Applied Physics A

Identification of wood specimens utilizing fs-LIBS and machine learning techniques

--Manuscript Draft--

Manuscript Number:	APYA-D-23-02434
Full Title:	Identification of wood specimens utilizing fs-LIBS and machine learning techniques
Article Type:	Regular papers
Corresponding Author:	Constantine Kosmidis University of Ioannina School of Sciences: Panepistemio Ioanninon Schole Thetikon Epistemon GREECE
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Ioannina School of Sciences: Panepistemio Ioanninon Schole Thetikon Epistemon
Corresponding Author's Secondary Institution:	
First Author:	Alexandros Sarafis
First Author Secondary Information:	
Order of Authors:	Alexandros Sarafis
	Theofanis Gerodimos
	Emmanouil Kechaoglou
	Dimitrios F. Anagnostopoulos
	Constantine Kosmidis
Order of Authors Secondary Information:	
Funding Information:	NSRF 2014–2020 — Prof Constantine Kosmidis
Abstract:	We report on the ability to identify wood specimens by utilizing 30 fs Laser Induced Breakdown Spectroscopy (LIBS) in conjunction with machine learning techniques. Ten different wood specimens have been studied. The spectral features were assigned to atomic / ionic and diatomic molecular transitions. The origin of the latter has been explored by investigating the dynamics of the created plume in ambient and argon atmosphere. Principal Component Analysis (PCA) was employed for dimensionality reduction based on the primary LIBS analysis. The principal components formation is grounded on the CN, Ca+, Ca, and Na, LIBS data. Furthermore, applying the weighted k nearest neighbor (kNN) algorithm led to an accurate identification of the investigated specimens, since the evaluation metrics of specificity value were found to be in the range of 0.96 - 1.00, while that of accuracy was within 0.93 - 1.00.

Identification of wood specimens utilizing fs-LIBS and machine learning techniques

A. Sarafis[1], T. Gerodimos[2], E. Kechaoglou[1], D. F. Anagnostopoulos[2], C. Kosmidis [1*]

[1]Department of Physics, University of Ioannina, Ioannina, 45110 Greece

[2]Department of Materials Science and Engineering, University of Ioannina, Ioannina, 45110 Greece

Abstract

We report on the ability to identify wood specimens by utilizing 30 fs Laser Induced Breakdown Spectroscopy (LIBS) in conjunction with machine learning techniques. Ten different wood specimens have been studied. The spectral features were assigned to atomic / ionic and diatomic molecular transitions. The origin of the latter has been explored by investigating the dynamics of the created plume in ambient and argon atmosphere. Principal Component Analysis (PCA) was employed for dimensionality reduction based on the primary LIBS analysis. The principal components formation is grounded on the CN, Ca^+, Ca, and Na, LIBS data. Furthermore, applying the weighted k nearest neighbor (kNN) algorithm led to an accurate identification of the investigated specimens, since the evaluation metrics of specificity value were found to be in the range of 0.96 - 1.00, while that of accuracy was within 0.93 - 1.00.

Keywords: LIBS, machine learning techniques, wood, sample identification, fs ablation.

*Author for correspondence: C. Kosmidis e-mail: kkosmid@uoi.gr

Introduction

The role of wood is intertwined with all aspects of human activity. Even from the prehistoric period, it has been established that its use as fuel was of vital importance for human activity and beyond that, it met the needs of manufacturing household utensils, tools, religious objects, houses, and bridges, to mention only a few of them.

Thus, there is a naturally occurring need for the characterization and criteria development for the appropriate, each time, wood selection for the construction of specific objects. The empirical evaluation criteria that were initially employed, were supplemented, and then advanced, following the newest scientific techniques and modern characterization methods that are based on the developments in Physics, Chemistry, and Materials Science (spectroscopy, spectrometry, strength tests, etc [1]–[8]).

One of the modern techniques that have been applied for wood characterization is Laser Induced Breakdown Spectroscopy (LIBS) [1], [9]–[13]. LIBS has been widely utilized in fields such as food chain quality control[14], materials characterization[15], [16], preservation of cultural heritage objects[17], [18], medical diagnostic methods [19], [20]and even in aerospace applications[21]. In most cases, the LIBS technique was performed using ns laser pulses. It is however known that with the use of ultrashort (fs) laser pulses the physical processes involved in the ablation process are differentiated and this is finally reflected both in the LIBS spectrum and in the irradiated material[22]–[24]. The differentiation is mainly related to the way that the laser pulse energy is delivered to the material. The few fs pulse duration is shorter than the time required for thermal dissipation of the absorbed energy, and since we are referring to organic materials, the result is an increased probability for multiphoton ionization i.e., climbing on the energy ladder is more probable compared to a ladder switching process that could be caused by thermal diffusion [25]. The released electrons are ejected from the surface while positively charged ions remain in the material. The strong repulsion between the ions leads to a Coulomb explosion and eventually to the formation of a plume. Thus, it is conceivable that the thermal affected area is dramatically reduced when fs laser pulses are employed, which is significant for many applications [26]. Furthermore, the plume formation is completed after the end of the fs laser pulse, and therefore the ejected material does not interact with the incident light beam, which cannot be avoided when the duration of the pulse is in the ns time scale[27]. The latter implies that the probability for observation of emission from molecular fragments is increased for fs LIBS [28].

The LIBS technique has already been applied to the characterization of wood and was evaluated as a very suitable one [1]–[4]. It is indicative that it has been a tool for forensic investigations, too [29]. Nevertheless, to the best of our knowledge, its implementation on wood studies so far has been applied by utilizing laser beams of 10 ns [30] and 470 ps [31] pulse duration. Thus, using short-lived pulses of the order of 30 fs presents a challenging issue.

In the present work, the fs-LIBS technique is applied to ten different wood samples. The aim is to achieve the most accurate identification possible. The procurement of the wood samples was done by a painter who specializes in hagiography following the Byzantine

tradition and style. Certainly, the characterization of wood samples is a difficult subject. However, it has been recently proven that using the XRF technique in conjunction with machine learning techniques produces safe results[6]. Machine learning techniques have been widely used to address similar problems. Brunnbauer et al. [32] contribute an insightful and comprehensive critical review that delves into contemporary trends surrounding the classification of samples through the utilization of LIBS. PCA is widely used in data analysis to transform complex, multi-dimensional datasets into a lower-dimensional space while preserving the most informative features [33]. By identifying principal components, which represent orthogonal directions capturing the maximum variance within the data, PCA facilitates the representation of the dataset in a lower-dimensional subspace without compromising its discriminatory power. This reduction simplifies the data visualization, analysis, and modeling processes. When applied to LIBS data, PCA identifies these spectral features that contribute the most to the variability among the samples.

Expanding on this, many supervised classification techniques, such as Support Vector Machine, k Nearest Neighbors, Random Forests and Artificial Neural Networks such as Multilayer Perceptron, have been proposed for the discrimination and identification of samples scanned by LIBS [14]. Thus, the utilization of machine learning techniques in the context of our work is expected to contribute to the effective identification of the wood specimens selected for study.

Materials and methods

Experimental set-up

The radiation source for the LIBS setup is a 30 fs, 5 mJ/pulse laser system (DUO Legend Elite) centered at 800 nm, which for these series of experiments was operating at 10 Hz. The laser beam was focused by an f=23cm convex lens. The laser intensity was continuously monitored, and its value was controlled by using a couple of stepwise neutral density filters. The back-reflection of these filters was utilized for triggering the set-up. A collimator/collector (Andor ME-OPT-0007) was used to collect light emitted only from the interaction region, which was connected via an optical fiber to a gated spectrograph (Andor iStar). The spectra recorded in the region of 230 - 630nm were observed by utilizing a 150 l/mm diffraction grating, and the spectral resolution was relatively low (~0.6nm). For the cases where higher resolution (0.04nm) was demanded (for instance the analysis of the emission from molecular fragments) another grating was used (1800 l/mm), and the spectral range was shorter.

A white light continuum component always exists in the LIBS experiments, lasting in the present case about 30ns, due to inverse bremsstrahlung and recombination processes taking place in the ejected plume. So, delayed gated operation of the spectrograph is imperative. The temporal width of the gate was decided to be 30ns, thus allowing us to study the emission dynamics of the plume. Each of the recorded spectra is the sum of those induced by a preset number of pulses, and this was controlled electronically by a homemade shutter. Moreover, an electronically controlled bespoke and a 2-axis stage was synchronized with the shutter, and the step precision in both x and y directions was 150μm.

For the experiments performed in Argon (Ar) atmosphere, a constant flow of Ar (6 lt/min) was presented over the interaction region.

Machine learning techniques

The k-NN [34] classifier, described as a non-parametric approach, emerges as a vital tool for sample classification and identification purposes. In the context of wood species classification through LIBS analysis, each wood spectrum, after PCA application, is considered a data point, representing a unique sample in the new feature space. The class label associated with each data point corresponds to the wood species from which the spectrum was obtained. This representation allows the algorithm to associate each wood spectrum with a specific wood species. To classify an unknown spectrum, the k-NN algorithm measures the Euclidean distance, as a metric of similarity, between the test spectrum and the training spectra that have been previously labeled. By comparing the test spectrum to its k nearest neighbors, where k is a user-defined parameter, the algorithm determines the prevalent class among these neighbors. The test spectrum is then assigned to that class. This study employed a variation of the conventional k-NN algorithm, known as weighted k-NN [35],[36]. This improved version includes a feature where nearby data points get different importance weights based on their proximity to the query point.

It is worth mentioning that a pivotal step in the above-mentioned process involves splitting the dataset into training and validation subsets for cross-validation [37]. The training set serves as the foundation for model development and hyperparameter tuning, while the validation set offers an independent set of samples to measure the classifier's performance, ensuring its robustness in identifying wood species through LIBS analysis. The validation set is typically smaller than the training set [32].

For visualizing the performance of the classifier, we utilize a confusion matrix [38]. From this, several metrics can be derived to assess the performance of the classification model, including true positives (TP) that represent the number of correctly classified positive instances (wood spectra belonging to a specific wood species), true negatives (TN) that correspond to the number of correctly classified negative instances (wood spectra not belonging to that specific wood species), false positives (FP) that are the instances that were incorrectly classified as positive (misclassified wood spectra), false negatives (FN) that refer to instances that were incorrectly classified as negative (wood spectra of the specific wood species that were mistakenly labeled as something else). Using these values, we can calculate the following evaluation metrics as shown in Table 1.

Table 1. Summary of the most common metrices for classifiers' evaluation

Metric	Formula	Explanation
Sensitivity (also known as true positive rate)	$\frac{TP}{TP+FN}$	Measures the proportion of actual positive instances correctly classified.
Specificity (also known as true negative rate)	$\frac{TN}{TN+FP}$	Measures the proportion of actual negative instances correctly classified.
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$	Measures the overall correctness of the classification model, considering both positive and negative instances.

As mentioned above, samples from 10 different wood species were investigated. These are: Acacia, Plum (Prunus Domestica), Berry (Morus), Wild plum (Prunus cerasifera), Rhododendron, Elm (Ulmus minor), Chestnut (Castanea), Walnut (Juglans), Beech (Fagus sylvatica) and Pine (Pinus Peuce).

Results and discussion

Spectra interpretation

In order to reveal the particular spectral characteristics of each wood specimen studied as well as the reliability of the applied identification techniques, it is important to record the LIBS spectra under the same experimental conditions. Thus, the pulse energy of the 30fs Ti:Sapphire laser beam (λ=800nm) was kept constant (0.99mJ/pulse) for all the recorded LIBS spectra, the laser operating frequency was 10Hz, while the gate width of the CCD camera was 120ns. Ten LIBS spectra were recorded from each irradiated spot of the samples. The spectra presented in Fig. 1 are the sum of the recordings from 100 different points.

The resolution of the spectra depicted in Fig. 1 is relatively low (0.6nm), thus, the assignment of the observed spectral features is based on a narrower wavelength range spectra of higher resolution (0.04nm) recorded utilizing a different diffraction grating. The NIST database for LIBS spectral analysis [39] was used to identify all the spectral peaks corresponding to atomic transitions. Spectral features that correspond to Carbon (C) atomic transition at ~248 nm (and its second-order diffraction at ~496nm) have been recorded, while spectral lines that could be attributed to atomic iron (Fe) transitions have not been observed, despite being previously reported [40]. In Table 2 the assignment of the observed transitions is presented in detail.

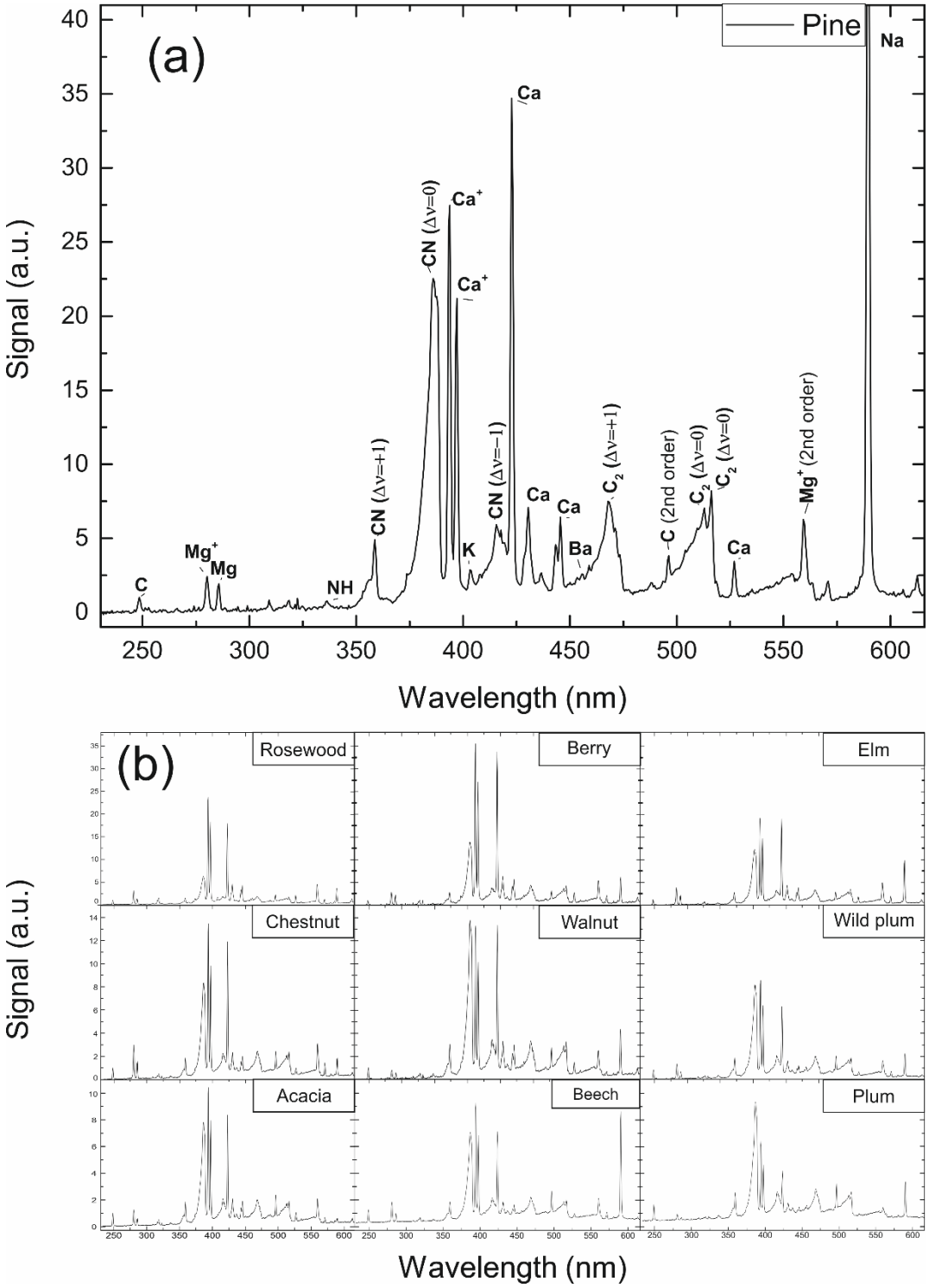


Figure 1. *a) the assigned transitions in LIBS spectra of pine samples, b) typical LIBS spectra obtained from the wood specimens studied.*

Table 2. Assignment of the recorded spectral features

Element/diatomic molecule	Wavelength (nm)	Transition
C	247.86	$^1P° \rightarrow {}^1S$ ($\Delta J=1$)
Mg^+	279.55 279.80	$^2P° \rightarrow {}^2S$ ($\Delta J=1$) $^2D \rightarrow {}^2P°$($\Delta J=0,1$)
Mg	285.21	$^1P° \rightarrow {}^1S$ ($\Delta J=1$)
Ca^+	393.37 396.85	$^2P° \rightarrow {}^2S$ ($\Delta J=1$) $^2P° \rightarrow {}^2S$ ($\Delta J=0$)
Ca	422.81 430.25 443.50, 443.57 445.48, 445.59, 445.66 526.42, 526.56	$^1P° \rightarrow {}^1S$ ($\Delta J=1$) $^3P \rightarrow {}^3P°$ ($\Delta J=0$) $^3D \rightarrow {}^3P°$ ($\Delta J=0,1$) $^3D \rightarrow {}^3P°$ ($\Delta J=0, \pm1$) $^3D \rightarrow {}^3P°$ ($\Delta J=0, -1$)
K	404.41	$^2P° \rightarrow {}^2S$ ($\Delta J=1$)
Sr^+	407.77 421.55	$^2P° \rightarrow {}^2S$ ($\Delta J=1$) $^2P° \rightarrow {}^2S$ ($\Delta J=1$)
Ba^+	455.40	$^2P° \rightarrow {}^2S$ ($\Delta J=1$)
Na	589.00, 589.59	$^2P° \rightarrow {}^2S$ ($\Delta J=1,0$)
NH	336.3	$A^3\pi_i \rightarrow X^3\Sigma^-$ ($\Delta v=0$)
CN	358.4-359.0 385.03-388.32 415.2-421.6	$B^2\Sigma^+ \rightarrow X^2\Sigma^+$ ($\Delta v=+1$) $B^2\Sigma^+ \rightarrow X^2\Sigma^+$ ($\Delta v=0$) $B^2\Sigma^+ \rightarrow X^2\Sigma^+$ ($\Delta v=-1$)
C_2	467.82-473.66 512.87, 516.52 544.61-563.48	$d^3\Pi_g \rightarrow a^3\Pi_u$ ($\Delta v=+1$) $d^3\Pi_g \rightarrow a^3\Pi_u$ ($\Delta v=0$) $d^3\Pi g \rightarrow a^3\Pi u$ ($\Delta v=-1$)

The spectral bands attributed to transitions of diatomic molecular species (C_2, CN, NH) are worth noting because their formation has been a subject of different interpretation scenarios [41]–[43]. It is characteristic that CN transitions have also been observed in the LIBS spectra of substances that do not contain nitrogen, so their formation is attributed to chemical reactions of the atmospheric nitrogen with C atoms and/or C_2 fragments ejected from the ablated sample [41]. Furthermore, a detailed analysis of these molecular bands can be exploited for studies of the isotope ratio $^{12}C/\ ^{13}C$ determination [44] and the temperature of the specific diatomic molecules, thus deriving information about as well as the temperature of the specific diatomic molecules, thus deriving information about both the material under study and the plasma produced.

In the present work, they were utilized to study the plasma temperature and its dynamic evolution. Recently, the non-trivial dependence of the fs-LIBS spectra characteristics on

the laser focus point with respect to the sample (Aluminum oxide, AlO) surface has been reported [45]. Hence, the laser beam focusing conditions on the wood specimens' surface were kept scrupulously constant. The temperature was determined using the Boltzmann Equilibrium Spectrum Program [46]. By simulating the corresponding spectral peaks (Fig.2) the temperature of the CN and C_2 molecules were estimated and found to be as high as ~8400K and ~6400K, respectively. Our findings align with previously reported measurements [47].

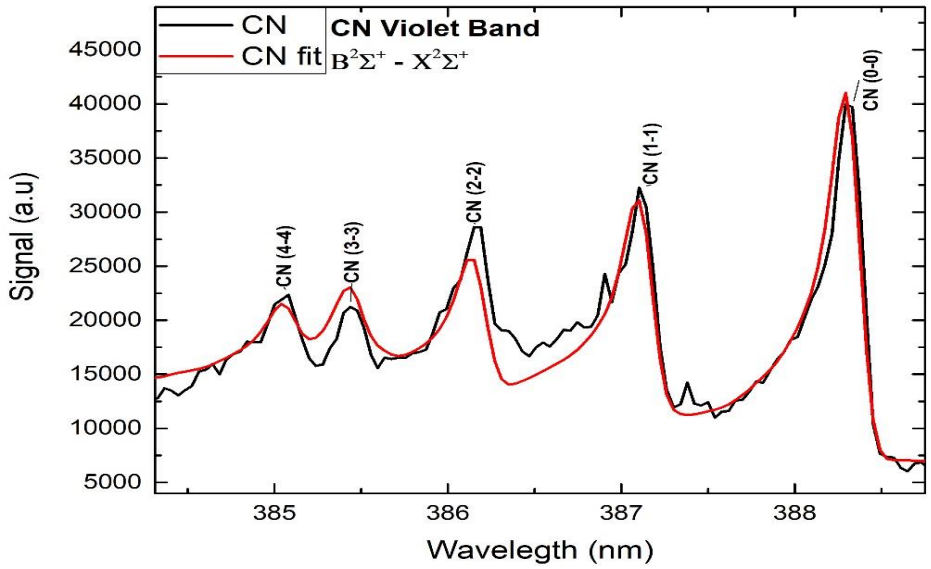


***Figure 2**: Fitting of the $B^2\Sigma^+ \rightarrow X^2\Sigma^+$ electronic decay in CN band for the case of pine. The accompanying vibrational transitions are also noted.*

For the investigation of the temperature decay the temporal width of the CCD gate was reduced to 30ns and the measurements was repeated by varying stepwise the time delay between the laser beam and the CCD gate.

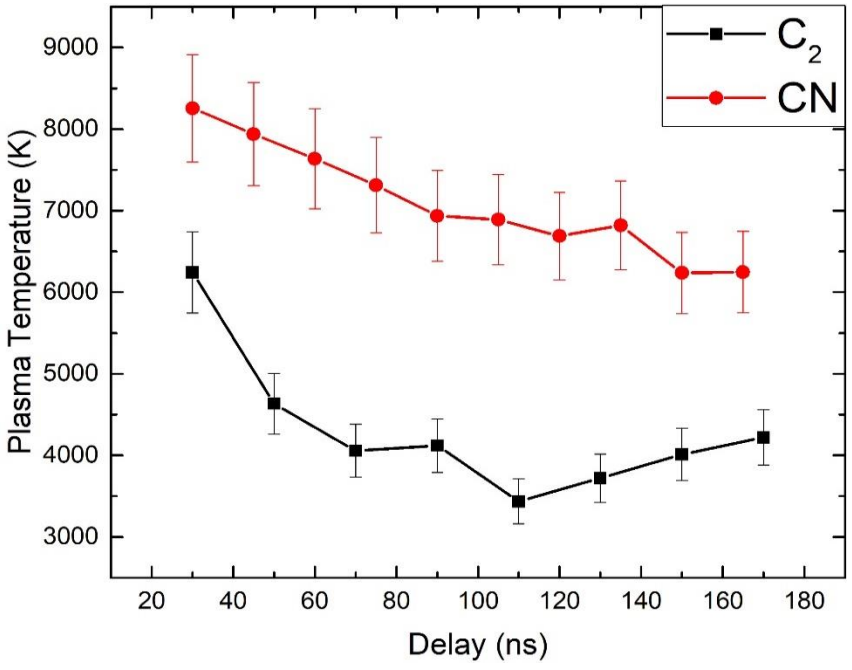


***Figure 3**: Vibronic temperature decay of the CN and C_2 species for the case of the wild plum. The error bars have been defined by executing repetitive measurements.*

The temperature decays of C_2 and CN, for the case of wild plum, are presented in Fig.3. Prima facie the temperature decay of C_2 seems to be faster than that of CN and this observation valid for all the wood specimens studied. This differentiation in the temperature decay time could be thought of as an indication that some processes are taking place within the plume.

Therefore, the emission dynamics of all species participating in plume have been investigated. The CCD gate width was reduced to 15 ns, and LIBS spectra have been recorded for a delay time of up to 500 ns. The emission intensity is reduced as the delay time increases (Fig. 4). The spectral intensity reduction rate of the species contributed to LIBS spectra has been evaluated by fitting the data with the exponential function:

$$y = A \times \exp\left(-\frac{x}{t}\right) + y_0 \qquad (1)$$

The determined values lie in the range of ~25 up to ~110ns.

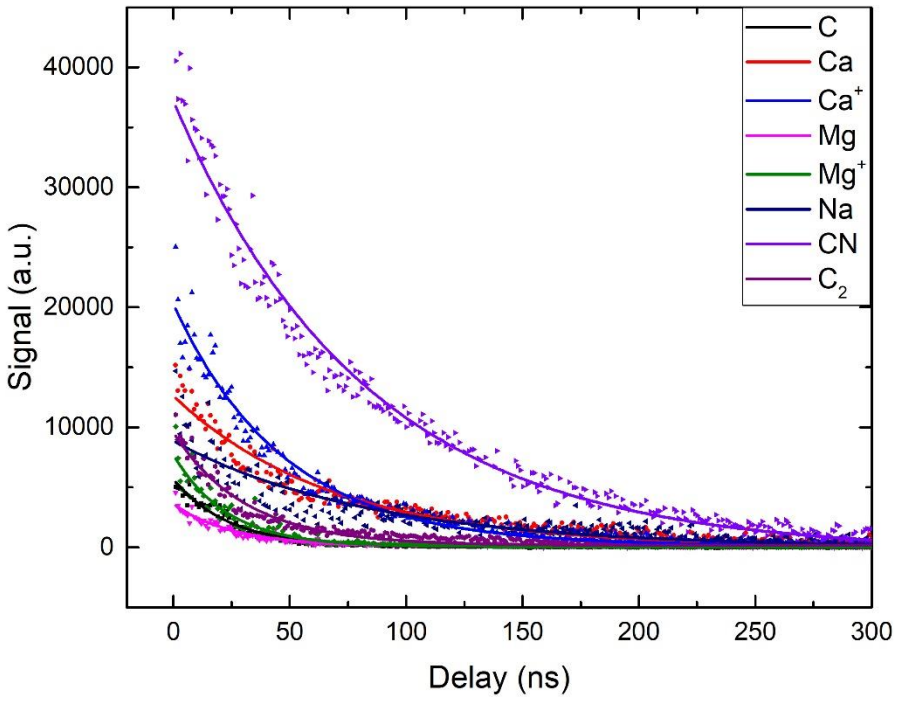


***Figure 4**: Decay rate of the intensity of some spectral features found in all of the spectra obtained from ablating the samples as a function of the time delay.*

From the analysis, it is concluded that the emission intensity reduction rate of CN (*103 ± 16 ns*) and Na (*85 ± 13 ns*) are the slowest. On the contrary, the evaluated rate for the rest of the atoms (t_c=24 ± 4 ns, t_{Ca}=68 ± 10 ns, t_{Ca^+}=43 ± 6 ns, t_{Mg}=26 ± 4 ns, t_{Mg^+}=23 ± 3 ns) and the C_2 (*31 ±5 ns*) fragment is much faster which means that the composition of plasma changes as time proceeds.

In order to explore the reason for this variation, LIBS spectra have been recorded in an argon (Ar) atmosphere (Fig. 5).

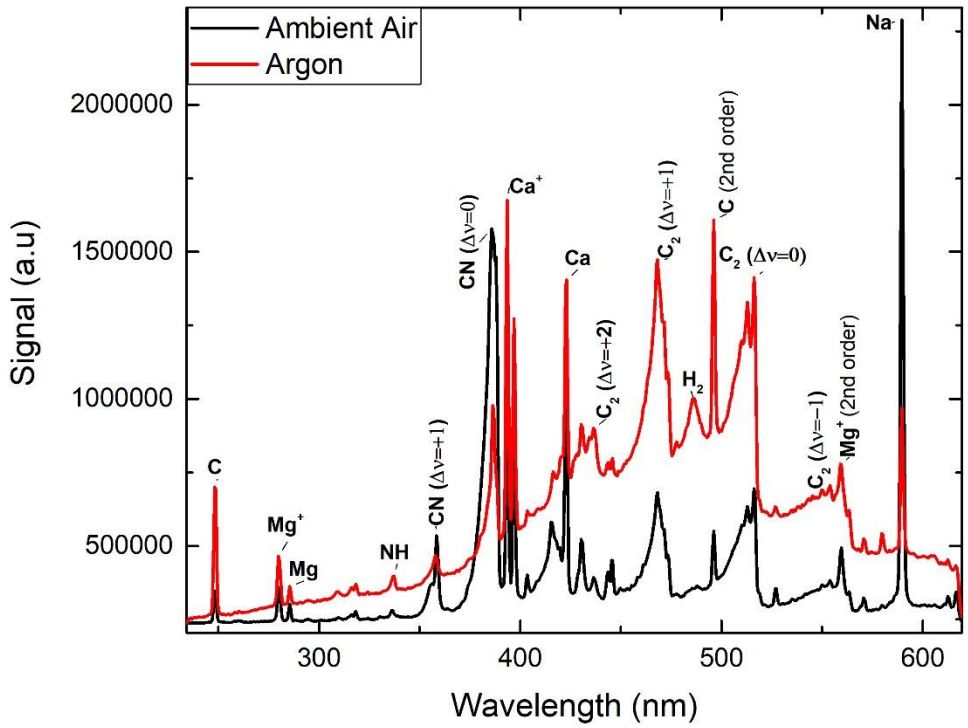


Figure 5*: LIBS spectra of pine tree sample in ambient atmosphere (black line) and under constant Ar flow (6 lt/min) (red line).*

From Fig. 5, it is evident for the spectra recorded in an Ar atmosphere, that the continuum background increases, along with the intensity of the spectral lines corresponding to the transitions of all atoms except those of Na. At the same time, the intensity of the bands assigned to CN transitions is reduced. The increase in the intensity of the carbon (C), calcium (Ca and Ca^+) and magnesium (Mg) peaks is conceivable because in the experiments performed in an Ar atmosphere, the oxidation processes (e.g. $2C + O_2 \rightarrow 2CO$) leading to the formation of CO, CaO and MgO [48], [49] are practically eliminated and therefore an intensity reduction of the corresponding atomic spectral peaks is avoided.

In addition, we notice that the increased signal of C is accompanied by a relative increase in C_2. This could imply that the C_2 population increases via the gas phase $C + C \rightarrow C_2$ reaction [42]. Nevertheless, it is more likely that C_2 is produced primarily by the irradiated sample rather than in the plume because in all the wood samples organic polymers with aromatic rings i.e., strong C=C double bonds, exist. This approach is also compatible with the fact that the intensity of the C_2 bands gains in amplitude in the Ar atmosphere, while those assigned to CN exhibit an apparent decrease. The latter is understandable by taking in account that the reaction leading to CN production in expense of C_2 ($C_2 + N_2 \rightarrow 2CN$) can

be efficient only in ambient atmosphere. Consequently, the activation of this reaction in the plume also accounts for the determined faster decay of C_2 as compared to that of CN. Another reason which could cause a reduction in the CN signal is the reactions involving atomic carbon (C + $N_2 \rightarrow$ CN + N and C + N$\rightarrow$ CN) which are unproductive in an Ar atmosphere experiments [50].

Thus, the reduced presence of CN in LIBS spectra in an Ar atmosphere is reasonable. However, even in an Ar atmosphere, CN as well as NH are clearly observed in the fs-LIBS spectra. This shows that these species can also be molecular fragments released by the dissociation of wood components and that they are not only reaction products in the gas phase. After all, it is known that in the wood composition the percentage participation of N ranges up to ~0.5%, while that of H ~6.5%.

As noted above, the only spectral atomic lines whose intensity decreases (like those corresponding to CN with which they show a common decay) are those corresponding to the characteristic double-line system of sodium (Na). The decrease in the intensity of the Na lines is understood as through a series of reactions, which are only possible in ambient atmosphere. The population of the $^2P_{1/2,\ 3/2}$ states of Na is increased, and deexcitation to the ground state is accompanied by the emission of photons with wavelength equal to 589.0/589.6nm [51]. This process is described by the following reactions:

$2O_2 + hf \rightarrow O_3 + O$

$Na + O_3 \rightarrow NaO + O_2$

$NaO + O \rightarrow Na\ (^2P_{1/2,\ 3/2}) + O_2$

$Na\ (^2P_{1/2,\ 3/2}) \rightarrow Na(^2S_{1/2}) + hf.$

Finally, regarding the increase of the spectral background, which is clearly apparent in the Swan transition of C_2 (Fig.6), it should be noted that these spectra have been recorded by keeping the same CCD gate delay and the temporal width of the gate for both the experiments in ambient and Ar atmosphere.

This observation has been attributed to the fact that for the same temperature gradient, the energy flow is lower in an Ar environment since [49]:

$$E = -\kappa \frac{dT}{dz} \qquad [2]$$

where κ is equal to 0.0163 J/K.m.s and 0.0241 J/K.m.s for Argon and ambient atmosphere, κ_{air} =respectively [52].

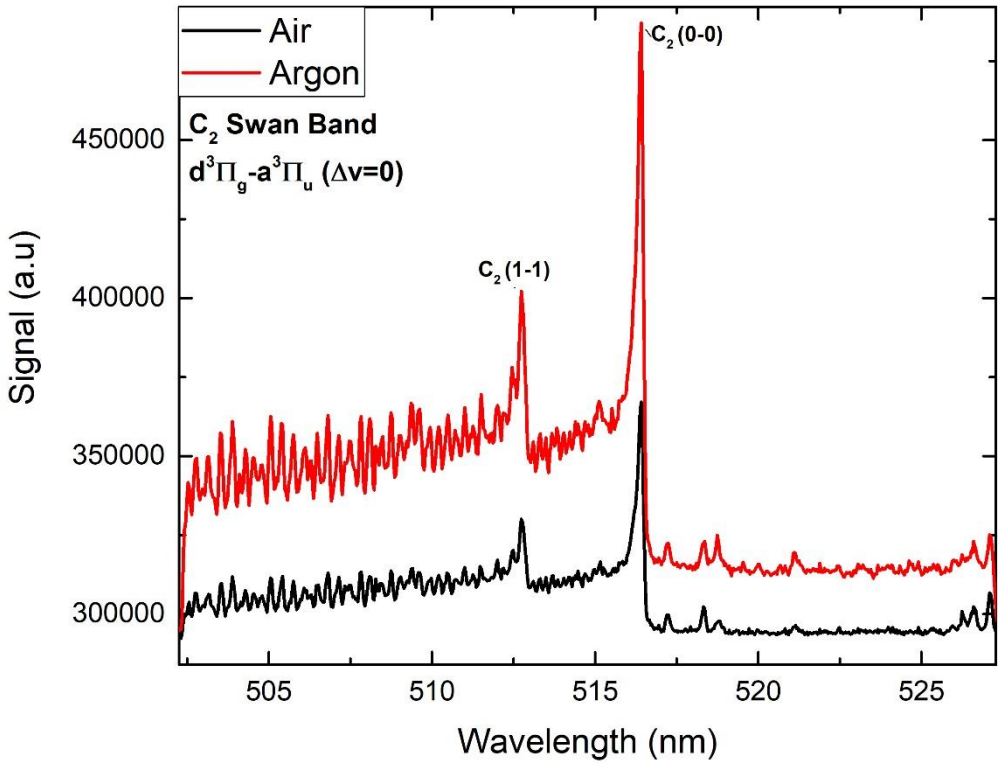


***Figure 6**. The Swan band ($d\,^3\Pi_g \rightarrow a^3\Pi_g$ (Δv =0)) of C_2 in LIBS spectra of pine tree sample recorded in ambient (black line) and Ar atmosphere (red line)*

Equivalently, the presence of a stronger background in the LIBS spectra in an Ar atmosphere, recorded for a specific CCD gate delay, could be attributed to the reduced scattering cross section as compared to that in the ambient atmosphere which is consisted of molecules (N_2, O_2, etc). This affects the processes leading to the emission of bremsstrahlung radiation and contributes to the differentiation of the time evolution of the background in LIBS spectra.

Furthermore, by repeating the emission dynamics measurements in an Ar atmosphere (Fig. 7), it is found that the reduction rate of the spectral features intensity is similar (within the experimental error) for all the species participating in the plume. This supports the above-presented approach about the processes taking place in ambient atmosphere experiments. In other words, indicates that the composition of the plasma is preserved for longer in the experiments performed in an Ar atmosphere. A similar study on nitroimidazoles has shown that the decay time for C and CN is relatively larger in Ar atmosphere experiments, while for C_2 there is no clear trend [53].

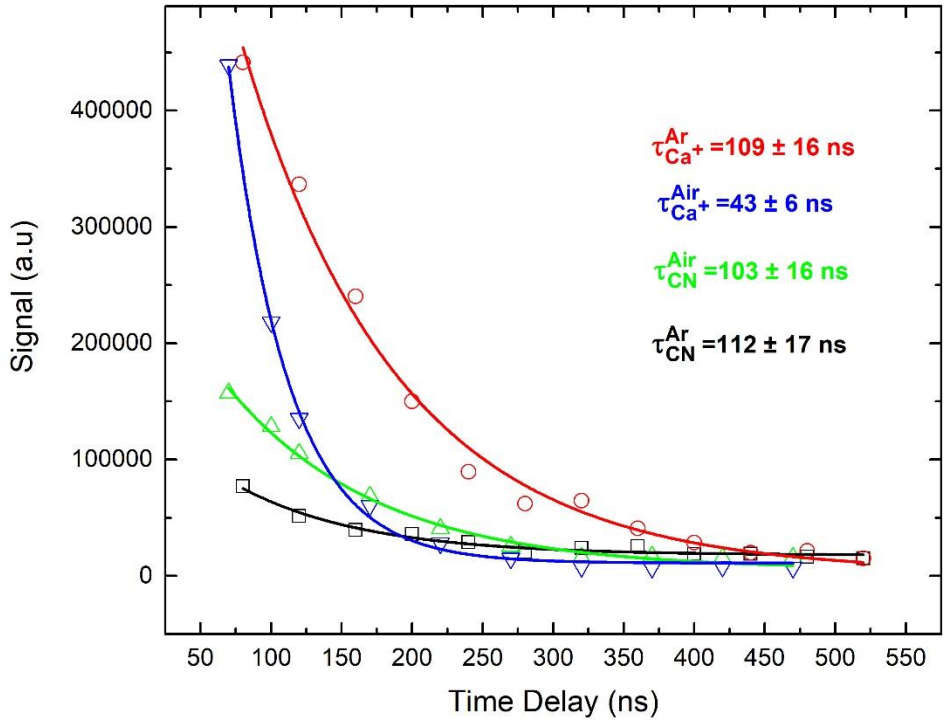


***Figure 7**: Comparison of the decay rates of Ca^+ (λ=393.5nm) and CN in ambient atmospheric conditions, as well as under constant Ar flow.*

The agreement of the conclusions based on the relative variation of the intensity of the spectral features and those obtained from the analysis of the plasma dynamics is apparent. Thus, having analyzed the fs-LIBS data, the next step is to exploit the ability for accurate, sensitive, and fast wood identification by employing data and machine learning techniques.

Data analysis

An extensive analysis was conducted to classify the LIBS data observed from experiments in the ambient atmosphere. The dataset consisted of 1000 collected spectra, with 100 spectra per tree. Eight features (C, Mg^+, Mg, CN, Ca^+, Ca, C_2, and Na) were carefully extracted from each spectrum (each spectrum corresponds to a vector). These features were selected based on their demonstrated relevance in characterizing the elemental composition of trees. Due to the high dimensionality of the dataset (1000 spectra X 8 feature/spectrum), PCA was firstly used as a well-known dimensionality reduction technique. The implementation of PCA in this study proved contributory in unraveling the underlying structure of the LIBS dataset, facilitating the exploration of key factors contributing to the classification of tree species.

Explained variance ratio index (EVR) was analyzed to understand the contribution of different numbers of principal components [54]. EVR represents the proportion of the total variance in the dataset explained by each principal component. After evaluation, it was found that the highest EVR was achieved by keeping three primary components, which captured approximately 99% of the overall information. Specifically, the first principal component explained about 72% of the variance, the second 20%, and the third 7%, respectively.

The contribution of each feature to the formation of each one of the three principal components is shown in Fig. 8. The x-axis represents the main feature produced after analysis, while the y-axis represents the % contribution values. Each feature is depicted as a bar on the x-axis, and the height of each bar corresponds to the magnitude of the feature's contribution to the principal component.

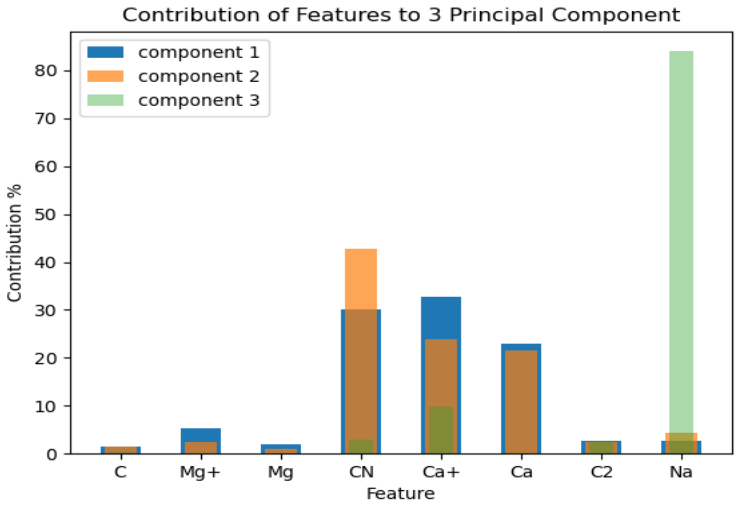


Figure 8: *The contribution (%) of each feature to the formation of each one of the three principal components.*

Based on the analysis of the contribution of features to the principal components, it is evident that CN, Ca^{+}, and Ca play a significant role in the formation of the first and second principal components, while Na is crucial for the formation of the third one. On the other hand, elements such as C and C_2 have minor contributions to the principal components. That mean that elements like C and C_2 occur in similar quantities across all types of wood and do not contribute significantly to the separation and differentiation between the wood species.

A three-dimensional scatter plot was created to visualize the interrelationships among the data points obtained through PCA (Fig. 9a). This visualization provided a comprehensive understanding of the correlations within the dataset and to identify distinct clusters. Arrows are incorporated in the plot to illustrate the directions of the principal components, aiding in the visualization of how the variables or features of the data contribute to the overall variance and their relative relationships.

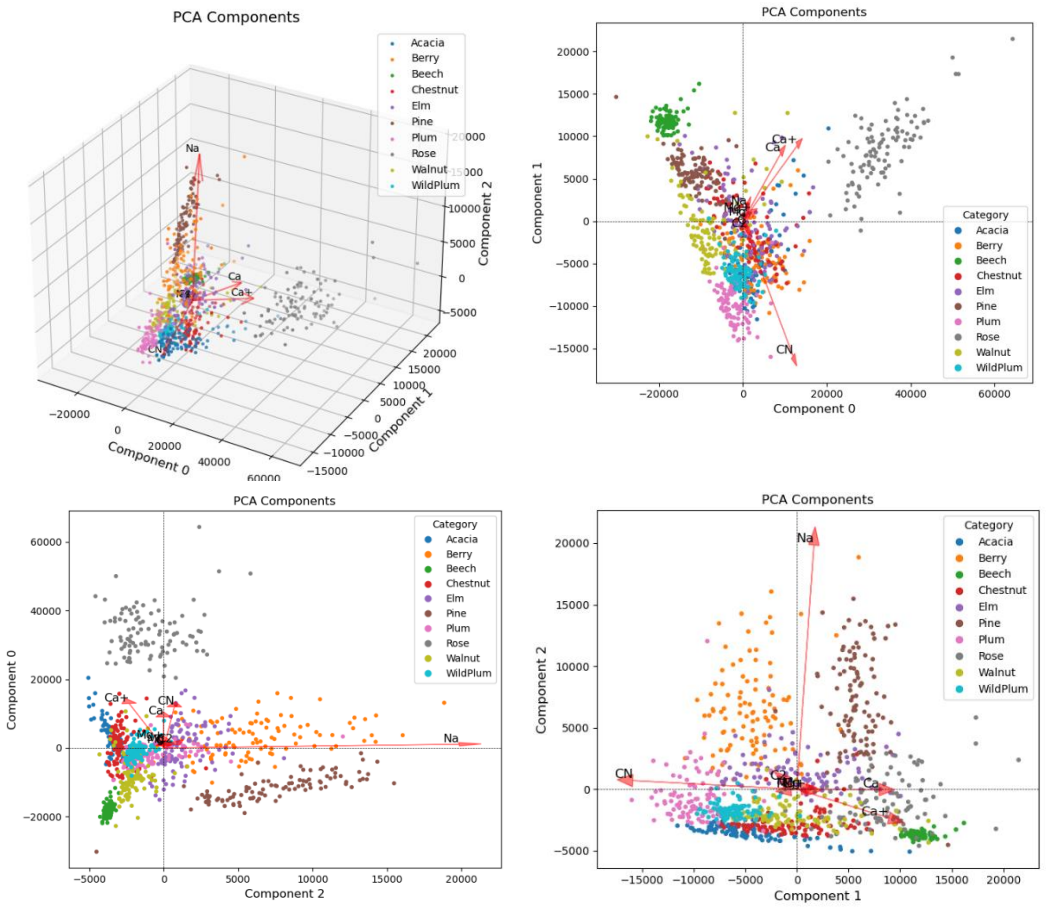


Figure 9: *Interrelationships among the data points obtained through PCA analysis:* ***(a)*** *3-D scatter plot with the three principal components.* ***(b)*** *2-D scatter plot representing components 0 and 1.* ***(c)*** *2-D scatter plot representing components 0 and 2.* ***(d)*** *2-D scatter plot representing components 1 and 2*.

A detailed examination of the scatter plots revealed the presence of well-defined clusters within the dataset. This observation carries important implications for understanding the underlying structure and patterns in the data. Specifically, certain wood species such as beech, pine, and rosewood exhibit distinct clusters that are clearly separated from the other species. These clusters are non-overlapping, indicating a strong differentiation between all wood species. On the other hand, some wood species, such as acacia, wild plum, and chestnut, exhibit overlapping clusters. This overlap suggests a level of confusion or difficulty in differentiating these specific wood species based on their chemical compositions. The lack of clear separation in these cases highlights the complexity and potential challenges involved in accurately classifying these wood species using the available features.

Recognizing and analyzing these patterns makes it apparent that certain wood species exhibit more distinguishable characteristics, while others may share similarities, making their classification more challenging.

After PCA analysis for dimensionality reduction and visualization purposes of the data, the weighted k-Nearest Neighbors (k-NN) algorithm was applied for wood species classification.

In this experiment, the dataset was repeatedly split into two equal parts: 50% for training and 50% for testing, utilizing a random approach. As previously mentioned, the training set is indeed larger than the validation set. Nevertheless, our experimental results, obtained through various partitioning approaches, demonstrate that this division also produces remarkable results. Subsequently, the k-NN classifier algorithm with weighted distances was applied, with k taking values ranging from 2 to 10. It was observed that the algorithm yielded similar results across the range of k values examined. Consequently, the accuracy of the method was evaluated using k=5 nearest neighbors.

Fig. 10 displays the confusion matrix obtained from the application of the weighted k-NN algorithm for wood species classification. By examining Fig. 10, one can gain a comprehensive understanding of the algorithm's effectiveness in accurately classifying wood species. The diagonal elements of the matrix represent the correctly classified instances, while the off-diagonal elements indicate misclassifications. This visual representation allows for the assessment of the classification model's overall accuracy both the overall accuracy of the classification model and the specific classes that may pose challenges for accurate classification.

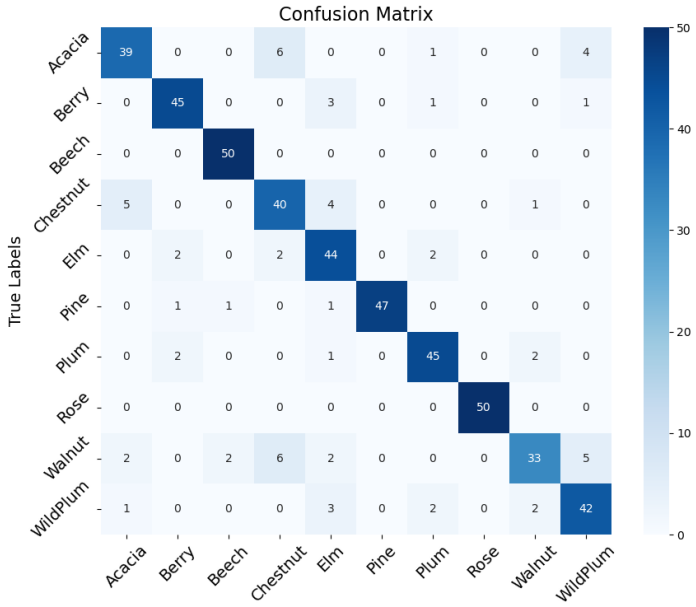


Figure 10: *Confusion matrix representing results of weighted k-NN classifiers' results.*

The results of the three-evaluation metrics (sensitivity, specificity, and accuracy) for each wood species are presented in Fig. 11. By computing these metrics, we can gain a more detailed understanding of the classification model's performance, including its ability to correctly classify positive instances (sensitivity), its ability to correctly classify negative instances (specificity), and the overall accuracy of the model in classifying wood species based on the provided features.

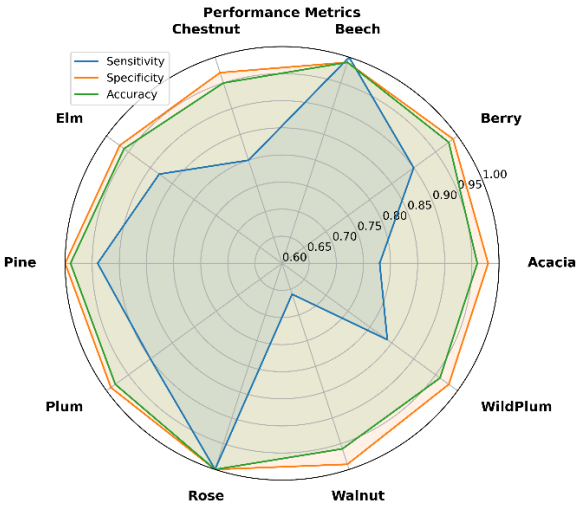


Figure 11: *Radar diagram representing sensitivity, specificity and accuracy of weighted k-NN classifiers' results.*

Sensitivity, representing the proportion of correctly identified positive instances, varied across the classes, with values ranging from 0.66 to 1.0. The k-NN classifier demonstrated high sensitivity for class 3 (Beech) and class 8 (Rose), correctly identifying all positive instances. Specificity, indicating the proportion of correctly identified negative instances, ranged from 0.96 to 1.0, suggesting a high accuracy in identifying negative instances in all classes. The overall accuracy of the classifier was 0.95, indicating that most of the predictions were correct.

Conclusions

The fs LIBS spectra of ten wood specimens have been recorded and analyzed. The observed spectral features have been attributed to atomic/ionic and diatomic molecular transitions. From the analysis of the dynamic characteristics of the created plasma it is concluded that its chemical composition is changing during its expansion. The recognition of the chemical reactions which result in variation of the plasma composition was facilitated by the comparative study of experiments carried out in Argon and ambient atmosphere. For the case of wood specimens, it is concluded that there is significant contribution in the CN and NH signals from transitions taking place within molecular fragments which are ejected during the ablation process and not only from products of reactions occurring in ambient atmosphere. Hence, CN's participation in forming the first and second principal component formation in PCA reflects immediate dissociation processes, too. Thus, the combination of LIBS analysis, PCA for dimensionality reduction, and the application of the weighted k-NN algorithm showcases its ability to classify wood samples based on their atomic/chemical compositions effectively. The evaluation metrics of specificity value were found to be in the range of 0.96-1.00, while that of accuracy was within 0.95-1.00. These results reaffirm the previously reported possibilities for classification purposes of the combined usage of LIBS and machine learning techniques, even for demanding cases like wood specimens.

Author contributions

A. Sarafis: Data Curation, Formal Analysis Investigation, Methodology, Formal Analysis, T. Gerodimos: Data Curation, Formal Analysis Investigation, Methodology, Formal Analysis, Writing –Original Draft, E. Kechaoglou: Data Curation, Formal Analysis Investigation, Methodology, Formal Analysis, D.F. Anagnostopoulos: Data Curation, Investigation, Methodology, Resources, Supervision, Writing – Review & Editing, C. Kosmidis: Conceptualization, Data Curation, Investigation, Methodology, Resources, Supervision, Project Administration, Writing – Original Draft

Conflicts of 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.

Acknowledgments

The LIBS experiments performed at the Central Laser Facility of University of Ioannina.
We acknowledge support of this work by the project ‘‘Center for research quality analysis of cultural heritage materials and communication of science (MIS 5047233) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure’’, funded by the Operational Programme ‘‘Competitiveness, Entrepreneurship and Innovation’’, Greece (NSRF 2014–2020) and co-financed by Greece and the European Union (European Regional Development Fund).

References

[1] X. Cui, Q. Wang, K. Wei, G. Teng, and X. Xu, “Laser-induced breakdown spectroscopy for the classification of wood materials using machine learning methods combined with feature selection,” *Plasma Science and Technology*, vol. 23, no. 5, May 2021, doi: 10.1088/2058-6272/abf1ac.

[2] W. Yu, Z. Sun, and Y. Liu, “Rapid detection and identification of objects using a self-designed methodology based on LIBS and PCA-DVSM – taking rosewood for example,” *Optik (Stuttg)*, vol. 248, Dec. 2021, doi: 10.1016/j.ijleo.2021.168069.

[3] L. Jiao *et al.*, “Discrimination of Salvia miltiorrhiza from Different Geographical Origins by Laser-Induced Breakdown Spectroscopy (LIBS) with Convolutional Neural Network (CNN),” *Anal Lett*, 2023, doi: 10.1080/00032719.2023.2180515.

[4] O. Ozgenc, S. Durmaz, B. Serdar, I. H. Boyaci, H. Eksi-Kocak, and M. Öztürk, “Characterization of fossil Sequoioxylon wood using analytical instrumental techniques,” *Vib Spectrosc*, vol. 96, pp. 10–18, May 2018, doi: 10.1016/j.vibspec.2018.02.006.

[5] K. Kobayashi, S.-W. Hwang, T. Okochi, W.-H. Lee, and J. Sugiyama, “Non-destructive method for wood identification using conventional X-ray

computed tomography data," *J Cult Herit*, vol. 38, pp. 88–93, 2019, doi: https://doi.org/10.1016/j.culher.2019.02.001.

[6] A. N. Shugar, B. L. Drake, and G. Kelley, "Rapid identification of wood species using XRF and neural network machine learning," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-96850-2.

[7] U. P. Agarwal and S. A. Ralph, "FT-Raman spectroscopy of wood: identifying contributions of lignin and carbohydrate polymers in the spectrum of black spruce (Picea mariana)," *Appl Spectrosc*, vol. 51, no. 11, pp. 1648–1655, 1997.

[8] M. Brunner, R. Eugster, E. Trenka, and L. Bergamin-Strotz, "FT-NIR Spectroscopy and Wood Identification," vol. 50, no. 2, pp. 130–134, 1996, doi: doi:10.1515/hfsg.1996.50.2.130.

[9] M. Z. Martin, N. Labbé, T. G. Rials, and S. D. Wullschleger, "Analysis of preservative-treated wood by multivariate analysis of laser-induced breakdown spectroscopy spectra," in *Spectrochimica Acta - Part B Atomic Spectroscopy*, Aug. 2005, pp. 1179–1185. doi: 10.1016/j.sab.2005.05.022.

[10] E. Képeš *et al.*, "Quantification of alloying elements in steel targets: The LIBS 2022 regression contest," *Spectrochim Acta Part B At Spectrosc*, vol. 206, Aug. 2023, doi: 10.1016/j.sab.2023.106710.

[11] P. Siozos, N. Hausmann, M. Holst, and D. Anglos, "Application of laser-induced breakdown spectroscopy and neural networks on archaeological human bones for the discrimination of distinct individuals," *J Archaeol Sci Rep*, vol. 35, Feb. 2021, doi: 10.1016/j.jasrep.2020.102769.

[12] P. Siozos, A. Philippidis, and D. Anglos, "Portable laser-induced breakdown spectroscopy/diffuse reflectance hybrid spectrometer for analysis of inorganic pigments," *Spectrochim Acta Part B At Spectrosc*, vol. 137, pp. 93–100, Nov. 2017, doi: 10.1016/j.sab.2017.09.005.

[13] N. Hausmann, P. Siozos, A. Lemonis, A. C. Colonese, H. K. Robson, and D. Anglos, "Elemental mapping of Mg/Ca intensity ratios in marine mollusc shells using laser-induced breakdown spectroscopy," *J Anal At Spectrom*, vol. 32, no. 8, pp. 1467–1472, Aug. 2017, doi: 10.1039/c7ja00131b.

[14] B. Sezer, G. Bilge, and I. H. Boyaci, "Capabilities and limitations of LIBS in food analysis," *TrAC - Trends in Analytical Chemistry*, vol. 97. Elsevier B.V., pp. 345–353, Dec. 01, 2017. doi: 10.1016/j.trac.2017.10.003.

[15] D. Stefas, N. Gyftokostas, E. Bellou, and S. Couris, "Laser-Induced Breakdown Spectroscopy Assisted by Machine Learning for Plastics/Polymers Identification," *Atoms*, vol. 7, no. 3, 2019, doi: 10.3390/atoms7030079.

[16] B. Praher, V. Palleschi, R. Viskup, J. Heitz, and J. D. Pedarnig, "Calibration free laser-induced breakdown spectroscopy of oxide materials," *Spectrochim Acta Part B At Spectrosc*, vol. 65, no. 8, pp. 671–679, 2010, doi: https://doi.org/10.1016/j.sab.2010.03.010.

[17] A. Botto *et al.*, “Applications of laser-induced breakdown spectroscopy in cultural heritage and archaeology: A critical review,” *Journal of Analytical Atomic Spectrometry*, vol. 34, no. 1. Royal Society of Chemistry, pp. 81–103, Jan. 01, 2019. doi: 10.1039/c8ja00319j.

[18] A. Giakoumaki, K. Melessanaki, and D. Anglos, “Laser-induced breakdown spectroscopy (LIBS) in archaeological science-applications and prospects,” *Anal Bioanal Chem*, vol. 387, no. 3, pp. 749–760, Feb. 2007, doi: 10.1007/s00216-006-0908-1.

[19] Y. Markushin, P. Sivakumar, D. Connolly, and N. Melikechi, “Tag-femtosecond laser-induced breakdown spectroscopy for the sensitive detection of cancer antigen 125 in blood plasma,” *Anal Bioanal Chem*, vol. 407, no. 7, pp. 1849–1855, Feb. 2015, doi: 10.1007/s00216-014-8433-0.

[20] J. H. Han, Y. Moon, J. J. Lee, S. Choi, Y.-C. Kim, and S. Jeong, “Differentiation of cutaneous melanoma from surrounding skin using laser-induced breakdown spectroscopy,” *Biomed Opt Express*, vol. 7, no. 1, pp. 57–66, 2016.

[21] Z. A. Arp, D. A. Cremers, R. D. Harris, D. M. Oschwald, G. R. Parker Jr, and D. M. Wayne, “Feasibility of generating a useful laser-induced breakdown spectroscopy plasma on rocks at high pressure: preliminary study for a Venus mission,” *Spectrochim Acta Part B At Spectrosc*, vol. 59, no. 7, pp. 987–999, 2004.

[22] S. Sunku, M. K. Gundawar, A. K. Myakalwar, P. P. Kiran, S. P. Tewari, and S. V. Rao, “Femtosecond and nanosecond laser induced breakdown spectroscopic studies of NTO, HMX, and RDX,” *Spectrochim Acta Part B At Spectrosc*, vol. 79–80, pp. 31–38, Jan. 2013, doi: 10.1016/j.sab.2012.11.002.

[23] J. B. Sirven, B. Bousquet, L. Canioni, and L. Sarger, “Time-resolved and time-integrated single-shot laser-induced plasma experiments using nanosecond and femtosecond laser pulses,” *Spectrochim Acta Part B At Spectrosc*, vol. 59, no. 7, pp. 1033–1039, Jul. 2004, doi: 10.1016/j.sab.2004.05.009.

[24] J. R. Freeman, S. S. Harilal, P. K. Diwakar, B. Verhoff, and A. Hassanein, “Comparison of optical emission from nanosecond and femtosecond laser produced plasma in atmosphere and vacuum conditions,” *Spectrochim Acta Part B At Spectrosc*, vol. 87, pp. 43–50, Sep. 2013, doi: 10.1016/j.sab.2013.05.011.

[25] S. S. Harilal, B. E. Brumfield, N. L. Lahaye, K. C. Hartig, and M. C. Phillips, “Optical spectroscopy of laser-produced plasmas for standoff isotopic analysis,” *Appl Phys Rev*, vol. 5, no. 2, Jun. 2018, doi: 10.1063/1.5016053.

[26] S. H. Chung and E. Mazur, “Surgical applications of femtosecond lasers,” *Journal of Biophotonics*, vol. 2, no. 10. pp. 557–572, Oct. 2009. doi: 10.1002/jbio.200910053.

[27] J. Serrano, J. Moros, and J. Javier Laserna, “Molecular signatures in femtosecond laser-induced organic plasmas: comparison with nanosecond

laser ablation," *Physical Chemistry Chemical Physics*, vol. 18, no. 4, pp. 2398–2408, Jan. 2016, doi: 10.1039/c5cp06456b.

[28] J. Moros and J. Laserna, "Laser-Induced Breakdown Spectroscopy (LIBS) of Organic Compounds: A Review," *Appl Spectrosc*, vol. 73, no. 9, pp. 963–1011, Sep. 2019, doi: 10.1177/0003702819853252.

[29] M. Z. Martin *et al.*, "High resolution applications of laser-induced breakdown spectroscopy for environmental and forensic applications," *Spectrochim Acta Part B At Spectrosc*, vol. 62, no. 12, pp. 1426–1432, Dec. 2007, doi: 10.1016/j.sab.2007.10.046.

[30] D. Holub, P. Pořízka, M. Kizovský, D. Prochazka, O. Samek, and J. Kaiser, "The potential of combining laser-induced breakdown spectroscopy and Raman spectroscopy data for the analysis of wood samples," *Spectrochim Acta Part B At Spectrosc*, vol. 195, Sep. 2022, doi: 10.1016/j.sab.2022.106487.

[31] J. G. R. Leandro, F. B. Gonzaga, and J. V. de F. Latorraca, "Discrimination of wood species using laser-induced breakdown spectroscopy and near-infrared reflectance spectroscopy," *Wood Sci Technol*, vol. 53, no. 5, pp. 1079–1091, Sep. 2019, doi: 10.1007/s00226-019-01119-7.

[32] L. Brunnbauer, Z. Gajarska, H. Lohninger, and A. Limbeck, "A critical review of recent trends in sample classification using Laser-Induced Breakdown Spectroscopy (LIBS)," *TrAC - Trends in Analytical Chemistry*, vol. 159. Elsevier B.V., Feb. 01, 2023. doi: 10.1016/j.trac.2022.116859.

[33] H. Abdi, L. J. Williams, and D. Valentin, "Wiley Interdiscip. Rev," *Comput. Stat*, vol. 2, no. 4, p. 433, 2010.

[34] E. Fix and J. L. Hodges, "Discriminatory analysis. Nonparametric discrimination: Consistency properties," *International Statistical Review/Revue Internationale de Statistique*, vol. 57, no. 3, pp. 238–247, 1989.

[35] H. Yigit, "A weighting approach for KNN classifier," in *2013 international conference on electronics, computer and computation (ICECCO)*, IEEE, 2013, pp. 228–231.

[36] W. Liu and S. Chawla, "Class confidence weighted k NN algorithms for imbalanced data sets," in *Advances in Knowledge Discovery and Data Mining: 15th Pacific-Asia Conference, PAKDD 2011, Shenzhen, China, May 24-27, 2011, Proceedings, Part II 15*, Springer, 2011, pp. 345–356.

[37] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, vol. 2. Springer, 2009.

[38] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit Lett*, vol. 27, no. 8, pp. 861–874, 2006.

[39] A. Kramida, Yu. Ralchenko, and J. Reader, “NIST Atomic Spectra Database (version 5.10), [Online],” *National Institute of Standards and Technology, Gaithersburg,MD*, 2022.

[40] X. Cui, Q. Wang, Y. Zhao, X. Qiao, and G. Teng, “Laser-induced breakdown spectroscopy (LIBS) for classification of wood species integrated with artificial neural network (ANN),” *Appl Phys B*, vol. 125, no. 4, Apr. 2019, doi: 10.1007/s00340-019-7166-3.

[41] S. J. Mousavi, M. Hemati Farsani, S. M. R. Darbani, A. Mousaviazar, M. Soltanolkotabi, and A. Eslami Majd, “CN and C2 vibrational spectra analysis in molecular LIBS of organic materials,” *Appl Phys B*, vol. 122, no. 5, May 2016, doi: 10.1007/s00340-016-6371-6.

[42] Y. L. Zhao, G. G. Li, H. M. Hou, J. C. Shi, and S. N. Luo, “CN and C2 formation mechanisms in fs-laser induced breakdown of nitromethane in Ar or N2 atmosphere,” *J Hazard Mater*, vol. 393, Jul. 2020, doi: 10.1016/j.jhazmat.2020.122396.

[43] F. Xu, S. Ma, C. Zhao, and D. Dong, “Application of Molecular Emissions in Laser-Induced Breakdown Spectroscopy: A Review,” *Frontiers in Physics*, vol. 10. Frontiers Media S.A., Jan. 27, 2022. doi: 10.3389/fphy.2022.821528.

[44] M. Dong, G. C. Y. Chan, X. Mao, J. J. Gonzalez, J. Lu, and R. E. Russo, “Elucidation of C2 and CN formation mechanisms in laser-induced plasmas through correlation analysis of carbon isotopic ratio,” *Spectrochim Acta Part B At Spectrosc*, vol. 100, pp. 62–69, Oct. 2014, doi: 10.1016/j.sab.2014.08.009.

[45] Q. Wang *et al.*, “Effect of lens focusing distance on AlO molecular emission from femtosecond laser-induced aluminum plasma in air,” *Opt Laser Technol*, vol. 122, Feb. 2020, doi: 10.1016/j.optlastec.2019.105862.

[46] C. G. Parigger, A. C. Woods, D. M. Surmick, G. Gautam, M. J. Witte, and J. O. Hornkohl, “Computation of diatomic molecular spectra for selected transitions of aluminum monoxide, cyanide, diatomic carbon, and titanium monoxide,” *Spectrochim Acta Part B At Spectrosc*, vol. 107, pp. 132–138, May 2015, doi: 10.1016/j.sab.2015.02.018.

[47] M. J. Witte and C. G. Parigger, “Measurement and Analysis of Carbon Swan Spectra Following Laser-induced Optical Breakdown in Air I R A M P Measurement and Analysis of Carbon Swan Spectra Following Laser-induced Optical Breakdown in Air,” *International Review of Atomic and Molecular Physics*, vol. 4, no. 1, pp. 63–67.

[48] F. C. De Lucia and J. L. Gottfried, “Influence of molecular structure on the laser-induced plasma emission of the explosive RDX and organic polymers,” *Journal of Physical Chemistry A*, vol. 117, no. 39, pp. 9555–9563, Oct. 2013, doi: 10.1021/jp312236h.

[49] S. J. Mousavi, M. Hemati Farsani, S. M. R. Darbani, A. Mousaviazar, M. Soltanolkotabi, and A. Eslami Majd, “CN and C2 vibrational spectra analysis

in molecular LIBS of organic materials," *Appl Phys B*, vol. 122, no. 5, May 2016, doi: 10.1007/s00340-016-6371-6.

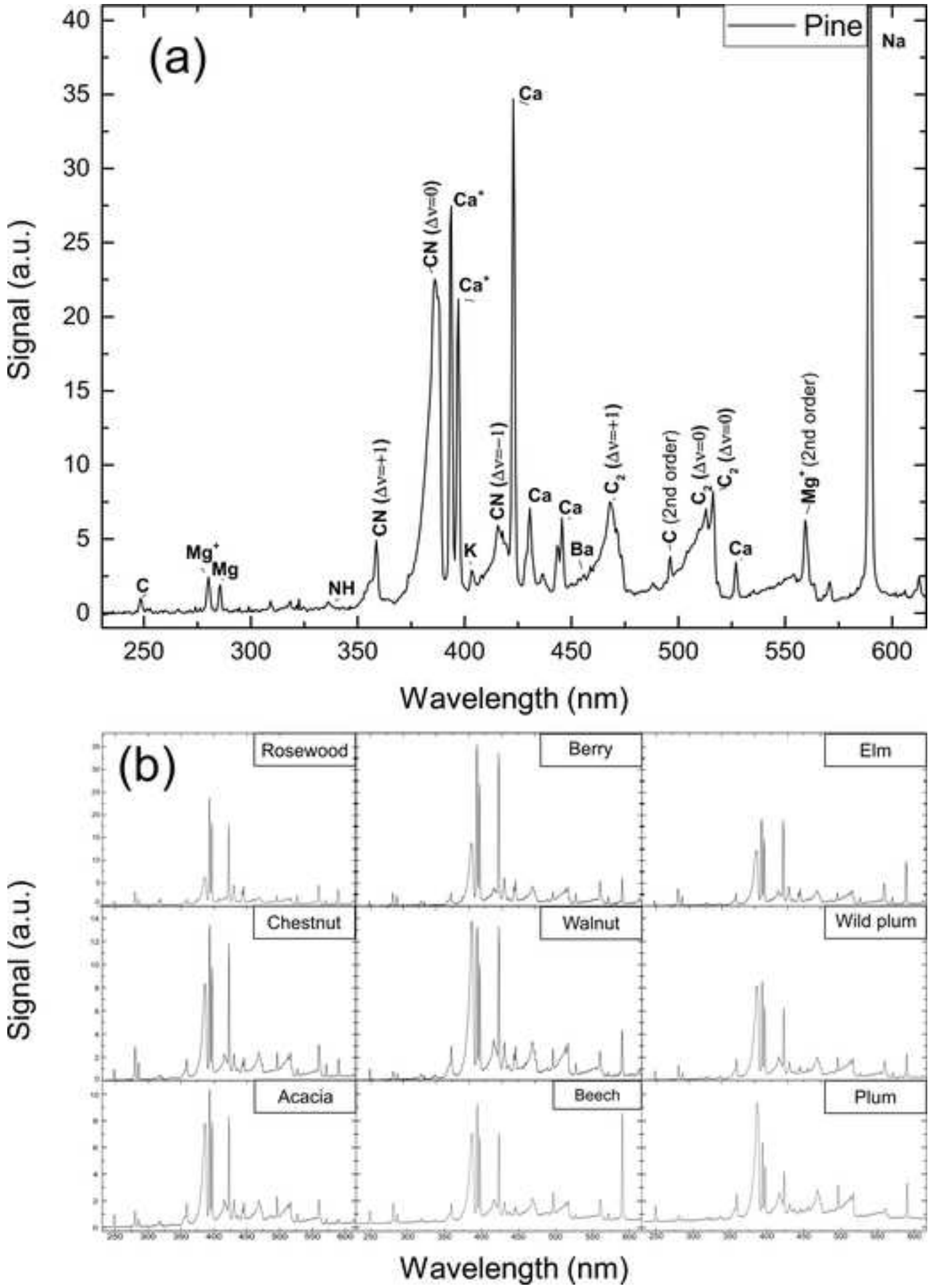
[50] M. Dong, X. Mao, J. J. Gonzalez, J. Lu, and R. E. Russo, "Time-resolved LIBS of atomic and molecular carbon from coal in air, argon and helium," *J Anal At Spectrom*, vol. 27, no. 12, pp. 2066–2075, Dec. 2012, doi: 10.1039/c2ja30222e.

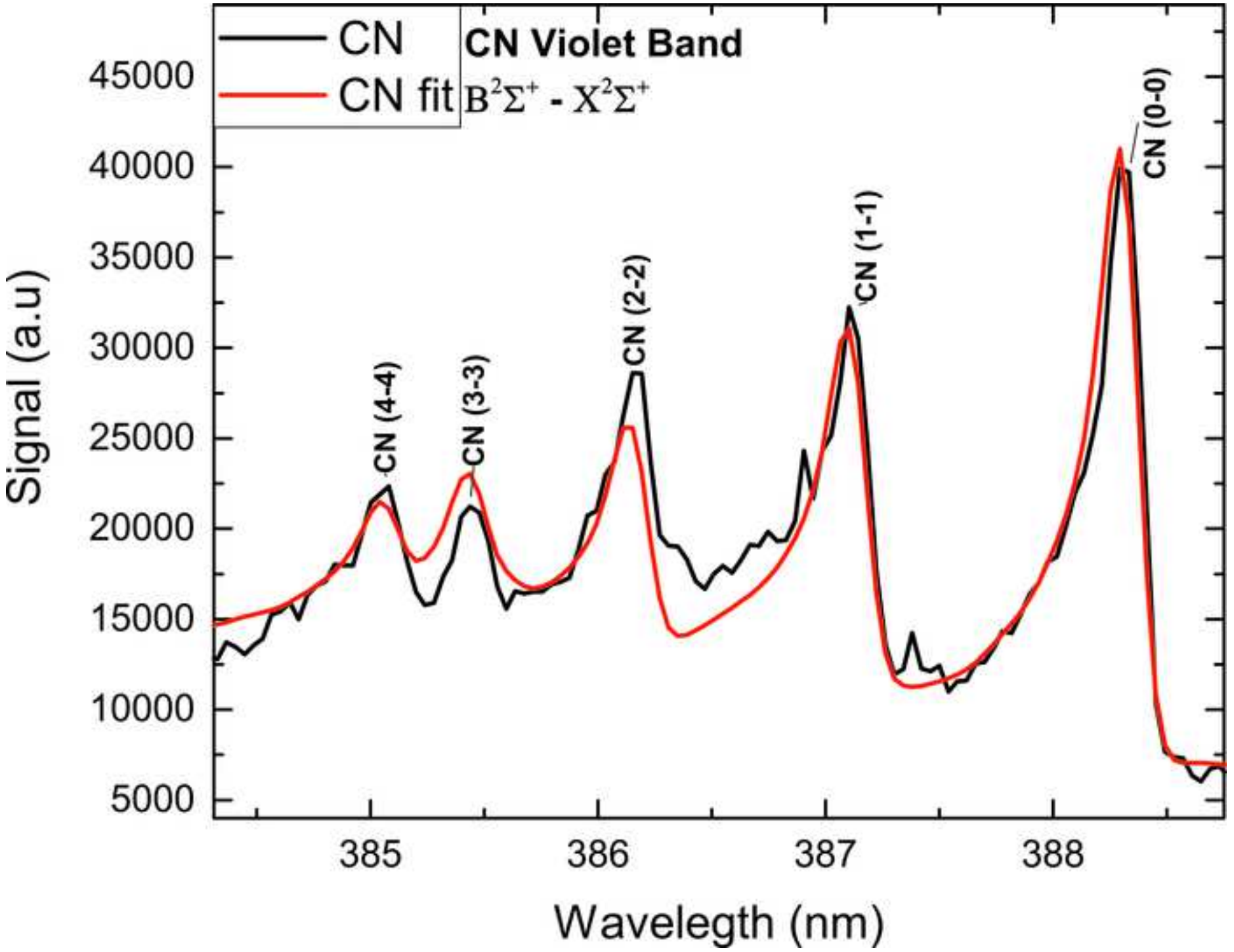
[51] J. Griffin, D. R. Worsnop, R. C. Brown, C. E. Kolb, and D. R. Herschbach, "Chemical kinetics of the NaO (A2∑+) + O(3P) reaction," *Journal of Physical Chemistry A*, vol. 105, no. 9, pp. 1643–1648, Mar. 2001, doi: 10.1021/jp002641m.

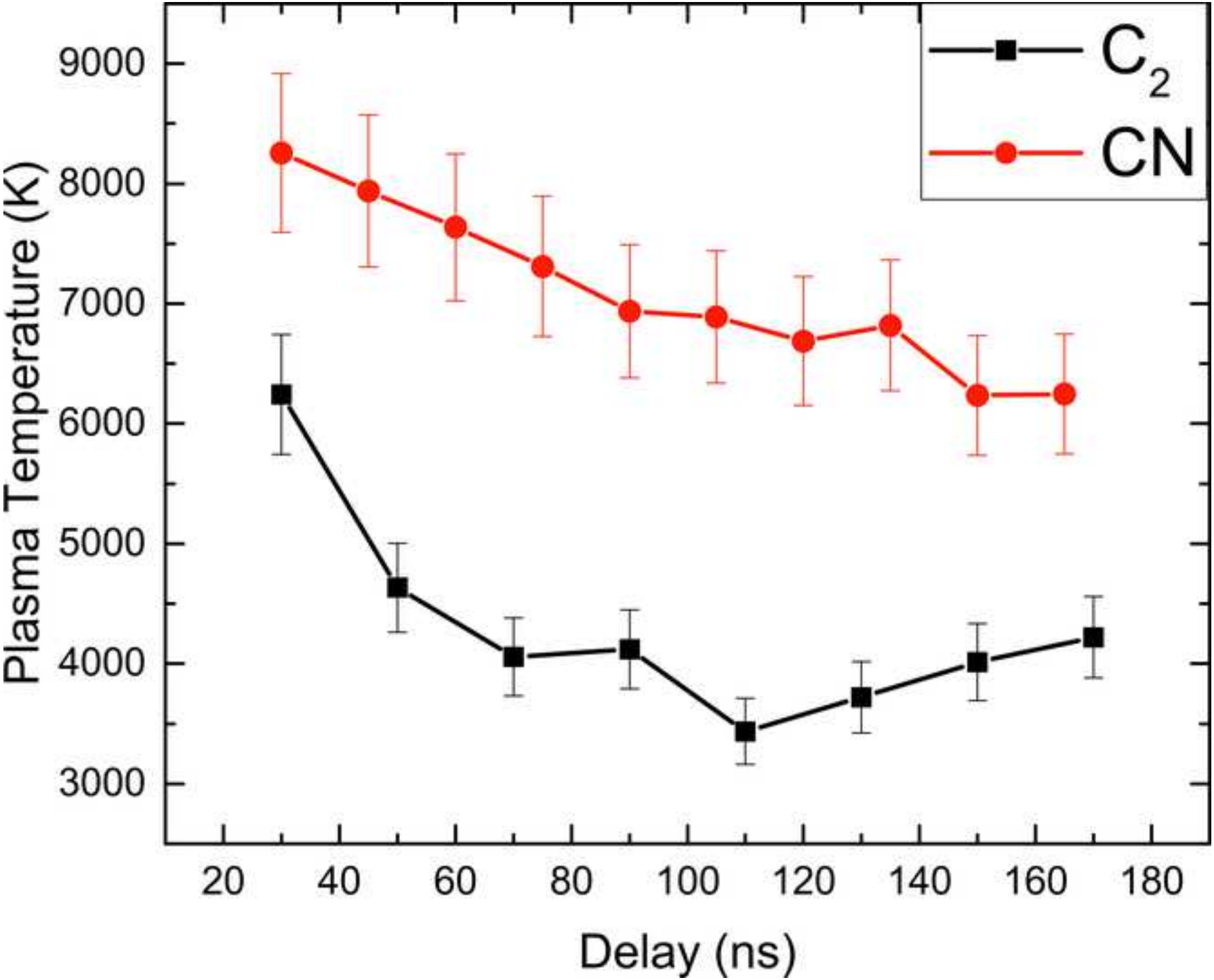
[52] P. W. Atkins and J. De Paula, *Atkins' physical chemistry*. Oxford university press, 2014.

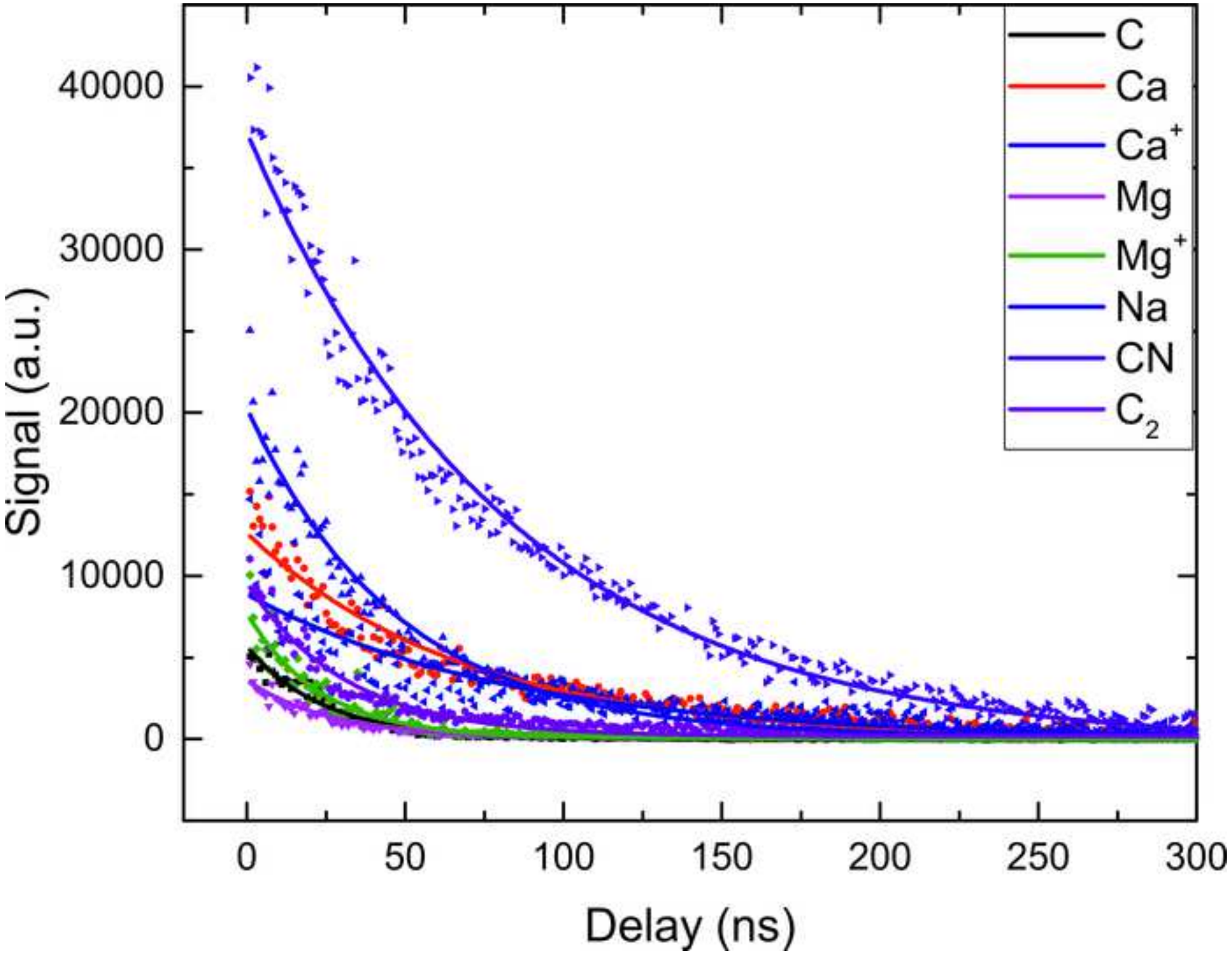
[53] E. N. Rao *et al.*, "Femtosecond and nanosecond LIBS studies of nitroimidazoles: Correlation between molecular structure and LIBS data," *J Anal At Spectrom*, vol. 31, no. 3, pp. 737–750, Mar. 2016, doi: 10.1039/c5ja00445d.

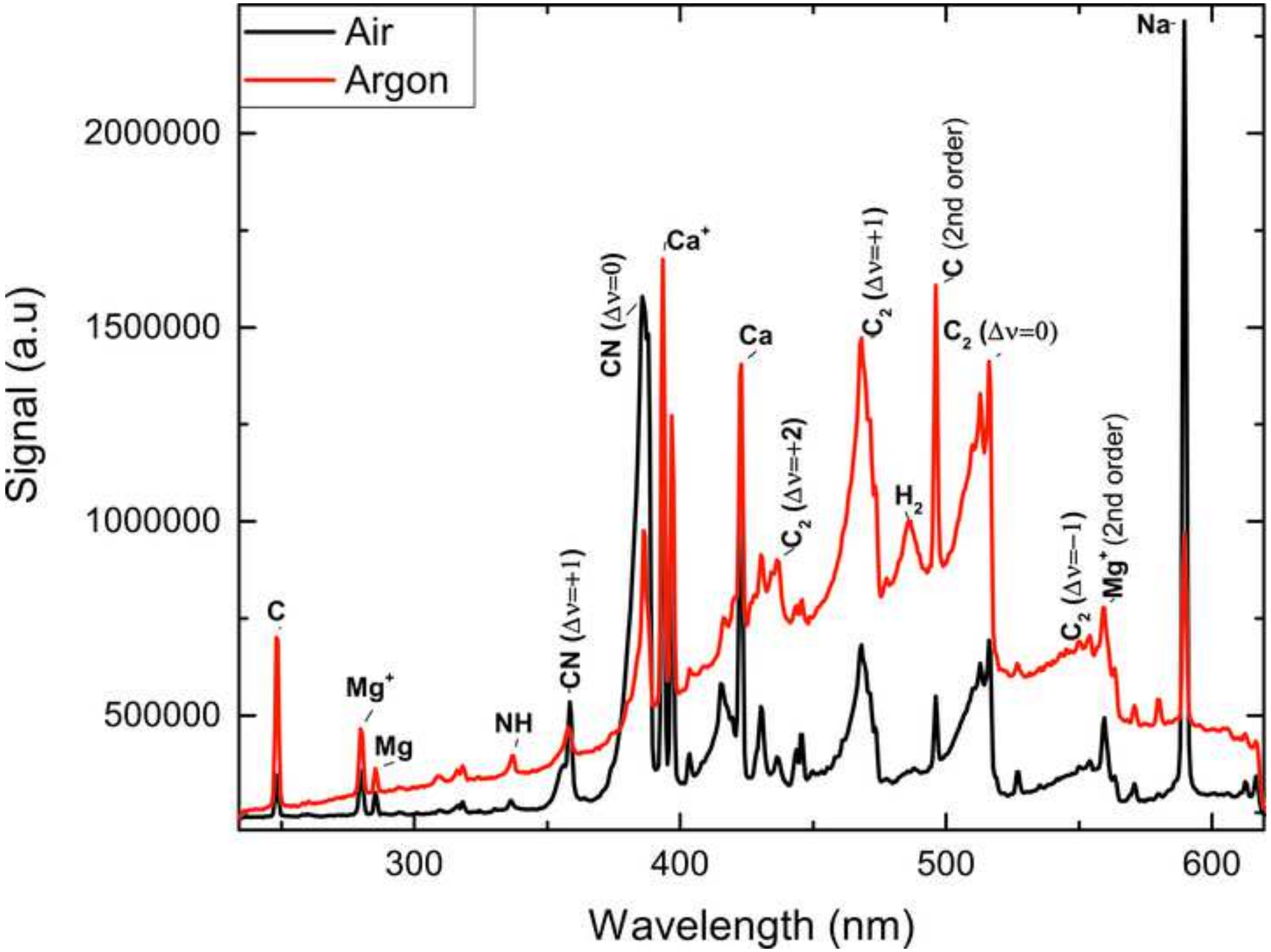
[54] Janžekovič, F., & Novak, T. (2012). PCA–a powerful method for analyze ecological niches. *Principal component analysis–multidisciplinary applications*, 127-142.

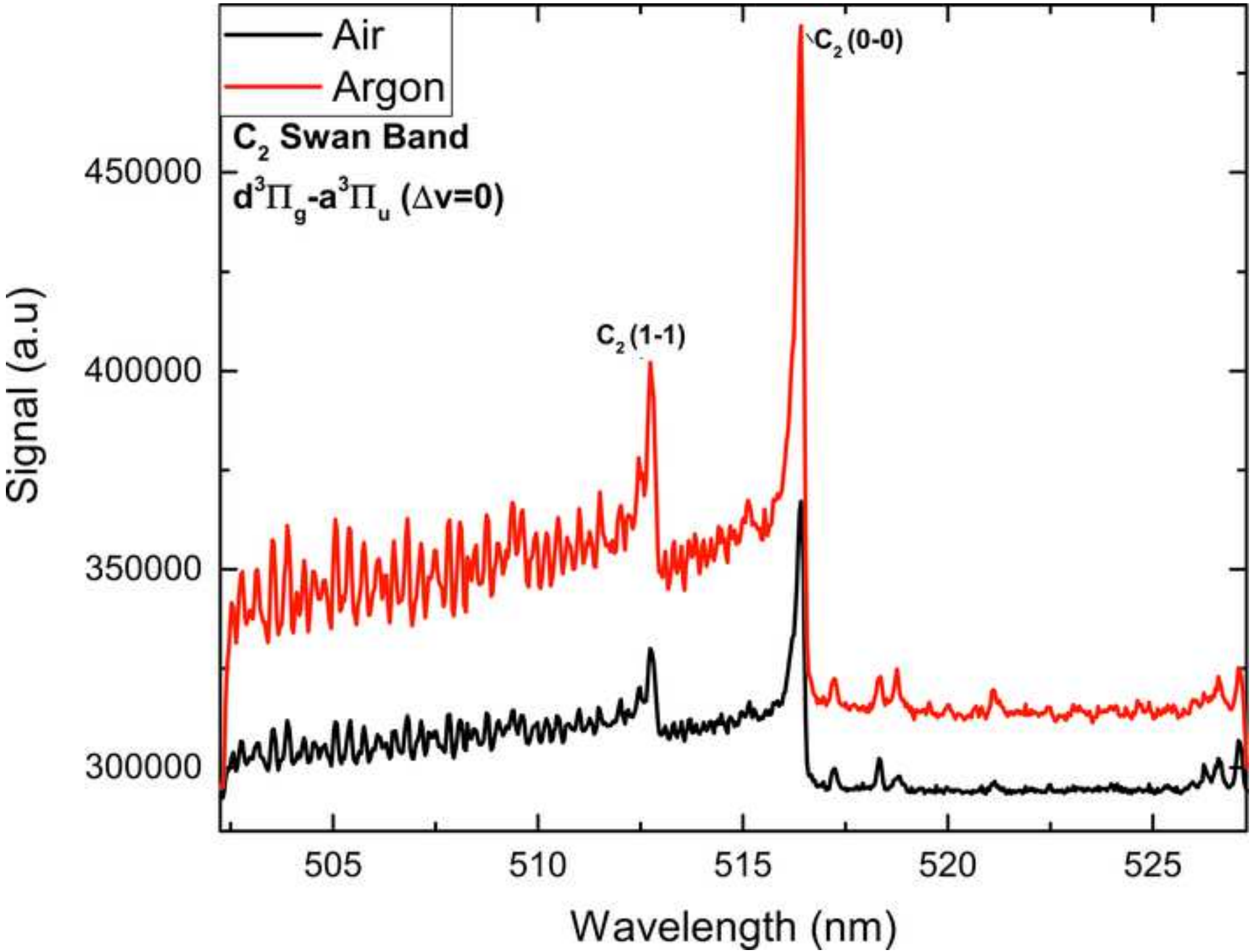


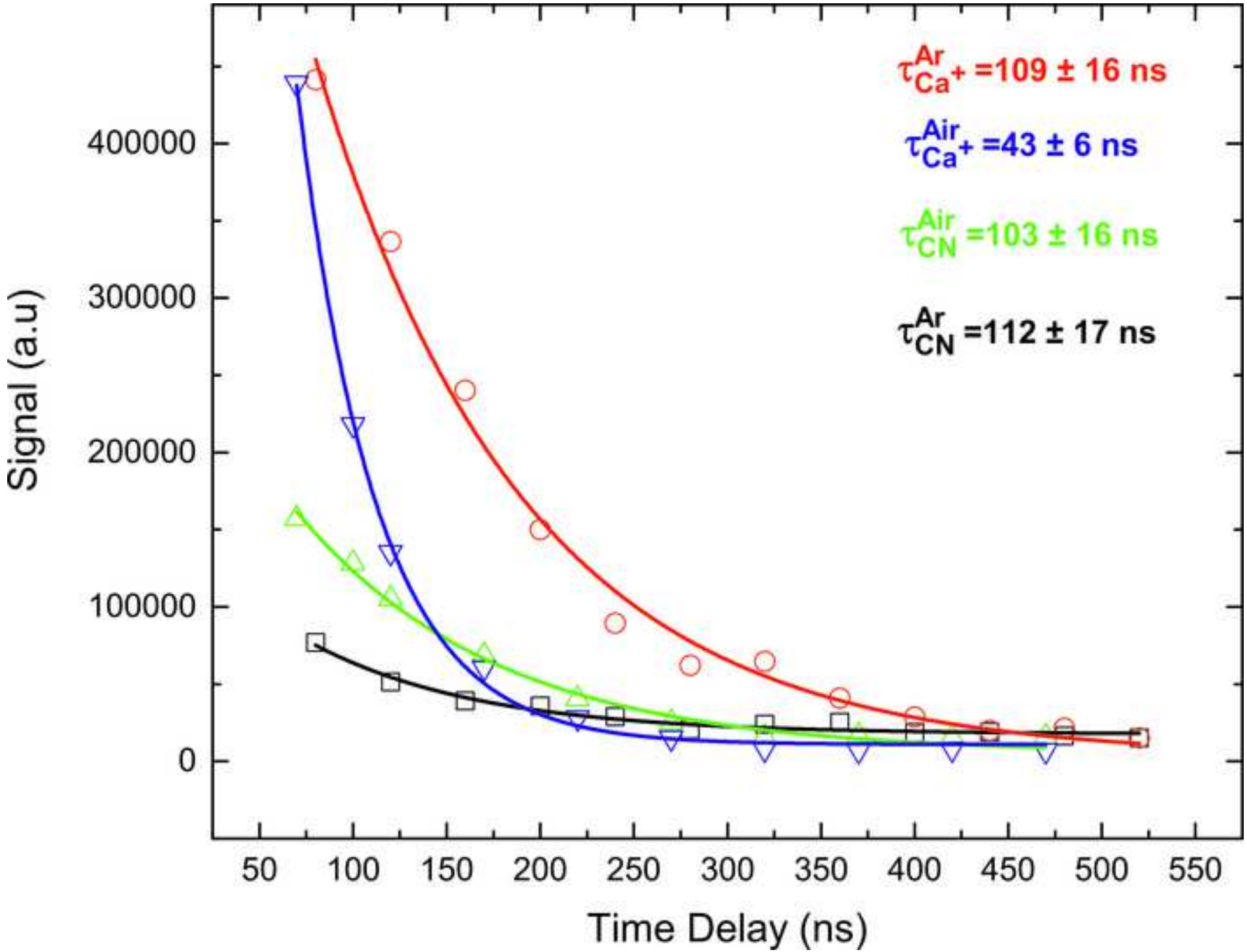


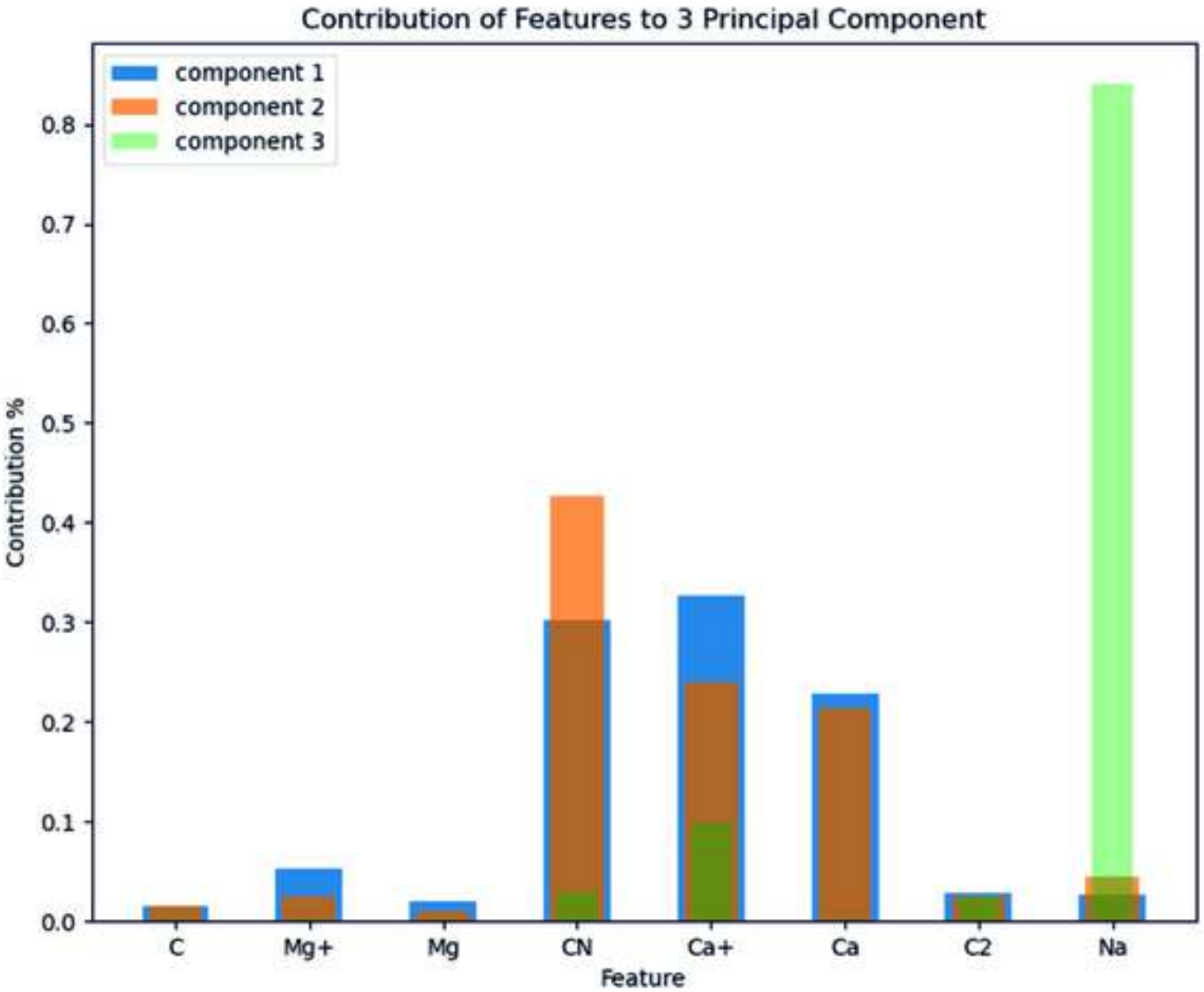


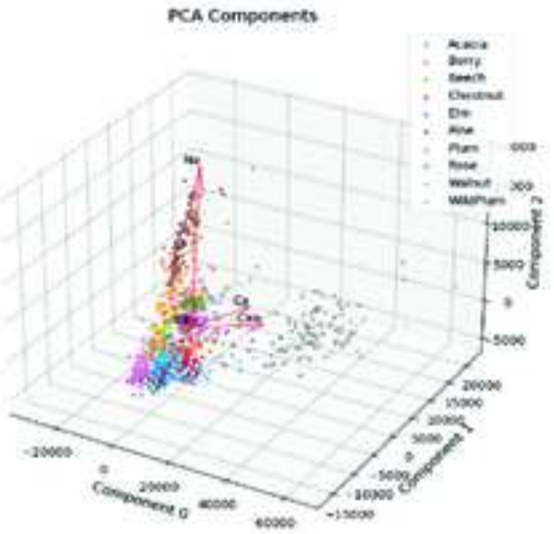


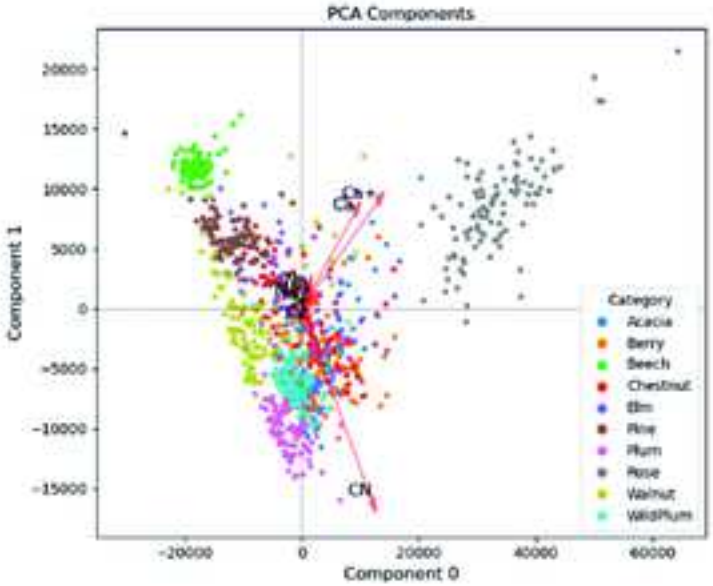


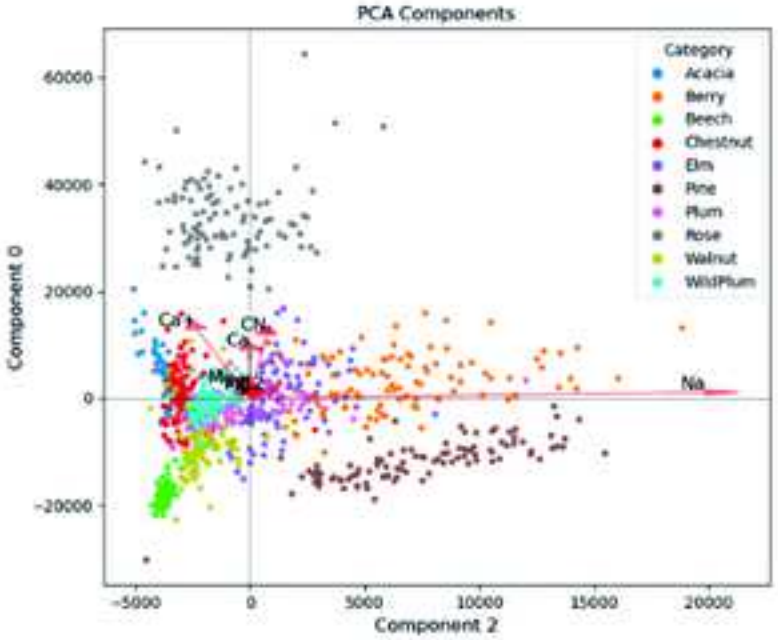


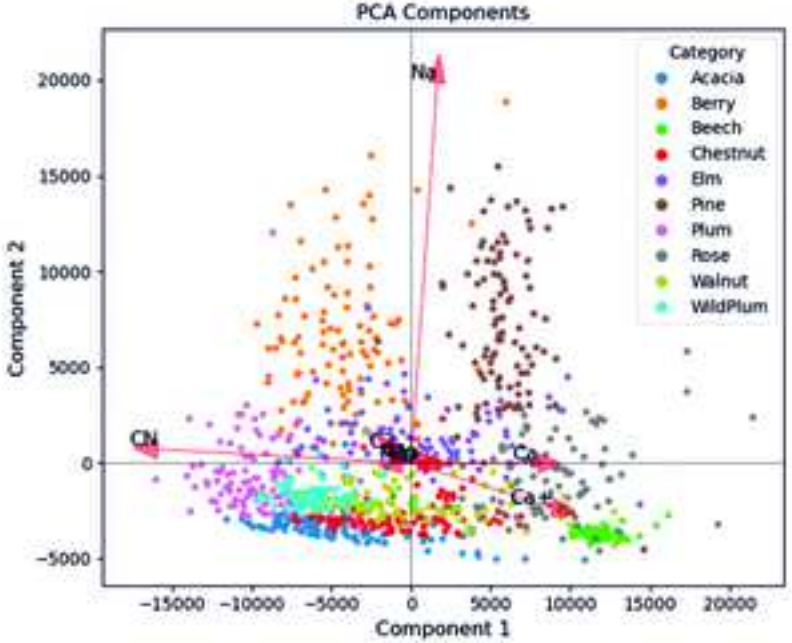


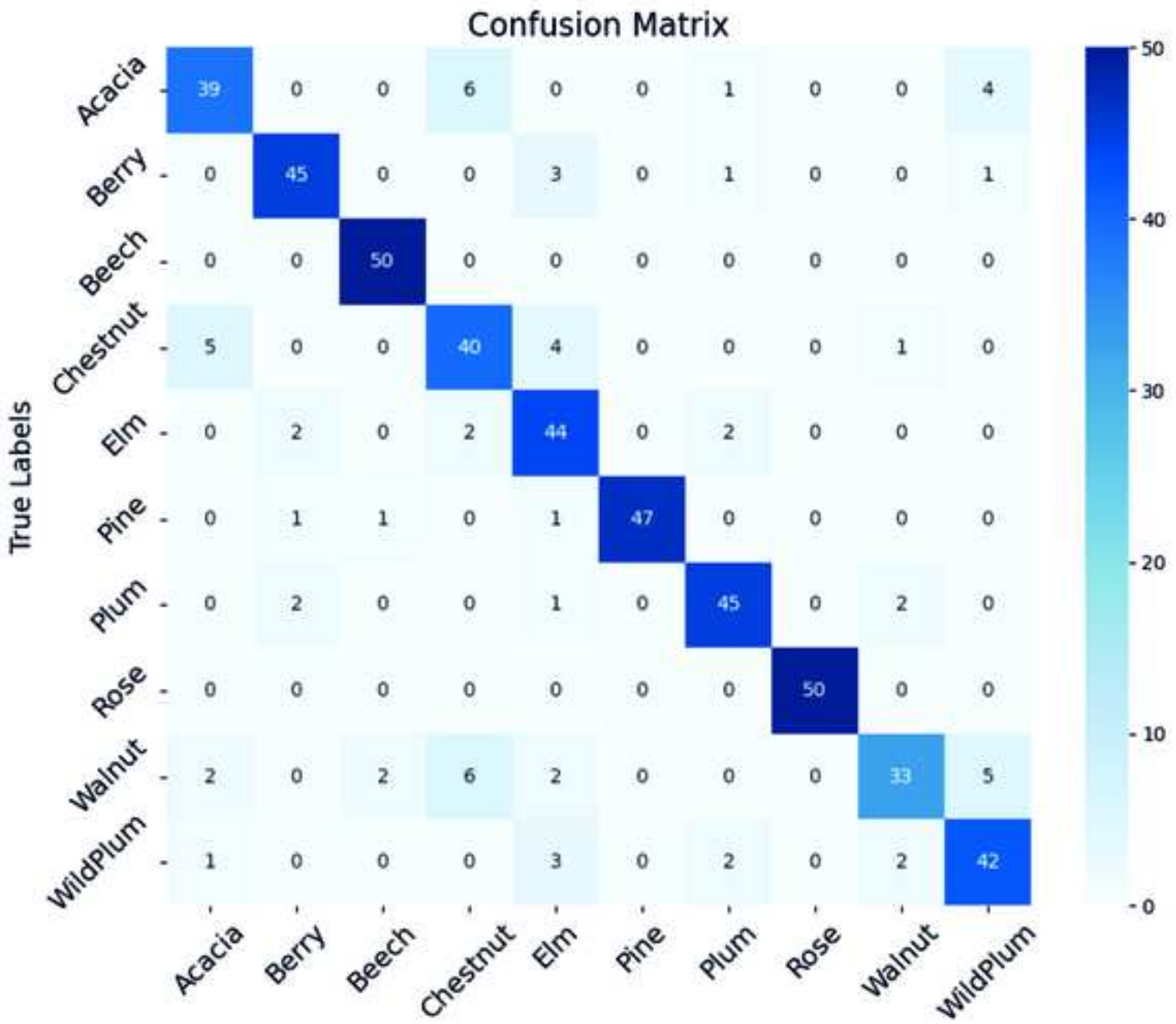


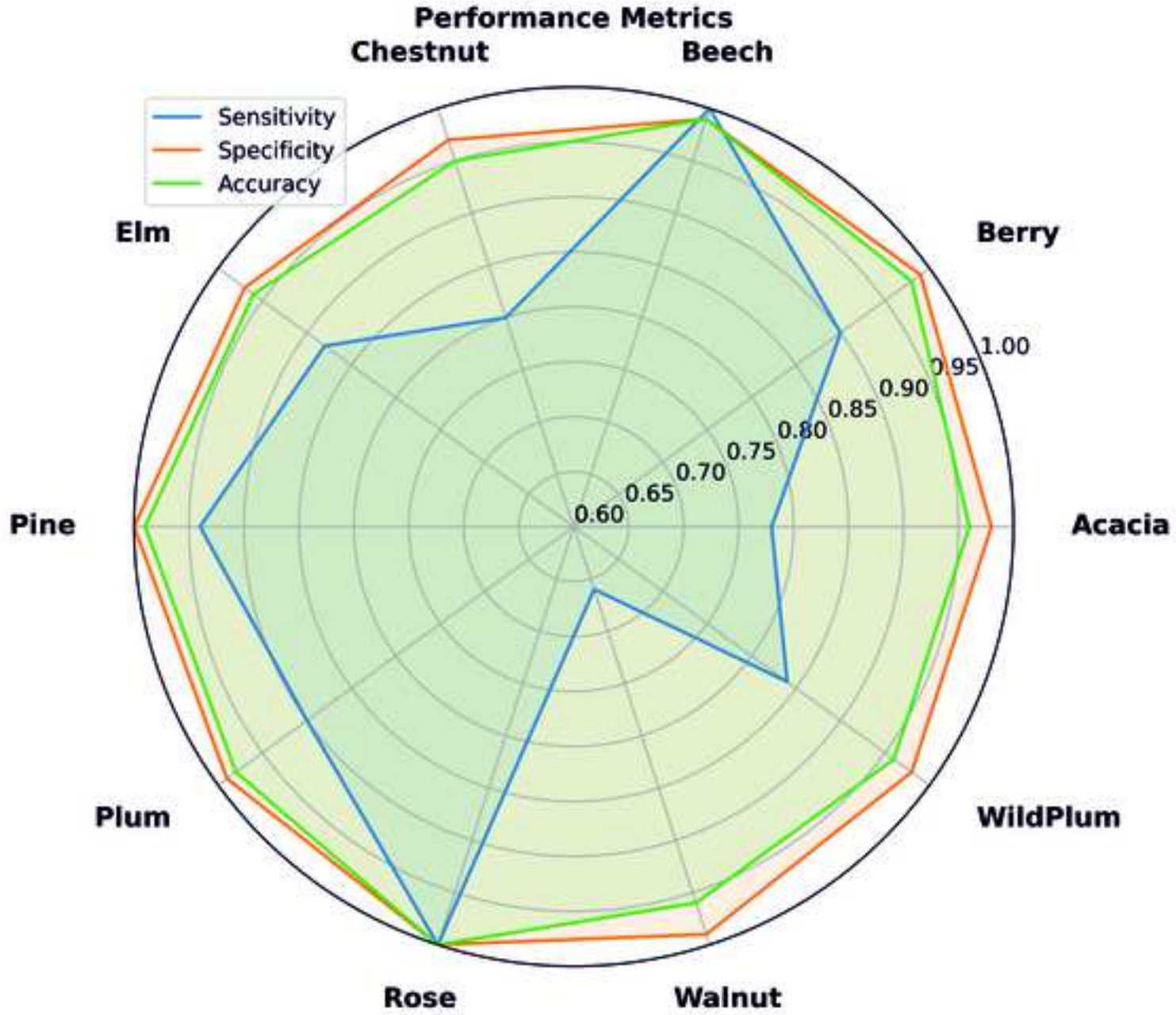












to the Editor of
Applied Physics A

Ioannina, September 12,2023

Dear Editor,

Please find attached our paper entitled: "**Identification of wood specimens utilizing fs-LIBS and machine learning techniques**" which we would like to be considered for publication in Applied Physics A.

Thank you in advance.

Sincerely

C. Kosmidis

Prof. Kosmidis Constantine
University of Ioannina, University Campus, Department of Physics,
Atomic and Molecular Physics Laboratory
GR-45110 Ioannina, Greece
Tel.: +30 26510 08537; fax: +30 26510 08695
E-mail addresses: kkosmid@uoi.gr