

Article

Towards Automated Classification of Zooplankton Using Combination of Laser Spectral Techniques and Advanced Chemometrics

Nikolai I. Sushkov ¹, Gábor Galbács ² , Patrick Janovszky ², Nikolay V. Lobus ³  and Timur A. Labutin ^{1,*} ¹ Department of Chemistry, Lomonosov Moscow State University, 119234 Moscow, Russia² Department of Inorganic and Analytical Chemistry, Faculty of Science and Informatics, University of Szeged, 6720 Szeged, Hungary³ Timiryazev Institute of Plant Physiology, Russian Academy of Sciences, 127276 Moscow, Russia

* Correspondence: timurla@laser.chem.msu.ru; Tel.: +7-495-939-36-35

Abstract: Zooplankton identification has been the subject of many studies. They are mainly based on the analysis of photographs (computer vision). However, spectroscopic techniques can be a good alternative due to the valuable additional information that they provide. We tested the performance of several chemometric techniques (principal component analysis (PCA), non-negative matrix factorisation (NMF), and common dimensions and specific weights analysis (CCSWA of ComDim)) for the unsupervised classification of zooplankton species based on their spectra. The spectra were obtained using laser-induced breakdown spectroscopy (LIBS) and Raman spectroscopy. It was convenient to assess the discriminative power in terms of silhouette metrics (*Sil*). The LIBS data were substantially more useful for the task than the Raman spectra, although the best results were achieved for the combined LIBS + Raman dataset (best *Sil* = 0.67). Although NMF (*Sil* = 0.63) and ComDim (*Sil* = 0.39) gave interesting information in the loadings, PCA was generally enough for the discrimination based on the score graphs. The distinguishing between *Calanoida* and *Euphausiacea* crustaceans and *Limacina helicina* sea snails has proved possible, probably because of their different mineral compositions. Conversely, arrow worms (*Parasagitta elegans*) usually fell into the same class with *Calanoida* despite the differences in their Raman spectra.

Keywords: zooplankton; LIBS; Raman; chemometrics; classification; PCA; NMF; ComDim

Citation: Sushkov, N.I.; Galbács, G.; Janovszky, P.; Lobus, N.V.; Labutin, T.A. Towards Automated Classification of Zooplankton Using Combination of Laser Spectral Techniques and Advanced Chemometrics. *Sensors* **2022**, *22*, 8234. <https://doi.org/10.3390/s22218234>

Academic Editor: Anna Chiara De Luca

Received: 8 September 2022

Accepted: 25 October 2022

Published: 27 October 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Analytical chemistry, especially in its applied branches, is largely related to decision making. For example, analytical measurements in environmental studies are needed to decide whether a sampled material is of interest for further work due to the presence of a specific marker or due to a high content of a hazardous substance [1]. Subsequent remedial or handling actions depend on the quantitative result of the measurement converted into categorical variables. The most evident example is the unsupervised classification [2] of unknown samples based on their analytical signals.

One of the fields where classification is extensively used is the analysis of biological materials—from crops and food to bacteria and human tissues [3–5]. Classification can be based on different types of information. The most evident way is to use optical images with a proper mathematical treatment (computer vision). This idea has been fruitful, e.g., in zooplankton identification, either in a supervised [6,7] or an unsupervised fashion [8–11]. However, spectroscopy can be used for individual tissues when simple or even hyperspectral imaging is not an ultimate solution. Rapid, sample-sparing, and information-rich analytical methods, such as laser-induced breakdown spectroscopy (LIBS) [12–16], Raman [17,18], and near-infrared (NIR) spectroscopies [19,20], supported by numerous mathematical techniques of multivariate data analysis, have come into question

during the past two decades. Many popular algorithms, such as discriminant analysis based on projection on latent structures (PLS-DA), discriminant function analysis (DFA), the k-nearest neighbour method (kNN), soft independent modelling of class analogies (SIMCA), support vector machine (SVM), classification trees (CTs), random forest (RF), and artificial neural networks (ANNs) [2,21–30], are supervised ones and thus require a training set of samples. Of course, this is not always possible, especially when there is no idea about how the classes should be defined.

The most common unsupervised technique is the principal component analysis (PCA) [31]. There are many other algorithms of this kind, e.g., non-negative matrix factorisation (NMF) [32,33] or independent component analysis (ICA) [34]. The common feature of these techniques is that the user obtains a set of scores which can be used for classification. NMF does not require its latent vectors to be orthogonal, which is a benefit for mixtures sharing some signals, but the solution is not unique. In other words, decomposition results in an infinite set of solutions due to the so-called rotational ambiguity [35,36]. On the other hand, the solution is always non-negative, which is convenient when working with many spectroscopic techniques (such as LIBS and Raman spectroscopy) where signals cannot be negative; NMF thus gives only those solutions that can be physically valid.

The independent component analysis (ICA) technique was initially used for blind source separation [37] and now finds many successful applications in fluorescence, nuclear magnetic resonance (NMR), infrared (IR) and Raman spectroscopy, chromatography and electrochemistry [34,38,39]. In addition to the non-orthogonality of the components, which brings the decomposition closer to physical reality [40–42] (the overlapping of signals of different compounds means that the spectra are not completely different and thus cannot be orthogonal—this is important for complex mixtures of natural compounds), ICA assumes that the components are non-Gaussian [40]. Thus, ICA not only tries to minimize correlation between the components, it also maximizes the deviation from normal distribution to obtain “independent” components.

Chemometrics can be applied to the detection of adulterations in the food industry, the detection and discrimination of pathogenic microorganisms or chemical warfare agents, the discrimination between diseased and healthy organisms or tissues, quality control, etc. [5]. The advantages of rapid spectroscopic techniques such as LIBS or Raman spectroscopy are the possible automations, the speeding up and scaling of analytical procedures [43] that, in principle, can be further applied for the sorting of large batches of organisms (e.g., fish, krill, or sea fruits), especially those which accumulate certain elements [44–47]. Sometimes, the very fact of sample clustering may bring valuable information about processes and their influence on a living organism, such as in studying the reaction of plants to different types of stress [48].

There is also a trend towards combining multiple measurement techniques. In order to use signals of different origin, data fusion techniques such as consensus PCA (CPCA [49]) or common components and specific weights analysis (CCSWA, or ComDim [50,51]) should be used. These techniques allow the description of the overall available information in order to extract the common and distinct information coming from different sources [50,52]. ComDim was specially designed and successfully used to analyse signals of different natures (expert ratings, fluorescence, NIR, MIR, NMR, etc.)—A feature which is appreciated in studying organic compound systems, such as in food quality control [51,53–55].

In [56,57], ComDim was used in combination with linear discriminant analysis to discriminate samples of instant coffee by means of an electronic nose. A more complicated combination of chemometric algorithms, which included PCA, predictive ComDim, parallel factor analysis, multivariate curve resolution, and PLS-DA, was used for the discrimination of vinegars based on their fluorescence, IR, and NMR spectra [58]. Gibbons et al. [59] combined LIBS and Raman spectroscopies, supported by PCA and LDA, to enhance the identification of clay minerals. The fusion of LIBS and UV-Vis absorption spectroscopy data resulted in very high classification accuracies, up to 100%, for the geographical provenance of Greek olive oils (PCA, LDA, and SVM were used) [60].

Thus, we can conclude that the fusion of LIBS and molecular spectroscopy data can result in high classification accuracy for organic samples. Advanced data fusion algorithms such as ComDim can further improve the results; at the same time, this technique, to the best of our knowledge, has almost never been used for atomic spectra (except in [61,62]). In the present study, we assess the potential of LIBS and Raman spectroscopy combined with PCA, NMF, and ComDim for the rapid classification of marine zooplankton animals without visual inspection of whole organisms, with an option of a classification based on the analysis of individual tissues, rather than of whole organisms.

2. Materials and Methods

Samples. Twenty-nine dried and pelletised zooplankton samples were involved in the study, namely 14 calanoid copepod (*Calanus spp.*) and 11 *Euphausiacea* crustaceans (totalling 25 crustacean samples), 2 arrow worm samples (*Parasagitta elegans*), 1 sea snail (*Limacina helicina*), and 1 mixed herbivorous zooplankton sample. The animals were caught by the research ships of the Shirshov Institute of Oceanology of the RAS during 2014–2017 expeditions in the Arctic seas and in the Black Sea. For details of the sample description and chemical composition, see Tables S1 and S2. The animals were washed with deionized water, dried at 50° C for 12 h, and pelletized under a moderate pressure ($d = 8$ mm, 20–1000 bar, typically 30 bar). Due to the inhomogeneity of the samples, we distinguished dark, light, and medium-coloured spots on the pellet surfaces. For each of the colours, LIBS and Raman spectroscopy were carried out in at least 3 spots of that colour, i.e., a total of at least 18 spots were interrogated on each pellet. In the case of LIBS, typically 10 laser shots per spot were delivered. After averaging, the final number of spectra was 29 samples \times 3 colours \times 2 methods = 174.

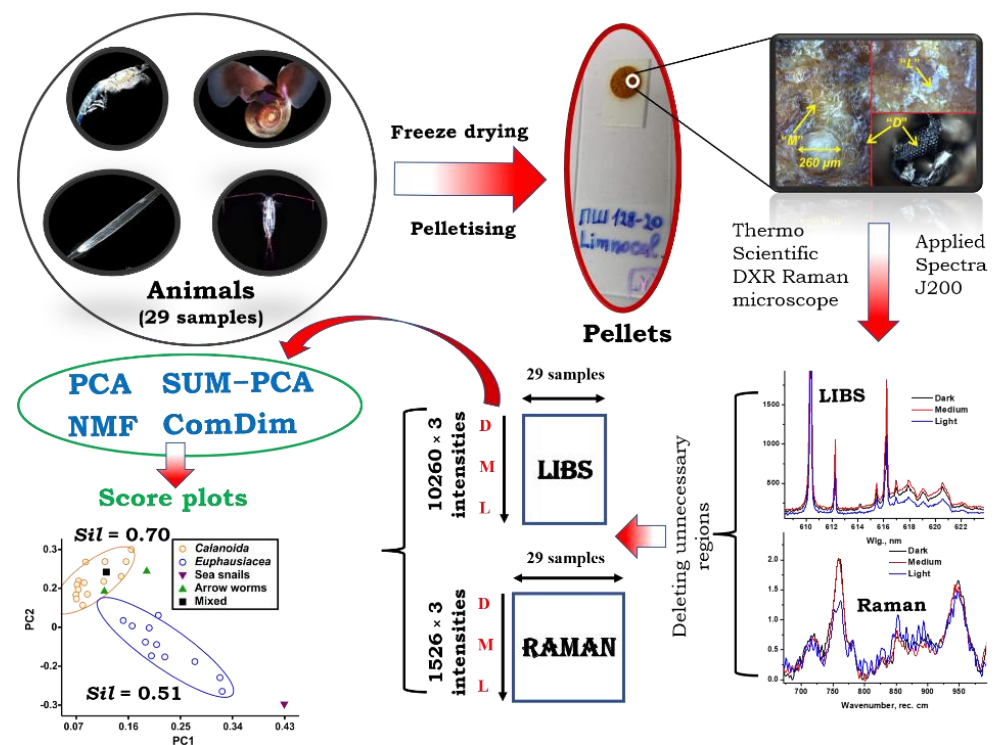
Equipment. Commercial instruments were employed. LIBS was performed using the J200 Tandem LA-LIBS instrument (Applied Spectra, USA; excitation laser: 266 nm Nd:YAG, 20 mJ/pulse, 10 Hz, spot diameter: 200 μ m; acquisition delay: 500 ns). As the instrument was equipped with a CCD detector, we obtained time-integrated spectra (integration time \sim 3 ms). The spectra covered the entire optical range (186–1049 nm, 12 275 points). Spectral resolution depended on the wavelength range and was typically around 3000 (pixel width \sim 70 pm with the spectral lines' full width at the half-maximum of 3 pixels). Automatic background correction was performed by the instrument software.

Raman scattering spectra were recorded using the DXR Raman Microscope (Thermo Scientific, USA; 780 nm laser, energy 1–14 mW, spot size 1–2 μ m, spectral resolution 4 cm^{-1}) in the range of 450–3150 cm^{-1} (2 801 points). We applied different energies depending on the fluorescent background intensity and the resistance of the samples to burning.

The aperture was selected individually for each sample to attain a sufficient signal-to-noise ratio and avoid CCD overflow at the same time. The following apertures were used: 25 and 50 μ m slits and 25 and 50 μ m pinholes. The acquisition time ranged from 30 s to 5 min. The spectra were instantaneously fluorescence corrected in the instrument software using a 6th order polynomial function.

The general flowchart of this study is shown in Scheme 1.

Data processing. In the emission spectra, the outliers were identified by the Grubbs' test applied sequentially to the peak intensities of the Li I 610.3 nm, C I 247.8 nm, Na I 568.3 nm, and Ca II 396.8 nm lines and the C_2 band head at 473.6 nm. After the removal of the outliers, the spectra were averaged for each sample. The Raman spectra were normalised to the total spectral intensity and averaged. The spectra from the spots of different colours were always chained without averaging. Thus, any LIBS or Raman spectrum consisted of three chained spectra.



Scheme 1. General flowchart of the study.

A number of spectral regions (mainly very intense resonance lines that are strongly self-absorbed in laser plasma [63] and non-characteristic signals) were discarded (see Table S3) to avoid their excessive influence on the decomposition results [62].

A number of matrix decomposition and data fusion algorithms were used for exploratory analysis. Data processing was performed in the GNU Octave, Microsoft Excel, Origin 8.5, Matlab 2020, and Wolfram Mathematica 8 software packages. When necessary, the data tables were divided by their Euclidean norms and again chained (LIBS data–Raman data). In this fashion, concatenated data were obtained (e.g., $29 \times 45,228$ matrices).

The GNU Octave *princomp* function was used for performing PCA. To determine the right number of principal components (PCs), we used conventional eigenvalue scree plots. This number of PCs was also considered in NMF decompositions (the *nmf_bpas* function). The PCA of concatenated LIBS + Raman datasets is called SUM-PCA. The ICA code made part of a ComDim algorithm (see below) which Emeritus Prof. D. N. Rutledge (AgroParisTech, France) kindly provided to us. In this algorithm, either ICA or PCA could be used for the successive computation of common components (for more details see [62]).

NMF and ICA may be considered as blind source separation (BSS) techniques, which give principal components with a physical sense [35]. The BSS problem is formulated in the following way: determine the N_s signals of unknown sources on the basis of the N_c combinations of these signals detected by N_c sensors. If we represent the problem by the relation $x = As$, where s and x denote the matrices of the source and the observed signals, respectively, and A is the so-called mixing matrix, then the goal is to obtain a “de-mixing” matrix A^{-1} which would convert x back to s without any a priori information on the sources. The rank of the A matrix is determined by the number of sources. The mixing and de-mixing phenomena can be described by different mathematical models; so, different BSS techniques are possible.

In ICA, the statistical independence and the non-Gaussianity of the source vectors are assumed; at the same time, ICA does not imply the orthogonality of the vectors [37,40,41]. Here, we used the ICA implementation based on the joint approximated diagonalization of eigen matrices (JADE) approach, which was a part of the above-mentioned ComDim code.

Conversely, NMF converts a data matrix X with m rows (descriptors) and n columns (samples) into matrices W ($m \times n$) and H ($n \times p$), where $p < \min(m, n)$. The p parameter determines the number of basis vectors and has to be chosen by the user. NMF works so that the resulting matrices will have no negative elements: $X = WH + E$, where E is the residual matrix. It is recommended to perform multiple runs of the algorithm with different initial approximations to avoid local minima of E [35,36], considering the decomposition with the highest explained variance as the final result of the factorization [64,65].

ComDim (also known as CCSWA) [51,53–55] is a method for the analysis of multi-block data. ComDim algorithms determine the latent common dimensions related to each data block and assess the contributions of each of the dimensions to the data blocks. The quantitative measure of this “relevance” is the so-called specific weight, or salience.

If there is a block of measurements X_k , with each of the N samples represented there by a line, it is possible to define a cross-product matrix $W_k = X_k \times X_k^T$. This matrix is then modelled as

$$W_k = Q\Lambda^{(k)}Q^T + E_k$$

Here, Q is an (N, N) orthogonal matrix, the columns of which are the common components. The specific weights are stored in $\Lambda^{(k)}$, which is a diagonal (N, N) matrix; $\Lambda^{(k)}$ can be different in the function of the block X_k , while Q is the same for all the data blocks X_k . The sum of the saliences in $\Lambda^{(k)}$ shows the extent to which a particular X_k contributes to Q . Saliences can be used to choose the appropriate number of common components, instead of the explained variance values. ComDim calculates the components one by one, and these steps may use PCA or ICA (see above), depending on the actual implementation.

3. Results and Discussion

The overview spectra of our samples can be seen in Figure 1. Our 29-sample dataset included four different animal taxa (see Table S1). The emission spectrum of the sea snail *Limacina helicina* is the richest in lines because of the presence of a mineralised shell. It is dominated by Ca lines together with strong signals of Na, K, Mg, and molecules such as CaOH and CaCl (Figure 1a). For the reliable identification of atomic emission, we have used the algorithm described elsewhere and earlier [63] (available at <https://libs.chem.msu.ru/>). The molecular bands were attributed with the help of the Pearse and Gaydon tables [66]. The tissues of arrow worms (*Parasagitta elegans*) are apparently the least mineralised; so, C_2 and CN molecular emission becomes important in their case. The spectra of *Calanoida* are peculiar for the presence of intense Li lines at 610.4 and 670.8 nm.

Conversely, the Raman spectra of *Limacina* (Figure 1b) are feeble and not very informative. *Parasagitta* present distinct amino acid features such as those at 755 cm^{-1} (probably corresponding to tryptophan) and 945 cm^{-1} (valine), accompanied by sharp peaks at 486, 503, and 3013 cm^{-1} (probably NH_2 groups). In *Euphausiacea*, relatively strong signals at 1609 cm^{-1} appear (phenyl rings) [67]. *Calanoida* show strong bands of carotenoid pigments at 1004, 1157, and 1518 cm^{-1} . Nearly all the animals show strong bands of saturated and unsaturated fatty acids at 1446 and 1656 cm^{-1} , respectively [67,68]. These spectral patterns, along with more subtle peculiarities, provide the classification basis. The signals at 1268 and 1302 cm^{-1} are also related to fatty acids.

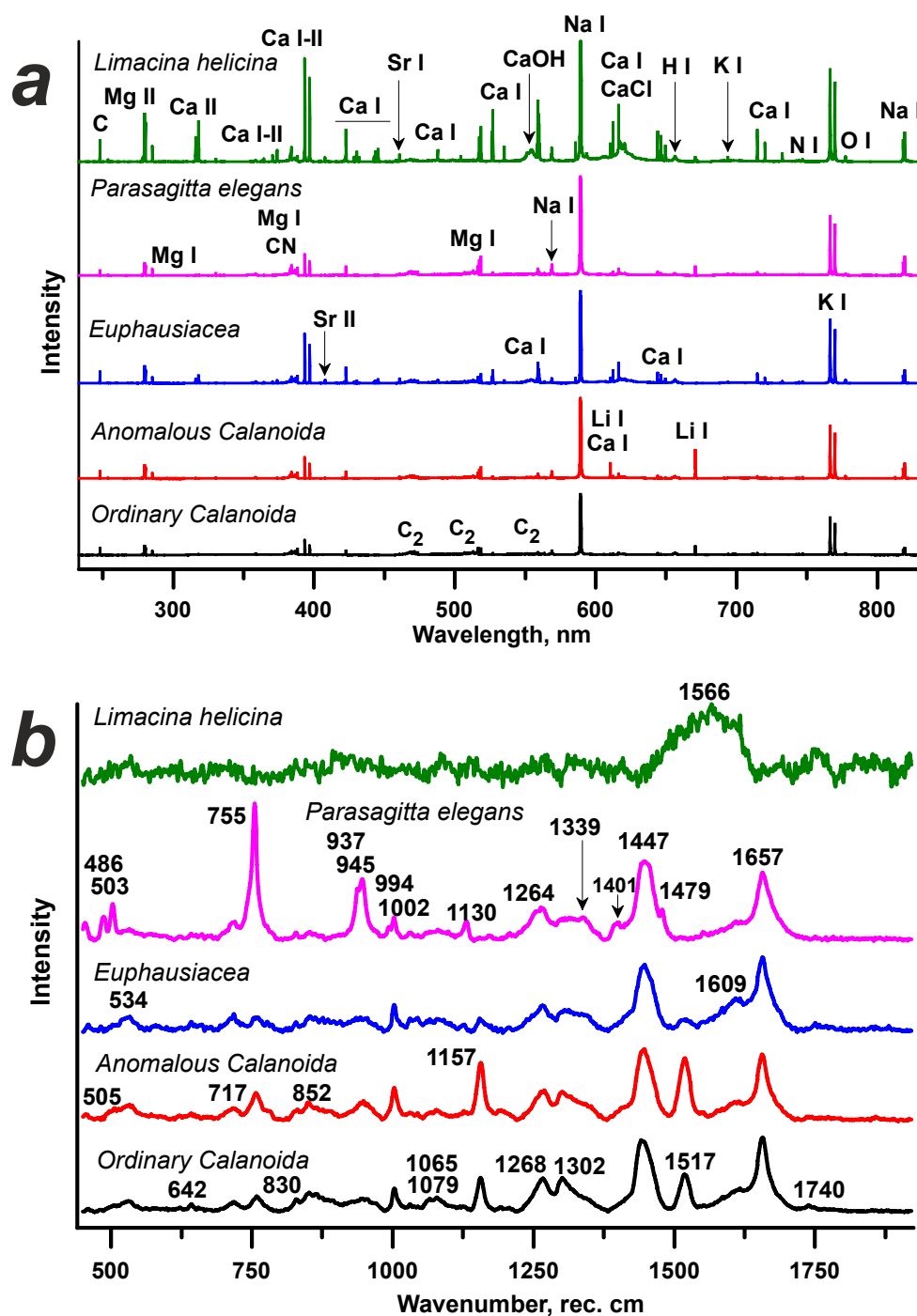


Figure 1. Overview (a) LIBS and (b) Raman spectra of the studied animals (all from medium-coloured spots). The spectra are averaged over the respective taxa.

The performance of chemometric techniques towards the classification of animals was tested using the whole dataset. In addition to visual inspection, we quantitatively assessed the classification performance of the models using the silhouette metric (*Sil*) [2]:

$$Sil = \sum_k \frac{1}{|C_k|} \sum_{i \in C_k} \frac{b_i - a_i}{\max(a_i, b_i)}$$

where a_i is the mean distance of a given point i from the other points within a cluster C_k , and b_i is the mean distance from i to all points outside that cluster. The silhouette values vary from -1 (the worst case, no clusters) through 0 (overlapping clusters) to $+1$ (distinct compact

clusters). In the considerations below, *Sil* will be multiplied by 100 and expressed in % for convenience (Table 1). The *Sil* values were computed for pairs of PCs, which are referred to as “planes” (e.g., a graph in the PC2 vs. PC1 coordinates will be called a “2-1 plane” for brevity). The last column of Table 1 shows multidimensional silhouettes computed using all the relevant subspace. As the relatively subtle discrimination between *Calanoida* and *Euphausiacea* crustaceans turned out to be impossible using the Raman spectra, we computed the Raman silhouettes for the consolidated *Crustaceans* cluster as opposed to the arrow worms and sea snails. Even at that, the silhouette values remained relatively low. Conversely, only *Calanoida* and *Euphausiacea* (a total of 25 samples) clusters were considered for the LIBS and fused spectral data.

Table 1. Silhouettes (in %) for all the score graphs involved in the present study.

No.	Method, Dataset	Plane	<i>Sil</i> , %	<i>Sil</i> (all dim.), %
Crustaceans–Arrow worms–Sea snails discrimination				
1	PCA—Raman only, full (5/97)	3-1	10	−10
2	PCA—Raman only, short (5/89)	5-1	65	39
3	NMF—Raman only, full (5/84)	4-2	11.5	−6
4	NMF—Raman only, short (5/75)	4-1	47	34
<i>Calanoida</i> – <i>Euphausiacea</i> discrimination				
5	PCA—LIBS only, full (5/99)	5-1	34	17
6	PCA—LIBS only, short (6/97)	2-1	62	22
7	NMF—LIBS only, full (5/88)	5-2	56	6
8	NMF—LIBS only, short (6/83)	6-4	60	37
9	SUM-PCA—LIBS + Raman (5/91)	2-1	67	21
10	NMF—LIBS + Raman (5/72)	5-3	63	28
11	ComDimPCA—LIBS + Raman (3/99.06)	2-1	39	30
12	ComDimICA—LIBS + Raman (3/96.89)	3-2	36	25

“Full” refers to intact spectra; “short” means that some signals were masked as described in the experiment section. The “*Sil*, %” column shows the maximum silhouette values among all the available 2D score graphs. Planes in which maximum values are obtained are indicated in the second column (e.g., 5-1 means PC5 vs. PC1 plane). Notations such as (5/99) mean that there are 5 PCs explaining 99% of variance. The “*Sil* (all dim.), %” column shows the multi-dimensional silhouette values computed using the whole available subspace.

First, note how different the silhouettes are for the raw and shortened spectra (i.e., with and without resonance lines—see the experiment section): 10 vs. 65% and 34 vs. 62% for the Raman and LIBS data, respectively (Figure 2a–d). For the LIBS, the “best plane” with the largest silhouette also shifts back from 5-1 to 2-1 due to the avoidance of the excessive influence of the resonance lines on the explained variance (in the 2-1 projection, the classification is driven by the C₂ and Li emission). Moreover, the confidence ellipses, which overlap in Figure 2a, become distinct in Figure 2b. Thus, in what follows we will imply the shortened datasets only.

As for the LIBS data, the PCA loadings are crowded with signals, most of which are correlated; it is convenient to choose a limited number of representative signals and show them on spider diagrams such as those in Figure 3. Here, the integrals of selected peaks of C, C₂, CN, H, P, Li, Na, K, Mg, Ca I, Ca II, Sr II, and Cu are shown (see details in Table S4); each one is normalized to the respective greatest absolute magnitudes among all the principal components.

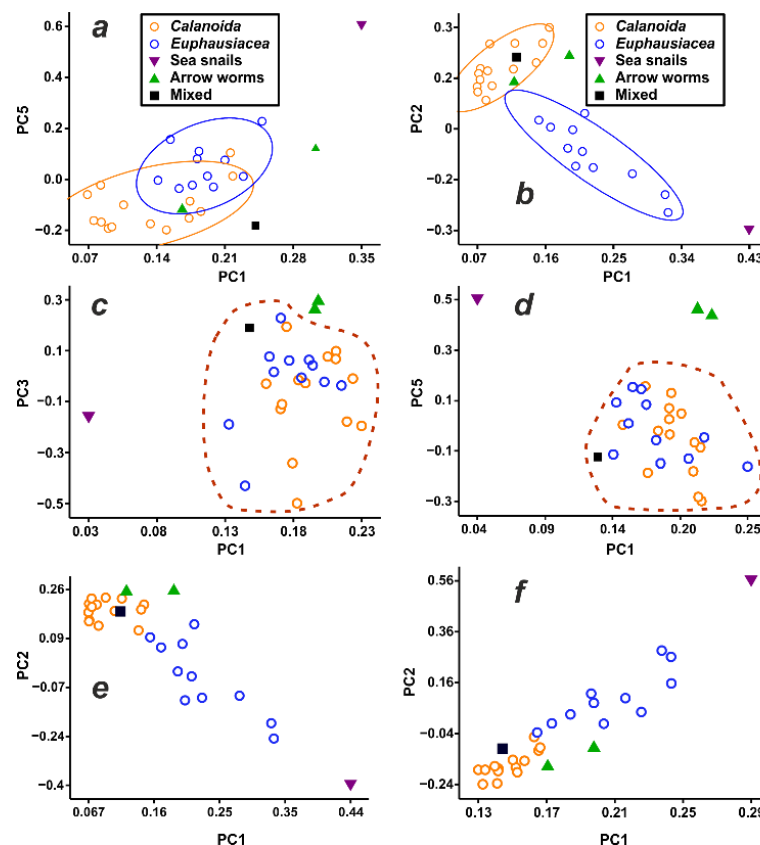


Figure 2. Score graphs for the PCA decompositions of LIBS (a,b) and Raman (c,d) spectra before (a,c) and after (b,d) masking the strongest signals. Dashed lines enclose the crustacean samples; (e) LIBS spectra without Li signals; (f) SUM-PCA for LIBS + Raman data without Li signals. In (a,b), confidence ellipses ($p = 0.95$) for the respective clusters are shown.

Surprisingly, the discrimination power of PCA with the Raman data is much worse than with the LIBS data. The cause of this may be the relatively low variability (and limited number) of the Raman signals. The PCA for the Raman data resulted in the separation of carotenoid and amino acid (tryptophan and valine) signals between different components (not shown). The best score graph is for the 5-1 plane (Figure 2d), which only involves the amino acid-related PC No. 5.

Emission spectra contain many sharp, distinct, and highly variable spectral features; this results in a better discriminative power. The PCA of the LIBS data yields two loadings (nos. 2 and 4 in Figure 3) with a prominent Li signal (with almost no other signals), which may suggest two different forms of Li in the tissues or two different accumulation patterns. PC1 apparently represents a basic composition which is common to all the samples. PC3 and PC5 highlight the importance of Ca I and Ca II emission in some samples. PC6 shows a correlation between magnesium and sodium. The loadings for the dark, medium-coloured, and light spots are obviously correlated to a high extent. One might think that the discrimination in Figure 2b is driven by the anomalously high Li content, which is characteristic of *Calanoida* [44,45]. It is, however, easy to see that the *Parasagitta* worms, which do not contain any lithium, fall close to *Calanoida* on the score graphs. Moreover, the masking of Li signals does not significantly alter the results of the decompositions. It is, rather, the calcium and strontium signals which drive the classification. Ca and P outbreaks in the PC3 can indicate the presence of calcium phosphate in the medium-coloured material.

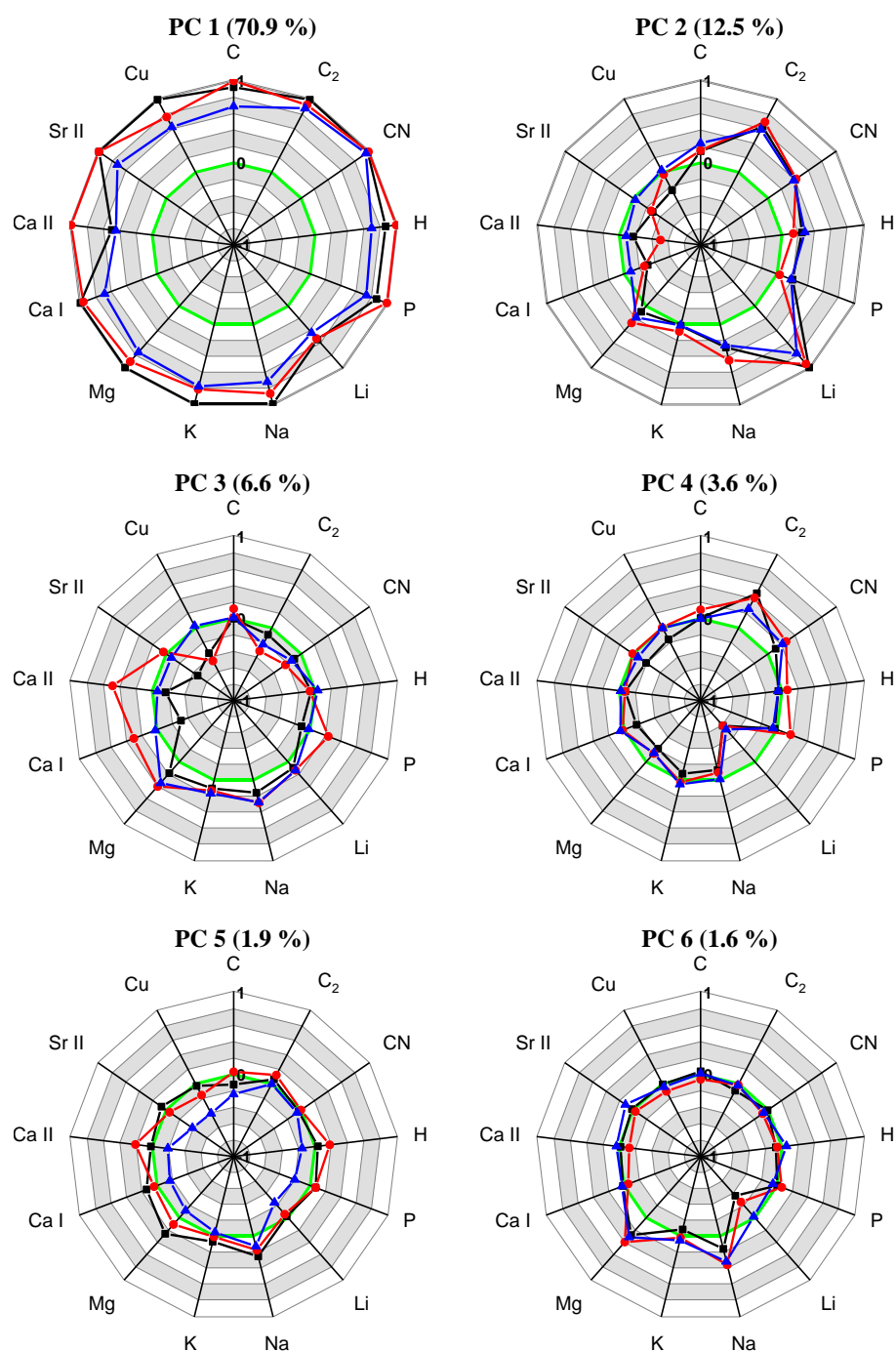


Figure 3. Selected features of PCA loadings for LIBS. Black, red, and blue lines correspond to dark, medium-coloured, and light spots, respectively. The width of the rings is 0.2. The green circle is at the zero level.

To get a deeper insight into the nature of the signals, we used NMF since it is likely to find physically meaningful components (e.g., signal sets related to a chemical compound or a group of compounds) [35,65]. We found that NMF loadings of the LIBS data do have different shapes for light, medium-coloured, and dark spots (Figure 4). The cause of this may be that the features cannot be easily smoothed by the factorization algorithm because of the non-negativity constraint.

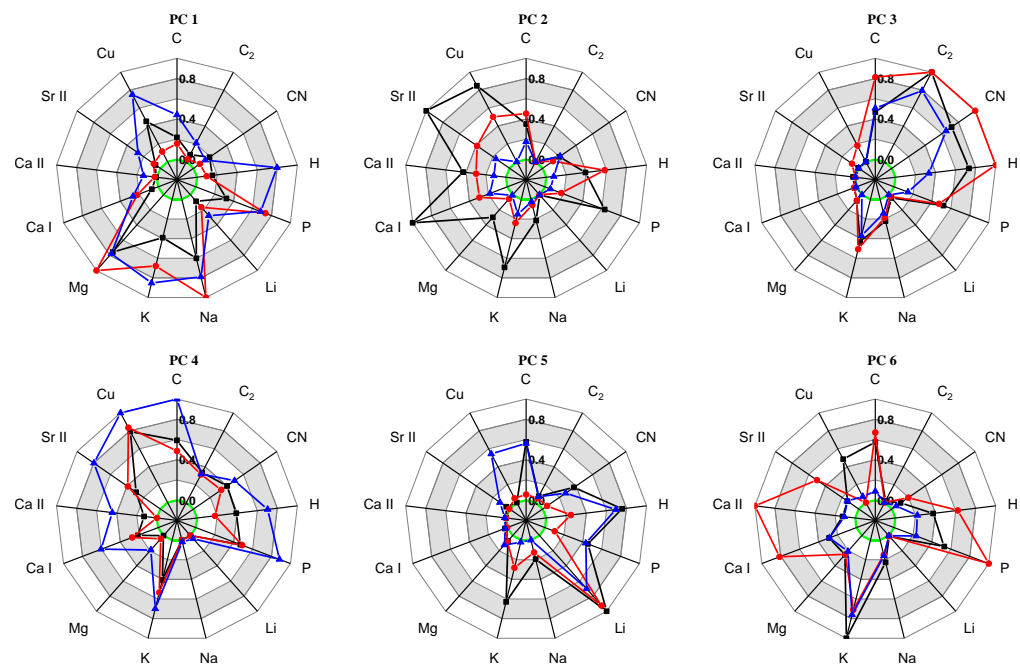


Figure 4. NMF decomposition of LIB spectra (without resonance lines). Shown are integrals of selected peaks on graphs of loadings, normalized to respective greatest absolute magnitudes among all the principal components. For wavelength positions of the peaks, please see the main text and supplementary tables. Black, red, and blue lines correspond to dark, medium-coloured, and light spots, respectively. The width of grey and white rings is 0.2.

The resonance line masking was quite efficient in improving the NMF-related silhouettes (see Table 1 and Figure 4). The Li signal is present together with K in one of the components and with Na, Mg, and K in another. It is noteworthy that the loadings for the dark, medium-coloured, and light spots are not so strongly correlated as in the PCA. Potassium occupies a special place, being present in all components and correlated with several groups of elements. As in the PCA, there is co-presence of Mg and Na in one of the components (PC1). The relation of Ca and P in PC6 is reminiscent of the phosphate minerals which can be present in the exoskeletons of crustaceans (this corroborates the above-discussed PCA results). As NMF gives non-orthogonal components, the respective score graphs are constructed in scalene rather than in rectangular coordinates. The inter-axis angle, α , is given by the relation: $\cos \alpha = (u, v) / \|u\| \|v\|$, where u and v are any two of the latent vectors obtained as a result of a decomposition. By applying standard trigonometry, it is possible to plot NMF scores (the H matrix) in these scalene coordinates.

The graph in Figure 5a shows a relatively good discrimination between *Calanoida* and *Euphausiacea*. It is again noteworthy that Li plays no role in this since the relevant latent vectors (nos. 6 and 4, cf. Figure 4) do not contain its signals. The difference between these vectors lies in the intensities of the Ca, Sr, Na, Cu and C₂ signals. For the *Euphausiacea* samples, high scores in both directions are observed, in contrast to *Calanoida* and *Parasagitta*. This just means that the latter are less mineralised. One of the *Calanoida* samples (No. 29) falls into the wrong cluster due to its anomalously high Sr content, which drives the point far to the right along the PC4 axis. This error, however, disappears when Raman data are also considered (NMF: see Figure 5b).

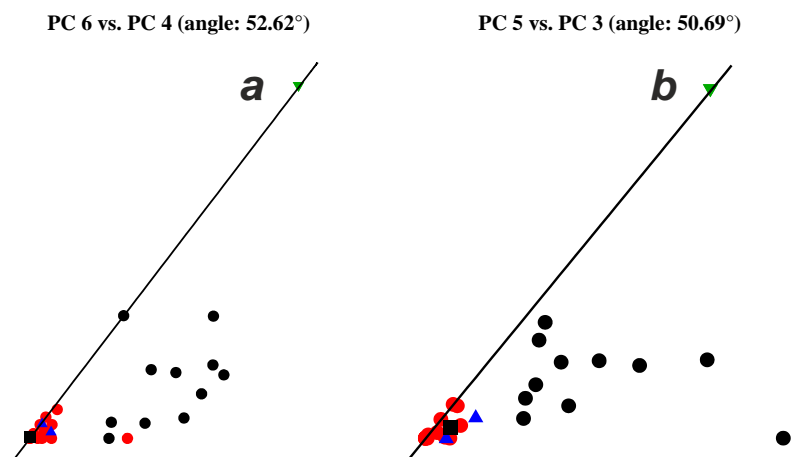


Figure 5. Score graphs of NMF decompositions: (a) LIBS only (cf. Figure 4); (b) fused Raman + LIBS dataset. Shown are projections with best silhouette metrics.

As with the PCA, the NMF factorization of the Raman spectra (Figure 6) did not provide a good sample clustering (the classification seems to be driven mainly by carotenoid bands). Carotenoid compounds and amino acids appear in two different components, though they are not free from the fatty acid bands, which themselves appear in a separate component. At the same time, the intensity of the carotenoid bands differs greatly as a function of spot colour, which is probably because these compounds are responsible for the tissue colour. PC4 looks quite like the *Limacina helicina* spectrum (Figure 1b). The explained variance is lower in the NMF than in the PCA for same datasets (see Table 1).

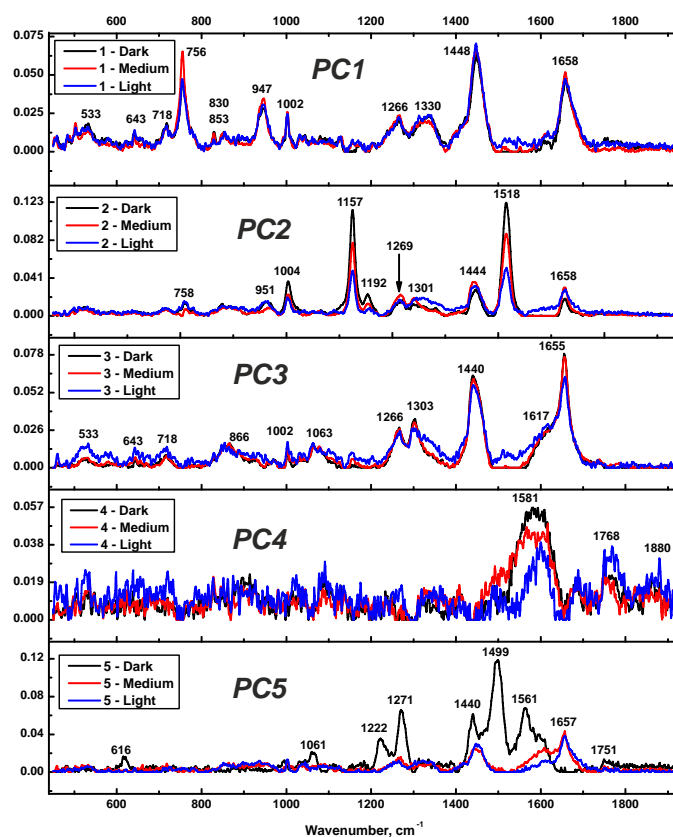


Figure 6. NMF decomposition of Raman spectra. Loadings obtained from spectra spanning the range of 449–1921 cm^{-1} . Black, red, and blue lines correspond to dark, medium-coloured, and light spots, respectively.

ComDim was tested in two variants: PCA-based and ICA-based. For both of them, the explained variance (EV) was very high. Despite the fact that the obtained loadings were easy to interpret and thus instrumental for finding correlations between the chemical components, the respective score graphs did not yield well-defined clusters, and the corresponding silhouette values were relatively modest. The best score graph (for ComDim-PCA) is shown in Figure 7 as an example.

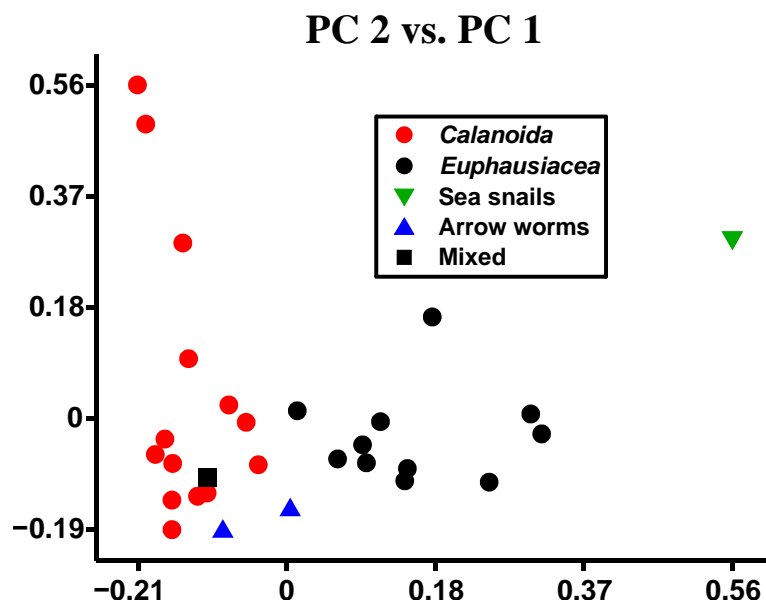


Figure 7. The best score graph for the ComDim-PCA decomposition.

Table 1 shows that the LIBS data were always modelled with higher explained variance (EV) than the Raman data. Although NMF yielded a much lower EV, its performance in discrimination was only slightly inferior to that of PCA. The SUM-PCA for the LIBS + Raman data was the best for the discrimination. Additionally, there was apparently no clustering of the *Calanoida* samples according to their geographical provenance (the Arctic Ocean or the Black Sea). This can be considered both as a strength and a limitation of the used protocol. On the one hand, it is not possible to distinguish the animals coming from the different regions, but on the other hand, this very fact makes the order-level discrimination more robust.

Critical Analysis and Discussion

In addition to visual inspection and computer vision [69], hyperspectral imaging [70] was also implemented for zooplankton species classification [71], but it is necessary to have an intact organism for reliable results. Conversely, LIBS and Raman spectroscopy can be adopted for microanalytical measurements, because they can deal both with individual millimetre-sized planktonic organisms and their parts (damaged bodies, etc.) since the achievable spatial resolution of LIBS and Raman is down to 5–10 μm [72,73]. In all cases, however, we should initially evaluate the classification performance of both techniques towards particular samples—this is why we have focused on pelleted samples and a relatively large sampling area in the current study.

Although the clusterisation quality obtained by us is better than in earlier studies [74,75], it would have been remiss to not investigate the capabilities of the chemometric techniques (PCA, NMF, and ComDim) in finding the latent characteristic features in the spectra of animals. Similar works do exist [5,76], but they usually either do not go further than conducting PCA or, on the contrary, use complicated algorithms requiring big training datasets (such as neural networks). We believe we have emphasized the fitness of NMF and ComDim for the classification and discrimination purposes and have also shown that PCA is still competitive, despite the existence of more sophisticated techniques.

4. Conclusions

We have tested the performance of several chemometric techniques (PCA, NMF, ComDim-PCA, and ComDim-ICA) for the unsupervised classification of zooplankton species based on their LIBS and Raman spectra. Given that these algorithms and the silhouette metric are rarely applied in analytical studies, we have emphasized their fitness for classification and discrimination purposes; this fitness is based on their capabilities in finding latent characteristic features in the spectra of animals. The LIBS data were substantially more useful for this task than the Raman spectra, although the best results were achieved for the combined LIBS + Raman dataset. Although NMF and ComDim gave interesting information in the loadings, the common SUM-PCA is the best solution for the discrimination based on the score graphs. It was not difficult to distinguish between the *Calanoida* and *Euphausiacea* crustaceans and the *Limacina helicina* sea snails because of their different mineral compositions. Conversely, the arrow worms (*Parasagitta elegans*) usually fell into the same class with *Calanoida* despite the differences in their Raman spectra. This issue could probably be solved in the future by using weighted mathematical algorithms. Apparently, the geographical provenance of the samples did not influence the results, which suggests that the found features belong to all the animals of the respective species, not only to a specific local population.

The results of the present study may be used for the automated LIBS + Raman-based sorting of large batches of organisms (e.g., fish, krill, or sea fruits). However, the most important result is that it has been possible to successfully perform a classification of animals from the presented set of samples because their chemical compositions turned out to be really different from each other. Further experiments with more samples and more species can provide a more detailed vision of the interrelations in the chemical compositions of sea creatures.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/s22218234/s1>, Table S1: collection conditions of zooplankton samples, Table S2: the results of elemental analysis of zooplankton by ICP-AES and ICP-MS, Table S3: list of spectral ranges eliminated from LIBS data to improve the performance of chemometric algorithms, Table S4: list of emission signals used to construct spider diagrams.

Author Contributions: Methodology, T.A.L. and N.I.S.; software, N.I.S. and P.J.; investigation, N.I.S., P.J. and N.V.L.; resources, G.G.; data curation, N.I.S. and P.J.; writing—original draft preparation, N.I.S.; writing—review and editing, T.A.L. and G.G.; conceptualization, T.A.L. All authors have read and agreed to the published version of the manuscript.

Funding: The reported study was funded by RFBR and BRFFR, project nos. 20-53-04036 and B21PM-085, respectively. Financial support of the research from the National Research, Development and Innovation Office of Hungary through project No. K_129063 is kindly acknowledged.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the Supplementary Materials.

Acknowledgments: The collaboration has become possible thanks to the Stipendium Hungaricum Programme of the Tempus Public Foundation (scholarship No. 175781 for N.S.). The authors are grateful to K. Fintor, Á. Béltéki, A. Kéri, D. Palásti (University of Szeged), R. Rajkó (University of Pécs, Hungary), and I. Krylov (Moscow State University) for their valuable assistance. The MATLAB code for performing ComDim, as well as valuable instructions thereon, were kindly provided by D.N. Rutledge (AgroParisTech, France).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Keith, L.H.; Crummett, W.; Deegan, J.; Libby, R.A.; Taylor, J.K.; Wentler, G. Principles of Environmental Analysis. *Anal. Chem.* **1983**, *55*, 2210–2218. [[CrossRef](#)]
2. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data: An Introduction to Cluster Analysis*; Wiley-Interscience: Hoboken, NJ, USA, 2005, ISBN 978-0-471-73578-7.
3. Gaudiuso, R.; Melikechi, N.; Abdel-Salam, Z.A.; Harith, M.A.; Palleschi, V.; Motto-Ros, V.; Busser, B. Laser-Induced Breakdown Spectroscopy for Human and Animal Health: A Review. *Spectrochim. Acta Part B At. Spectrosc.* **2019**, *152*, 123–148. [[CrossRef](#)]
4. Rehse, S.J. A Review of the Use of Laser-Induced Breakdown Spectroscopy for Bacterial Classification, Quantification, and Identification. *Spectrochim. Acta Part B At. Spectrosc.* **2019**, *154*, 50–69. [[CrossRef](#)]
5. Sushkov, N.I. Qualitative Classification of Biological Materials. In *Laser-Induced Breakdown Spectroscopy in Biological, Forensic and Materials Sciences*; Galbács, G., Ed.; Springer International Publishing: Cham, Switzerland, 2022; ISBN 9783031145018.
6. González, P.; Álvarez, E.; Díez, J.; López-Urrutia, Á.; del Coz, J.J. Validation Methods for Plankton Image Classification Systems. *Limnol. Oceanogr. Methods* **2017**, *15*, 221–237. [[CrossRef](#)]
7. Lumini, A.; Nanni, L. Ocean Ecosystems Plankton Classification BT—Recent Advances in Computer Vision: Theories and Applications. In *Recent Advances in Computer Vision. Theory and Applications*; Hassaballah, M., Hosny, K.M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 261–280. ISBN 978-3-030-03000-1.
8. Pastore, V.P.; Zimmerman, T.G.; Biswas, S.K.; Bianco, S. Annotation-Free Learning of Plankton for Classification and Anomaly Detection. *Sci. Rep.* **2020**, *10*, 12142. [[CrossRef](#)]
9. Kuzminykh, D.; Polykovskiy, D.; Zhebrak, A. Extracting Invariant Features from Images Using an Equivariant Autoencoder. *Proc. Mach. Learn. Res.* **2018**, *95*, 438–453.
10. Wang, C.; Yu, Z.; Zheng, H.; Wang, N.; Zheng, B. CGAN-Plankton: Towards Large-Scale Imbalanced Class Generation and Fine-Grained Classification. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 855–859.
11. Salvesen, E.; Saad, A.; Stahl, A. Robust Deep Unsupervised Learning Framework to Discover Unseen Plankton Species. In Proceedings of the SPIE, Fourteenth International Conference on Machine Vision (ICMV 2021 Rome, Italy), Hangzhou, China, 4 March 2022; Volume 12084, p. 120840V.
12. Zorov, N.B.; Popov, A.M.; Zaytsev, S.M.; Labutin, T.A. Qualitative and Quantitative Analysis of Environmental Samples by Laser-Induced Breakdown Spectrometry. *Russ. Chem. Rev.* **2015**, *84*, 1021–1050. [[CrossRef](#)]
13. Wang, Q.; Xiangli, W.; Teng, G.; Cui, X.; Wei, K. A Brief Review of Laser-Induced Breakdown Spectroscopy for Human and Animal Soft Tissues: Pathological Diagnosis and Physiological Detection. *Appl. Spectrosc. Rev.* **2021**, *56*, 221–241. [[CrossRef](#)]
14. Hu, Z.; Zhang, D.; Wang, W.; Chen, F.; Xu, Y.; Nie, J.; Chu, Y.; Guo, L. A Review of Calibration-Free Laser-Induced Breakdown Spectroscopy. *TrAC Trends Anal. Chem.* **2022**, *152*, 116618. [[CrossRef](#)]
15. Gunawan, R.; Imran, A.; Ahmed, I.; Liu, Y.; Chu, Y.; Guo, L.; Yang, M.; Lau, C. FROZEN! Intracellular Multi-Electrolyte Analysis Measures Millimolar Lithium in Mammalian Cells. *Analyst* **2021**, *146*, 5186–5197. [[CrossRef](#)]
16. Luarte, D.; Myakalwar, A.K.; Velásquez, M.; Álvarez, J.; Sandoval, C.; Fuentes, R.; Yañez, J.; Sbarbaro, D. Combining Prior Knowledge with Input Selection Algorithms for Quantitative Analysis Using Neural Networks in Laser Induced Breakdown Spectroscopy. *Anal. Methods* **2021**, *13*, 1181–1190. [[CrossRef](#)] [[PubMed](#)]
17. Teng, G.; Wang, Q.; Cui, X.; Wei, K.; Xiangli, W.; Chen, G. Feature-Level Fusion of Laser-Induced Breakdown Spectroscopy and Raman Spectroscopy for Improving Support Vector Machine in Clinical Bacteria Identification. *J. Raman Spectrosc.* **2021**, *52*, 805–814. [[CrossRef](#)]
18. Khan, M.N.; Wang, Q.; Idrees, B.S.; Teng, G.; Xiangli, W.; Cui, X.; Wei, K. Evaluation of Human Melanoma and Normal Formalin Paraffin-Fixed Samples Using Raman and LIBS Fused Data. *Lasers Med. Sci.* **2022**, *37*, 2489–2499. [[CrossRef](#)] [[PubMed](#)]
19. Eum, C.; Jang, D.; Kim, J.; Choi, S.; Cha, K.; Chung, H. Improving the Accuracy of Spectroscopic Identification of Geographical Origins of Agricultural Samples through Cooperative Combination of Near-Infrared and Laser-Induced Breakdown Spectroscopy. *Spectrochim. Acta Part B At. Spectrosc.* **2018**, *149*, 281–287. [[CrossRef](#)]
20. Mishra, P.; Roger, J.M.; Marini, F.; Biancolillo, A.; Rutledge, D.N. Parallel Pre-Processing through Orthogonalization (PORTO) and Its Application to near-Infrared Spectroscopy. *Chemom. Intell. Lab. Syst.* **2021**, *212*, 104190. [[CrossRef](#)]
21. Vandeginste, B.G.M.; Massart, D.L.; Buydens, L.M.C.; De Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J. Chapter 33—Supervised Pattern Recognition. In *Handbook of Chemometrics and Qualimetrics: Part B*; Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J.B.T.-D.H., Eds.; Elsevier: Amsterdam, The Netherlands, 1998; Volume 20, pp. 207–241; ISBN 0922-3487.
22. Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. *Classification and Regression Trees*, 1st ed.; Routledge: New York, NY, USA, 1984.
23. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
24. Vandeginste, B.G.M.; Massart, D.L.; Buydens, L.M.C.; De Jong, S.; Lewi, P.J.; Smeyers-Verbeke, J. Chapter 35—Relations between Measurement Tables. In *Handbook of Chemometrics and Qualimetrics: Part B*; Vandeginste, B.G.M., Massart, D.L., Buydens, L.M.C., De Jong, S., Lewi, P.J., Smeyers-Verbeke, J.B.T.-D.H., Eds.; Elsevier: Amsterdam, The Netherlands, 1998; Volume 20, pp. 307–347; ISBN 0922-3487.

25. Putnam, R.A.; Mohaidat, Q.I.; Daabous, A.; Rehse, S.J. A Comparison of Multivariate Analysis Techniques and Variable Selection Strategies in a Laser-Induced Breakdown Spectroscopy Bacterial Classification. *Spectrochim. Acta Part B At. Spectrosc.* **2013**, *87*, 161–167. [[CrossRef](#)]
26. Rehse, S.J.; Diedrich, J.; Palchadhuri, S. Identification and Discrimination of *Pseudomonas Aeruginosa* Bacteria Grown in Blood and Bile by Laser-Induced Breakdown Spectroscopy. *Spectrochim. Acta Part B At. Spectrosc.* **2007**, *62*, 1169–1176. [[CrossRef](#)]
27. Malenfant, D.J.; Gillies, D.J.; Rehse, S.J. Bacterial Suspensions Deposited on Microbiological Filter Material for Rapid Laser-Induced Breakdown Spectroscopy Identification. *Appl. Spectrosc.* **2016**, *70*, 485–493. [[CrossRef](#)]
28. Cisewski, J.; Snyder, E.; Hannig, J.; Oudejans, L. Support Vector Machine Classification of Suspect Powders Using Laser-Induced Breakdown Spectroscopy (LIBS) Spectral Data. *J. Chemom.* **2012**, *26*, 143–149. [[CrossRef](#)]
29. Burges, C.J.C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]
30. Metzinger, A.; Rajkó, R.; Galbács, G. Discrimination of Paper and Print Types Based on Their Laser Induced Breakdown Spectra. *Spectrochim. Acta Part B At. Spectrosc.* **2014**, *94*, 48–57. [[CrossRef](#)]
31. Pořízka, P.; Klus, J.; Képeš, E.; Prochazka, D.; Hahn, D.W.; Kaiser, J. On the Utilization of Principal Component Analysis in Laser-Induced Breakdown Spectroscopy Data Analysis, a Review. *Spectrochim. Acta Part B At. Spectrosc.* **2018**, *148*, 65–82. [[CrossRef](#)]
32. Choi, S.; Cichocki, A.; Park, H.-M.; Lee, S.-Y. Blind Source Separation and Independent Component Analysis: A Review. *Neural Inf. Process. Lett. Rev.* **2005**, *6*, 1–57.
33. Lee, D.D.; Seung, H.S. Learning the Parts of Objects by Non-Negative Matrix Factorization. *Nature* **1999**, *401*, 788–791. [[CrossRef](#)]
34. Monakhova, Y.B.; Rutledge, D.N. Independent Components Analysis (ICA) at the “Cocktail-Party” in Analytical Chemistry. *Talanta* **2020**, *208*, 120451. [[CrossRef](#)]
35. Cichocki, A.; Zdunek, R.; Phan, A.H.; Amari, S. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*; Wiley Publishing: Chichester, UK, 2009; ISBN 0470746661.
36. Tavakkoli, E.; Rajkó, R.; Abdollahi, H. Duality Based Direct Resolution of Unique Profiles Using Zero Concentration Region Information. *Talanta* **2018**, *184*, 557–564. [[CrossRef](#)]
37. Héroult, J.; Jutten, C.; Ans, B. Détection de Grandeurs Primitives Dans Un Message Composite Par Une Architecture de Calcul Neuromimétique En Apprentissage Non Supervisé. In *Dixième Colloque sur le Traitement du Signal et ses Applications, Actes du Xème Colloque*; GRETSI: Nice, France, 1985; pp. 1017–1022.
38. Kassouf, A.; Rakwe, M.E.; Chebib, H.; Ducruet, V.; Rutledge, D.N.; Maalouly, J.; Matn, J. El Independent Components Analysis Coupled with 3D-Front-Face Fluorescence Spectroscopy to Study the Interaction between Plastic Food Packaging and Olive Oil. *Anal. Chim. Acta* **2014**, *839*, 14–25. [[CrossRef](#)]
39. Meksiarun, P.; Ishigaki, M.; Huck-Pezzei, V.A.C.; Huck, C.W.; Wongravee, K.; Sato, H.; Ozaki, Y. Comparison of Multivariate Analysis Methods for Extracting the Paraffin Component from the Paraffin-Embedded Cancer Tissue Spectra for Raman Imaging. *Sci. Rep.* **2017**, *7*, 44890. [[CrossRef](#)]
40. Hyvärinen, A. Independent Component Analysis: Recent Advances. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2013**, *371*, 20110534. [[CrossRef](#)]
41. Hyvärinen, A. Survey on Independent Component Analysis. *Neural Comput. Surv.* **1999**, *2*, 94–128.
42. Khlaifi, A. Estimation Des Sources de Pollution Par Modélisation Inverse. Thèse présentée pour l’obtention du Doctorat de l’Université Paris XII, Créteil, France, 2007; p. 366.
43. Werheit, P.; Fricke-Begemann, C.; Gesing, M.; Noll, R. Fast Single Piece Identification with a 3D Scanning LIBS for Aluminium Cast and Wrought Alloys Recycling. *J. Anal. At. Spectrom.* **2011**, *26*, 2166–2174. [[CrossRef](#)]
44. Lobus, N.V.; Drits, A.V.; Flint, M. V Accumulation of Chemical Elements in the Dominant Species of Copepods in the Ob Estuary and the Adjacent Shelf of the Kara Sea. *Oceanology* **2018**, *58*, 405–415. [[CrossRef](#)]
45. Lobus, N.V.; Arashkevich, E.G.; Flerova, E.A. Major, Trace, and Rare-Earth Elements in the Zooplankton of the Laptev Sea in Relation to Community Composition. *Environ. Sci. Pollut. Res.* **2019**, *26*, 23044–23060. [[CrossRef](#)]
46. Lobus, N. V Elemental Composition of Zooplankton in the Kara Sea and the Bays on the Eastern Side of Novaya Zemlya. *Oceanology* **2016**, *56*, 809–818. [[CrossRef](#)]
47. Freese, D.; Niehoff, B.; Søreide, J.E.; Sartoris, F.J. Seasonal Patterns in Extracellular Ion Concentrations and PH of the Arctic Copepod *Calanus Glacialis*. *Limnol. Oceanogr.* **2015**, *60*, 2121–2129. [[CrossRef](#)]
48. Martin, M.Z.; Labbé, N.; André, N.; Harris, R.; Ebinger, M.; Wullschleger, S.D.; Vass, A.A. High Resolution Applications of Laser-Induced Breakdown Spectroscopy for Environmental and Forensic Applications. *Spectrochim. Acta Part B At. Spectrosc.* **2007**, *62*, 1426–1432. [[CrossRef](#)]
49. Westerhuis, J.A.; Kourti, T.; MacGregor, J.F. Analysis of Multiblock and Hierarchical PCA and PLS Models. *J. Chemom.* **1998**, *12*, 301–321. [[CrossRef](#)]
50. Mishra, P.; Roger, J.M.; Jouan-Rimbaud-Bouveresse, D.; Biancolillo, A.; Marini, F.; Nordon, A.; Rutledge, D.N. Recent Trends in Multi-Block Data Analysis in Chemometrics for Multi-Source Data Integration. *TrAC Trends Anal. Chem.* **2021**, *137*, 116206. [[CrossRef](#)]
51. Cariou, V.; Jouan-Rimbaud Bouveresse, D.; Qannari, E.M.; Rutledge, D.N. ComDim Methods for the Analysis of Multiblock Data in a Data Fusion Perspective. In *Data Handling in Science and Technology*; Elsevier Ltd.: Amsterdam, The Netherlands, 2019; Volume 31, pp. 179–204.

52. Mazerolles, G.; Devaux, M.F.; Dufour, E.; Qannari, E.M.; Courcoux, P. Chemometric Methods for the Coupling of Spectroscopic Techniques and for the Extraction of the Relevant Information Contained in the Spectral Data Tables. *Chemom. Intell. Lab. Syst.* **2002**, *63*, 57–68. [[CrossRef](#)]
53. Cariou, V.; Qannari, E.M.; Rutledge, D.N.; Vigneau, E. ComDim: From Multiblock Data Analysis to Path Modeling. *Food Qual. Prefer.* **2018**, *67*, 27–34. [[CrossRef](#)]
54. Qannari, E.M.; Wakeling, I.; MacFie, H.J.H. A Hierarchy of Models for Analysing Sensory Data. *Food Qual. Prefer.* **1995**, *6*, 309–314. [[CrossRef](#)]
55. Qannari, E.M.; Wakeling, I.; Courcoux, P.; MacFie, H.J.H. Defining the Underlying Sensory Dimensions. *Food Qual. Prefer.* **2000**, *11*, 151–154. [[CrossRef](#)]
56. Makimori, G.Y.F.; Bona, E. Commercial Instant Coffee Classification Using an Electronic Nose in Tandem with the ComDim-LDA Approach. *Food Anal. Methods* **2019**, *12*, 1067–1076. [[CrossRef](#)]
57. Vieira, T.F.; Makimori, G.Y.F.; dos Santos Scholz, M.B.; Zielinski, A.A.F.; Bona, E. Chemometric Approach Using ComDim and PLS-DA for Discrimination and Classification of Commercial Yerba Mate (*Ilex Paraguariensis* St. Hil.). *Food Anal. Methods* **2020**, *13*, 97–107. [[CrossRef](#)]
58. Ríos-Reina, R.; Callejón, R.M.; Savorani, F.; Amigo, J.M.; Cocchi, M. Data Fusion Approaches in Spectroscopic Characterization and Classification of PDO Wine Vinegars. *Talanta* **2019**, *198*, 560–572. [[CrossRef](#)]
59. Gibbons, E.; Léveillé, R.; Berlo, K. Data Fusion of Laser-Induced Breakdown and Raman Spectroscopies: Enhancing Clay Mineral Identification. *Spectrochim. Acta Part B At. Spectrosc.* **2020**, *170*, 105905. [[CrossRef](#)]
60. Gyftokostas, N.; Nanou, E.; Stefas, D.; Kokkinos, V.; Bouras, C.; Couris, S. Classification of Greek Olive Oils from Different Regions by Machine Learning-Aided Laser-Induced Breakdown Spectroscopy and Absorption Spectroscopy. *Molecules* **2021**, *26*, 1241. [[CrossRef](#)]
61. Breitwieser, M.; Vigneau, E.; Viricel, A.; Becquet, V.; Lacroix, C.; Erb, M.; Huet, V.; Churlaud, C.; Le Floch, S.; Guillot, B.; et al. What Is the Relationship between the Bioaccumulation of Chemical Contaminants in the Variegated Scallop *Mimachlamys varia* and Its Health Status? A Study Carried out on the French Atlantic Coast Using the Path ComDim Model. *Sci. Total Environ.* **2018**, *640*, 662–670. [[CrossRef](#)]
62. Sushkov, N.I.; Galbács, G.; Fintor, K.; Lobus, N.V.; Labutin, T.A. A Novel Approach for Discovering Correlations between Elemental and Molecular Composition Using Laser-Based Spectroscopic Techniques. *Analyst* **2022**, *147*, 3248–3257. [[CrossRef](#)]
63. Zaytsev, S.M.; Popov, A.M.; Labutin, T.A. Stationary Model of Laser-Induced Plasma: Critical Evaluation and Applications. *Spectrochim. Acta Part B At. Spectrosc.* **2019**, *158*, 105632. [[CrossRef](#)]
64. Brunet, J.-P.; Tamayo, P.; Golub, T.R.; Mesirov, J.P. Metagenes and Molecular Pattern Discovery Using Matrix Factorization. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4164–4169. [[CrossRef](#)] [[PubMed](#)]
65. Buciu, I. Non-Negative Matrix Factorization, a New Tool for Feature Extraction: Theory and Applications. *Int. J. Comput. Commun. Control* **2008**, *3*, 67–74.
66. Pearse, R.W.B.; Gaydon, A.G. *The Identification of Molecular Spectra*; Chapman & Hall: London, UK, 1963.
67. De Gelder, J.; De Gussem, K.; Vandenebeele, P.; Moens, L. Reference Database of Raman Spectra of Biological Molecules. *J. Raman Spectrosc.* **2007**, *38*, 1133–1147. [[CrossRef](#)]
68. Lin-Vien, D.; Colthup, N.B.; Fateley, W.G.; Grasselli, J.G. *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*, 1st ed.; Academic Press: Cambridge, MA, USA, 1991.
69. Dias, P.A.; Dunkel, T.; Fajado, D.A.S.; Gallegos, E. de L.; Denecke, M.; Wiedemann, P.; Schneider, F.K.; Suhr, H. Image Processing for Identification and Quantification of Filamentous Bacteria in in Situ Acquired Images. *Biomed. Eng. Online* **2016**, *15*, 64. [[CrossRef](#)] [[PubMed](#)]
70. Baltussen, E.J.M.; Kok, E.N.D.; Koning, S.G.B.d.; Sanders, J.; Aalbers, A.G.J.; Kok, N.F.M.; Beets, G.L.; Flohil, C.C.; Bruin, S.C.; Kuhlmann, K.F.D.; et al. Hyperspectral Imaging for Tissue Classification, a Way toward Smart Laparoscopic Colorectal Surgery. *J. Biomed. Opt.* **2019**, *24*, 16002. [[CrossRef](#)]
71. Nielsen, J.H.; Pedersen, C.; Kiørboe, T.; Nikolajsen, T.; Brydegaard, M.; Rodrigo, P.J. Investigation of autofluorescence in zooplankton for use in classification of larval salmon lice. *Appl. Opt.* **2019**, *26*, 7022–7027. [[CrossRef](#)]
72. Limbeck, A.; Brunnbauer, L.; Lohninger, H.; Pořízka, P.; Modlitbová, P.; Kaiser, J.; Janovszky, P.; Kéri, A.; Galbács, G. Methodology and Applications of Elemental Mapping by Laser Induced Breakdown Spectroscopy. *Anal. Chim. Acta* **2021**, *1147*, 72–98. [[CrossRef](#)]
73. ThermoFisherScientific Website. Available online: <https://www.thermofisher.com/order/catalog/product/IQLAADGABFFAHCMAPB> (accessed on 17 October 2022).
74. Zhao, Z.; Chen, L.; Liu, F.; Zhou, F.; Peng, J.; Sun, M. Fast Classification of Geographical Origins of Honey Based on Laser-Induced Breakdown Spectroscopy and Multivariate Analysis. *Sensors* **2020**, *20*, 1878. [[CrossRef](#)]
75. Yang, Y.; Hao, X.; Zhang, L.; Ren, L. Application of Scikit and Keras Libraries for the Classification of Iron Ore Data Acquired by Laser-Induced Breakdown Spectroscopy (LIBS). *Sensors* **2020**, *20*, 1393. [[CrossRef](#)]
76. Bilge, G.; Velioglu, H.M.; Sezer, B.; Eseller, K.E.; Boyaci, I.H. Identification of Meat Species by Using Laser-Induced Breakdown Spectroscopy. *Meat Sci.* **2016**, *119*, 118–122. [[CrossRef](#)]