 6 ns Nd:YAG laser source and a six-channel CCD spectrometer with a resolution of 0.07 nm. For every laser shot, the full LIBS spectrum over the wavelength range of 190–1040 nm was recorded in the Axiom data acquisition software (Applied Spectra, USA), using a 0.5 μs gate delay and 1 ms gate width. No specific atmosphere was used, the experiment was carried out in air at atmospheric pressure.

The LIBS experiment settings were as follows. The energy of 14 mJ was used with a spot size setting of 60 μm. On each sample, 50 (25 for color additives) spots were ablated with two laser shots at each location resulting in 100 (50) spectra per sample. A number of two shots at the same location were chosen because of no visible contamination on the surface of the sample. An increased number of laser shots per spot could be employed to tackle visible surface contamination. Acquired spectra were background corrected by moving minimum method, and further used by custom script in Python programming language.

2.3. Raman spectroscopy

Raman experiments were performed on a DXR3 Raman microscope (Thermofisher scientific, USA) using a laser of 780 nm wavelength with a power output of 14 mW. The investigated spectral range was 100–2000 cm^{-1}. The experiments were done with an objective of 50 × magnification (Leica, GER).

Each spectrum consisted of 10 × 1 s spectral acquisition accumulated together. On every sample, 10 different spots were measured. Acquired spectra were background corrected to tackle any possible fluorescence by the OMNIC software (Thermofisher scientific, USA) via iterative polynomial fitting. The resulting data set was uploaded to the custom Python script for further data treatment.

2.4. Labelling options and settings for classification algorithms

To provide accurate and comparable results, different labeling options will be used throughout the document. The main reason for this is to use all possible points of view of the data set. The labeling options are as follows.

- **Matrix labeling** – All data is used, but only the matrix information is considered. For example, the yellow PE sample will be labeled as "PE" and the clear PP sample as "PP".
- **Matrix labeling without additives** – Same labeling as "Matrix labeling", but all color additive data are omitted from the data set.
- **Matrix + color labeling** – All data is used, and complete information is included in the label meaning yellow PE is labeled as "PE yellow" and clear PP as "PP clear".
- **Matrix + color labeling without additives** – Same labeling as "Matrix + color labeling", but all color additive data are omitted from the data set.

By making these four labeling options it is possible to evaluate all aspects of the classification capabilities of the chosen spectroscopic methods as well as the performance of selected classification methods. The color additives are omitted from half of the labeling options because their addition does not provide additional polymer type but only makes the data set more challenging for the classification algorithms.

In this work, two classification algorithms will be used, namely Linear Discriminant Analysis (LDA) and Random Forest (RF). LDA is a projection-based method that reduces the dimensionality of the dataset by projecting it into a new space created by the linear combination of the variables in the original dataset, where the variables are not correlated. The aim of the projection is to find the planes along which the separation of the groups is the best. It is generally similar to the Principal Component Analysis (PCA), but it needs a supervised input to define the groups to separate [41].

RF on the other hand is based on the classification and regression tree method, which utilizes binary conditions (yes or no questions) in a tree structure. It is an easy method to visualize and help identify the important variables, that differentiate between the groups, but due to its hierarchical nature, this method can be very sensitive to the input. To make this method more robust Breiman created the Random Forest version of the method [42], which is based on growing multiple decision trees, all of them using a little bit different input (different portions of the dataset are excluded from the model). The categorization of the

Table 1
Summary of all available samples. The number of dots determines the number of available samples. Four dots at PC/PS and PP/PE groups mean the availability of four different percentages of the polymer mixtures. The red dot at orange color additive means that the sample was omitted from the sample set as is explained in the discussion.

Matrix → Color ↓	PC	PE	PP	PS	PC/PS	PP/PE	Color additive
Clear	●	●	●	●	●●●●	●●●●	×
White	×	●	●	●	×	×	●
Yellow	×	●	●	●	×	×	●
Orange	×	●	●	●	×	×	●
Red	×	●	●	●	×	×	●
Blue	×	●	●	●	×	×	●

unknown samples is decided by the majority vote of members of the forest. This version of the algorithm is also used in this work.

To use the classification methods properly, the data has to be divided into training and testing data sets. A 30/70 split was chosen for this work as we try to simulate real-life scenarios, where the availability of training data or samples will be problematic. The expectation is that some model samples of different kinds will be chosen to serve as a training sample set and the testing data set will be formed by a much larger sample size.

Additionally, both methods are separated by the 30/70 split separately and fused data sets are formed separately for training and testing data. This way, no data leak is possible, and the algorithm cannot be trained on spectra, which would be present in the testing data set.

Both used classification algorithms are part of a scikit package for the Python programming language [43]. The Linear Discriminant Analysis (LDA) algorithm can be controlled only by a maximum number of used Linear Discriminants (LD), but in our study, all LDs were used for all data sets as the LD number reduction did not have a considerable influence on the computation time or classification accuracy of the method. RF can be controlled in more variables and the final settings were: number of estimators = 100, max features = square root of the number of features, max depth = 10 (there is a maximum depth of 10 splits for the nodes), other settings were not restricted or changed from default. Both algorithms are evaluated for each data set option by three metrics: time, training accuracy, and classification accuracy.

- **Time [s]**: time needed for the classification model to perform.
- **Training accuracy [%]**: cross-validation score of the training model for five folds and three repeats of the scoring calculations.
- **Classification accuracy [%]**: percentage of the correctly classified spectra of the data set. Calculated from the confusion matrix as $\frac{TP}{Tot} = CA$, where TP are true positive spectra – spectra which were correctly classified, Tot is the total number of the testing spectra and CA is the classification accuracy.

3. Results and discussion

3.1. Classification based on LIBS data

To fully understand the benefits of LIBS + Raman data fusion, it is mandatory to first consider both of these methods individually. As has been described in the Introduction section, the data processing in the individual sections is minimal. This enables consideration of the individual method's limitations and benefits. For the first outlook of the data set, Principal Component Analysis (PCA) is used for both Matrix and Matrix + color labeling options. The PCA is a valuable tool for the detection of variables, which provides the separation between data. The variables that contribute to the separation of samples the most in the PC space can then be selected as important features in further analysis. This feature selection and sequential dimensionality reduction will be discussed in detail later in the manuscript (a typical LIBS spectrum is available in the supplementary material – Fig. S1).

From the PCA score plot for the first two principal components (Fig. S2), it is obvious, that not many distinct groups of data can be separated from each other. One outlying sample group is the data from the white color additive, which has a very rich LIBS spectrum as the colorant contains titanium dioxide. The other samples form a more or less uniform group. The PC space separation does not get better even if we consider the data set without the color additives, but without the color additives, the loading plots show more than only titanium lines and can be considered for the spectral feature selection. Regions from the first three PCs, which explained over 80 % of the total data set variance, were considered. Together with polymer-related spectral areas, which were added in agreement with the previous work [20], the selected spectral features consisted of C I 247.86 nm, CN violet transition

area around 388 nm, Ca II 393.37 nm, C_2 transition $v = -1$ and $v = 0$ areas (473.70 nm and 516.50 nm), Na I doublet around 589 nm area, H I 656.28 nm and K I doublet around 766.49 nm area. This spectral area selection will be introduced to the data fusion in the section *LIBS + Raman data fusion*.

The background corrected LIBS data was then introduced to the classification algorithms. The results confirmed the outcome suspected from the PCA. With a calculation time of over a minute (76 s on average), the LIBS data were challenging to classify. The classification accuracy for the Matrix labeling did not exceed 75 % on average. It improved by 5 % for the Matrix without additives data set (see Fig. 1), however, if the Matrix + color labeling was used, the classification accuracy was really poor at around 55 %.

The main problem observed with the LIBS data is the very similar spectral response of different plastics in the investigated UV–Vis range. All polymer matrices used in this study contain only carbon, hydrogen, and oxygen. Consequently, the color additives may overshadow the differences between the plastic samples by their own spectral responses and also different matrix effects because of the different light-matter interactions between clear, light, and dark samples.

3.2. Classification based on Raman spectroscopy data

The Raman and Near Infrared (NIR) spectroscopy are both structure sensitive methods, and they are well known for their ability to discriminate between different polymers, the latter is already used in several recycling plants and polymer production companies to determine the polymer type. The discrimination accuracy of both methods suffers when the plastic sample contains a lot of additives, especially dark colorants. NIR cannot acquire spectra from dark samples at all, and Raman spectroscopy usually fails because of the fluorescence caused by the presence of these additives.

Out of all the samples, only the orange color additive showed severe fluorescence. The fluorescence could not be overcome in any way and the Raman spectrum was not obtained. The orange color additive sample was omitted from all the data sets because of this. Nonetheless, the plastics colored with the orange additive had a sufficient Raman spectral response and were included in the data set (A typical Raman spectrum is available in the supplementary material – Fig. S3).

The PCA analysis showed better discrimination with some groups being separated nicely from the rest of the data groups (Fig. S4). However, the majority of data groups formed a big cluster of data, which was not possible to separate via PCA. This big cluster of data was formed mainly from the colored plastic samples, but some clear samples were also present. The PC loadings cannot be used in the same sense as in the LIBS analysis, because the occurrence of the Raman spectral bands is not uniform across different samples. On the other hand, the fingerprint part of the Raman spectrum (between 150 and 2000 cm^{-1}) which is used in this study is narrow enough and can be used as a whole with a simple dimensionality reduction method. The dimensionality reduction method exchanges every two data points by the mean value of these two data points. This method of dimensionality reduction is not very sensitive, as it can erase some very thin spectral lines. On the other hand, Raman spectra usually consist of wide spectral bands, which are almost immune to this kind of dimensionality reduction as they lose no information content by it. This dimensionality reduction is introduced to unify the dimensions of Raman and LIBS parts in the fused pseudo-spectra as will be discussed later in the manuscript.

The methodology for the evaluation of classification was the same as for the LIBS data. Background corrected Raman data was used with LDA and RF algorithms and the results confirm the prior assumptions. The computation time is much shorter, but the data set is also 10 times smaller as well as the number of points of each spectrum is lower (12288 for LIBS and 3715 for Raman). The Raman classification accuracy for Matrix labeled data reaches 84.3 % for the testing data set with 82.7 ± 5.7 % accuracy for the cross-validation of the training data set for the
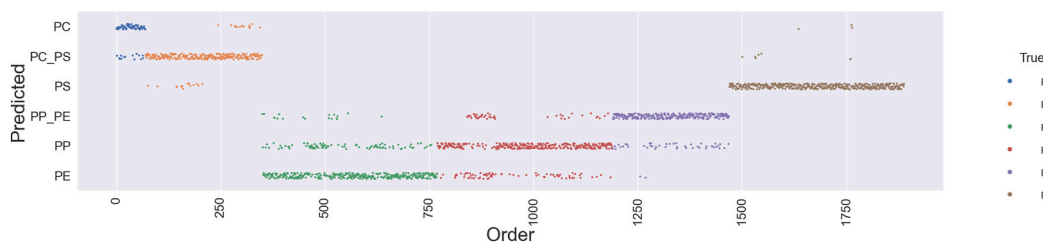
Fig. 1. A classification accuracy visualization for the Matrix without additives labeling. Each point is one spectrum from the test data set. Color depicts the true class, location in a row shows how the spectrum was classified. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

LDA algorithm and 89.4 % testing accuracy with 86.4 ± 2.7 % training accuracy for the RF algorithm (see Fig. 2). The removal of the color additive data from the data set improves the classification accuracy by 10 % (94.7 %) and 7 % (96.2 %) for LDA and RF respectively. The classification accuracy of more than 95 % is already very good and the confusion mainly occurs between PP, PE, and the mixed PP/PE classes. However, the addition of color information influences the classification accuracy significantly. The training accuracy drops by 5–10 % for both classifiers and LDA test accuracy stays above 90 % (92.1 % for the Matrix + color data set, 92.5 % for the Matrix + color without additives data set). On the other hand, the test accuracy of the RF classification drops drastically and does not reach more than 66 % for either of the two Matrix + color data sets. This indicates that the algorithm may need a bigger data set. Training the data by only three spectra does not provide the desired results for non-linear methods such as RF. The simplicity and linearity of the LDA algorithm help in this rare case, where the training data set is small and the class number is big.

3.3. LIBS and Raman data fusion

The data fusion was done by adding together the LIBS and Raman spectrum of a sample to form a new pseudo-spectrum (low level data fusion) that contains information from both methods. The two connected spectra do not have matching parameters like dimension and intensity range. To suppress the effect of these dissimilarities, both the dimension and the intensity range should be unified between the two data parts (middle level data fusion). If the dimension is not unified, the spectra of one method have a bigger influence on the resulting behavior than the dimensionally smaller spectrum. To tackle this issue, several areas of the LIBS spectrum were selected based on prior knowledge and

the PC loading vectors, as discussed in the LIBS section. The Raman spectra dimension was reduced to half by exchanging every two points of the spectrum by their mean value, as was discussed in the Raman section of the manuscript. This was necessary to unify the dimensions of the two methods as the Raman spectrum contained around 3700 dimensions and feature selected LIBS parts only around 1900 dimensions.

If the intensity range of the two methods is not controlled, the scale of the methods and the difference between the setups will have an effect on the outcome of the data analysis. Additionally, the matrix effect can be reduced by scaling the data sets as well. Every spectrum for both methods was min-max scaled (0–1) to unify the intensity ranges. After preprocessing the spectra from both measurement methods, the data could be added together and thus fused. The separation for training and testing sets was preserved by fusing only training LIBS and training Raman data to form training fused data set. To show the influence of the data preprocessing on the outcome of the classification calculations, three data sets were prepared.

- **Raw data fused data set** – the LIBS and Raman spectra were only background corrected and fused. No additional data treatment was done (low level data fusion).
- **Peak selected fused data set** – the data with dimensionality reduction were fused together. The LIBS part consists of the selected spectral areas specified in the LIBS section and the Raman part consists of the spectra with the reduced dimension specified in the Raman section. Both spectra sets were left with original intensity values.
- **Processed fused data set** – in addition to the dimensionality reduction, these spectra were also min-max scaled before the data fusion. The dimension and the intensity range of the LIBS and Raman spectra in this data set are synchronized (middle level data fusion).

The data fusion time changes with the reduction of the dimension and intensity ranges from 35 to 10 s, but the data preprocessing takes considerably longer time (around 5 min), so the benefit of dealing with preprocessed data is not significant in this case.

The raw data fused data set consists of all the information available in this study. Raw data fused pseudo-spectra for clear PC and blue PS are shown in Fig. 3A. The biggest disadvantage is that the fused data are very big and it takes a lot of time and memory to deal with them. In addition, the data contains a lot of redundant and unimportant pieces of spectra such as multiple spectral lines of the same element and noise. As the LIBS part of the fused pseudo-spectra is much bigger, the PC score plot (Fig. S5) resembles the individual LIBS PC score plot quite accurately. In addition to this, the size of the data set makes the computations of the classification methods very slow. The LDA calculation takes from 8 to 10 min and the RF one takes 18–21 s. But as the pseudo-spectra contain all the information from the two spectroscopic methods and there is no restriction on the variance used for the calculation of LDA and RF algorithms, the classification accuracy result increases dramatically. The test classification accuracy for the Matrix labeling reaches 87.6 % and 98.5 % for LDA and RF respectively, Matrix labeling without
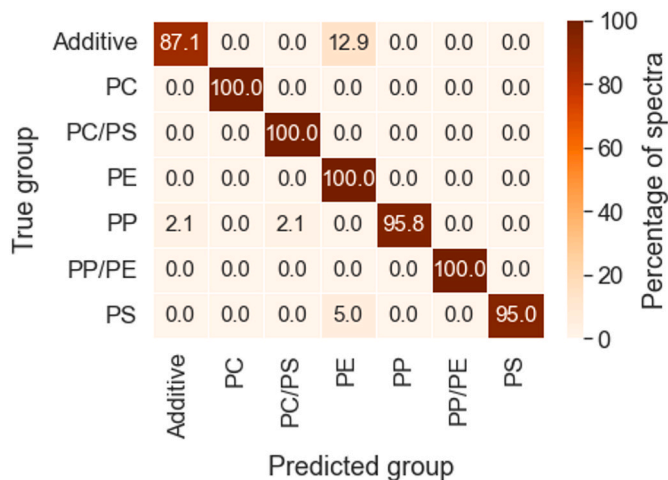

Fig. 2. Confusion matrix of RF results on categorical Raman data. Numbers on the main diagonal are the percentage of correctly classified spectra. Numbers in a row stand for spectra from one sample.
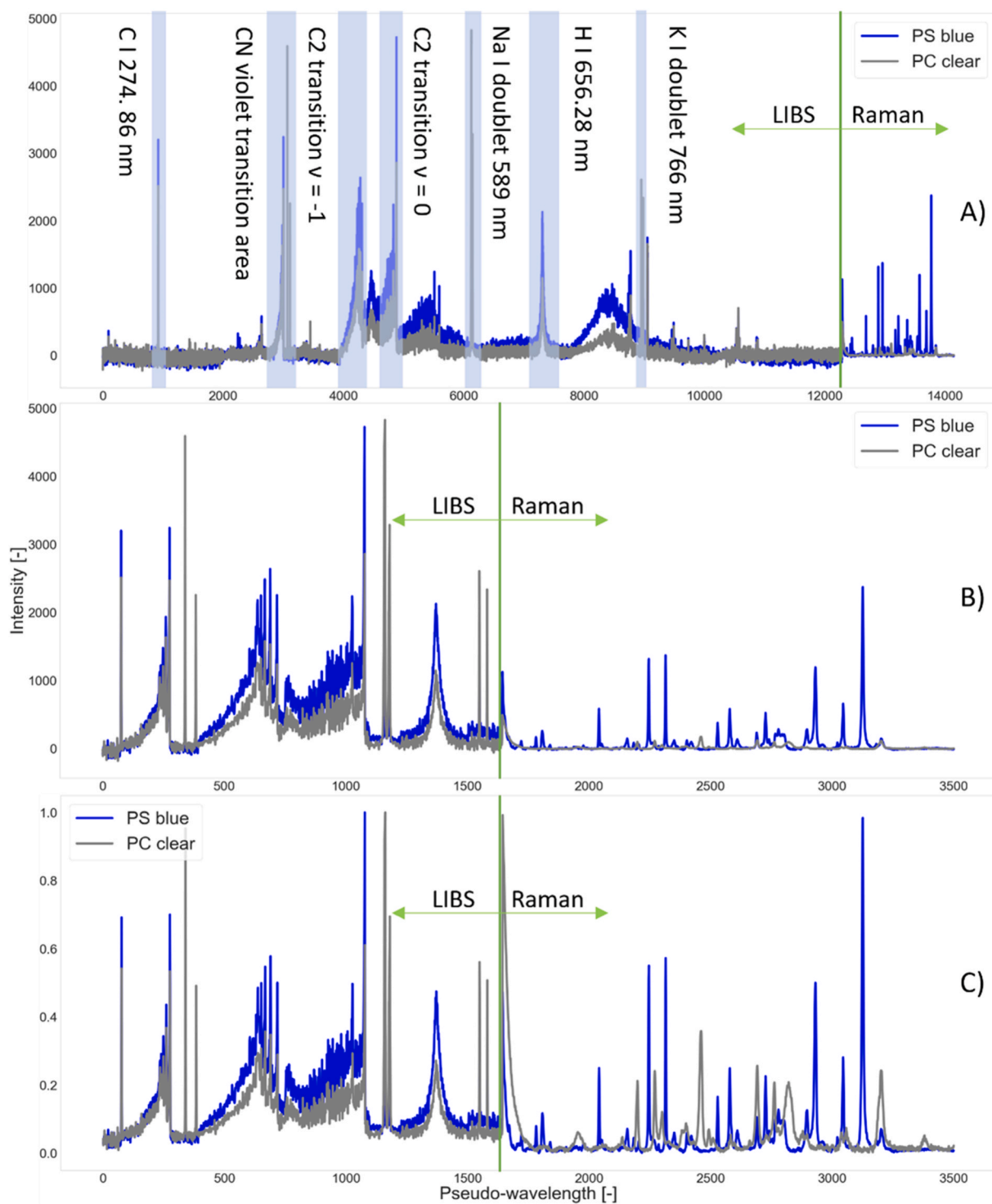
Fig. 3. The three versions of the LIBS and Raman data fusion used in this work. A) Raw fused data set, B) Peak selected fused data set, C) Processed fused data set. In the A) section, spectral areas chosen for LIBS dimensionality reduction are highlighted. The Raman data would not benefit from such selection as is described in section 3.2.

additives results in a perfect (100 %) classification by the RF method and 96.2 % test classification accuracy for the LDA algorithm. The under 90 % result for the LDA algorithm at the Matrix labeling with additives is influenced by the LIBS part of the spectrum and its importance in the linear decision method. The Matrix + color labeling does not have a bigger effect anymore, as the test classification accuracy reaches 97.7 % for the LDA algorithm and 96.1 % for the RF algorithm, and for the Matrix + color labeling without color additives, the results are almost identical (97.7 % for LDA and 96.2 % for RF). These results prove that the combination of LIBS and Raman spectroscopies can be used for a very wide range of plastic samples including clear polymers, mixtures of polymers, and colored plastics. However, the biggest advantage of the suggested methods – speed – is reduced considerably by operating with raw data.

The peak selected fused data set should contain only important spectral information gained by LIBS and Raman spectroscopy. The fused

pseudo-spectra with reduced dimensions for clear PC and blue PS are shown in Fig. 3B. By matching the dimensions of LIBS and Raman parts of the pseudo-spectra, the automatic algorithms should take advantage of the information of both methods simultaneously. In the PCA score plot the distribution of the data is still similar to the distribution of the data in the PCA score plot for LIBS data only, however, better discrimination between several groups of data can be observed. A big difference can be noticed in the computation time of the classification algorithms. The LDA algorithm calculation time dropped to under 2 min and the RF algorithm took 7 s to calculate. As all the important spectroscopic information was preserved in the data set, the classification accuracy is comparable to the classification accuracy of the raw fused data set. LDA test classification accuracy for Matrix labeled data reached 90.3 % with 99.7 ± 0.3 % training accuracy and RF test classification accuracy reached 98.9 % with 98.5 ± 1.0 % training accuracy. These results were replicated for all other labeling options with a maximum of 2 % difference. The composure and the high level of these results exceeding 90 % on average for LDA and 97.5 % on average for RF are the results, which were sought by this study. The robustness and speed of this method show very good potential for further laboratory and in-situ tests of joint LIBS and Raman plastic classification systems with data fusion algorithms.

The processed fused data set carries only the spectroscopic information as all the other aspects of the spectra are diminished. The majority of the noise does not play a role because of the dimensionality reduction and the differences between systems and matrices are accounted for by the standardization of every spectrum. The resulting pseudo-spectrum has matching dimensions and intensity ranges of the LIBS and the Raman part as seen in Fig. 3C. The PCA score plot does not resemble any other PCA score plot anymore, but the PCA algorithm has problems dividing the data, as the data points overlap in the PC space of the first three PCs and it takes 9 PCs to reach 90 % of the explained variance of the data set. This suggests that the linear method is not suitable for a preprocessed data set. The computation time of the classification algorithms is comparable to the peak selected data set as the number of dimensions stays the same. However, the results show a critical decrease in the classification accuracy for the LDA algorithm. As was already visible in the PCA space, a linear method is not enough to discriminate the data and even with training accuracy exceeding 97 %, the test classification accuracy does not reach more than 50 %. On the other hand, the non-linear method – RF – keeps the performance even on this preprocessed data set as the spectroscopic information is still present, but it is not accompanied by matrix effects and differences between systems.

The summary of the classification results for both LDA and RF (Table 2) confirms prior expectations of the manuscript. The classification accuracy increases by the data fusion of LIBS and Raman spectroscopies and the joint dataset can be used to classify even challenging samples like colored plastics and mixtures of polymers. Surprisingly, preprocessing the data does not increase the classification accuracy, as the spectral signal stays similar, but due to matrix effects and different light-matter interactions between light and dark samples, the intensity of the signal can help with the classification. The non-linear classifying method – Random Forest – yields better results than the linear method LDA. The calculations of the RF are also faster, but need more data for the training, as it failed to train properly on the smallest data set (individual Raman data set). Overall, reaching almost perfect classification accuracy with a diverse sample set of plastics confirms the robustness and universality of suggested LIBS and Raman data fusion.

4. Conclusion

The presented work focused on showing the capabilities of data fusion of LIBS and Raman spectroscopy data for plastic sample sorting. This work included a very wide range of plastic samples to contain as many real-life scenarios as possible – clear versions, colored versions, and even mixtures of polymer types. Some of the most colored samples

Table 2
Summary of test classification accuracy results for LDA and RF.

Linear Discriminant Analysis (LDA)				
Labeling option →Data set ↓	Matrix (7 classes)	Matrix without additives (6 classes)	Matrix + color (31 classes)	Matrix + color without additives (27 classes)
LIBS only	73.6 %	82.9 %	58.2 %	63.2 %
Raman only	84.3 %	94.7 %	92.1 %	92.5 %
Raw data fusion	87.6 %	**96.2 %**	**97.7 %**	**97.7 %**
Peak selected data fusion	**90.3 %**	88.4 %	90.8 %	92.4 %
Processed data fusion	36.9 %	48.0 %	31.7 %	31.1 %

Random Forest (RF)				
Labeling option →Data set ↓	Matrix (7 classes)	Matrix without additives (6 classes)	Matrix + color (31 classes)	Matrix + color without additives (27 classes)
LIBS only	74.4 %	78.0 %	53.1 %	55.4 %
Raman only	89.4 %	96.2 %	63.1 %	65.6 %
Raw data fusion	98.5 %	**100 %**	96.1 %	96.2 %
Peak selected data fusion	**98.9 %**	99.5 %	**97.0 %**	**96.3 %**
Processed data fusion	93.8 %	95.8 %	92.6 %	92.1 %

from this set would be impossible to classify via NIR spectroscopy, which is sometimes used in plastic sorting companies. The complementarity of LIBS and Raman spectroscopy methods was exploited by the data fusion of these two techniques with several stages of data preprocessing. Afterward, chemometric methods were used to determine the accuracy of the classification of all data sets – individual LIBS, Raman, and fused LIBS-Raman data sets with several stages of data preprocessing. The classification accuracy differed between the applied linear (LDA) and nonlinear (RF) classification methods. While the LDA took a long time to be calculated (up to 10 min), it needed fewer data to be trained to provide good results on small data sets (Raman only data set), but it was not robust enough to deal with the dimensionally reduced and scaled fused data set, where the classification accuracy dropped under 50 %. On the other hand, RF as the nonlinear classification method needed more data to be trained but was much faster (maximum 20 s for calculation) and reached perfect or almost perfect classification accuracy for the raw data fusion and peak selected data fusion data sets. Additionally, it was also possible to obtain very good results from the processed data fusion data set by the RF algorithm. In conclusion, data fusion of LIBS and Raman spectroscopy clearly benefit from the strong suits of both methods simultaneously and reach better results than the aforementioned methods individually. This was proven on a challenging data set of plastics, which contained pure, colored, and also mixture polymer samples. On this data set, test classification accuracy reached more than 95 % using both linear and nonlinear classification methods, when LIBS and Raman data fusion was used.

CRediT authorship contribution statement

Daniel Holub: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft. **Dávid Jenő Palásti:** Conceptualization, Investigation, Methodology, Writing – review & editing. **Krisztian Fintor:** Investigation, Resources. **Pavel Pořízka:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing. **Gábor Galbács:** Writing – review & editing, Validation, Supervision, Software, Resources, Project administration, Methodology, Conceptualization, Funding

acquisition. **Jozef Kaiser:** Supervision, Resources, Funding acquisition.

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.

Data availability

Data will be made available on request.

Acknowledgment

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.polymertesting.2024.108414.

References

[1] Sameh Samir Ali, Tamer Elsamahy, Eleni Koutra, Michael Kornaros, Mostafa El-Sheekh, Esraa A. Abdelkarim, Daochen Zhu A Jianzhong Sun, Degradation of conventional plastic wastes in the environment: a review on current status of knowledge and future perspectives of disposal [online]. 771 [cit. 2022-04-26], Sci. Total Environ. (2021), https://doi.org/10.1016/j.scitotenv.2020.144719. ISSN 00489697.

[2] A.J. Jamieson, L.S.R. Brooks, W.D.K. Reid, S.B. Piertney, B.E. Narayanaswamy, T. D. Linley, Microplastics and synthetic particles ingested by deep-sea amphipods in six of the deepest marine ecosystems on Earth, R. Soc. Open Sci. 6 (2) (2019), https://doi.org/10.1098/rsos.180667.

[3] Philipp Schwabl, Sebastian Köppel, Philipp Königshofer, Theresa Bucsics, Michael Trauner, Thomas Reiberger, Bettina Liebmann, Detection of various microplastics in human stool, Ann. Intern. Med. 171 (7) (2019), https://doi.org/10.7326/M19-0618.

[4] Jean-Paul Lange, Managing Plastic Waste—Sorting, Recycling, Disposal, and Product Redesign, vol. 9, 2021, pp. 15722–15738, https://doi.org/10.1021/acssuschemeng.1c05013, 47.

[5] G. Galbács, Laser-Induced breakdown spectroscopy, Laser-Induced Breakdown Spectroscopy in Biological, Forensic and Materials Sciences (2022) 3–23, https://doi.org/10.1007/978-3-031-14502-5_1.

[6] J. Laserna, J.M. Vadillo, P. Purohit, Laser-Induced breakdown spectroscopy (LIBS): fast, effective, and agile leading edge analytical Technology, Appl. Spectrosc. 72 (1_suppl) (2018) 35–50, https://doi.org/10.1177/0003702818791926.

[7] R. Noll, Laser-Induced Breakdown Spectroscopy: Fundamentals and Applications, vol. 1, Springer-Verlag Berlin Heidelberg, Berlin, 2012. ISBN 978-3-642-20667-2.

[8] Frances Rivera-Hernández, Dawn Y. Sumner, Nicolas Mangold, Kathryn M. Stack, Olivier Forni, Newsom Horton, Amy Williams, et al., Using ChemCam LIBS data to constrain grain size in rocks on Mars: proof of concept and application to rocks at Yellowknife Bay and Pahrump Hills, Gale crater, Icarus 321 (2019) 82–98, https://doi.org/10.1016/j.icarus.2018.10.023.

[9] Ayumu Matsumoto, Tetsuo Sakka, A review of underwater laser-induced breakdown spectroscopy of submerged solids, Anal. Sci. 37 (8) (2021) 1061–1072, https://doi.org/10.2116/analsci.20R007.

[10] Georg Rombach, Nils Bauerschlag, LIBS based sorting—a solution for automotive scrap, Light Met. (2019) 1351–1357, https://doi.org/10.1007/978-3-030-05864-7_167, 2019.

[11] Mikio Kuzuya, A sorting system for aluminum alloy scrap based on laser-induced breakdown spectroscopy, Sci. J. Anal. Chem. 7 (3) (2019), https://doi.org/10.11648/j.sjac.20190703.11.

[12] Reinhard Noll, Cord Fricke-Begemann, Frederik Schreckenberg, Laser-induced breakdown spectroscopy as enabling key methodology for inverse production of end-of-life electronics, Spectrochim. Acta B Atom Spectrosc. 181 (2021), https://doi.org/10.1016/j.sab.2021.106213.

[13] Romppanen, Sari, Ilkka Pölönen, Heikki Häkkänen, and Saara Kaski. "Optimization of spodumene identification by statistical approach for laser-induced breakdown spectroscopy data of lithium pegmatite ores". Appl. Spectrosc. Rev.: 1-21. https://doi.org/10.1080/05704928.2021.1963977.

[14] Jesús Anzano, Roberto-Jesús Lasheras, Beatriz Bonilla, Justiniano Casas, Classification of polymers by determining of C1:C2:CN:H:N:O ratios by laser-induced plasma spectroscopy (LIPS): C2, Polym. Test. 27 (6) (2008) 705–710, https://doi.org/10.1016/j.polymertesting.2008.05.012.

[15] R. Sattmann, I. Mönch, H. Krause, R. Noll, S. Couris, A. Hatziapostolou, A. Mavromanolakis, C. Fotakis, E. Larrauri, R. Miguel, Laser-Induced breakdown spectroscopy for polymer identification, Appl. Spectrosc. 52 (3) (1998) 456–461, https://doi.org/10.1366/0003702981943680.

[16] Jesús Anzano, María-Esther Casanova, María-Soledad Bermúdez, Roberto-Jesús Lasheras, Rapid characterization of plastics using laser-induced plasma spectroscopy (LIPS), Polym. Test. 25 (5) (2006) 623–627, https://doi.org/10.1016/j.polymertesting.2006.04.005.

[17] R. Viskup, B. Praher, T. Linsmeyer, H. Scherndl, J.D. Pedarnig, J. Heitz, Influence of pulse-to-pulse delay for 532nm double-pulse laser-induced breakdown spectroscopy of technical polymers, Spectrochim. Acta B Atom Spectrosc. 65 (11) (2010) 935–942, https://doi.org/10.1016/j.sab.2010.09.003.

[18] V.K. Unnikrishnan, K.S. Choudhari, Suresh D. Kulkarni, Rajesh Nayak, V.B. Kartha, C. Santhosh, Analytical predictive capabilities of laser induced breakdown spectroscopy (LIBS) with principal component analysis (PCA) for plastic classification, RSC Adv. 3 (48) (2013), https://doi.org/10.1039/c3ra44946g.

[19] Myriam Boueri, Vincent Motto-Ros, Wen-Qi Lei, Qain-Li Ma, Li-Juan Zheng, He-Ping Zeng, Yu Jin, Identification of polymer materials using laser-induced breakdown spectroscopy combined with artificial neural networks, Appl. Spectrosc. 65 (3) (2011) 307–314, https://doi.org/10.1366/10-06079a.

[20] Ivana Chamradová, Pavel Pořízka, Jozef Kaiser, Laser-Induced Breakdown Spectroscopy analysis of polymers in three different atmospheres, Polym. Test. 96 (2021), https://doi.org/10.1016/j.polymertesting.2021.107079.

[21] Zuzana Gajarska, Lukas Brunnbauer, Hans Lohninger, Andreas Limbeck, Identification of 20 polymer types by means of laser-induced breakdown spectroscopy (LIBS) and chemometrics, Anal. Bioanal. Chem. 413 (26) (2021) 6581–6594, https://doi.org/10.1007/s00216-021-03622-y.

[22] Ruchita S. Das, Y.K. Agrawal, Raman spectroscopy: recent advancements, techniques and applications, Vib. Spectrosc. 57 (2) (2011) 163–176, https://doi.org/10.1016/j.vibspec.2011.08.003.

[23] Ota Samek, Silvie Bernatová, Jan Ježek, Martin Šiler, Mojmír Šerý, Vladislav Krzyžánek, Kamila Hrubanová, Pavel Zemánek, Veronika Holá, Filip Růžička, Identification of individual biofilm-forming bacterial cells using Raman tweezers, J. Biomed. Opt. 20 (5) (2015), https://doi.org/10.1117/1.JBO.20.5.051038.

[24] Valerie Allen, John H. Kalivas, René G. Rodriguez, Post-consumer plastic identification using Raman spectroscopy, Appl. Spectrosc. 53 (6) (2016) 672–681, https://doi.org/10.1366/0003702991947324.

[25] Steven E.J. Bell, Gaëlle Charron, Emiliano Cortés, Janina Kneipp, Marc Lamy Chapelle, Judith Langer, Marek Procházka, Vi Tran, Sebastian Schlücker, Towards reliable and quantitative surface-enhanced Raman scattering (SERS): from key parameters to good analytical practice, Angew. Chem. Int. Ed. 59 (14) (2020) 5454–5462, https://doi.org/10.1002/anie.201908154.

[26] Tiril Aurora Lintvedt, Petter V. Andersen, Nils Kristian Afseth, Brian Marquardt, Lars Gidskehaug, Jens Petter Wold, Feasibility of in-line Raman spectroscopy for quality assessment in food industry: how fast can we go? Appl. Spectrosc. 76 (5) (2022) 559–568, https://doi.org/10.1177/00037028211056931.

[27] H.G.M. Edwards, A.F. Johnson, I.R. Lewis, Applications of Raman spectroscopy to the study of polymers and polymerization processes, J. Raman Spectrosc. 24 (8) (1993) 475–483, https://doi.org/10.1002/jrs.1250240803.

[28] Sari Romppanen, Heikki Häkkänen, Jere Kekkonen, Jan Nissinen, Ilkka Nissinen, Juha Kostamovaara, Saara Kaski, Time-gated Raman and laser-induced breakdown spectroscopy in mapping of eudialyte and catapleiite, J. Raman Spectrosc. (May. 2019), https://doi.org/10.1002/jrs.5622.

[29] Hequn Wang, Jianhua Zhao, Anthony M.D. Lee, Lui Harvey, Haishan Zeng, Improving skin Raman spectral quality by fluorescence photobleaching, Photodiagnosis Photodyn. Ther. 9 (4) (2012) 299–302, https://doi.org/10.1016/j.pdpdt.2012.02.001.

[30] Daniel Holub, Pavel Pořízka, Martin Kizovský, David Prochazka, Ota Samek, Josef Kaiser, The potential of combining laser-induced breakdown spectroscopy and Raman spectroscopy data for the analysis of wood samples, Spectrochim. Acta B Atom Spectrosc. 195 (2022), https://doi.org/10.1016/j.sab.2022.106487.

[31] Javier Moros, Mostafa ElFaham Mohamed, J. Javier Laserna, Dual-spectroscopy platform for the surveillance of Mars mineralogy using a decisions fusion architecture on simultaneous LIBS-Raman data, Anal. Chem. 90 (3) (2018) 2079–2087, https://doi.org/10.1021/acs.analchem.7b04124.

[32] U.K. Adarsh, E. Bhoje Gowd, Aseefhali Bankapur, V.B. Kartha, Santhosh Chidangil, V.K. Unnikrishnan, Development of an inter-confirmatory plastic characterization system using spectroscopic techniques for waste management, Waste Manag. 150 (2022) 339–351, https://doi.org/10.1016/j.wasman.2022.07.025.

[33] Shameem, K.M. Muhammed, Khoobaram S. Choudhari, Aseefhali Bankapur, Suresh D. Kulkarni, V.K. Unnikrishnan, Sajan D. George, V.B. Kartha, C. Santhosh, A hybrid LIBS–Raman system combined with chemometrics: an efficient tool for plastic identification and sorting, Anal. Bioanal. Chem. 409 (13) (2017) 3299–3308, https://doi.org/10.1007/s00216-017-0268-z.

[34] P. Pořízka, L. Brunnbauer, M. Porkert, U. Rozman, G. Marolt, D. Holub, M. Kizovský, M. Benešová, O. Samek, A. Limbeck, J. Kaiser, G. Kalčíková, Laser-based techniques: novel tools for the identification and characterization of aged microplastics with developed biofilm, Chemosphere 313 (2023), https://doi.org/10.1016/j.chemosphere.2022.137373.

[35] Marek Hoehse, Andrea Paul, Igor Gornushkin, Panne Ulrich, Multivariate classification of pigments and inks using combined Raman spectroscopy and LIBS,

Anal. Bioanal. Chem. 402 (4) (2012) 1443–1450, https://doi.org/10.1007/s00216-011-5287-6.

[36] D. Prochazka, M. Mazura, O. Samek, K. Rebrošová, P. Pořízka, J. Klus, P. Prochazková, J. Novotný, K. Novotný, J. Kaiser, Combination of laser-induced breakdown spectroscopy and Raman spectroscopy for multivariate classification of bacteria, Spectrochim. Acta B Atom Spectrosc. 139 (2018) 6–12, https://doi.org/10.1016/j.sab.2017.11.004.

[37] Kristin Rammelkamp, Susanne Schröder, Kubitza Simon, David S. Vogt, Sven Frohmann, Peder B. Hansen, Ute Böttger, Franziska Hanke, Heinz-Wilhelm Hübers, Low-level LIBS and Raman data fusion in the context of in situ Mars exploration, J. Raman Spectrosc. 51 (9) (2020) 1682–1701, https://doi.org/10.1002/jrs.5615.

[38] Erin Gibbons, Richard Léveillé, Berlo Kim, Data fusion of laser-induced breakdown and Raman spectroscopies: enhancing clay mineral identification, Spectrochim. Acta B Atom Spectrosc. 170 (2020), https://doi.org/10.1016/j.sab.2020.105905.

[39] Lihui Ren, Ye Tian, Xiaoying Yang, Qi Wang, Leshan Wang, Xin Geng, Kaiqiang Wang, Zengfeng Du, Ying Li, Hong Lin, Rapid identification of fish species

by laser-induced breakdown spectroscopy and Raman spectroscopy coupled with machine learning methods, Food Chem. 400 (2023), https://doi.org/10.1016/j.foodchem.2022.134043.

[40] X. Xu, C. Du, F. Ma, Y. Shen, K. Wu, D. Liang, J. Zhou, Detection of soil organic matter from laser-induced breakdown spectroscopy (LIBS) and mid-infrared spectroscopy (FTIR-ATR) coupled with multivariate techniques, Geoderma 355 (2019), https://doi.org/10.1016/j.geoderma.2019.113905.

[41] K. Varmuza, P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, 2016, https://doi.org/10.1201/9781420059496.

[42] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, https://doi.org/10.1023/A:1010933404324.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al., Scikit-learn: machine learning in Python, J. Mach. Learn. Res. (12) (2011) 2825–2830, 2011, https://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html.