Applied Science and Biotechnology Journal for Advanced Research
e-ISSN: 2583-553X
Peer Reviewed Journal
Volume-2, Issue-1 (January 2023), Page: 29-35
https://doi.org/10.31033/abjar.2.1.5

Analysis of Environmental Context using Multivariate Techniques

Rohit Sikdar
Guest Faculty, Department of Chemistry, National Institute of Technology (NIT) Durgapur, West Bengal, India

Corresponding Author: rohitsikdar22@gmail.com

Received: 29-11-2022 Revised: 19-01-2023 Accepted: 30-01-2023

ABSTRACT
Under ideal analysis settings, the measurement of emission from laser-induced plasma offers a unique capacity for quantifying the major and minor elements present in any kind of sample. Chemometric techniques are extremely efficient and trustworthy tools for determining the quantities of several components in complicated matrices. For the investigation of environmental reference materials, the viability of laser-induced breakdown spectroscopy (LIBS) in conjunction with multivariate analysis (RMs) was examined. Several (Certified/Standard) References Materials of plant and soil origin were examined using LIBS in the current work, and the presence of Al, Ca, Mg, Fe, K, Mn, and Si was detected in the LIBS spectra of these materials. Using the LIBS spectral data, partial least square regression and partial least square discriminant analysis were used as multivariate statistical techniques for the quantitative study of the constituent elements. To verify the accuracy of the calibration models, the concentrations of the various components in test samples were predicted using the calibration models, and the predicted concentrations were then compared to the certified concentrations. The results of two RMs by LIBS were also compared using the non-destructive analytical technique known as Instrumental Neutron Activation Analysis (INAA), which makes use of high-flux reactor neutrons and high-resolution gamma-ray spectroscopy.

Keywords: techniques, environmental, spectroscopy, microplasma, chemometric

I. INTRODUCTION

To validate an analytical method, certified/standard reference materials (CRMs/SRMs) analysis is necessary. CRMs can be used to assess the technique's and methodology's correctness, boosting confidence when the same methodology is applied to unidentified samples. Since CRMs for many matrices in numerous laboratories are not readily available, it is vital to use appropriate procedures to assess the accuracy of analytical methods. The analytical approaches are crucial for identifying crucial components in a range of materials that are crucial for a direct investigation of materials in a variety of circumstances. For the compositional analysis, a variety of analytical methods have been thoroughly investigated, including Inductively Coupled Plasma Emission Spectrometry (ICPAES), Inductively Coupled Plasma Mass Spectrometry (ICP-MS), X-ray Fluorescence (XRF), Instrumental Neutron Activation Analysis (INAA), etc. These procedures function excellently analytically, but the most of them are destructive because a sizable portion of the samples are used. Additionally, the procedures for sample preparation take a lot of time. On the other hand, nondestructive nuclear analytical methods, such as INAA and IBA, are capable of determining many elements. However, in order to do routine material analysis, INAA requires a high flux research reactor, while IBA requires a tandem particle accelerator.

A multielement analytical method based on emission spectroscopy called laser induced breakdown spectroscopy (LIBS) is used to collect data on multiple elements at once. Compared to more traditional methods that call for pretreatment, this technique may have some advantages. The emission spectroscopy method requires small, minimally intrusive experimental equipment and does not require sample pretreatment. Through the use of appropriate experimental setups, it can also deliver remote in-situ analysis in hostile contexts or on inaccessible targets.

This technique involves focusing a strong laser pulse on the sample surface, which causes a plasma plume to form, dissociating all molecules and tiny particles inside the extremely intense microplasma. A complete elemental fingerprint of the material under analysis can be obtained from the LIBS spectra and the associated plasma emission, which can be connected to the elemental content of the sample surface. You can evaluate the method's accuracy by analyzing proper matrix matching certified reference materials. To do a quantitative analysis, standards that are similar to the matrix of the unknown (test) sample must be prepared, and a calibration curve for each element must be obtained.

While the spectral information in LIBS spectra often consists of hundreds of data points that may be gathered in less than a second, they are frequently quite complicated and include considerable information. It is never possible to compare the analytical results for elements using every potential emission line for that element in such a chemically complicated matrix for complex spectra, such as those of soils and plants. So, a good plan is needed to deal with these problems and make sure that valuable spectral data doesn't get lost.

A well-known chemometric analytical technique called multivariate analysis (MVA) makes full use of the LIBS spectral data while accounting for all potential variables and eliminating redundant and associated ones. Numerous studies have used MVA on LIBS to take advantage of the wealth of spectra data from the sample's elemental compositions. To adjust the data points for various deviations, multivariate approaches such as partial least square regression (PLSR) and partial least square discriminant analysis (PLSDA) are helpful. These techniques can be used to forecast the concentrations of various elements in unidentified samples and are linked with LIBS to generate the calibration models. LIBS are more promising for the qualitative and quantitative analysis of different elements in a variety of samples as a result of the creation and optimization of numerous strong statistical analytical methods. In this study, the chemometric analysis was carried out using PLSR and PLSDA, and LIBS spectra of environmental samples (plant and soil RMs) were produced. When using the higher neutron flux irradiation position of the Dhruva Research reactor, the results of IAEA RM SL-1 and NIST SRM 1632a were compared with the results obtained from INAA. These techniques' estimated concentrations for test samples are compared to their certified values.

II. RESOURCES AND TECHNIQUES

2.1. Setup of the Experiment

To record the LIBS spectra, pellets of each sample and RM were created using a hydraulic press machine (H-Br Press MODEL M-15).One gram of each sample was added to a pellet die (20 mm in diameter and 3 mm high) to create the pellets, which were then compressed under six tons of pressure for one minute. With a maximum deliverable laser energy of 425 mJ per pulse, a frequency-doubled Q-switched pulsed Nd:YAG (continuum Surelite III-10) laser source was employed. The pulse width FWHM (full width at half maximum) was 4 ns (variable repetition rate 1 e-10 Hz). The surface of the sample was targeted by a 532 nm laser beam that was focused using a converging lens (f/14, 15 cm). The laser pulse energy and pulse repetition rate were tuned, and the best signal-to-background ratio was found at an energy of 20 mJ at a repetition rate of 4 Hz (measured with an energy meter, Genetec-e model UP19K-30 H-VM-DO). Calculated by [D 14 4LF/pd], where l is the wavelength, f is the lens's focal length, and d is the aperture, or unfocused (initial) beam diameter, the focal point for a laser beam with a diameter of 9 mm is around 11 mm.

The power density of the laser beam flux is 5.24 1012 Wcm2 (laser pulse with energy of 20 mJ and pulse width of 4 ns).First, a lens (diameter 5 mm, f/2) fixed to the end of an optical fiber bundle was used to collect the emission from plasma. This lens was positioned at around 45 degrees to the laser beam to best collect the emission signal. The other end of the optical fiber was at the Czernye-Turner (CeZ) spectrometer's entry slit (Ocean Optics LIBS 2000; fixed gate delay: 1.5 ms). A spectrometer consists of four parts. The fourth module, which covers the spectral range 200–900 nm, has a low resolution of 0.75 nm, while the first three modules, covering the 200–510 nm range, have a high resolution of 0.1 nm (FWHM). The spectra were obtained using a gated charge-coupled device (CCD) paired with a spectrometer that has 14,336 pixels as a detector.

All of the samples in the experiment were analyzed using the optimized experimental parameters. Three soil samples and two plant samples, totaling five environmental samples, have been examined in this work. To improve the signal-to-noise ratio, an average of 20 laser shots was recorded. To eliminate statistical inaccuracy caused by laser shot-to-shot fluctuations and to take into account sample heterogeneity, seven spectra of each sample are recorded. The resulting spectra were examined with the aid of the program OOI LIBS 2000.

In addition to a reference standard of approximately 10 mg, powder samples (RMs) were sealed in polythene pouches and exposed to radiation for one minute in the Dhruva reactor's pneumatic carrier facility (PCF), located at BARC's Trombay facility in Mumbai. This radiation made it easier to extract samples of Al, Ca, K, and Mn. To determine Fe concentration, a different set of samples weighing 100 mg each were exposed to radiation for one day in the tray rod facility of the Dhruva reactor. Using a 30% relative efficiency HPGe detector, samples were checked for gamma activity after being exposed to radiation. Peak-fit software called PHAST was used to identify the peak areas, and the standard relative technique was used to calculate concentrations.

2.2 Statistical Analysis

For statistical comparison, the LIBS spectra of diverse materials were organized into a matrix with multiple variables (spectral emission lines corresponding to different wavelengths). This LIBS spectral data matrix was utilized to run the multivariate procedures using the Unscrambler-X program (CAMO Software India Pvt. Ltd.). Data sets were then separated into two matrices, one for plant samples (14 5855) and the other for soil samples (21 5855), as we have two different types of samples (plants and soils). For both matrices, multivariate approaches are utilized. The most popular multivariate methodologies for data analysis are PLSDA and PLSR. These methods are based on the partial least squares (PLS) method, which is frequently used for large-scale data analysis. PLS is used to lower the calibration model's residuals in order to efficiently employ the spectral data, decrease the risk of overfitting, and ultimately increase the model's accuracy. In its simplest form, PLS is a method for simulating a linear connection between variables used as inputs and outputs. The primary drawback of using the PLS technique to analyze spectral data is that it ignores any underlying physical concepts in favor of concentrating solely on the mathematical connections in the data. The data decomposition technique known as PLSR uses a mathematical process to create latent variables (factors). A large number of correlated independent variables are orthogonally transformed into a smaller number of uncorrelated independent variables using latent variables. PLSR can understand the correlations between many variables and identify hidden trends. When independent variables include shared information, such as correlations, PLSR works particularly well since it linearly links the fluctuations of the dependent variables to the independent variables. It makes it easier to understand how independent variables and dependent variables are related. A set of unknown samples are then used to validate the PLSR model's performance. The PLSR calibration model is used to generate the PLSDA model in this case. Classes of unknown samples can be predicted using this technique. It establishes the ideal variance for each class.

Comparisons of the PLSDA model's effectiveness are made using validation test sets. This work's major goal is to build reliable calibration models that relate the concentration of various elements in a range of samples and to use those models to forecast the concentrations of those elements in unidentified samples. The use of MVA in extracting and evaluating the spectrum information will unquestionably improve the quantitative analytical capability of LIBS, making it more promising.

III. RESULTS AND ANALYSIS

The sample descriptions used by LIBS for the multivariate analysis are displayed in Table 1. The typical LIBS spectrum of a cabbage leaf CRM is shown in Fig. 1 and spans the spectral range of 200 nm to 500 nm. This spectrum clearly demonstrates the presence of Mg, Ca, Fe, K, Al, Si, and other notable lines. All other examined standard RMs have similar Mg, Ca, Fe, K, Al, Mn, etc. spectral lines in their LIBS spectra. Using the NIST database for atomic spectroscopy and chemical spectroscopy developed by W. R. Brode, the wavelengths of various atomic and ionic species found in the spectra were determined. Tables 2 show the certified values of the concentrations of different elements in both plant and soil RMs.

In this case, the calibration model for RMs is built using PLSR. These are the predicted vs. reference charts for the PLSR model. The typical wavelength regions that comprise the majority of the emission lines of the relevant elements are selected for each sample in order to develop the calibration models of various elements with variable concentrations. Typical Al PLSR calibration models for plant and soil samples are shown in Figs. 2(a) and (b). All other elements are plotted using PLSR models in a similar manner. The model's performance is

Table 1: List of the various analysis-related references.

S. No.	RM code	Matrix	Sample code
1	NCSZC73012	Cabbage leaf	P1
2	NCSZC73014	Tea Leaf	P2
3	SRM 2704	River Sediment	S1
4	SRM 1632	a Bituminous Coal	S2
5	RM SL-1	Buffalo River Sediment	S3

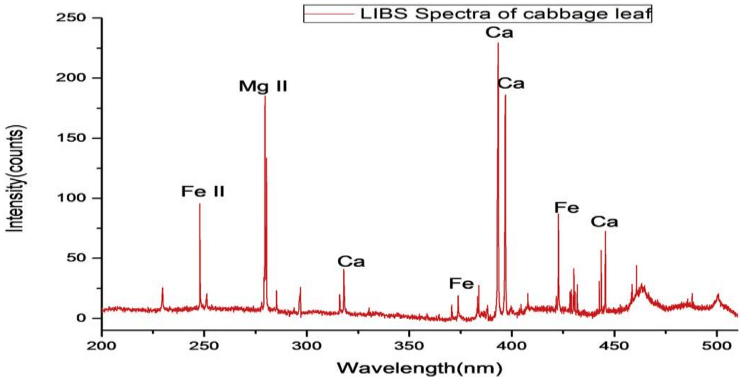

Figure 1: Cabbage leaf LIBS spectra in the 200- to 500-nm range

Table 2: Certified concentrations (weight percentages) of several elements for soil CRMs with anticipated concentrations of unidentified soil samples

Element	SRM 2704 (S1)		SRM 1632a (S2)		IAEA-RM-SL-1 (S3)	
	Certified	**Predicted**	**Certified***	**Predicted**	**Certified****	**Predicted**
Al	6.11 ± 0.16	6.14 ± 0.26	(3.1)	2.9 ± 0.20	(8.9)	9.02 ± 0.50
Ca	2.60 ± 0.03	2.36 ± 0.16	0.23 ± 0.03	0.29 ± 0.11	(0.25)	0.26 ± 0.01
Fe	4.11 ± 0.10	3.74 ± 0.39	1.11 ± 0.02	1.45 ± 0.40	6.74 ± 0.20	6.75 ± 0.75
K	2.00 ± 0.04	1.93 ± 0.15	0.42 ± 0.02	0.33 ± 0.18	(1.5)	1.87 ± 0.30
Mg	1.20 ± 0.02	1.10 ± 0.11	(0.1)	0.12 ± 0.02	(2.9)	2.82 ± 0.22
Mn	0.0555 ± 0.0019	0.0307 ± 0.0127	0.0028 ± 0.0002	0.0029 ± 0.0001	0.346 ± 0.017	0.339 ± 0.031

By using the coefficient of determination and the root mean square error to measure (R2). The R2 values are seen to be virtually equal to 1, while the RMSE is seen to be very close to zero, indicating a significant correlation between the predictions and references. The reference values and the anticipated values have to match perfectly. This model can be considered adequate for use in the regression if the slopes tend to 1. Seven spectra from each of the two plant samples make up the 14 spectra we recorded. Ten of the 14 spectra are selected as the training set for the PLSR model, while the remaining four are selected as the test set. Three soil samples totaling 21 spectra (7 spectra for each sample) are acquired. Out of the total of 21 spectra, 15 are selected as the training (known) set, and the remaining 6 are the test (unknown) set. The PLS model is validated using the cross-validation technique. Test sets provide information about the model's strengths and weaknesses. To build the PLSR model, we used the concentrations of various elements listed in Tables 1 and 2. Furthermore, the close agreement between the calibration's (Cal) and validation's (Val) best fits (blue and red, respectively) demonstrates that the R2 for calibration (Cal) is near to the validation's (Val). The model cannot be believed if Cal and Val have significant differences. For all elements in both matrices, the model's root-mean-square error of calibration and prediction (RMSEC&P) is near 0. The PLSR calibration model is assessed using RMSEC for the model, and test sets are forecasted using these calibration models. These PLSR calibration models' predictive abilities are checked using RMSEP. If the matrix of the measured samples does not deviate from that of the calibration sample set, we can assess the correctness of the determination based on the predictions from a PLS model. The RMSE is low while R2 remains high, demonstrating the proposed model's general resilience and indicating that the PLSR model is more accurate and reliable for all samples (Fig. 2(a) and (b). This method can fix the problems with traditional internal standard calibration techniques and compensate for matrix effects.

The three-dimensional scatter plot of the three specified components for both types of samples is displayed in Fig. 3 The closer the samples are to one another on the score plot, the more similar they are, giving us a map of the samples. In the first matrix, samples are separated into two groups, and in the second matrix, they are divided into three clusters. Although they are grouped independently, it is clear that the majority of the elements are the same across all samples in both matrices. Interestingly, this demonstrates the

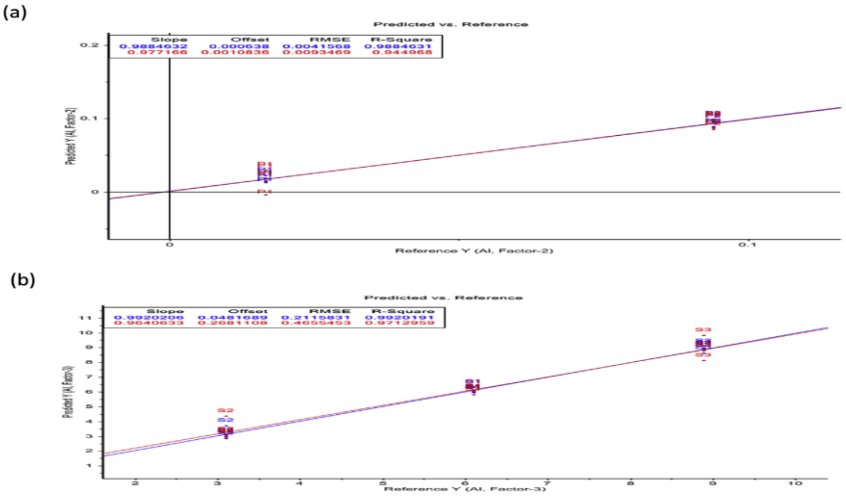

Figure 2: (a & b). Al concentration prediction vs. reference plots for plant and soil RMs

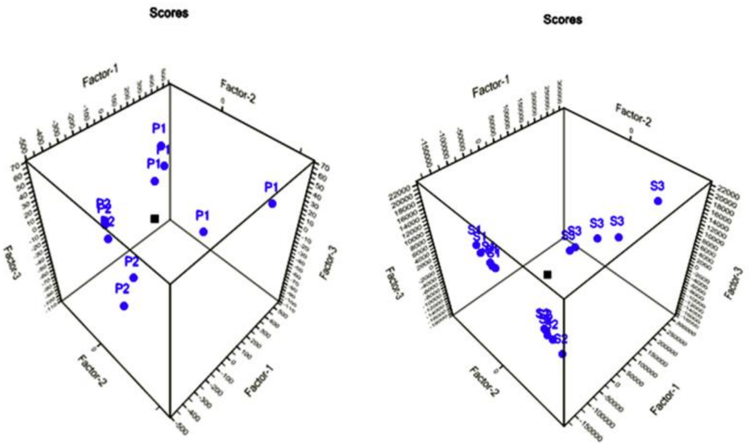

Figure 3: Plant and soil CRMs assess plots in three dimensions.

As can be observed from the concentration information in Tables 1 and 2 of samples, there is a difference in the concentration of the elements contained in the samples. By computing the factors, these graphs (Fig. 3) provide information on trends in samples.

The peaks in Fig. 4 are utilized to create the PLS calibration models and have a significant impact on the variance of Mg concentrations. The regression coefficients show a direct linear link between the elemental concentrations and the LIBS spectra. The peaks that have been allocated here display the emission lines that are pertinent to variations in elemental concentrations. A restricted spectral range is chosen to generate the PLS models, where the most intense peaks of a given element are exhibited based on the regression coefficient plot. This reduces the interference of irrelevant emission lines and makes the PLS models practically feasible.

The relationship between the cumulative explained variance and the number of components is seen in Fig. 5. This variance curve shows the evolution of the explained variance of a certain variable.

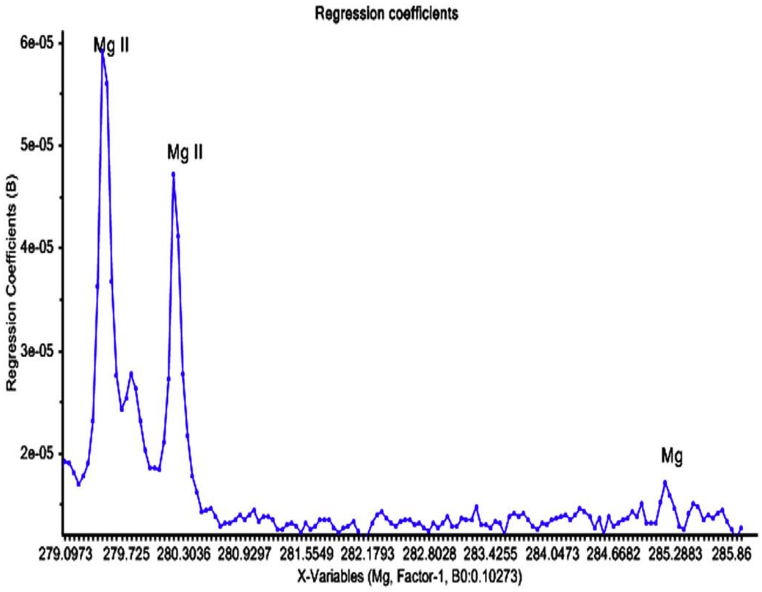

Figure 4: Shows a plot of the regression coefficient versus wavelength for the Mg element in soil samples.

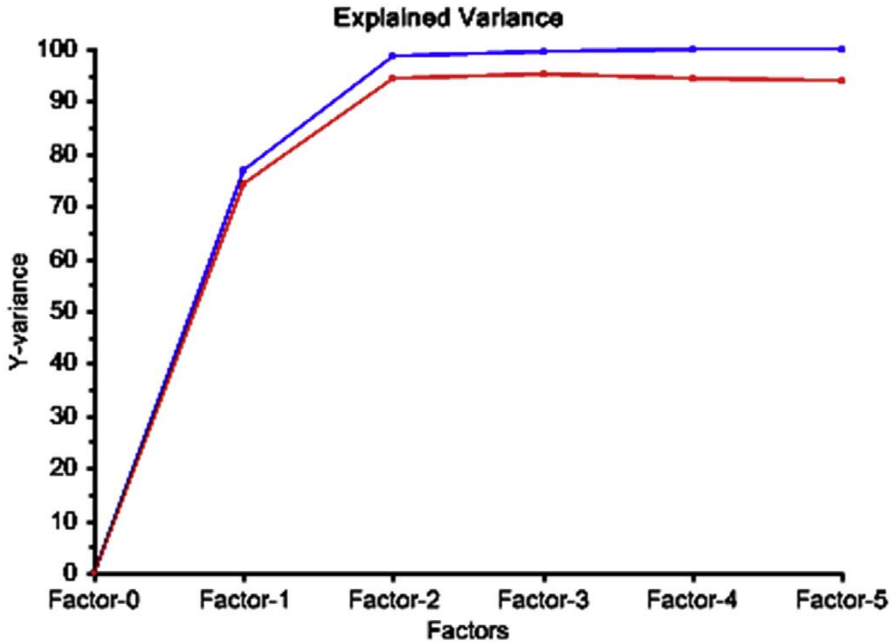

Figure 5: Shows the PLSR model's explained variance plot.

With the model's multiplicity of factors. According to this graph, only two variables are required to produce a calibration model with the highest level of predictability. From this figure, we can determine the variation in responses that each component requires to produce a calibration model with the highest level of predictability. From this figure, we can determine the variation in the responses that each component describes. The calibration variance is shown by the blue line, while the validation variance is represented by the red line. The basis for calculating calibration variance is fitting the calibration data to the model that was created from the calibration data. The model is put to the test on the data to calculate the validation variance. The model does not adequately represent fresh data if the validation variance differs from or is much smaller than the calibration variance. On the other hand, the model is representative if these curves are close together. The plot shows that the calibration variance and validation variance are very similar.

As a result of the PLSR model's reliability, it is used to forecast the concentrations of test samples. Figures 6 (a) and (b) show the Al test set results for plant and soil samples, respectively. These models are also plotted similarly for all other components. The expected concentrations for all test (unknown) samples are displayed in these figures. Here, test sets are utilized to calculate the concentrations of various elements before evaluating the efficacy of the calibration model. Horizontal lines represent the anticipated values, while the boxes surrounding the values represent departures from

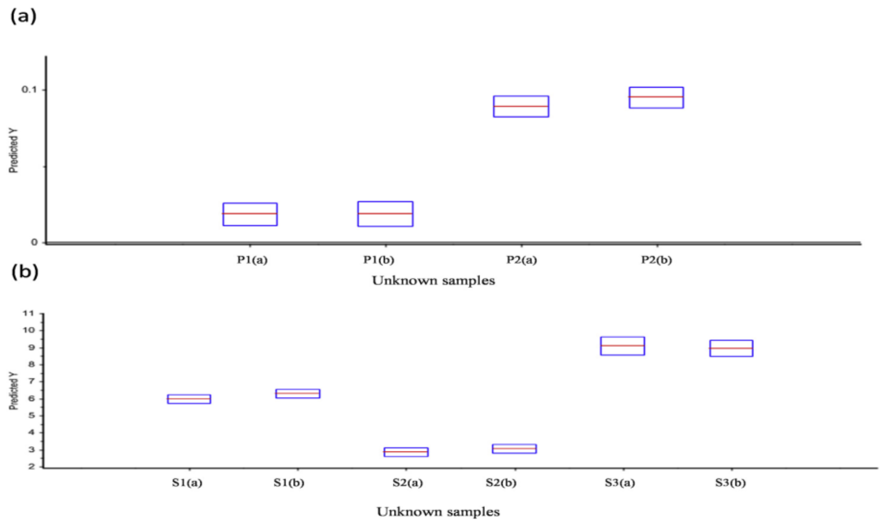

Figure 6: (a & b), PLSDA models for the Al test set for samples of soil and plants

the data training set is depicted in Figures 6(a) and (b).Boxes show the high variance if the calibration model's training sets and the test sets on which the predictions are applied are not very similar. The estimated concentrations of all elements with RSD for plant and soil samples are displayed in Tables 2 and 3, respectively. Here, it should be noted that, with only minor differences in values, the anticipated concentration values of several elements for unidentified samples are almost identical to the certified concentrations shown in Tables 2 and 3. The fact that unknown samples are remarkably similar to known samples demonstrates the proposed model's robustness for unknown samples.

The range of 3.0–9.5% applies to the data from four separate samples that INAA evaluated. As can be observed, the INAA results for the IAEA RM SL-1 and NIST SRM 1632a are in good agreement with the certified values as well as with respect to each other (within 10% except for a few elements of 1632a) (within 3.3% for SL-1 and within 10% for 1632a). A high-flux neutron source, such as a nuclear reactor, is required for the experiments since INAA is capable of delivering data on numerous elements with widely different concentration ranges simultaneously. LIBS is useful because it lets scientists get information without damaging the object being studied.

IV. CONCLUSION

The potential of LIBS to analyze RMs of environmental samples quickly using multivariate statistical analysis of data is examined in this work. Authenticated reference materials were used to construct the model for the multivariate analysis. The created model was then used on unidentified samples. The findings unequivocally show that PLSR and PLSDA are effective tools for applying multivariate techniques to the analysis of LIBS spectral data. Here, calibration models for relating the concentrations of various elements are constructed. The LIBS-predicted concentrations for the RMs are similar to the certified concentrations. The usual calibration curve method makes it very difficult to determine the element concentrations in a range of samples. This methodology offers a workable solution. Even in the absence of RMs for the same matrix, the methodology is useful for assessing how resilient the strategy is. These strategies produce better outcomes because multivariate methods have robust and potent analytical capabilities.

REFERENCES

1. N.S. Rajurkar, & M.M. Damame. (1997). Elemental analysis of some herbal plants using in the treatment of cardiovascular diseases by NAA and AAS. *J. Radio Anal. Nucl. Chem.*, 77e80.
2. C.G. Ryan. (2000). Quantitative trace element imaging using PIXE and the nuclear microscope. *Int. J. Imaging Syst. Technol. 11*, 219e230.
3. Jie Wang, Tetsuya Nakazato, Kinya Sakanishi, Osamu Yamada, Hiroaki Tao, & Ikuo Saito. (2004). Microwave digestion with HNO3/H2O2 mixture at high temperatures for determination of trace elements in coal by ICP-OES and ICP-MS. *Anal. Chim. Acta*, *514*(2004), 115e124.
4. N. Civici, Sh. Gjongecaj, F. Stamati, T. Dilo, E. Pavlidou, E.K. Polychroniadis, & Z. Smit. (2007). Compositional study of IIIrd century BC silver coins from Kreshpan hoard (Albania) using EDXRF spectrometry. *Nucl. Instrum. Methods Phys. Res., B 258*, 414e420.
5. Sarah C. Jantzi, & Jose R. Almirall. (2011). Characterization and forensic analysis of soil samples using laser-induced breakdown spectroscopy (LIBS). *Anal. Bioanal. Chem., 400*, 3341e3351.
6. Gulab Singh Maurya, Aradhana Jyotsana, Rohit Kumar, Ajai Kumar, & A.K. Rai. (2014). Analysis of deposited impurity material on the surface of the optical window of the Tokamak using LIBS. *Phys. Scr., 89*, 75601.
7. Rohit Kumar, Awadhesh K. Rai, Devanathan Alamelu, & Suresh K. Aggarwal. (2013). Monitoring of toxic elements present in sludge of industrial waste using CFLIBS. *Environ. Monit. Assess., 185*, 171e180.
8. Gabriel Gustinelli Arantes de Carvalho, Javier Moros, Dario Santos Jr., Francisco Jose Krug, & J. Javier Laserna. (2015). Direct determination of the nutrient profile in plant materials by femtosecond laser-induced breakdown spectroscopy. *Anal. Chim. Acta, 876*, 26e38.
9. Jennifer L. Gottfried, Russell S. Harmon Jr., Frank C. De Lucia, & Andrzej W. Miziolek.(2009). Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification. *Spectrochim Acta Part B, 64*, 1009e1019.
10. K. Ayyalasomayajula, Vivek Dikshit, Fang Yu Yueh, Jagdish P. Singh, & Laura T. Smith. (2011). Quantitative analysis of slurry sample by laser-induced breakdown spectroscopy. *Anal. Bioanal. Chem., 400*, 3315e3322.
11. Jie Feng, Zhe Wang, Logan West, Zheng Li, & Weidou Ni. (2011). A PLS model based on dominant factor for coal analysis using laser-induced breakdown spectroscopy. *Anal. Bioanal. Chem., 400*, 3261e3271.