



Application of laser-induced breakdown spectroscopy (LIBS) coupled with PCA for rapid classification of soil samples in geothermal areas

Sitangshu Chatterjee^{1,2} · Manjeet Singh^{2,3} · Bishnu Prasad Biswal⁴ · Uday Kumar Sinha¹ · Suraj Patbhaje⁴ · Arnab Sarkar^{2,3}

Received: 13 November 2018 / Revised: 21 February 2019 / Accepted: 27 February 2019 / Published online: 16 March 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The Manuguru geothermal area, located in the Telangana state, is one of the least explored geothermal fields in India. In this study, characterization of the soil samples is carried out by laser-induced breakdown spectroscopy (LIBS) coupled with analytical spectral-dependent principal component analysis. A total of 20 soil samples were collected both from near the thermal discharges as well as away from the thermal manifestations. LIBS spectra were recorded for all the collected soil samples and principal component analysis (PCA) was applied to easily identify the emission lines majorly responsible for variety classification of the soil samples. In this submission, a modified PCA was developed which is based on the spectral truncation method to reduce the huge number of spectral data obtained from LIBS. The PCA bi-plot on the LIBS data reveals the presence of two different clusters. One cluster represents the soil samples collected from the close vicinity of the thermal manifestations whereas the other cluster contains the soil samples collected away from the thermal sprouts. PCA performed on the chemical dataset of the soil samples also reveals the same clustering of the soil samples. Both LIBS and chemical analysis data shows that soil samples near the thermal waters are found to be enriched in B, Sr, Cs, Rb, Fe, Co, Al, Si, Ti, Ru, Mn, Mg, Cu, and Eu concentrations compared to the soil samples located away from thermal manifestations. This study demonstrates the potential use of LIBS coupled with PCA as a tool for variety discrimination of soil samples in a geothermal area. LIBS is shown to be a viable real-time elemental characterization technology for these samples, avoiding the rigorous dissolution required by other analytical techniques.

Keywords LIBS · PCA · Soil samples · Geothermal area · India

Introduction

Water and gas sampling of the natural thermal discharges are the most commonly used techniques for exploring the prospective geothermal sources. On the other hand, soil characterization is probably the least adopted survey method employed over

the geothermal areas. However, in the geothermal areas where surface discharges are relatively few and the extent of the geothermal field is not known, soil sampling has been proven to be a very effective tool [25]. The elevated temperature of the geothermal reservoirs increases the mobility of volatile vapor-borne species, e.g., mercury, arsenic, antimony, boron, and ammonia; as a result, these elements migrate upward through the permeable zone and get adsorbed in the upper soil matrix. In fact, the presence of an anomalous concentration of Hg and its mineralization in various high-temperature geothermal areas around the world are well documented and is an effective tool for exploring potential geothermal fields [12, 22, 26, 27, 39]. Although there are abundant literatures available about mapping of Hg in high-temperature geothermal areas, there is no study being carried out about the overall classification of soils near the thermal manifestations and its comparison with the soils away from the thermal discharges.

Analytical techniques like atomic absorption spectroscopy (AAS), X-ray fluorescence spectroscopy (XRFS), inductively

✉ Sitangshu Chatterjee
sitangshujuchem@gmail.com

✉ Arnab Sarkar
amab@barc.gov.in

¹ Isotope & Radiation Application Division, Bhabha Atomic Research Centre, Mumbai 400085, India

² Homi Bhabha National Institute, Anushaktinagar, Mumbai 400094, India

³ Fuel Chemistry Division, Bhabha Atomic Research Centre, Mumbai 400085, India

⁴ Geological Survey of India, Central Region, Nagpur 44006, India

coupled plasma atomic emission spectroscopy (ICP-AES), and gravimetric analysis are often used for qualitative and quantitative variety discrimination of soil [40]. However, these techniques are costly and time consuming and cannot be applied as an on-field technique. Laser-induced breakdown spectroscopy (LIBS) is an elemental emission-based spectroscopic technique that has inherent capability for both qualitative and quantitative study of materials in solids, liquids or gaseous samples and can be used in-situ / on-field and also if required remotely. In LIBS, a short laser pulse of high energy is focussed on the sample surface to produce a micro-plasma. The resulting emission lines from the atomic, ionic, and molecular fragments, created by the plasma, is then resolved both optically and spectrally to produce a spectrum of intensity against wavelength (Harmon, 2009). LIBS is being considered as a front runner in green chemical analysis due to its unique features, such as, real-time analysis, pseudo non-destructive technique, minimal to no sample preparation protocol, high sensitivity to low atomic weight elements, and capability to carry out close-in as well as stand-off detection. As a consequence, over the last two decades, LIBS has been widely applied in a variety of fields, like environmental monitoring [1, 2], biomedical applications [11, 19], archeological investigations [4], pharmaceutical industries [23, 24], extra-terrestrial explorations [6, 7, 20, 28, 29], hazardous materials identification [9, 10, 14], nuclear fuel characterization [17, 30, 32, 33], and geological material characterization [3, 5, 15, 16].

The application of multivariate chemometric methods in conjunction with LIBS data has recently shown tremendous potential in the field of soil analysis. Principal component analysis (PCA) is one of the most widely used chemometric procedure for multivariate data system. Chemometrics basically reduce the dimension of the input data to describe the complete information with considerably fewer variables than was originally present, thereby revealing the simple underlying structure that is present within a complex input dataset [8]. A chemometrics-LIBS couple can be utilized both for qualitative and quantitative assessment of samples. Qualitative applications include a study by Zhang et al. [42] for classification of slag samples using partial least squares discriminant analysis (PLS-DA). Fink et al. [13] and Unnikrishnan et al. [38] used PCA for identification of polymers and related materials. Yueh et al. [41] used hierarchical cluster analysis (HCA) for tissue classification, Sirven et al. [35] studied feasibility of rock identification in the Mars surface by applying PCA, PLS-DA, and soft independent modeling of class analogy (SIMCA) to the LIBS data. However, there are relatively few numbers of literature studies available on the classification of soil samples using PCA-LIBS coupled methods [34, 40].

In the present work, we have employed PCA methodology on the LIBS spectrum data to classify the soil samples collected from the Manuguru geothermal area of the Telangana state,

India. The soil samples consist of several samples collected close to thermal discharges and these were compared with the samples taken away from the discharge areas to observe any distinguishable characteristics of any specific mineral (major, trace, and rare earth elements) concentrations. This work addressed the use of PCA in an indigenously developed modified form to identify emission lines significantly responsible for variety classification in LIBS spectral analysis. The developed method employed a relatively simple spectral truncation-based PCA technique. Several PCA score classification plots representing different sets of wavelength ranges were evaluated. These plots were based on truncated spectra obtained by applying a threshold radius on the corresponding PCA loading plot. Traditional PCA has been applied to the chemical data set to further corroborate the results obtained from the LIBS data.

General geology of the study area

The soil samples were collected from the Manuguru geothermal area, located in the Khammam district of the Telangana state, India. The district forms a part of the Godavari river basin. The study area consists of several thermal water manifestations having a temperature in the range of 36–76 °C. The Godavari basin, a NNW–SSE trending graben on a Precambrian platform, is filled with Gondwana sedimentary formations. This area is known for coal exploration and there are several opencast coal mines operated by Singareni Collieries Company Ltd. (SCCL). Almost all of the thermal manifestations are from the bore wells drilled for coal exploration. These thermal discharges are located near the Pagaderu, Gollakatur, ST colony, Shantinagar, and Kodichenkuntala villages of the Manuguru administrative division. Geological mapping in 1:25000 scales was carried out over an area of 80 km² in and around the study area. The geological map along with the sample location points are shown in Fig. 1. Lower Gondwana sedimentary formations of the Permian period rest unconformably over Proterozoic Pakhal meta-sediments, which form the basement rock of the study area. The Talchir formation, the Barakar formation, the Barren Measures formation, and the Kamthi formation collectively form the lower Gondwana subgroup. The simple lithostratigraphy of the study area is given in Table 1. Among different structures, well-defined bedding planes and cross beddings are observed in the sandstone. There are two sets of joints observed in the sandstone. One strikes N30°W with a dip towards the east and the other E–W with a dip towards the south area. Apart from the joints, faults are the main diastrophic structure which controls the hot water movements within the basin. The fault system is characterized by a prominence of dip faults whereas some are oblique in the study area. Some faults extend beyond the limit of Gondwana basin

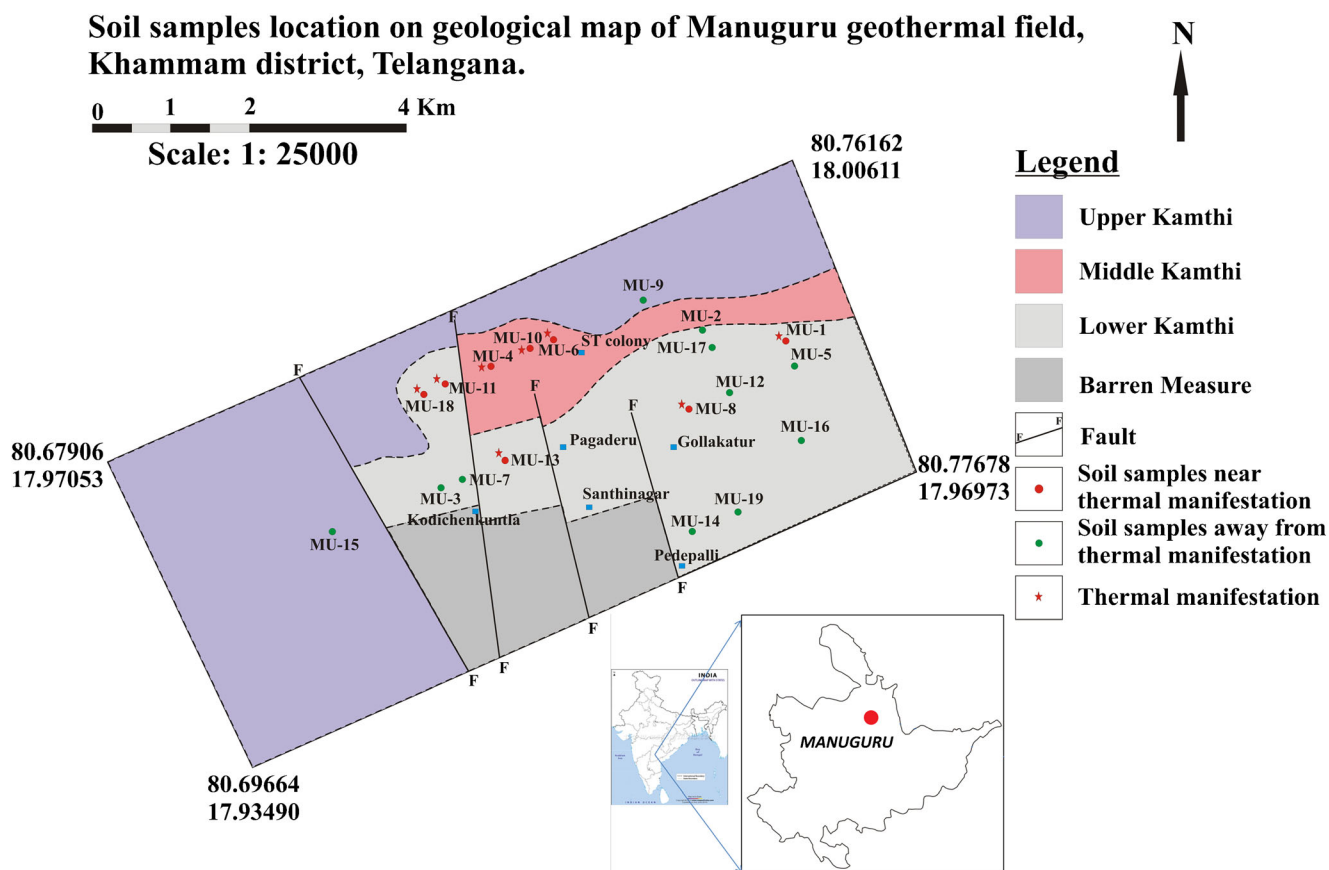


Fig. 1 Soil sample location points along with the geological map in the Manuguru geothermal field. The red circle indicates the location of thermal manifestations

into Pakhal Group, whereas others are restricted within the basin. The geothermal manifestations seem to be confined to the NE–SW trending fault. A total of 20 samples were collected from the B horizon (subsurface layer, 9–15-cm depth) of the soil.

Methods

To ensure the homogenous nature of the soil composition, the samples were dried, grounded, and sieved to a particle size of 80 mesh. The samples were divided into three sets, one for LIBS spectra recording and the other two for chemical composition analysis by inductively coupled plasma optical emission spectroscopy (ICP-OES) and instrumental neutron activation analysis (INAA).

LIBS

A common configuration of the LIBS system was used in this study [33]. Figure 2 presents a schematic diagram of the experimental setup. Laser pulses from an Nd: YAG laser

(Brilliant B, Quantel, France) with a 532-nm wavelength (6-ns pulse duration) and having a maximum energy of 440 mJ were focussed on to the sample surface using a plano-convex lens ($f = 10$ cm) to produce a laser-induced micro-plasma. The samples were translated using a motorized linear XYZ translator (Velmex, USA) to ensure that each laser pulse hits a fresh portion of the sample. The emission light was collected at 45° with respect to the laser pulse propagation, through a collimator (CC52, Andor, UK). The collected light was fed to an echelle spectrometer (Mechelle, ME5000, Andor, UK) through an optical fiber. The echelle spectrometer covers 200–975-nm wavelength regions and has a spectral resolution of ~ 4750 CSR ($\lambda/\Delta\lambda$) at a 50- μm entrance slit width. The spectrograph is equipped with an ICCD (iStar, Andor, UK, 1024×1024 pixels), which is synchronized with the Q-switch of laser pulse to control acquisition time delay (t_d) and detector gate width (t_g). The digital spectra recording and controlling the delay generation were carried out with data acquisition software (Solis 4.28). For the present study, a laser energy of 50 mJ, $t_d = 1.2$ μs , and $t_g = 50$ μs was used. The soil samples were initially pelletized to 3-cm diameter pellets by applying a pressure of 2×10^9 Pa for 5 min using a pellet

Table 1 Litho-stratigraphy of the study area

Formation		Lithology	Age
Quaternary alluvium		Soil and alluvium	Quaternary to recent
Lower Gondwana	Kamthi formation	Upper Kamthi Coarse grain ferruginous sandstone with cherty siltstone and pebble bed	Up. Permian to Low. Triassic
		Middle Kamthi Alternate sequence of sandstone and clay bed	Up. Permian to Low. Triassic
		Lower Kamthi Medium to coarse-grained calcareous sandstone with few coal seams	Up. Permian to Low. Triassic
Unconformity			
	Barren measure	Medium to coarse grained felspathic sandstone with subordinate micaceous siltstone and clay	Up. Permian
	Barakar formation	Felspathic sandstone, carbonaceous shale, and coal	Low. Permian
	Talchir formation	Fine sandstone, greenish and chocolate shale, pebble bed, and tillites	Up. Carboniferous to Low. Permian
Unconformity			
	Pakhal formation (basement)	Quartzite, chert, phyllite, and marble	

presser. Each LIBS analysis consists of an accumulation of 60 laser shots taken in scanning mode over a 15-mm straight line and three replicate spectra for every individual sample were recorded. These three replicate spectrums were then averaged to get a spectrum representing the overall sample surface.

Chemical analysis

Thermal waters contain variable amounts of trace metals (i.e., Li, Rb, Cs, B, Sb, Cu, Pb, Mn, Co, etc.), rare earth elements (i.e., Sc, Ce, Eu, Tb, etc.) and transition metals (i.e., Hf, Ta, etc.) which are usually not present in detectable amounts in natural groundwater systems. High-temperature thermal waters show elevated concentrations of these trace and rare earth elements depending upon the extent of rock-water interaction at the high reservoir temperature ([18, 31] and the references cited therein). The elevated concentration of trace and rare earth elements in the thermal waters prompted us to analyze the concentrations of some of these elements in the soil samples collected from the geothermal area to check their preferential distribution.

Quantitative analysis of Fe, Na, K, Mn, Cu, B, As, Hg, Sb, Li, Pb, and Co concentrations in the soil samples were measured by the ICP-OES technique (Model: ACTIVA, M/S HORIBA Scientific). Prior to the analysis by ICP-OES, digestion of soil samples was carried out as per the methodology described by Kumar et al. [21]. Silica content of the soil samples was analyzed by the conventional gravimetric method [36]. The concentrations of Sc, Rb, Cs, Ce, Eu, Tb, Hf, Ta, and Th were determined by instrumental neutron activation analysis (INAA). For INAA analysis, 50 g of each soil sample was dried in an oven and carefully ground using mortar and agar from which 20 mg of the powdered sample was sealed

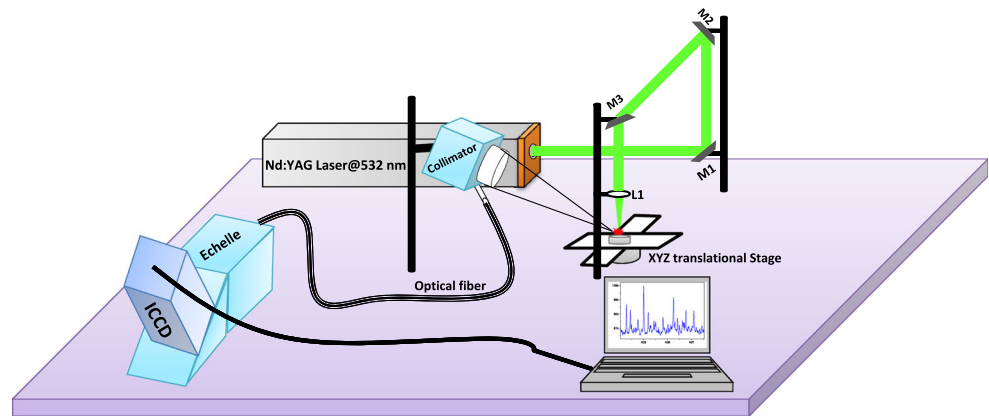
doubly in aluminum foil and irradiated in the self-serve facility of the DHRUVA reactor, Mumbai, with a neutron flux of $10^{13} \text{ cm}^{-2} \text{ s}^{-1}$ for 6 h. IAEA RMs (reference materials) SL-1 and Soil-7 were used as reference and control standards, respectively. Gamma-ray measurements were carried out after the appropriate cooling time by using an HPGe detector coupled with a computer-assisted multichannel analyzer [37].

Results and discussion

Characterization and PCA of LIBS spectra

PCA on the LIBS data was employed to identify the variations in the spectral data and to interpret the data relative to the subset of the spectral variations. On applying a linear mathematical transformation, these variations were reduced to a smaller set of principal components. The first PC (PC1) contained the largest variance of data set followed by second PC (PC2), third PC (PC3), and so on. In this current study, PCA was applied to transform the LIBS spectra of all the 20 samples into several principal components (PCs). It was seen that that PC1, PC2, and PC3 explain 73%, 18%, and 8.9% of the variance respectively; implying three PCs collectively could explain 99.9% of the total variance of the original dataset. Figure 3 b to d show the comparison of principle components (PCs) after applying the PCA on the whole spectra with a typical LIBS spectra obtained from the soil sample (Fig. 3 a). One of the drawbacks of the abovementioned PCA procedure was the frequent crashing of the software due to excessive load of data in the algorithm. The spectrograph-detector system generates > 16,000 wavelength or pixel data per spectrum covering a 200–975-nm region. Hence, applying the PCA model on this LIBS

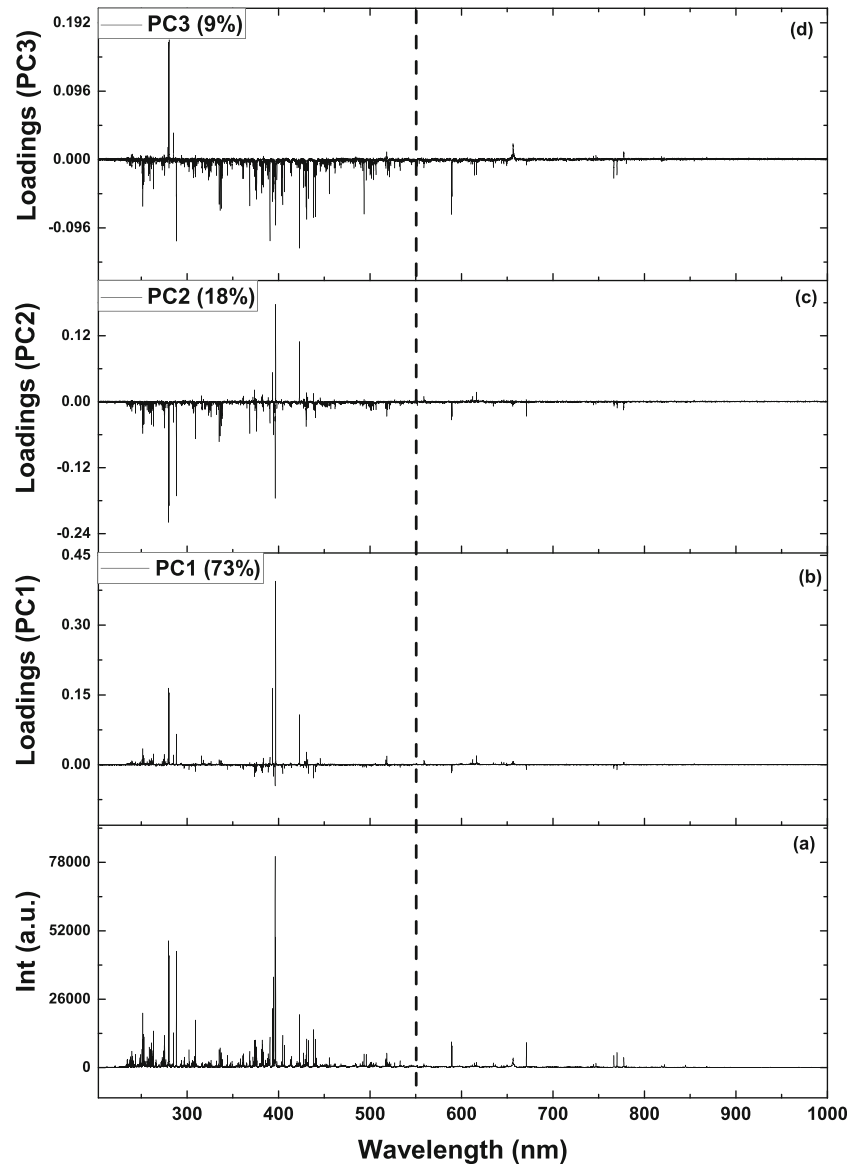
Fig. 2 The experimental setup of LIBS used for the present study



whole spectrum meant introducing $20 \text{ (sample)} \times 16000$ pixel data points, which was equal to 320,000 variables. This caused a heavy load on the program. To reduce this

load, regions with no or minimal contribution to the overall analytical results were needed to be removed. Figure 3 clearly depicts the lack of a significant amount of emission

Fig. 3 **a** A complete LIBS spectra of the soil sample showing the 232–1000-nm region. **b–d** A comparison of different PCs on the whole LIBS spectra



lines after 550 nm. The same thing was reflected from the PC1, PC2, and PC3 with a lack of a significant amount of contribution in the 550–975-nm regions, and hence, this region was irrelevant in the variety classification exercise and was omitted. Figure 4 a and b show the comparison of the PCA loading plot using whole a 232–1000-nm region and 232–550-nm region respectively and were found to be almost identical. Both Fig. 4 a and b show the presence of two different clusters. Soil samples very near to the thermal manifestations (MU-1, MU-4, MU-8, MU-10, MU-11, MU-13, MU-18) fall in one cluster (square symbols) whereas samples (MU-3, MU-12, MU-14, MU-15, MU-17, MU-20) collected away from the thermal manifestations fall in a different cluster (circle symbols). The MU-6 sample although collected near the thermal manifestations did not fall in the thermal cluster. Anomalies or wrong clustering were shown by only three samples (MU-7, MU-2, and MU-16). These three samples although collected away from the thermal manifestations but fell in the proximity of the thermal cluster in the PCA plot.

Ascertaining the reason behind the differential clustering of soil samples depending upon the proximity to the thermal waters, the PCA on the loadings of PC1 and PC2 were carried out (Fig. 5). Since the echellogram consisted of a huge number of pixels, even after eliminating 550–975 nm, the total number of pixels in the PCA model was ~ 9000 . Due to this huge number of the pixel data, it became virtually impossible to identify which emission lines were majorly influencing the clustering in Fig. 4 b. To make the model more efficient in

distinguishing major influencing variables or pixels, the irrelevant pixel input data to the program were needed to be reduced. When the LIBS spectrum is used as the input data set in a PCA model, a weakly intense emission line can be an effective variable for classification but a strong one can be irrelevant. Singh et al. [33] developed and successfully used a modified algorithm of PLSR known as analytical spectral-dependent PLSR (ASD-PLSR) for qualitative analysis of glass samples. The core idea of the algorithm was to delete the irrelevant spectral region. Based on a similar principle, in the present work, we have developed a spectral truncation method known as analytical spectral-dependent truncation, which was guided by the loading's Euclidean distance in the loading plot (Fig. 5).

The loading plot is simply the Cartesian plot of the individual scalar values of the loading vector in the XY-plane made by the respective principle components. The direction and distance of these scalar points (i.e., the representation of the corresponding variable) of the loading vector indicate the direction and magnitude of that particular variable contributing to the variety classification. In other words, the points close to the origin of the loading plot are irrelevant in variety classification. Hence, variables or pixels having a near-zero Cartesian distance from the origin can be removed. To select only those wavelength regions or pixels which are at a significant distance from the origin and significantly contributing in variety classification, an arbitrary parameter named the PCA coefficient threshold (T^c) was introduced, whose value varied from 0 to the maximum Cartesian distance in the loading plot. A

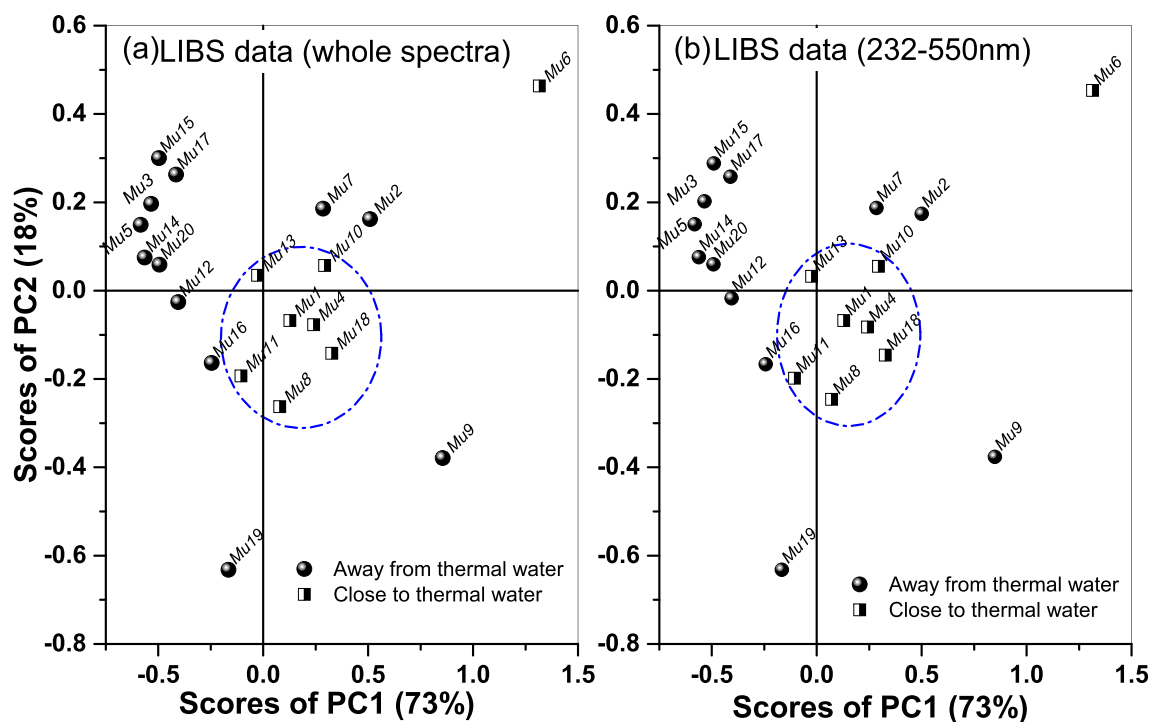


Fig. 4 a PCA bi-plot covering all spectral regions 232–1000 nm. b PCA bi-plot covering the 232–550-nm region

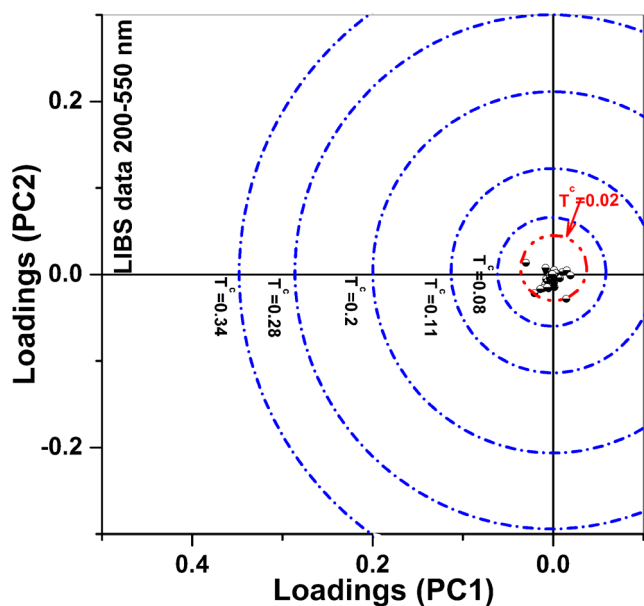


Fig. 5 PCA bi-plot of loading factors along with different T^c values

program was written in such a way that it had converted the LIBS spectrum to a modified spectrum by replacing all emission intensities to 0 whose Cartesian distance in the loading plot was less than a selected T^c whereas the remaining intensities were left unaltered. Basically, this procedure removes all the intensities of the pixels which are irrelevant for the classification purpose and helps to retain only relevant intensities in the spectra. Figure 5 shows the six arbitrarily chosen T^c circles. These circles indicate that if say $T^c = 0.08$ is chosen, then all the pixels having a loading Euclidean distance less than the value of 0.08 will be replaced by the 0 value. A T^c value equal to 0 indicates no spectral region or pixel deletion, i.e., the PCA plot of the whole spectra as it is. The T^c value was changed from 0 to 0.4 with an increment of 0.005, and for each T^c value, separate PCA plots were constructed. Figure 6 shows PCA plots at four arbitrary T^c values. With the increase in the T^c value from 0 to 0.02, 0.08, and 0.2, the change in the PCA classification was observed. At $T^c = 0$, only MU-6 sample was the outlier from the thermal soil grouping. But at $T^c = 0.02$, when some amount of irrelevant signal or intensities were removed, a very similar grouping is observed, but now, two non-thermal samples gave false positive results. At $T^c = 0.08$, this clustering became less obvious and the degree of separation between two groups of soil samples (collected from near the thermal manifestations and away from thermal manifestations) became less, indicating a significant loss of relevant information or intensities. At $T^c > 0.08$, the separate clustering of the soil samples was not at all there. These observations indicated the requirement of the variables or pixels having a T^c value ≤ 0.02 for proper variety classification, without which PCA classification would fail. Although, with the increase in the T^c value, the chance of false positive results increased, but

this methodology helped us to separate the relevant pixel from irrelevant pixels thereby enabling us to identify the emission lines majorly responsible for the variety classification.

Figure 7 shows the PC1 and PC2 loadings obtained from PCA of truncated spectral data using a T^c of 0.02. For identification purposes, the truncated spectrum of an arbitrary chosen sample (MU-11) was also shown in Fig. 7. By identifying the different scalar values in the PC1 and PC2 and comparing with truncated spectra, the variables or emission lines majorly responsible for the PCA variety classification were identified. Table 2 tabulated the list of emission lines in the LIBS spectra responsible for the variety classification of the PCA plot shown in Fig. 4 b. It was interesting to note that, Table 2 mainly contained ionic lines rather than atomic lines. The occurrence of these ionic lines was not an indication of the plasma's ionic nature; rather, the spectrally pure lines (in heavy spectrally impure emission spectra) which play a major role in variety classification were accidentally ionic in nature. The use of a spectrograph with high resolution might solve this dilemma by resolving many of these emission lines. It was observed that PC1 and PC2 had greater loadings for the emission lines of the few elements, i.e., Fe, Ca, Co, Sr, Al, Si, Ti, Ru, Mn, Mg, Cu, and Eu which caused the preferential clustering shown in the PCA plot (Fig. 4). However, a higher loading of these emission lines did not necessarily imply the higher intensities (higher intensity is proportion to the higher concentration of the elements) of the abovementioned elements in the soil samples near the thermal manifestations. Rather, the concentration of all these elements as a whole affected the emission pattern of soil samples near the thermal manifestation to be different from the other samples resulting in differential clustering in PCA. Simply, the overall distribution of the Fe, Ca, Co, Sr, Al, Si, Ti, Ru, Mn, Mg, Cu, and Eu compared to the other elements present in the soil samples near the thermal water has resulted in this alternate grouping.

PCA of the chemical data

To correlate the findings obtained from the LIBS analysis, PCA on the chemical analysis data was performed. The concentrations of the corresponding elements are given in Table 3. The PCA bi-plot (Fig. 8) using PC1 and PC2 showed the existence of two different clusters. In Fig. 8, the soil samples close to the thermal water manifestations are shown in square blocks whereas the circle symbol denotes the soil samples collected away from the thermal manifestations. The tip of the arrow in Fig. 8 indicates the loading value of the respective elemental concentration. Figure 8 basically shows the PCA plot of the sample as well as the PCA plot of the elemental concentration in a single graph. The higher loadings of boron, cesium, rubidium, and cobalt as a whole made the soil samples close to the thermal water very distinct compared to other samples. Although the separate grouping of soil samples near

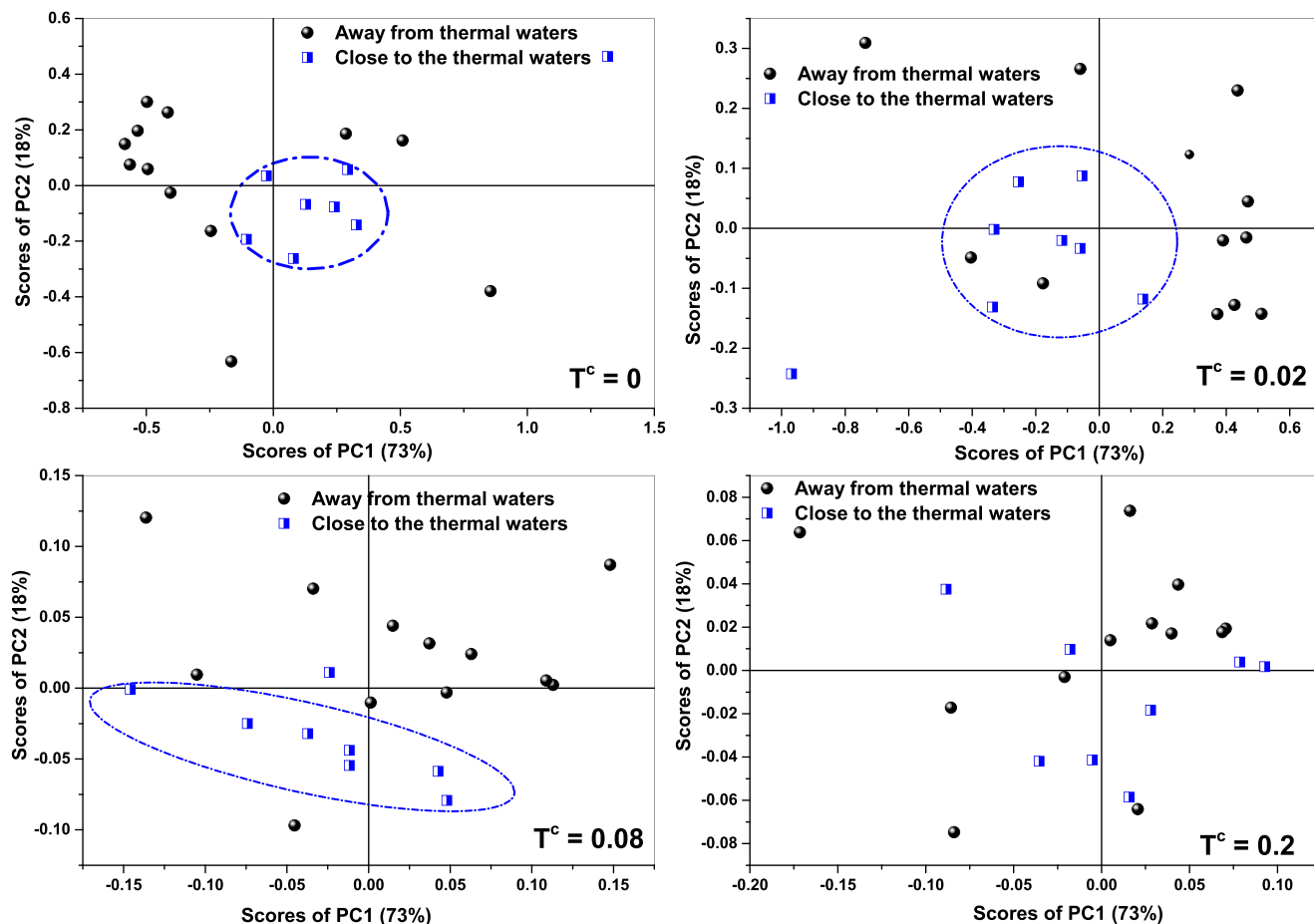


Fig. 6 PCA bi-plots of loading factors having different T^c values

the thermal manifestations was strikingly similar in both LIBS and chemical analysis methods, there were a few mismatches in the clustered samples. In Fig. 8, the MU-10 sample did not fall into the thermal cluster whereas MU-14 and MU-15 fall within the thermal cluster although they represented the soil samples away from the thermal manifestations. The MU-10 sample had nearly twice the concentration of Na and K, along with abnormally low Cu concentration than the fellow thermal soil sample. These unusual elemental patterns had forced the MU-10 sample to move away from thermal soil clustering. On the other hand, the MU-14 and MU-15 samples possessed a concentration range of few elements similar to thermal ones and few elements having a concentration range similar to non-thermal samples. This characteristic pattern had caused the PCA modeling to misinterpret the two samples as a part of the thermal cluster. A complete analysis of other traces and bulk elements will probably be able to correct this PCA misinterpretation.

It is interesting to note that the number of elements, assumed to play a significant role in the clustering process, is found to be a little bit different as per LIBS and chemical analysis data. This happens due to some intrinsic shortcomings of LIBS compared to traditional chemical analysis

technique and vice versa. In chemical analysis, the matrix is generally separated and the elements are measured sequentially, whereas in LIBS, the spectrum of the whole matrix is taken so overall composition plays a very vital role in the LIBS technique. The spectral intensities of major elements (i.e., Fe, Mg, Si, etc.) were so high compared to the spectral intensities of the trace elements that in the PCA, these small intensities were unable to play any significant role in the clustering process. Not only that, some of the elements like boron (B) which was present in a significant amount (~ 100 ppm) and was a known influencer in the chemical analysis was found to be insignificant in the LIBS analysis due to the heavy spectral interference of iron (which is present in percentage amounts). The B (I) 249.677-nm line got heavily interfered by the Fe (II) 249.913-nm emission line. Cs and Ru being alkali metals are an emission-poor system, and hence, they were not seen in the LIBS spectrum with significant intensity. In short, Cs and Ru also could not play any significant role in clustering processes. It is true that the significant elements to distinguish between soil samples from close to the thermal manifestations and samples far away from thermal manifestations cannot depend on the analytical techniques employed in the study. But the LIBS system (resolution = 0.05 nm) used in the present study

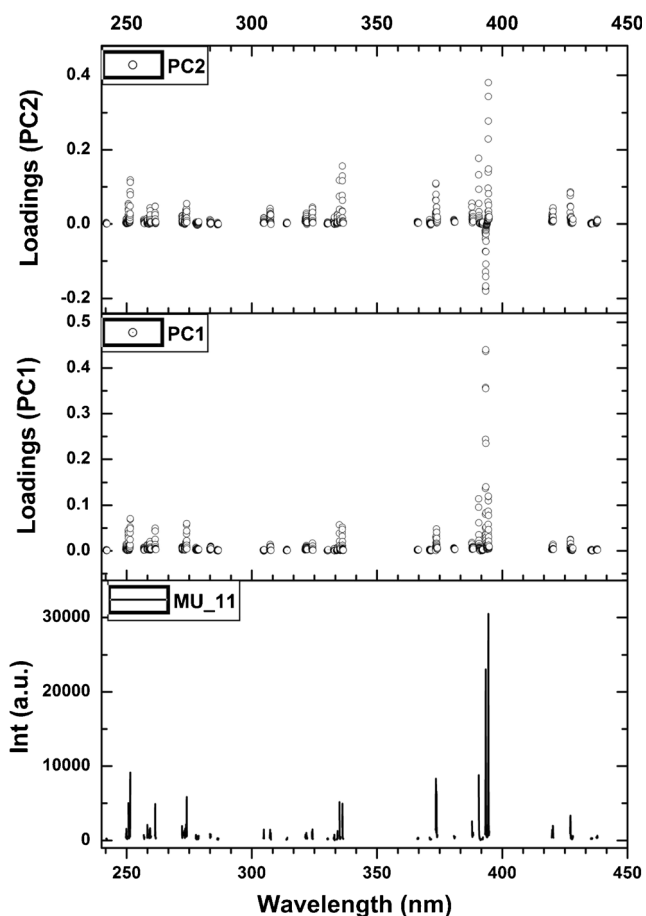


Fig. 7 PC1 and PC2 loadings obtained from PCA of truncated spectral data using the $T^c = 0.02$ parameter along with the truncated spectra of the MU_11 sample in the 230–450-nm region

is unable to detect some of the signal from trace elements due to the instrumental limitation and intrinsic nature of the emission lines. Moreover, when the idea of the current study was conceived, the authors were completely in the dark regarding the elements which were to be analyzed chemically as there

was no prior study regarding the soil analysis in the geothermal area. The only reference available was the analysis of the thermal waters which were found to be enriched in certain trace metals (i.e., Li, Rb, Cs, B, Sb, Cu, Pb, Mn, Co, etc.), rare earth elements (i.e., Sc, Ce, Eu, Tb, etc.) and transition metals (i.e., Hf, Ta, etc.) compared to the non-thermal groundwater ([18, 31], and the references cited therein). The elevated concentration of trace and rare earth elements in the thermal waters prompted us to analyze the concentrations of some of these elements in the soil samples collected from the geothermal area to check their preferential distribution, this being the main reason that some of the elements found to be important in LIBS analysis (Ca, Al, Ru, Mg, and Sr) were excluded from the chemical analysis. However, the available chemical data was found to be sufficient for the classification purpose. We are very sure that analysis of some more elements by chemical methods would have been able to dispel this concern. Thus combining the results obtained from both the LIBS and chemical analysis, it can be concluded that the relative difference in concentrations of few elements (i.e., B, Sr, Cs, Rb, Fe, Co, Sr, Al, Si, Ti, Ru, Mn, Mg, Cu and Eu) in the soil samples near the thermal manifestations has made them characteristically different from the soil samples situated away from the thermal springs.

Conclusion

This study for the very first time demonstrates the potential use of LIBS coupled with PCA as a tool for identification and discrimination of soil samples in a geothermal area through a geochemical fingerprinting approach. This submission also shows that indigenously developed analytical spectral-dependent truncation based PCA method applied on the LIBS spectra utilizing only relevant pixels helped to understand the majorly influencing emission lines. Combining the results obtained from both the LIBS and chemical analysis, it

Table 2 The emission lines and the corresponding spectral regions of the majorly influencing elements in the soil samples near the thermal manifestations

Co	II 427.222, II 272.103, II 251.454
Ca	I 428.301, II 393.366
Sr	I 437.947, I 420.221
Fe	I 419.909, I 387.872, II 373.532, I 334.191, II 307.567, II 304.779, II 283.571, I 273.977, II 261.399, II 259.154, II 258.273, II 250.702, II 249.913, II 249.808
Al	I 394.400, II 373.391, I 396.152
Si	I 390.552, I 413.09, I 288.158
Ti	I 388.289, II 374.164
Ru	I 373.740
Mn	II 334.956, I 307.313, II 273.389, II 259.372
Mg	I 332.992, I 517.27, II 279.553
Cu	II 324.243, I 323.916, II 321.731, II 273.826, II 273.977
Eu	I 272.395

Table 3 Chemical analysis results of the soil samples collected from the study area

Sample ID	Fe (%)	Cu (ppm)	B (ppm)	Co (ppm)	Sc (ppm)	Rb (ppm)	Cs (ppm)	Ce (ppm)	Eu (ppm)	As (ppm)	Sb (ppm)
MU-1	2.59±0.1	23±0.9	121±6	11.7±0.4	11.8±0.3	75±10	4.4±0.7	92±14	1.07±0.16	<5	<5
MU-2	1.69±0.18	9±0.5	120±5	10.1±2.3	7.5±0.8	74±11	2.1±0.3	149±35	1.1±0.18	<5	<5
MU-3	4.1±0.30	38±1.5	137±7	20.9±1.2	16.1±0.8	113±16	7.6±0.9	106±15	1.35±0.12	<5	<5
MU-4	3.53±0.31	25±0.8	104±5	12.5±1.0	14.8±1.5	104±14	7.3±0.4	103±16	1.04±0.15	<5	<5
MU-5	2.46±0.02	50±2.3	73±3	14.6±1.5	10.4±0.3	69±8	3.8±0.5	88±12	1.19±0.20	<5	<5
MU-6	4.89±0.16	20±0.6	64±4	14.3±0.8	14.9±0.4	87±15	3.6±0.6	70±9	1.45±0.13	<5	<5
MU-7	1.51±0.02	193±8.5	89±5	9.8±0.7	8.6±0.8	47±7	2.5±0.3	230±45	1.72±0.12	<5	<5
MU-8	1.76±0.25	20±0.8	148±9	7.1±0.3	8.2±0.9	75±11	3.5±0.5	71±12	0.88±0.16	<5	<5
MU-9	1.43±0.08	15±0.7	105±6	9.1±0.3	6.9±0.2	71±2	1.8±0.3	99±4	1.39±0.08	<5	<5
MU-10	2.62±0.46	9±0.2	94±4	10.5±1.0	12±1.4	122±20	3.3±0.2	145±38	1.7±0.16	<5	<5
MU-11	4.8±0.64	38±1.5	99±6	12.5±1.6	16.8±0.8	330±27	3.2±0.2	69±5	1.12±0.10	<5	<5
MU-12	1.8±0.20	18±0.6	97±6	8.9±0.5	8.8±0.4	76±8	2.8±0.2	174±40	1.39±0.08	<5	<5
MU-13	3.24±0.10	29±1.5	84±5	16.3±2.0	14.3±0.2	73±10	6.5±0.6	105±12	1.19±0.20	<5	<5
MU-14	2.85±0.30	24±0.86	77±5	11.9±0.7	11.3±0.8	81±12	7.9±0.3	82±2	1.14±0.08	<5	<5
MU-15	2.74±0.14	25±0.9	128±8	13.3±1.3	11.1±0.6	95±9	4.8±0.5	59±8	1.12±0.11	<5	<5
MU-16	3.4±0.36	25±0.7	89±4	10.6±0.6	13.4±1.5	125±25	6.3±0.8	95±16	1.08±0.06	<5	<5
MU-17	1.62±0.10	19±0.8	68±4	5.8±0.4	7.1±1.0	55±8	2.5±0.7	329±31	0.86±0.28	<5	<5
MU-18	2.53±0.41	19±0.6	128±9	12.8±1.4	11.1±0.9	103±18	5.5±0.9	141±10	1.36±0.15	<5	<5
MU-19	1.96±0.21	14±0.45	10±6	5.9±0.4	11.2±0.7	76±8	4.3±0.7	79±7	1.39±0.10	<5	<5
MU-20	1.8±0.3	15±0.5	95±4	6.8±0.2	10.8±0.6	74±6	4.5±0.8	75±8	1.7±0.2	<5	<5
Sample ID	Tb (ppm)	Hf (ppm)	Ta (ppm)	Th (ppm)	Mn (ppm)	Na (ppm)	K (ppm)	SiO ₂ (%)	Pb (ppm)	Hg (ppm)	Li (ppm)
MU-1	1.4±0.6	20±1	1.1±0.2	23±3	400±10	680±50	15,900±600	66.05±2.5	<5	<5	<5
MU-2	1.2±0.7	18±4	1.1±0.3	36±4	370±11	1470±80	25,300±800	76.4±3.0	<5	<5	<5
MU-3	1.6±0.7	16±2	1.3±0.3	42±10	890±20	700±60	8000±200	43.43±1.8	<5	<5	<5
MU-4	1.6±0.5	19±5	1.5±0.3	46±8	380±10	730±75	16,300±560	48.05±1.5	<5	<5	<5
MU-5	1.5±0.3	20±1	1.4±0.2	21±3	560±12	1860±110	24,200±850	62.02±2.7	<5	<5	<5
MU-6	1.7±0.4	8±1	1.4±0.1	19±2	770±19	700±56	17,800±750	51.54±2.0	59	<5	<5
MU-7	1.4±0.3	24±5	1.3±0.1	105±12	450±15	670±48	12,600±450	66.39±2.9	<5	<5	<5
MU-8	1.2±0.2	13±1	1±0.2	33±7	210±16	960±52	14,700±520	73.76±2.5	<5	<5	<5
MU-9	1.5±0.4	14±2	1.7±0.3	49±1	470±18	2160±80	29,400±950	74.95±2.3	<5	<5	<5
MU-10	2.3±0.6	28±1	1.3±0.1	68±9	360±15	2000±92	31,900±900	58.39±1.8	<5	<5	<5
MU-11	2±0.5	9±2	1.6±0.6	21±4	360±12	490±30	17,400±600	53.34±1.7	56	<5	<5
MU-12	1.6±0.3	22±6	1.5±0.3	51±3	410±11	990±50	15,200±550	78.02±3.5	<5	<5	<5
MU-13	1.7±0.4	15±3	1.7±0.4	30±2	400±12	940±45	9600±350	62.73±2.6	<5	<5	<5
MU-14	1.3±0.3	15±2	1.3±0.2	23±1	360±10	670±50	16,100±560	53.44±1.9	<5	<5	<5
MU-15	1.6±0.4	15±2	1.1±0.3	23±3	510±16	710±53	8700±300	61.04±2.85	<5	<5	<5
MU-16	1.6±0.5	17±2	1.6±0.4	32±5	400±18	1920±100	32,900±980	59.72±2.0	<5	<5	<5
MU-17	1.1±0.3	21±2	1±0.3	111±17	200±10	890±74	16,600±880	71.91±3.2	<5	<5	<5
MU-18	1.1±0.3	19±4	1.2±0.2	47±5	420±12	980±64	18,600±820	68.93±3.1	<5	<5	<5
MU-19	2.0±0.2	16±2	1.4±0.1	24±4	500±16	1890±80	41,300±1100	61.07±2.8	<5	<5	<5
MU-20	1.8±0.3	15±2	1.7±0.2	23±2	700±16	1500±95	15,400±570	75±3.4	<5	<5	<5

can be concluded that soil samples near the thermal manifestations had a distinctly different concentration pattern for some trace and bulk elements (B, Sr, Cs, Rb, Fe, Co, Al, Si, Ti, Ru, Mn, Mg, Cu, and Eu) compared to the soil samples located away from thermal manifestations. But this PCA-based LIBS method was not completely robust as it misinterpreted four samples out of 20 samples implying

80% accuracy. On the other hand, the chemical analysis-based PCA method also was not completely robust as a total of 17 samples out of 20 samples were clearly differentiated implying 85% accuracy. Yet, considering the time consumption and cost involved in multiple analytical techniques to get the chemical data, the LIBS method with a comparatively lower success rate seems to be a simple and good alternative

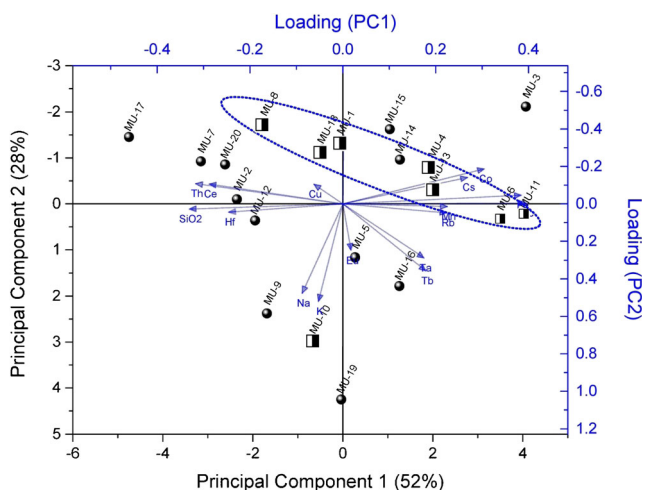


Fig. 8 A PCA bi-plot using chemical analysis results shows the clustering of soil samples near the thermal manifestations

technique which can be field adopted and generate real-time data. Due to the limited number of samples with known geothermal linkage, the validation of this method using an unknown sample was not possible to carry out in the present work but the obtained results demonstrate the successful application of the LIBS-PCA combination for fast classification of the geothermal soil samples avoiding the rigorous dissolution required by other analytical techniques.

Acknowledgments The authors (SC and UKS) wish to acknowledge Dr. P.K. Pujari, AD, RC&I group and the officers of GSI for their help and encouragement during the study. The authors (MS and AS) are thankful to Dr. S. Kannan, Head, FCD, for their constant support and encouragement in the LIBS work. The authors greatly acknowledge the in-depth and insightful reviews of two anonymous reviewers which have immensely contributed to the improvement of the quality of the manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Awan MA, Ahmed SH, Aslam MR, Qazi IA, Baig MA. Determination of heavy metals in ambient air particulate matter using laser-induced breakdown spectroscopy. *Arab J Sci Eng*. 2013;38:1655–61.
2. Badday MA, Bidin N, Rizvi ZH, Hosseinian R. Determination of environmental safety level with laser-induced breakdown spectroscopy technique. *Chem Ecol*. 2015;31:379–87.
3. Boucher TF, Ozanne MV, Carmosino ML, Dyar MD, Mahadevan S, Breves EA, et al. A study of machine learning regression methods for major elemental analysis of rocks using laser-induced breakdown spectroscopy. *Spectrochim Acta B*. 2015;107:1–10.
4. Caneve L, Diamanti A, Grimaldi F, Palleschi G, Spizzichino V, Valentini F. Analysis of fresco by laser induced breakdown spectroscopy. *Spectrochim Acta B*. 2010;65:702–6.
5. Clegg SM, Sklute E, Dyar MD, Barefield JE, Wiens RC. Multivariate analysis of remote laser-induced breakdown spectroscopy spectra using partial least squares, principal component analysis, and related techniques. *Spectrochim Acta B*. 2009;64:79–88.
6. Clegg SM, Wiens R, Misra AK, Sharma SK, Lambert J, Bender S, et al. Planetary geochemical investigations using Raman and laser-induced breakdown spectroscopy. *Appl Spectrosc*. 2014;68(9):925–36. <https://doi.org/10.1366/13-07386>.
7. Cousin A, Forni O, Maurice S, Gasnault O, Fabre C, Sautter V, et al. Laser induced breakdown spectroscopy library for the Martian environment. *Spectrochim. Acta Part B*. 2011;66:805–14.
8. Davis JC. *Statistics and data analysis in geology*. New York: John Wiley & Sons Inc; 1986.
9. De Lucia FC Jr, Gottfried JL, Munson CA, Miziolek AW. Double pulse laser induced breakdown spectroscopy of explosives: initial study towards improved discrimination. *Spectrochim. Acta Part B*. 2007;62:1399–404.
10. De Lucia FC, Samuels AJ, Harmon RS, Walters RA, McNesby KL, LaPointe A, et al. Laser-induced breakdown spectroscopy (LIBS): a promising versatile chemical sensor technology for hazardous material detection. *IEEE Sensors J*. 2005;5:681–9.
11. Diedrich J, Rehse SJ, Palchadhuri S. Pathogenic *Escherichia coli* strain discrimination using laser-induced breakdown spectroscopy. *J Appl Phys*. 2007;102:014702. <https://doi.org/10.1063/1.2752784>.
12. Fang SC. Sorption and transformation of mercury vapor by dry soil. *Environ Sci Technol*. 1978;12:285–8.
13. Fink H, Panne U, Niessner R. Process analysis of recycled thermoplasts from consumer electronics by laser-induced plasma spectroscopy. *Anal Chem*. 2002;74:4334–42.
14. Gottfried JL, De Lucia FC Jr, Munson CA, Miziolek AW. Double-pulse standoff laser-induced breakdown spectroscopy for versatile hazardous materials detection. *Spectrochim Acta B*. 2007;62:1405–11.
15. Gottfried JL, Harmon RS, De Lucia FC Jr, Miziolek AW. Multivariate analysis of laser-induced breakdown spectroscopy chemical signatures for geomaterial classification. *Spectrochim Acta B*. 2009;64:1009–19.
16. Harmon RS, Remus J, McMillan NJ, McManus C, Collins L, Gottfried JL Jr, et al. LIBS analysis of geomaterials: geochemical fingerprinting for the rapid analysis and discrimination of minerals. *Appl Geochem*. 2009;24:1125–41.
17. Jung EC, Lee DH, Yun JI, Kim JG, Yeon JW, Song K. Quantitative determination of uranium and europium in glass matrix by laser-induced breakdown spectroscopy. *Spectrochim Acta B At Spectrosc*. 2011;66(9–10):761–4.
18. Kaasalainen H, Stefánsson A. The chemistry of trace elements in surface geothermal waters and steam, Iceland. *Chem Geol*. 2012;330–331:60–85.
19. Kaiser J, Novotny K, Martin MZ, Hrdlicka A, Malina R, Hartl M, et al. Trace elemental analysis by laser-induced breakdown spectroscopy-biological applications. *Surf Sci Rep*. 2012;67:233–43.
20. Knight AK, Scherbarth NL, Cremers DA, Ferris MJ. Characterization of laser induced breakdown spectroscopy (LIBS) for application to space exploration. *Appl Spectrosc*. 2000;54:331–40.
21. Kumar A, Singhal RK, Rout S, Ravi PM. Spatial geochemical variation of major and trace elements in the marine sediments of Mumbai Harbor Bay. *Environ Earth Sci*. 2013. <https://doi.org/10.1007/s12665-013-2366-3>.
22. Landa ER. The retention of metallic mercury vapour by soils. *Geochim Cosmochim Acta*. 1978;42:1407–11.
23. Mowery MD, Sing R, Kirsch J, Razaghi A, Bechard S, Reed RA. Rapid at-line analysis of coating thickness and uniformity on tablets using laser induced breakdown spectroscopy. *J Pharm Biomed*

- Anal. 2002;28:935–43. [https://doi.org/10.1016/S0731-7085\(01\)00705-1](https://doi.org/10.1016/S0731-7085(01)00705-1).
24. Myakalwar AK, Sreedhar S, Barman I, Dingari NC, Rao VS, Kiran PP, et al. Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis. *Talanta*. 2011;87:53–9.
 25. Nicholson K. *Geothermal fluids: chemistry and exploration techniques*, ISBN 3-540-56017-3. Berlin: Springer-Verlag; 1993. p. 209–10.
 26. Phelps DW, Buseck PR. Natural concentration of Hg in the Yellowstone and Coso geothermal fields, vol. 2: *Geotherm. Resources Council Trans*; 1978. p. 521–2.
 27. Phelps D, Buseck PR. Distribution of soil mercury and the development of soil mercury anomalies in the Yellowstone geothermal area, Wyoming. *Econ Geol*. 1980;75:730–41.
 28. Sallé B, Lacour JL, Vors E, Fichet P, Maurice S, Cremers DA, et al. Laser induced breakdown spectroscopy for Mars surface analysis: capabilities at stand-off distances and detection of chlorine and sulfur elements. *Spectrochim Acta B*. 2004;59:1413–22.
 29. Sallé B, Lacour JL, Mauchien P, Fichet P, Maurice S, Manhès G. Comparative study of different methodologies for quantitative rock analysis by laser-induced breakdown spectroscopy in a simulated Martian atmosphere. *Spectrochim Acta B*. 2006;61:301–13.
 30. Sarkar A, Mishra RK, Kaushik CP, Wattal PK, Alamelu D, Aggarwal SK. Analysis of barium borosilicate glass matrix for uranium determination by using ns-IR-LIBS in air and Ar atmosphere. *Radiochimica Acta*. 2014;102(9):805–12.
 31. Shakeri A, Ghoreyshinia S, Mehrabi B, Delavari M. Rare earth elements geochemistry in springs from Taftan geothermal area SE Iran. *J Volcanol Geotherm Res*. 2015;304:49–61.
 32. Shen XK, Lu YF. Detection of uranium in solids by using laser-induced breakdown spectroscopy combined with laser-induced fluorescence. *Appl Opt*. 2008;47(11):1810–5.
 33. Singh M, Karki V, Mishra RK, Kumar A, Kaushik CP, Mao X, et al. Analytical spectral dependent partial least squares regression: a study of nuclear waste glass from thorium based fuel using LIBS. *J Anal At Spectrom*. 2015;30:2507–15.
 34. Sirven JB, Bousquet B, Canioni L, Sarger L, Tellier S, Gautier MP, et al. Qualitative and quantitative investigation of chromium-polluted soils by laser-induced breakdown spectroscopy combined with neural networks analysis. *Anal Bioanal Chem*. 2006;385:256–62.
 35. Sirven JB, Sallé B, Mauchien P, Lacour JL, Maurice S, Manhès G. Feasibility study of rock identification at the surface of Mars by remote laser-induced breakdown spectroscopy and three chemometric methods. *J Anal At Spectrom*. 2007;22:1471–80.
 36. Snyder GH. Methods for silicon analysis in plants, soils, and fertilizers. *Stud Plant Sci*. 2001;8:185–96.
 37. Tirumalesh K, Ramakumar KL, Chidambaram S, Pethaperumal S, Singh G. Rare earth elements distribution in clay zones of sedimentary formation, Pondicherry, south India. *J Radioanal Nucl Chem*. 2012;294:303–8.
 38. Unnikrishnan VK, Choudhari KS, Kulkarni SD, Nayak R, Karthaa VB, Santhosh C. Analytical predictive capabilities of laser induced breakdown spectroscopy (LIBS) with principal component analysis (PCA) for plastic classification. *RSC Adv*. 2013;3:25872–80.
 39. White DE. Mercury and base-metal deposits with associated thermal and mineral waters. In: Barnes HL, editor. *Geochemistry of hydrothermal ore deposits*. New York: Holt, Rinehart and Winston; 1967. p. 575–631.
 40. Yu KQ, Zhao YR, Liu F, Yong H. Laser-induced breakdown spectroscopy coupled with multivariate chemometrics for variety discrimination of soil. *Sci Rep*. 2016;6:27574. <https://doi.org/10.1038/srep27574>.
 41. Yueh FY, Zheng HB, Singh JP, Burgess S. Preliminary evaluation of laser-induced breakdown spectroscopy for tissue classification. *Spectrochim Acta B*. 2009;64:1059–67.
 42. Zhang T, Wu S, Dong J, Wei J, Wang K, Tang H, et al. Quantitative and classification analysis of slag samples by laser induced breakdown spectroscopy (LIBS) coupled with support vector machine (SVM) and partial least square (PLS) methods. *J Anal Atom Spectrom*. 2015;30:368–74.

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