Spectrochimica Acta Part B: Atomic Spectroscopy 206 (2023) 106708

Contents lists available at ScienceDirect

Spectrochimica Acta Part B: Atomic Spectroscopy

journal homepage: www.elsevier.com/locate/sab

Sensitive and accurate determination of nitrogen in simulated Martian soil and environment with LIBS spectrum fusion and regression based on neural network

Beiyi Zhang [a], Chen Sun [a,*], Xiaowen Yu [c], Fengye Chen [a], Li Wang [a], Yunfei Rao [a], Tianyang Sun [a], Yu-Yan Sara Zhao [b,c,d,**], Jin Yu [a,*]

[a] School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China
[b] Research Center for Planetary Science, College of Earth Science, Chengdu University of Technology, Chengdu, China
[c] Center for Lunar and Planetary Sciences, Institute of Geochemistry, Chinese Academy of Sciences, Guiyang, China
[d] CAS Center for Excellence in Comparative Planetology, Hefei, China

ARTICLE INFO

Keywords:
Nitrogen
Simulated Martian environment
Synthesized Martian soil
Chemical matrix effect
Spectrum fusion
Generalized spectrum
Neural network

ABSTRACT

Nitrogen is a key element necessary for the emergence and development of life. It is one of the elements targeted by the landed missions on Mars in accordance with their scientific goal of investigation of habitability and search for traces of life. A gas chromatography mass spectrometer (GCMS) instrument on board the Mars Science Laboratory Curiosity rover has revealed the existence of oxidized nitrogen-bearing compounds on Martian surface with an equivalent nitrogen concentration up to 0.01 wt%. Although the detection with laser-induced breakdown spectroscopy (LIBS) also on board the Curiosity rover is desirable, the current performance of LIBS for nitrogen analysis does not show the capacity in terms of limit of detection (LOD) and accuracy. Research on a suitable method should therefore be first engaged in laboratory in order to guide further improvements of LIBS instrument on board Mars rover, as well as the data treatment method. Beyond the sensitivity issue, matrix effect also affects LIBS determination of nitrogen, especially due to its various chemical speciation in geological materials. Method research should answer thus double requirements of improving the sensitivity and reinforcing the robustness with respect to different nitrogen-bearing compounds. An experimental configuration of double detections with a narrow bandwidth Czerny-Tuner (CT) spectrometer and a broad bandwidth Echelle spectrometer, was implemented in this work, in such way that the first ensured a sensitive detection of emission lines from nitrogen and the second complemented with those from major elements in the sample. The fusion of the simultaneously acquired spectra took into account the emission characteristics of the both two types of elements, necessary for an effective and robust multivariate regression based on a neural network. In addition, for a better treatment of different chemical speciation of nitrogen in samples, generalized spectrum was used for training of regression models, after an unsupervised clustering having assigned a type label to each training spectrum. The trained model was tested by collections of independent test samples, resulting in a limit of detection (LOD) of 0.18 wt%, and a root mean square error of prediction (RMSEP) of 0.041 wt%, representing a step forward to nitrogen detection using LIBS on Mars.

1. Introduction

Laser-induced breakdown spectroscopy (LIBS) has been widely used for analysis of geological materials [1], not only on Earth but also on Mars for its *in situ* exploration [2]. After the first successive demonstration with the Mars Science Laboratory (MSL) Curiosity rover, two new missions, Perseverance rover and Zhurong rover, have landed on Mars with a LIBS payload on board. Investigation of habitability and search for past and present traces of life on Mars feature as major scientific goals of these exploration missions [3]. Organic compounds have been thus actively searched with the instruments on board the rovers. Up to now, only one positive detection, concerning chlorobenzene and

* Corresponding authors at: School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China.
** Corresponding author at: Research Center for Planetary Science, College of Earth Science, Chengdu University of Technology, Chengdu, China.
E-mail addresses: sc11297@sjtu.edu.cn (C. Sun), zhaoyuyan@cdut.edu.cn (Y.-Y.S. Zhao), jin.yu@sjtu.edu.cn (J. Yu).

https://doi.org/10.1016/j.sab.2023.106708
Received 21 November 2022; Received in revised form 16 May 2023; Accepted 16 May 2023
Available online 18 May 2023

C_2 to C_4 dichloroalkanes, has been reported using the gas chromatography mass spectrometer (GCMS) of the Sample Analysis at Mars (SAM) instrument on board the Curiosity rover [4]. A closely related discovery made using SAM concerns the evidence of oxidized nitrogen-bearing compounds in sedimentary and aeolian deposits [5], with an equivalent nitrogen concentration up to 0.01 wt%. The presence of indigenous nitrogen in Mars surface materials has important implications for habitability and, specifically, for the potential evolution of a nitrogen cycle at some points in Martian history [5]. In spite of the importance of the above discoveries, the results have not yet been repeated in other circumstances. This can be explained by (i) the difficulties SAM must overcome to access the sites to be probed, and (ii) the used thermal and chemical sample processing, which makes the detection of organic compounds difficult [6]. LIBS is expected to provide an additional approach to detect and quantify nitrogen on Mars, with a higher sampling rate and a larger space coverage, especially in rugged areas, in such way to bring information to guide the rover and the dedicated instrument, such as SAM, toward preselected samples potentially containing organics [6]. Such approach is up to now precluded by the incapacity of LIBS to satisfy the requirement of *in situ* detection of nitrogen on Mars.

Early research works on LIBS determination of nitrogen in soil on Earth showed the influences of nitrogen contained in the atmosphere and other elements contained in soil, Ti for example [7,8]. Such influences were systematically studied shortly later at atmospheric and reduced pressures using sea sands as samples, which have a composition simpler than soil to minimize spectral interferences from minor and trace elements [8]. It was demonstrated that the influence of the atmospheric nitrogen can be controlled by reducing the ambient pressure of a LIBS measurement. Performing the experiment under an ambient pressure of 0.04 Torr, with 1064 nm and 60 mJ laser pulses, and recoding the spectra using a Czerny-Turner (CT) type spectrometer tuned to the 740 nm region and equipped with an intensified CCD detector, the N I 746.83 nm line was correctly detected. Its peak intensity was then used to build a univariate calibration model, which allowed determining a limit of detection (LOD) of 0.8 wt% [8]. A more recent approach to fix the issue of atmospheric nitrogen consisted in blowing the surface of a farmland soil sample with an argon gas jet, while ablating it with high energy 180 mJ laser pulses [9]. A compact CCD spectrometer was used to detect the emissions from the plasmas with a large spectral range of 200–1100 nm. Relying on the characteristics of the N I 744.23 nm and N I 746.83 nm lines, the sensitivity of the detection can be estimated at 1 wt% level. More specifically for the purpose of detection and determination of nitrogen-bearing compounds with LIBS *in situ* on Mars, research has been conducted within a laboratory simulated Mars environment, and using a replica of the ChemCam instrument on board the Curiosity rover [6], equipped in particular with a pulsed Nd:KGW laser at 1067 nm, operating at 3 Hz with a pulse duration of 5 ns and an energy of ~10 mJ, and a detection system with three CCD spectrometers covering respectively 240.1–342.2 nm, 382.1–469.3 nm and 474.0–906.5 nm, wavelength ranges [10,11]. The studied samples were a mixture between nontronite, a clay mineral representative of Mars environment, and adenine, a nitrogen-bearing organic material [6]. The Mars-like environment solve naturally the problem of ambient gas interference for nitrogen detection, and the absence of Ti in nontronite allowed the N II 500.5 nm line being suitable for a univariate regression model for N, showing a LOD of 10 wt% as a tracer of adenine under Martian atmospheric condition [6]. It is therefore clear that the current performance of LIBS for nitrogen determination as demonstrated in different laboratory studies, does not yet ensure the capacity of *in situ* detection and analysis of nitrogen-bearing materials on Mars.

This work intends to improve the capacity of LIBS for the detection and analysis of nitrogen in Mars environment and for Mars surface materials. The experiment was performed in a Mars atmosphere simulation chamber, and using synthetic Mars soil as the matrix of the samples. Different nitrogen-containing compounds were mixed into the matrix powder, resulting in samples with a nitrogen concentration gradient from blank matrix to 2.0 wt% nitrogen. The objective was on one hand, to improve the sensitivity of the method, and on the other hand, to control the matrix effect due to different chemical speciation of nitrogen-containing compounds. The approach to reach such double objectives consists in spectrum fusion between a narrow band and sensitive detection of nitrogen emissions around the N I 746.8 nm line, and a large band detection of emission lines of other elements in the sample. A fused spectrum contains high quality information about nitrogen and its chemical environment in the sample for an effective training of a multivariate regression model based on neural network. In addition, unsupervised clustering first labels a sample with the type of the nitrogen-bearing compound. The label is then used as the addition dimension in the generalized spectrum [12], which further helps to correct the chemical matrix effect. Testing the trained model with independent test samples results in analytical performances of the model with a limit of detection (LOD) of 0.18 wt% and a root mean square error of prediction (RMSEP) of 0.041 wt%, which are much improved with respect to the reported performances and get a step closer to the requirements of *in situ* detection and analysis of nitrogen on Mars.

2. Sample preparation and experimental setup

2.1. Samples

A synthesized Martian soil, Jining Martian Soil Simulant (JMSS-1), was used as matrix to prepare the samples, in order to simulate Martian surface materials in terms of chemical composition, mineralogy, and physical properties [13]. The concentrations of the major elements of JMSS-1 are shown in Table 1 [13] with the contents of the corresponding oxides [13]. The soil was available in a powder of grain size below 50 μm.

Potassium nitrate (KNO_3), sodium nitrate ($NaNO_3$), and ammonium dihydrogen phosphate ($NH_4H_2PO_4$) were purchased as analytically pure compounds (purity ≥ 99.7%) in the form of powders from General Reagent. These nitrogen-bearing compounds are related to nitrate salts found on Mars by SAM and considered as playing an important role in the nitrogen cycle [5]. Potassium and sodium nitrates were further manually ground in an agate mortar into a powder of grain size below 100 μm. Purchased ammonium dihydrogen phosphate powder had a grain size below 150 μm. Taking into account the reported nitrogen concentrations in Mars surface materials [5], the above pure nitrogen-bearing compounds were mixed with the Martian soil simulant in such way that series of sample powders with a gradient of nitrogen concentration were prepared as shown in Table 2. In the table, the mixture rate represents the weight percentage of pure nitrogen-bearing compound or unknown farmland soil in a sample powder. For example, the mixture

Table 1
Contents of the major oxides of JMSS-1 in wt% [13].

Oxides	SiO_2	TiO_2	Al_2O_3	Fe_2O_3	MnO
Content (wt%)	49.28 ± 0.24	1.78 ± 0.01	13.64 ± 0.33	16.00 ± 0.07	0.14 ± 0.01
Oxides	MgO	CaO	Na_2O	K_2O	P_2O_5
Content (wt%)	6.35 ± 0.08	7.56 ± 0.06	2.92 ± 0.09	1.02 ± 0.03	0.30 ± 0.01

Table 2
Information about the preparation of sample powders by mixing nitrogen-bearing compounds with the Martian soil simulant, and their roles in this work.

Sample			Nitrogen-bearing compounds		Mixture rate (wt%)	Nitrogen concentration (wt%)
Set	Type	ID				
Training	Tr 1	S1-1	Pure N-bearing compounds	Potassium nitrate	2.00	0.29
		S1-2			5.00	0.69
		S1-3			7.00	0.97
		S1-4			9.00	1.25
	Tr 2	S2-1		Sodium nitrate	2.00	0.33
		S2-2			5.00	0.83
		S2-3			7.00	1.15
		S2-4			10.00	1.65
	Tr 3	S3-1		Ammonium dihydrogen phosphate	1.00	0.13
		S3-2			5.00	0.61
		S3-3			9.00	1.11
		S3-4			15.00	1.83
Test	Te 1	S4-1	Unknown farmland soils	Soil 1	80.00	0.43
		S4-2			100.00	0.54
	Te 2	S5-1		Soil 2	80.00	0.45
		S5-2			100.00	0.57

rate of the sample S4–2 is 100%, meaning that it is a farmland soil without Martian soil simulant. These samples were used as training samples in the following studies of regression models. Two series of samples were further prepared using two kinds of farmland soil powders, as nitrogen-bearing compounds to mix with the Martian soil simulant, treated and characterized for their nitrogen concentration by a qualified analytical laboratory (Zhejiang Geology and Mineral Resources Research Institute) in the form of a powder of 50 μm grain size. The two series of additional samples were used as independent test samples, which simulates a real application case where the trained models are used to predict with LIBS spectra acquired from samples containing nitrogen with an unknown chemical speciation. For the preparation of the sample powders, the mixture powders were homogenized in a rotary mixer for 12 h. An amount of 0.5 g of each sample powder was weighed and pressed into a pellet under a pressure of 12 tons for 10 min, resulting in a pellet of 15 mm dimeter and 2 mm thickness.

2.2. Experimental setup and measurement protocol

The used experiment setup is schematically shown in Fig. 1. The inset shows a plasma generated on the surface of an aluminum sample and a soil pellet sample after a LIBS measurement with a matrix of ablation craters visible on the surface. A Mars atmosphere simulation chamber contained the sample together with its 3D displacement stage, in such way that the LIBS measurements took place in a Mars-like ambient gas. For each experiment, the chamber was first evacuated using a

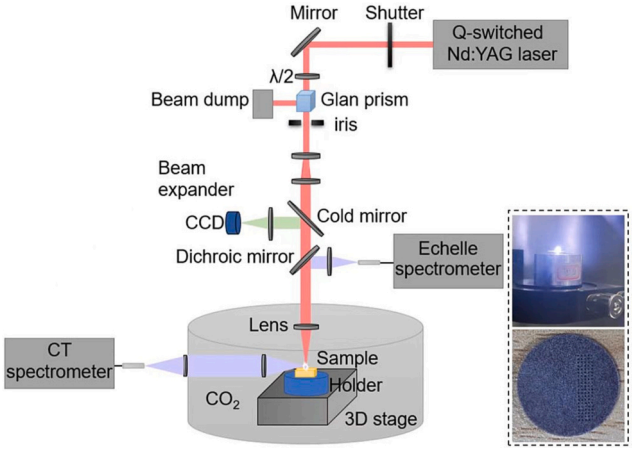

Fig. 1. Schematic presentation of the experiment setup.

mechanical pump to its limit pressure of 10 Pa. It was then filled with a CO_2 gas (99.9% pure grade) from a high-pressure cylindrical container (Air Liquide), up to a pressure of 690 Pa. During an experiment, the pressure inside of the chamber was kept bellow 710 Pa, ensuring a quasi-stable simulant of Martian atmosphere. The ablation source was a Q-switched Nd:YAG laser operating at 1064 nm, with a repetition rate of 10 Hz, a pulse energy of 96 mJ, and a pulse width of 10 ns. A mechanical shutter was used to control the delivery of laser pulses to a sample in synchronization with its displacement ensured by the 3D stage. Such synchronization allowed to program, for a given sample, the sites on its surface to be ablated by laser pulses and the number of the successive laser pulses hitting a same site. Laser pulses propagating beyond the mechanical shutter were attenuated by a combination of a half-wave plate and a Glan prim. In the experiment, the half-wave plate was finely tuned to deliver pulses of 20 mJ energy to the sample. A combination of a divergent ($f = -50$ mm) and a convergent ($f = 100$ mm) lenses, enlarged an initial beam section of 5 mm diameter to a final beam section of 10 mm diameter. Laser pulses were finally focused by a quartz plano-convex lens of 50 mm focus length. For a given height of the sample, the actual position of the focus was finely adjusted by the inter-lens distance of the beam expander. With a focus 0.5 mm under the sample surface, a stable plasma was generated as shown in the inset of Fig. 1. In this focusing condition, the spot size of the focused laser on the sample surface was estimated to be 300 μm diameter, resulting in a fluence and irradiance of respectively 28 J/cm^2 and 2.8 GW/cm^2, delivered on the sample surface. The lens to sample distance was kept constant during an experiment by precisely setting the sample surface in the horizontal plane using a laser pointer shining on the sample surface with a tilted angle (not shown in Fig. 1). The spot of the laser pointer on the sample surface was monitored by a CCD camera aiming the sample surface with the help of a cold mirror installed in the main laser beam path and reflecting a part of visible light from the sample surface toward the camera. The transversal position of the laser pointer spot on the sample surface correlated therefore to its height.

The emission from an induced plasma was simultaneously recorded by two detection systems respectively in the lateral and backward axial directions as shown in Fig. 1. The lateral optical collimation system was composed by a pair of quartz lenses of a focus length of 100 mm and an optical fiber of 50 μm core diameter, receiving a part of the emitted light collimated by the quartz lenses and connected to the entrance of a Czerny-Turner spectrometer (Shamrock 500i, Andor Technology), calibrated for working with the first order of the diffraction. The spectrometer was equipped with an intensified CMOS camera (ICMOS, iStar, Andor Technology). It was used with a grating of 1200 line/mm, blazed at 300 nm and tuned to the spectral range of 730–748 nm covering the intense N I lines at 742.36 nm, 744.23 nm, and 746.83 nm [8], and an

entrance slit opening of 10 μm, which resulted in a spectral resolution power $\lambda/\Delta\lambda = 7000$. In the experiment, the gain of the amplifier was specifically optimized (to be a value of 3500 over a maximum of 4096) for the lines detected in the covered spectral range, offering a sensitive detection of the N I lines. In addition, since along the vertical axis of the plasma propagation, the plasma exhibited an obvious inhomogeneity, the input extremity of the fiber was adjusted laterally on the middle of the image of the plasma, and vertically at a height of about two fifth of the size of the plasma image from the sample surface, where the plasma emission was observed strong over the detection time interval.

The backward optical collimation system was composed by the ablation laser pulse focusing lens, a dichroic mirror presenting a wide range of high reflectivity from 230 nm to 850 nm (reflectivity >90%, Xin Xiang Bai He O. E. CO. LTD), an addition focusing quartz lens of 50 mm focus length, and an optical fiber of 50 μm core diameter connected to an Echelle spectrometer (Mechelle 5000, Andor Technology). The latter was equipped with an intensified CCD camera (ICCD, iStar, Andor Technology), with a wide spectral range of 220–900 nm and resolution power of $\lambda/\Delta\lambda = 5000$. The input extremity of the fiber was centered on the image of the plasma formed by the collection optical system. The gain of the intensifier was adjusted to be a value of 2000 over a maximum of 4096 in such way that the strongest lines (in blue and near UV ranges) in the covered spectral range did not saturate the ICCD camera. It was therefore not optimized for a sensitive detection of the weak emission lines from nitrogen. A fast photodiode (not shown in Fig. 1) detecting scattered lights of laser pulses was used to trigger the both ICMOS and ICCD cameras. The both CT and Echelle spectrometers were calibrated for wavelengths with a mercury-argon lamp. The Echelle spectrometer was in addition, calibrated for a relative radiometric response with a deuterium-tungsten halogen lamp.

During the experiment, the delays and the gate widths of the both ICMOS and ICCD cameras were identically set to 50 ns and 2000 ns. For each sample, a matrix of 5 × 20 ablation sites was applied on the surface as shown in the inset of Fig. 1, resulting in 100 replicate spectra per sample. A distance of 0.5 mm was left between two neighbor sites to avoid their overlapping. Each site was sequentially fired by 20 laser pulses. The induced plasma emissions were accumulated on CMOS or CCD to provide a replicate spectrum.

3. Results and discussions

3.1. Univariate analysis

Typical average replicate spectra (over 40 individual replicate spectra) from the sample S2–4 respectively and simultaneously recorded by the CT (a) and the Echelle (b) spectrometers are shown in Fig. 2. The spectra are presented within two specific and corresponding spectral ranges. In Fig. 2 (a) within the spectral range from 740 nm to 748 nm covered by the CT spectrometer, three emission lines from nitrogen can be identified using the NIST Atomic Spectra Database [14] as indicated in the figure for N I 742.36 nm, N I 744.23 nm, and N I 746.83 nm lines. Other prominent lines can be detected and numbered from 1 to 10 in Fig. 2 (a). Some of them severely interfere with N I lines as the case of the line at 746.96 nm. These lines cannot be directly identified using the NIST database. A check in the Echelle spectrum in Fig. 2 (b), within the corresponding spectral range from 370 nm to 374 nm at the half of the wavelengths covered by the CT spectrometer, reveals the existence of well detected lines with a similar structure of the relative intensities. These lines can be identified using the NIST database as indicated in Fig. 2 (b). By multiplying by a factor of two the wavelengths of these lines, we find the positions of the lines in the CT spectrum, except the nitrogen lines. Such coincidence confirms us the fact that some lines detected by the CT spectrometer in the spectral range from 740 nm to 748 nm, correspond to the ghosts (order two of the diffraction) of the lines in the spectral range from 370 nm to 374 nm, also simultaneously detected by the Echelle spectrometer. We can therefore interpret the

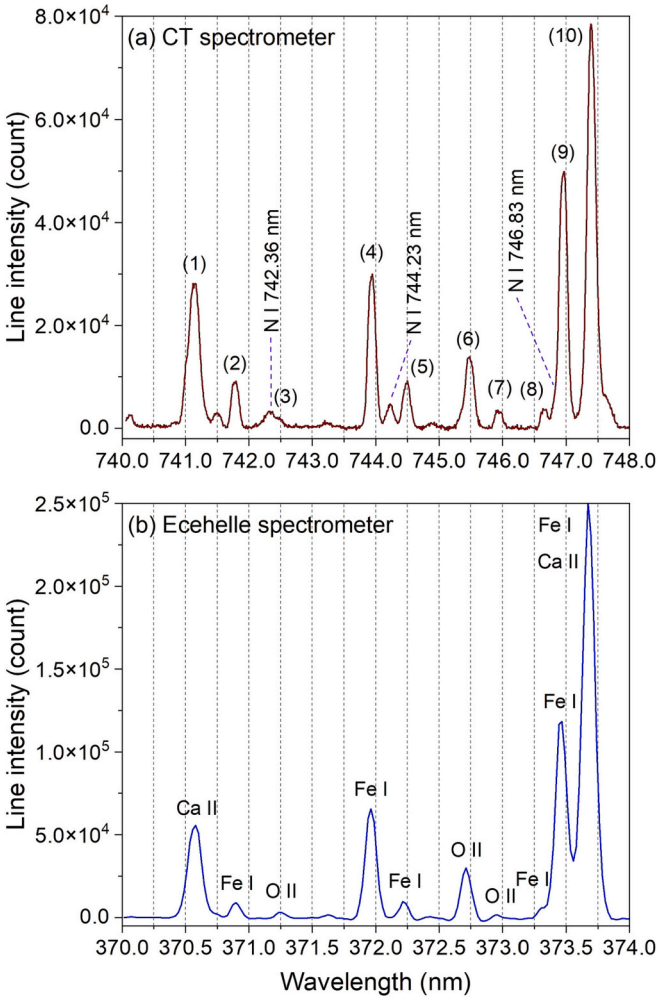

Fig. 2. Average spectra recorded with the sample S2–4, respectively and simultaneously by the CT spectrometer (a), and the Echelle spectrometer (b). The spectra are presented within two specific and corresponding spectral ranges, the Echelle spectrum covers a spectral range of wavelengths which are the halves of the wavelengths covered by the CT spectrometer.

lines numbered from 1 to 10 in the CT spectrum by respectively the emission lines of Ca II, Fe I, O II, Fe I, Fe I, O II, O II, Fe I, Fe I, and Fe I/Ca II, as also can be seen by comparing Fig. 2 (a) and Fig. 2 (b).

In order to assess the analytical performance of the detected nitrogen lines by the CT spectrometer, they were used to establish univariate nitrogen concentration calibration curve with baseline-corrected raw spectra and with baseline-corrected and normalized spectra. The baseline correction was performed using a method based on continuous wavelet transform-based peak detection combined with derivative peak width estimation and polynomial fitting of the spectral background [15]. The normalization was performed with the total spectral intensity.

The performances of models were assessed in this work using usual calibration and prediction performance assessment parameters, including coefficient of determination R^2, limit of detection (LOD), average relative error of calibration (REC), average root mean square error of calibration (RMSEC), and relative standard deviation of calibration (RSDC) for the calibration performance; average relative error of prediction (REP), average root mean square error of prediction (RMSEP), and relative standard deviation of prediction (RSDP) for the prediction performance. Detailed definitions of these parameters can be found elsewhere [12]. We recall below their mathematic expressions:

$$REC/P(\%) = \frac{100}{K_{(C/P)}} \sum_{k=1}^{K_{(C/P)}} \frac{1}{L_k} \sum_{l=1}^{L_k} \left| \frac{\hat{c}_{kl(C/P)} - c_{k(C/P)}}{c_{k(C/P)}} \right|, \quad (1)$$

where $K_{(C/P)}$ are respectively the numbers of samples in the training (for calibration C) and test (for prediction P) sample sets, L_k is the number of replicate spectra of sample k, $\hat{c}_{kl(C/P)}$ is the concentration predicted by a model for the replicate spectrum l of the sample k in the training or test set, $c_{k(C/P)}$ is the label concentration of the sample k.

$$RMSEC/P = \sqrt{\frac{1}{K_{(C/P)}} \sum_{k=1}^{K_{(C/P)}} \frac{1}{L_k} \sum_{l=1}^{L_k} \left(\hat{c}_{kl(C/P)} - c_{k(C/P)} \right)^2}, \quad (2)$$

$$RSDC/P = \frac{100}{K_{(C/P)}} \sum_{k=1}^{K_{(C/P)}} \frac{\sigma_{k(C/P)}}{\bar{c}_{k(C/P)}} = \frac{100}{K_{(C/P)}} \sum_{k=1}^{K_{(C/P)}} \frac{1}{\bar{c}_{k(C/P)}} \sqrt{\frac{\sum_{l=1}^{L_k} \left(\hat{c}_{kl(C/P)} - \bar{c}_{k(C/P)} \right)^2}{L_k - 1}}, \quad (3)$$

where $\sigma_{k(C/P)}$ and $\bar{c}_{k(C/P)} = \frac{1}{L_k} \sum_{l=1}^{L_k} \hat{c}_{kl(C/P)}$ are respectively the standard deviation and the mean of the concentrations predicted by the model for the replicate spectra of the sample k in the training or test set.

$$LOD = \frac{3\sigma_n}{s}, \quad (4)$$

where σ_n is the standard deviation of the noise of the spectra, calculated in this work by the average value of the standard deviations of model-predicted concentrations respectively of the three samples with the smallest label concentrations in the training sample set (one for each type of training samples), and s is the slope of the calibration curve.

Fig. 3 shows the behaviors of the calibration models respectively established with the three detected nitrogen lines (N I 742.36 nm, N I 744.23 nm, and N I 746.83 nm lines) with baseline-corrected spectra (a, b, c), and with baseline-corrected and total spectral intensity-normalized spectra (d, e, f). Notice that before the normalization, the spectra were first truncated to keep only the spectral region from 400 nm to 900 nm. The reason is that the spectral region before 400 nm is dominated by the lines from the major elements of the Martian soil simulant, Si, Fe and Al. But the contents of these elements are not constant in the samples, due to the mixture with nitrogen-bearing

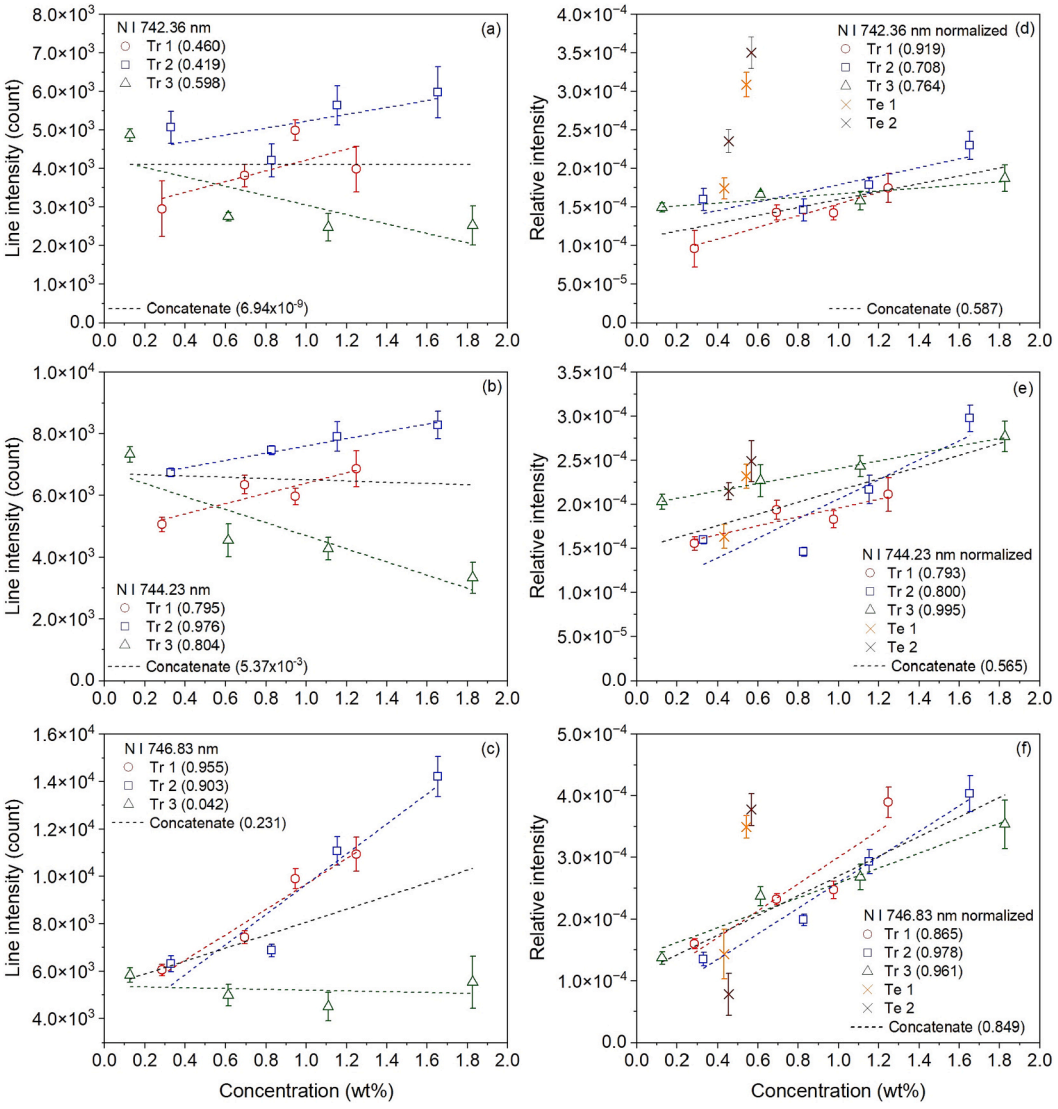

Fig. 3. Behavior of the univariate models established with N I 742.36 nm (a, d), N I 744.23 nm (b, e), and N I 746.83 nm (c, f) lines, using the baseline-corrected spectra (a, b, c), and baseline-corrected and normalized spectra (d, e, f). Are presented for each model, the concentration-intensity data points of the three types of training samples, their respective linear fittings, and the linear fitting of the concatenated training data, together with their respective R^2 values in brackets. For the models with baseline-corrected and normalized spectra (d, e, f), the concentration-intensity data points of the test samples are also presented (Te 1 and Te 2).

compounds at various concentrations. This leads to a notable variation of the intensity of the spectral range before 400 nm, from one sample to another, invalidating it as an internal standard [16]. For each model, are presented in the figure the concentration-intensity data points of the three types of training samples, their respective linear fitting together with the R^2 value, the linear fitting of the concatenated three types of training data, and for the models with normalized spectra, the concentration-intensity data points of the test samples to show the relevance for them to be analyzed by a calibration model (established with one type of the training data or with the concatenated training data). The intensity of a data point corresponds to the average over the 100 replicate spectra of a sample. The error bar corresponds to the standard deviation among the replicates ($\pm SD$). Performance assessment parameters of the models trained with the concatenated normalized data in Fig. 2 (d, e, f) are given in detail in Table 3.

We can see in Fig. 3 (a - c) that raw spectra exhibit in general, a low correlation between the line intensity and the concentration of nitrogen for the three considered lines. A better behavior can be found for N I 746.83 nm line, when the Tr 1 (KNO_3) and Tr 2 ($NaNO_3$) samples are considered. While for the Tr 3 ($NH_4H_2PO_4$) samples, a global trend of negative correlation can be observed for the three lines, with especially a significantly higher intensity for the sample of the lowest N concentration (S3–1). The above behaviors can be explained first by a higher relative intensity of N I 746.83 nm line with respect to two other lines [14]. The presence of alkali metals, K and Na in the Tr 1 and Tr 2 samples favorites absorption of ablation laser pulse, since these elements can be easily dissociated and then ionized, leading to an enhanced laser energy coupling into the sample. On the other hand, the addition of $NH_4H_2PO_4$ in a sample decreases the amount of matrix material of the used Mars simulant soil, which is rich in metals such as iron. The coupling efficiency of laser into a sample thus decreases with the presence of $NH_4H_2PO_4$, leading to a lower plasma temperature. In order to support the explanations, we have calculated the plasma temperature using the Boltzmann plot of a set of Fe I lines shown in Table 4 with the corresponding spectroscopic parameters from the NIST Atomic Spectra Database [14], under the assumption of a plasma in local thermodynamic equilibrium (LTE). The Boltzmann plots for the 4 samples of the Tr 3 type are shown in Fig. 4, which allowed to extract the corresponding temperatures of 7310±460 K, 6240±230 K, 6160±340 K, and 6090±200 K respectively for S3–1, S3–2, S3–3 and S3–4. In the figure, we can see the plot for the sample S3–1 with a clearly smaller (in absolute value) slope than the other plots. This corresponds to the significantly higher temperature deduced for this sample as presented above. On the other hand, the plots corresponding to the other three samples appear quite similar, with a slight increase of the slop for an increasing N concentration. The change is however within the error bars deduced by taking into account the uncertainties on the experimental line intensities. The above observed behaviors qualitatively correspond well to the behaviors of lines intensities observed in Fig. 3. Another reason for the observed trend of a globally negative correlation between the N line intensity and the N concentration would be a less efficient ablation with an increasing maxing rate for the samples of the Tr 3 type. But we had not available experimental data to check this second possibility. The difference in molecular structure of the nitrogen-bearing compounds used for sample preparation leads thus to chemical matrix effect, which makes a clear distinction between the samples of Tr 1 and Tr 2 on one hand, and those of Tr 3 on the other hand, as can been seen in Fig. 3 (a - c). As a consequence, the linear fit of the ensemble of the data of the three types of training samples exhibits a very low R^2 of 0.231, excluding such data to be valid for training of a correct calibration model.

Normalization is known as able to mitigate the influence of matrix effect in quantitative analysis with LIBS [17]. In Fig. 3 (d - f), we can see the behavior of the calibration models trained with normalized spectra. A better R^2 is in general observed for an individual type of training samples, and especially for the models trained with the ensemble of the three types of training data. We also plot the concentration-intensity data points of the two types of test samples in Fig. 3 (d - f), in order to observe their relevance to be taken into account by the calibration models based on the nitrogen lines. We can see in Fig. 3 (d - f) that although the models are improved, their ability to analyze a sample with a different nitrogen-bearing compound in a different matrix, is very limited. In Table 3, we present the calibration and prediction

Table 3
Performance assessment parameters of the univariate models trained with the normalized spectra of the ensemble of the three types of training samples.

Model performance parameters		Used nitrogen lines		
		N I 742.36 nm	N I 744.23 nm	N I 746.83 nm
Calibration	R^2	0.587	0.565	0.849
	LOD (wt%)	1.08	0.539	0.331
	RMSEC (wt%)	0.12	0.11	0.048
	REC (%)	96.9	84.6	21.6
	RSDC (%)	26.6	16.8	16.6
Prediction	RMSEP (wt%)	1.40	0.297	0.324
	REP (%)	537.0	114.9	125.3
	RSDP (%)	14.2	37.8	4.09

Table 4
Fe I lines together with the corresponding spectroscopic parameters used for the Boltzmann plots [14].

Wavelength λ_{ki} (nm)	Transition probability A_{ki} (s^{-1})	Upper and lower level energies E_i (eV) - E_k (eV)	Degeneracy of the upper level g_k
349.78	3.08 × 106	0.11–3.65	2
356.53	4.29 × 107	0.96–4.43	4
360.67	8.29 × 107	2.69–6.13	6
360.89	8.13 × 107	1.01–4.45	2
361.88	7.22 × 107	0.99–4.42	3
372.26	4.97 × 106	0.087–3.42	2
373.49	9.01 × 107	0.86–4.18	5
373.71	1.41 × 107	0.052–3.37	4
374.56	1.15 × 107	0.087–3.40	3
375.82	6.34 × 107	0.96–4.26	3
376.38	5.44 × 107	0.99–4.28	2
376.55	9.51 × 107	3.24–6.53	7
381.58	1.12 × 108	1.48–4.73	3
382.04	6.67 × 107	0.86–4.10	4
404.58	8.62 × 107	1.48–4.55	4
406.36	6.65 × 107	1.56–4.61	3

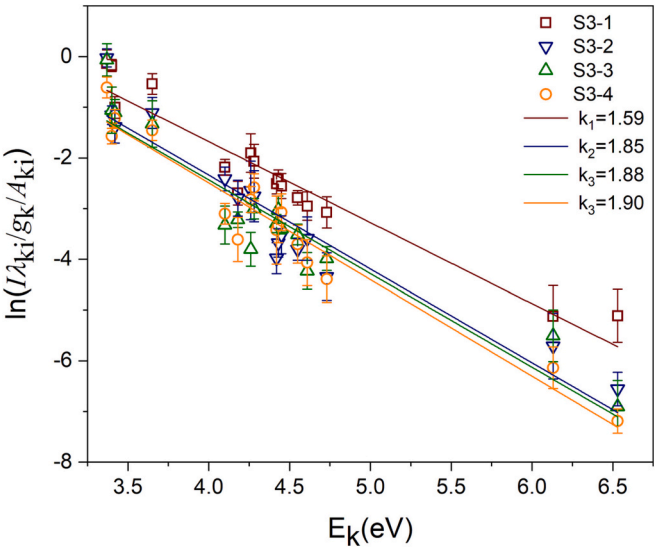

Fig. 4. Boltzmann plots for the 4 samples of the type Tr 3 (S3-1, S3-2, S3-3, S3-4). k_1, k_2, k_3, and k_4 are respectively the absolute values the slopes of the four Boltzmann plots.

performances of the models respectively trained with the three nitrogen lines of the ensemble of the three types of training samples. The models were then tested with the data of the two types of test samples.

3.2. Implementation of multivariate regression

The results in the previous section clearly indicate the insufficiency of the univariate models, with or without normalization, to take into account chemical matrix effect in order to provide a suitable performance to satisfy the targeted application. In our work, a multivariate regression based on machine learning was studied to effectively correct the matrix effect, as the method has been demonstrated capable of such task in our previous works [12,18–21]. In particular we used the configuration of generalized spectrum [12], more specifically optimized to treatment obvious matrix effect due to a clear difference in chemical composition of the samples. The implementation of such method needs two prerequisites: a wide range spectrum including emission lines from the element to be determined and those from other elements contained in the matrix in order to provide information about the element and its chemical environment, which was realized by a fusion of the simultaneously acquired spectra by the CT and the Echelle spectrometers; and a labelling of the spectra according to the chemical speciation of the nitrogen-bearing compound in the corresponding sample, which was realized by a clustering procedure of the spectra using principal component analysis (PCA). The choice of such data treatment method leads to the flowchart for training of the multivariate regression model as shown in Fig. 5.

3.2.1. Spectrum pretreatment

Spectrum pretreatment consisted in successive operations applied to the 200 raw replicate spectra of each sample, a group of 100 from the CT spectrometer and another group of 100 from the Echelle spectrometer. Notice that the spectra with the Echelle spectrometer were truncated to keep the spectral range from 400 nm to 900 nm, as we mentioned above. A spectrum was first baseline-corrected and then normalized with its total intensity as mentioned above. For the two groups spectra respectively, normalized baseline-corrected replicate spectra of a sample were further averaged using the moving average method, where a first ensemble of 40 randomly chosen replicating spectra were first averaged generating a mean spectrum. One of the spectra of the first ensemble was then replaced by a new spectrum from the rest of 60 replicate spectra, leading to a new mean spectrum. Such operation repeated until all the replicate spectra of a sample were involved in the averaging process, yielding a total number of 61 average normalized baseline-corrected replicate spectra for the two spectrum groups of each sample. In the following, they are called pretreated spectra.

3.2.2. Spectrum fusion

Spectrum fusion consisted in replacing the segment from 740 nm to 748 nm of a pretreated Echelle spectrum by the simultaneously recorded corresponding CT pretreated spectrum, generating a fused pretreated spectrum. Such process is illustrated in Fig. 6 with pretreated spectra of the sample S2–4. Fused spectra were separated into two sets of training date with the fused spectra of the training samples, and test data with the fused spectra of the test samples. We can see in the figure that the replaced segment in the Echelle spectrum does not contain any significant lines from nitrogen distinguishable above the noise.

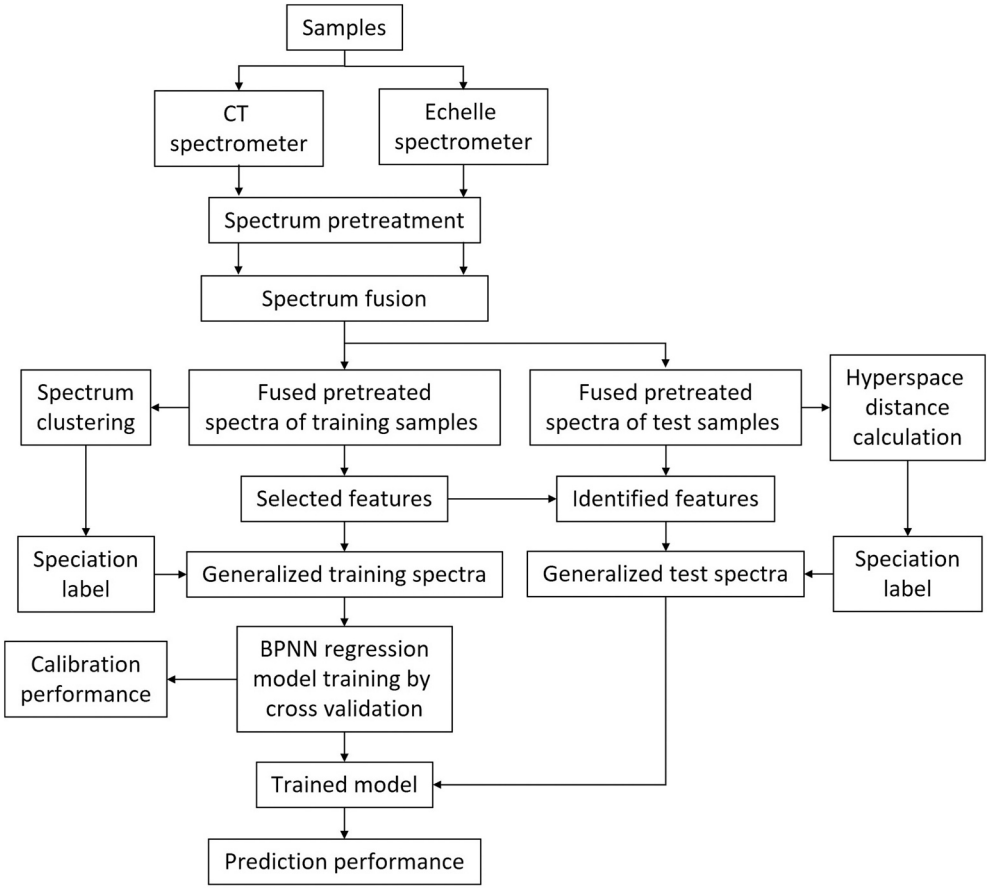

Fig. 5. Flowchart for the buildup of the multivariate calibration model based on back-propagation neural network (BPNN) associated with LIBS spectrum fusion and generalized spectrum.

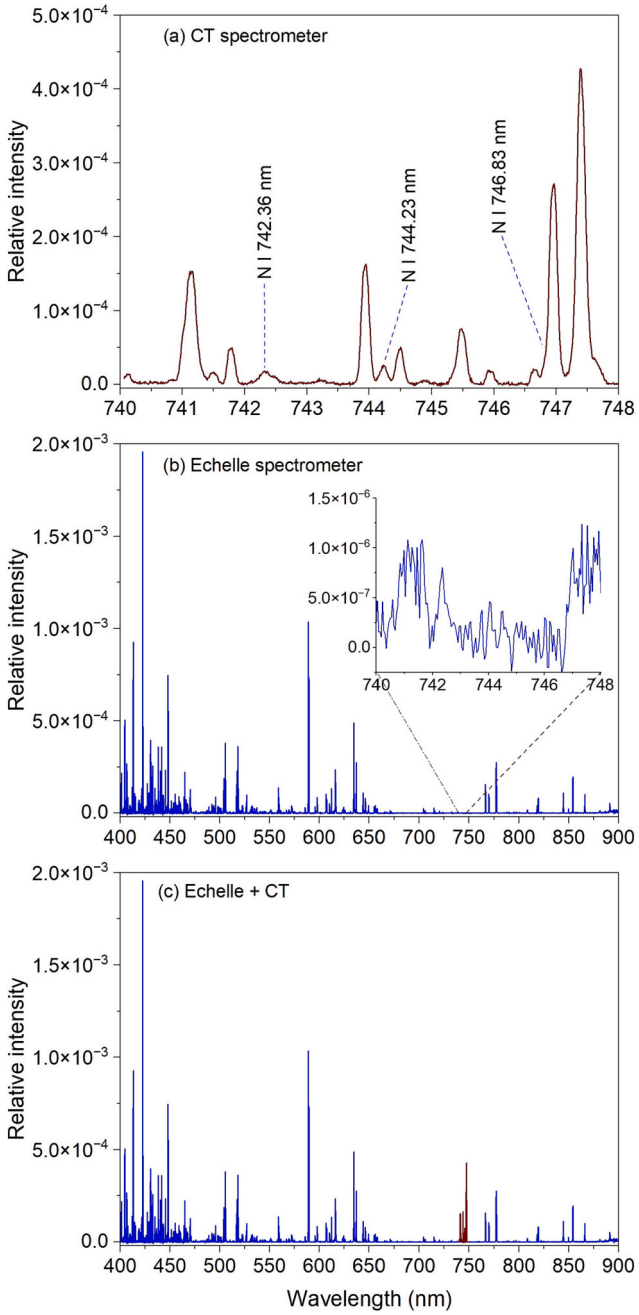

Fig. 6. Spectrum fusion for the sample S2–4, (a) a pretreated CT spectrum, (b) a pretreated Echelle spectrum, (c) the fused spectrum.

3.2.3. Feature selection for the training data and feature identification for the test data

Feature selection was performed for the ensemble of fused spectra of the training samples with SelectKBest algorithm [22], which assigns a score value to each spectral channel according to the covariance of the spectral intensities of the channel over all the spectra of the considered training data set and the nitrogen concentrations of the corresponding training samples. Such score values allow a classification of the spectral channels from the most important ones for the determination of nitrogen, with the highest scores, to the least important ones, with the lowest scores. The selected top 200 important features are shown in Fig. 7. In the figure, we can see that a majority (59.6%) of the selected features correspond to emission lines from elements or molecules. In particular, a part of them, with the highest scores, belongs to the three emission lines of nitrogen detected by the CT spectrometer in the spectral range of 740

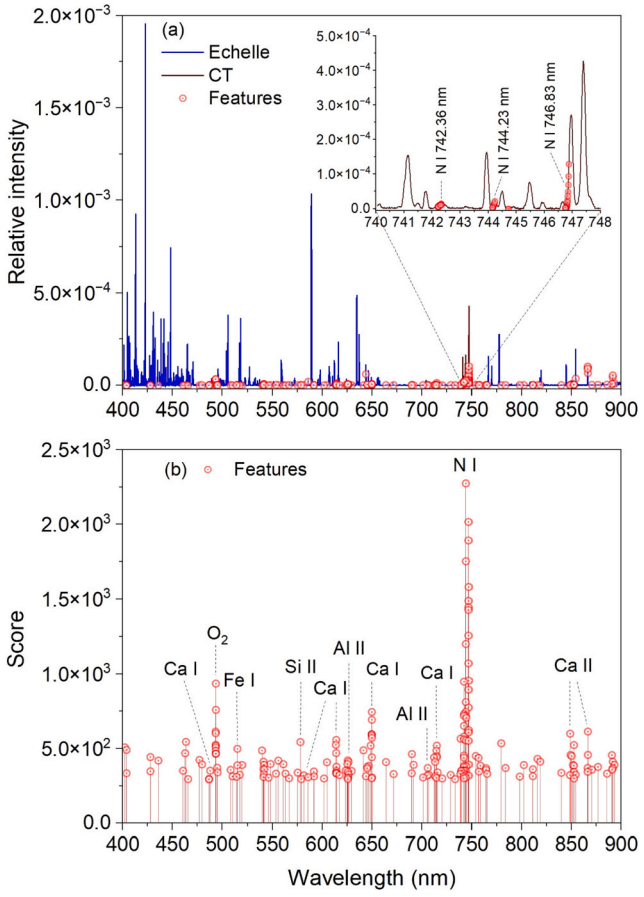

Fig. 7. Results of the feature selection. (a) The 200 selected features in red circles indicated in a fused spectrum of the sample S2–4 in blue (for the Echelle spectrum) and red (for the CT spectrum). The inset shows in detail the selected features for the nitrogen lines. (b) The scores of the 200 selected features with indications of the corresponding elements and molecule. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

nm to 748 nm, as shown in the inset of Fig. 7 (a). The ghost lines in the same spectral range are not included in the selected features in spite of their higher intensities compared to the N I ones, limiting therefore their influence in the data processing. Notice that without spectrum fusion, the spectral range of the Echelle spectrum from 740 nm to 748 nm does not provide any important feature, which underlines the effectiveness of the contribution of the CT spectrum in the selected features. Beyond the emission lines from nitrogen, spectral channels belonging to the oxygen molecule emission band around 493.3 nm [23], received also high scores just below the nitrogen lines as shown in Fig. 7 (b). Oxygen molecules can be formed in a plasma through dissociation of CO_2 followed by O—O recombination [24]. The detected emission band can be thus correlated to the excitation of the ambient gas. Its inclusion in the selected feature brings the information about the temperature of the plasma, which is dependent on the coupling efficiency of the laser pulse to the sample, therefore indicative of the chemical matrix effect. Emission lines from other elements such as Ca, Si, Fe and Al contribute also to the selected features. These elements originate from the matrix (Martian soil simulant) [13]. Their inclusion within the selected features helps the model to learn from the influence of the matrix effects and thereby to correct them. The spectral channels selected as the important features for model training were used to identify the features of the fused pretreated spectra of the test samples for the test of the models. Such feature identification will also be applied to the spectra of unknown samples in future real applications after the spectrum pretreatment and fusion

processes applied to those spectra.

3.2.4. Spectrum labelling with clustering

Univariate models in Fig. 3 show an obvious matrix effect related to different nitrogen-bearing compounds mixed into Martian soil simulant to prepare the samples. Such strong matrix effect can be effectively taken into account with the approach of generalized spectrum [12], where the type of a sample according to its chemical specificity is explicitly integrated in its spectrum in the form of an additional feature appended to those selected in the above section, in such way that the model can directly learn from the sample type of a spectrum. In this work, the type of a sample is assigned by an unsupervised clustering process by looking at the structure of the data without any *a priori* knowledge. We performed therefore a PCA of the ensemble of the fused pretreated spectra of the training samples, and used the two first PCs to present the spectral data in a two-dimensional map as shown in Fig. 8. In the figure we can see a clear clustering of the spectra into two clusters that we can labeled as "a" and "b". The cluster "a" includes the spectra of the Tr 1 and Tr 2 training samples, while the cluster "b" those of the Tr 3 type of training samples. Such clustering result is consistent with the behaviors observed in Fig. 3 with univariate regressions. We therefore appended a label "a" to the features of the fused pretreated spectra of the training samples of types Tr 1 and Tr 2, as the additional feature in the generalized spectrum, while a label "b" was appended to the features of the fused pretreated spectra of the training samples of type Tr 3, as the additional feature. We notice that any other data dimensionality reduction algorithms, tSNE (t-distributed stochastic neighbor embedding) for example [25], can also be used for the clustering of the training data, similar results can be obtained, due to the obvious differences between the spectra of the different types of training samples.

3.2.5. Model training by cross-validation and model calibration performance assessment

The generalized spectra of the ensemble of training samples, with their 200 selected features and an additional one of sample type label "a" or "b", were fed into a back-propagated neural network (BPNN) to train the model under the supervision of the ensemble of the reference nitrogen concentration values of the training samples, to establish the mapping between generalized spectra and nitrogen concentrations [12]. The used neural network had 3 layers, with an input layer of 201 neurons receiving the 200 + 1 features of the input spectra, a hidden layer of 5 neurons, and an output layer of one neuron providing the nitrogen concentration predicted by the model for the input spectrum. The connection weights between the neurons of the successive layers were optimized by gradient descent within a five-fold cross-validation loop performed with a dynamic separation of the training spectra into training and validation data sets [12,26]. Such cross-validation loop generated the calibration assessment parameters of the model. The learning rate, batch size and epoch of the BPNN training are selected as externally adjustable parameters to optimize the performance of model, which were set to 0.001, 32 and 100 respectively in this work. During the training, the model exhibited a fats convergence until the 50th epoch. Reduction of the validation error slowed down afterward.

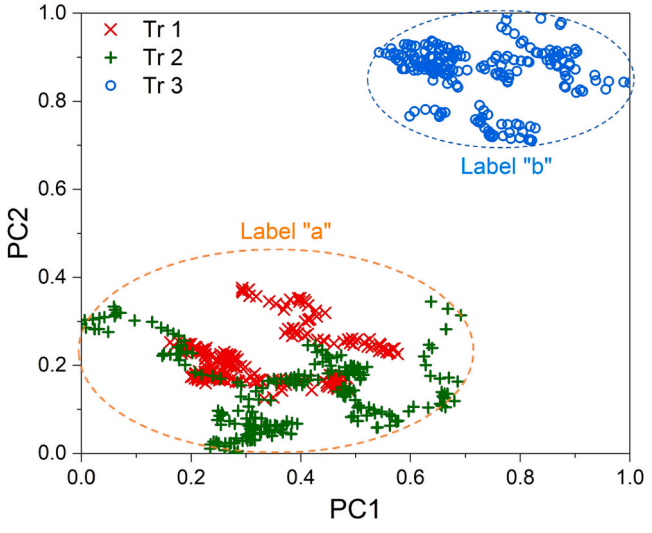

Fig. 8. PCA map of the fused pretreated spectra of the three types of training samples, showing their clustering into two distinct clusters that we labeled as cluster "a" and cluster "b".

3.2.6. Model testing assisted by sample type labelling and model prediction performance assessment

The trained model was tested with the fused pretreated spectra from the test samples not involved in the model training process. A sample type label was first assigned to each of the test spectra. Such label was determined by calculating the distances in the hyperspace between the fused pretreated spectra of the test samples and the training ones. More specifically, note by I_{jklm}, spectral intensity of m^{th} channel of l^{th} pretreated spectrum of k^{th} sample of j^{th} type of samples, then the distance in the hyperspace between two spectra jkl and $j'k'l'$ can be expressed by

$$d_{jkl,j'k'l'} = \sqrt{\sum_{m=1}^{M} \left(I_{jklm} - I_{j'k'l'm}\right)^2}, \tag{5}$$

where M is the dimension of the fused pretreated spectra. The distance in the hyperspace between two samples jk and $j'k'$ can be expressed by

$$d_{jk,j'k'} = \frac{1}{LL'} \sum_{l=1}^{L} \sum_{l'=1}^{L'} d_{jkl,j'k'l'}, \tag{6}$$

where L and L' are the number of pretreated spectra in the samples jk and $j'k'$ respectively. The distance in the hyperspace between two types of samples j and j' can be expressed by

$$d_{j,j'} = \frac{1}{KK'} \sum_{k=1}^{K} \sum_{k'=1}^{K'} d_{jk,j'k'}, \tag{7}$$

where K and K are the number of samples in the type j and j' respectively. The resulted $d_{j,j'}$ are shown in Table 5, where the cases $j \neq j'$ represent the distances between different types of samples, while that of $j = j'$ the dispersion of the data within a given type of samples. In the table, we can see first that the dispersion of a given type of samples is always smaller than the distance between two different types of samples. Then the distance between the two types of test samples Te 1 and Te 2 is the smallest distance among those between different types, indicating a similarity between the two types of test samples. Finally, we find that they are clearly closer to the cluster "a" including the Tr 1 and Tr 2 types of training samples than to the type Tr 3. A label "a" was thus assigned to all the fused pretreated spectra of the test samples.

As mentioned above, spectral features were identified by correspondence of spectral channel with the spectra of the training samples. These spectral features were then concatenated with the sample type

Table 5
Distance in the hyperspace between different types of samples and data dispersion within a given type of samples.

	Tr 1 + Tr 2	Tr 3	Te 1	Te 2
Tr 1 + Tr 2	5.9×10^{-3}			
Tr 3	1.59×10^{-2}	3.2×10^{-3}		
Te 1	1.02×10^{-2}	2.42×10^{-2}	2.9×10^{-3}	
Te 2	1.19×10^{-2}	2.67×10^{-2}	8.2×10^{-3}	2.4×10^{-3}

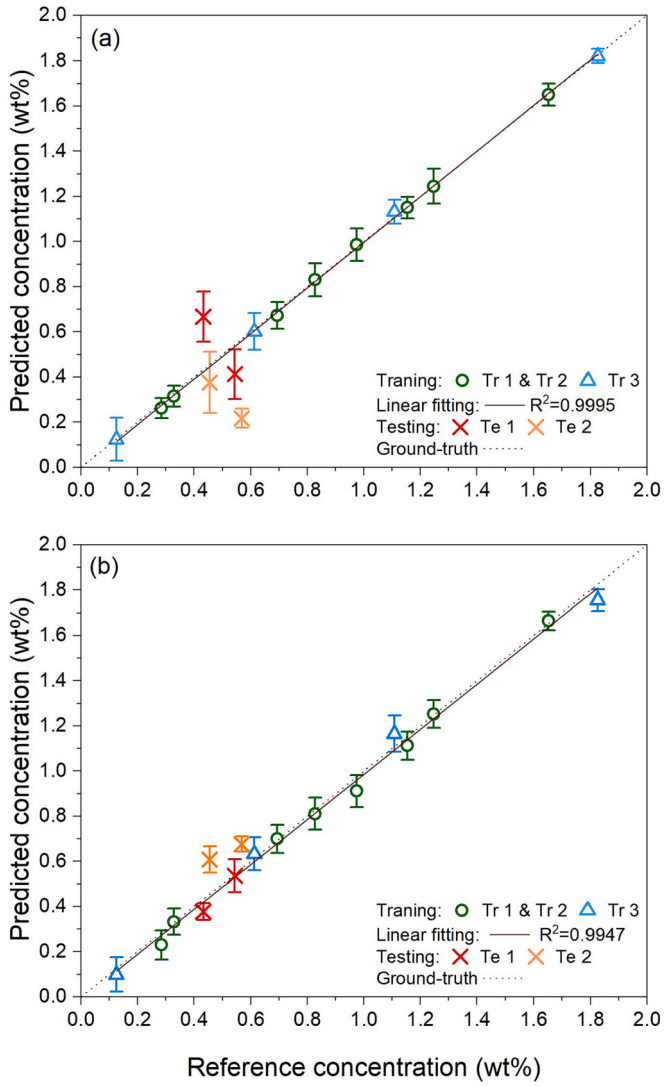

Fig. 9. Behaviors of the multivariate models based on BPNN, trained with the selected features of the fused pretreated spectra of the training samples (a), and with the generalized spectra by including the sample type label in the selected features (b).

label to form a generalized test spectrum. They were fed into the trained model to assess its prediction performance.

3.3. Performances of multivariate regressions

Fig. 9 shows the behavior of trained models. Are presented for each model, the reference concentration *vs* predicted concentration data points of the three types of training samples, the linear fitting of the ensemble of the three types of training data, together with the associated R^2 value, and the reference concentration *vs* predicted concentration data points of the test samples. The predicted concentration of a data point corresponds to the average over the 61 pretreated spectra of a sample. The error bar corresponds to the standard deviation among the pretreated spectra ($\pm SD$). The ground-truth is also plotted in the figures to compare with the linear fitting of the concatenated training data. For a direct observation of the effect of the additional feature in the generalized spectrum, we show in Fig. 9 (a) the behavior of the model trained without additional sample type feature, while in Fig. 9 (b), the behavior of the model trained with the additional feature is presented. The two models were trained with a similar optimization procedure. The calibration and prediction performance parameters extracted from the models are presented in detail in Table 6.

Table 6
Performance assessment parameters of the multivariate models trained with selected features and generalized spectra respectively.

Model performance parameters		Type of the model	
		BPNN with selected features	BPNN with generalized spectrum
Calibration	R^2	0.9995	0.9947
	LOD (wt%)	0.19	0.18
	RMSEC (wt%)	0.0058	0.0093
	REC (%)	5.4	6.2
	RSDC (%)	5.9	6.8
Prediction	RMSEP (wt%)	0.099	0.041
	REP (%)	39.4	16.7
	RSDP (%)	24.7	3.6

In Fig. 9 and Table 6, we can see first that the calibration performance of the models is significantly improved with respect to the univariate models presented in Fig. 3 and Table 4, showing a strong nonlinear fitting ability of a model based on neural network. Such fitting ability is able to take into account the chemical matrix effect due to various chemical speciation of the nitrogen-bearing compounds. A slight weakness of the calibration performance of the model trained with generalized spectra compared to that trained with selected features can be remarked. Look at now the prediction performance of the models, we can observe a general degradation with respect to the calibration performances for the both models. This corresponds to a usual behavior of a machine learning model, because of the differences that can exist between the training and the testing data. It is exactly the situation occurring in our case, the experiment was designed to test the models with independent samples in order for the resulted assessment parameters to be representative of the performance of the models in real applications. As expected, a better prediction performance is observed for the generalized spectrum model with respect to the selected feature model, which represents the capacity of the first to provide more accurate nitrogen concentrations in a real application. This improvement contrasts with the above-mentioned weakness of the calibration performance of the generalized spectrum model compared to the selected feature model. We can understand such contrast by the fact that an intervention in the features based on a physical *a priori* knowledge can perturb the inherent structure of the data, leading thus to a less efficient fitting of them. At the same time, the introduced physical knowledge, being an indication of the physical relation existing between the spectra and the concentrations, reenforces the robustness and the generalizability of the model. We have obtained finally a LOD of 0.18 wt% and a RMSEP of 0.041 wt% with the model trained with generalized spectra, which is much improved compared to the previously reported values, and represents one step forward to detect nitrogen with LIBS on Mars.

4. Conclusion

In conclusion, this work intends to improve the sensitivity and the accuracy of LIBS detection of nitrogen for quantitative analysis of nitrogen in Mars surface materials. The method developed can also be applied to other circumstances where a sensitive and accurate nitrogen determination is required, analysis of farmland soils for example. These applications can meet common difficulties of a weak nitrogen emission and a chemical matrix effect due to various chemical speciation of nitrogen-bearing compounds in analyzed materials. The developed method tackles the problem in the experimental level with a sensitive detection using a CT spectrometer tuned to the nitrogen emission spectral range. Univariate models based on the nitrogen lines present a poor performance, with an obvious matrix effect. A simultaneous detection with an Echelle spectrometer offers the possibility of spectrum fusion, where the spectral range of nitrogen emissions of an Echelle spectrum is replaced by the corresponding CT spectrum. Such fused

spectrum contains optimized information of nitrogen emissions with the CT spectrum, and those of other elements of the matrix with the Echelle spectrum, allowing an effective training of a multivariate model based on machine learning. Notice that the combination of two spectrometers equipped with respectively an ICMOS and an ICCD cameras used in our experiment does not represent the unique way to implement of the method. Any practical solution with a lighter equipment can be a good one provided its equivalent capacity. The merit of our work consists in demonstrating that spectra with different properties can be combined to provide information, directly and indirectly correlated to the element to be determined, to a neural network, in order to ensure the effectiveness and the robustness of a model. The approach of generalized spectrum, particularly relevant for the case of obvious matrix effect was implemented with the help of unsupervised data clustering and hyperspace distance calculation, to further improve the prediction performance of the trained model. Final results show a LOD of 0.18 wt% and a RMSEP of 0.041 wt%, representing a step forward to a sensitive and accurate determination of nitrogen in a complex matrix as Mars soil, required for Mars explorations as well as other applications such as farmland soil analysis for agriculture.

Author statement

Beiyi Zhang: investigation, experiment, data curation, original draft writing. Xiaowen Yu: investigation, sample preparation. Fengye Chen, Li Wang, Yunfei Rao, and Tianyang Sun: investigation, experiment preparation. Chen Sun: data analysis, funding acquisition. Yu-Yan Zhao: conceptualization, supervision, manuscript writing, review and editing. Jin Yu: conceptualization, project administration, supervision, manuscript writing, review and editing. All the co-authors: revision of the paper.

Declaration of Competing Interest

The authors have no conflicts of interest to declare.

Data availability

Data will be made available on request.

Acknowledgements

We acknowledge funding from the Startup Fund for Youngman Research at SJTU, the National Natural Science Foundation of China (11574209, 11805126, 61975190), the Pre-research project on Civil Aerospace Technologies funded by China National Space Administration (CNSA) (No. D020102), the Key Research Program of the Chinese Academy of Sciences (ZDBS-SSW-TLC001).

References

[1] R.S. Harmon, G.S. Senesi, Laser-induced breakdown spectroscopy–a geochemical tool for the 21st century, Appl. Geochem. 128 (2021), 104929.
[2] S. Maurice, S.M. Clegg, R.C. Wiens, et al., ChemCam activities and discoveries during the nominal mission of the Mars Science Laboratory in Gale crater, Mars, J. Anal. At. Spectrom. 31 (2016) 823–1050.
[3] B.L. Ehlmann, F.S. Anderson, J. Andrews-Hanna, et al., The sustainability of habitability on terrestrial planets: insights, questions, and needed measurements from Mars for understanding the evolution of earth-like worlds, J. Geophys. Res. Planets 121 (2016) 1927–1961.
[4] C. Freissinet, D.P. Glavin, P.R. Mahaffy, et al., Organic molecules in the Sheepbed Mudstone, Gale Crater, Mars, J. Geophys. Res. Planets 120 (2015) 495–514.
[5] J.C. Stern, B. Sutter, C. Fressinet, et al., Evidence for indigenous nitrogen in sedimentary and seolian deposits from the Curiosity rover investigations at Gale crater, Mars. PNAS 112 (2015) 4245–4250.
[6] T. Dequaire, P.Y. Meslin, P. Beck, et al., Analysis of carbon and nitrogen signatures with laser-induced breakdown spectroscopy; the quest for organics under Mars-like conditions, Spectrochim. Acta B 131 (2017) 8–17.
[7] M.Z. Martin, S.D. Wullschleger, C.T. Garten Jr., A.V. Palumbo, Laser-induced breakdown spectroscopy for the environmental determination of total carbon and nitrogen in soils, Appl. Opt. 42 (2003) 2072–2077.
[8] R.D. Harris, D.A. Cremers, M.H. Ebinger, B.K. Bluhm, Determination of nitrogen in sand using laser-induced breakdown spectroscopy, Appl. Spectrosc. 58 (2004) 770–775.
[9] D.M. Dong, C.J. Zhao, W.G. Zheng, X.D. Zhao, L.Z. Jiao, Spectral characterization of nitrogen in farmland soil by laser-induced breakdown spectroscopy, Spectrosc. Lett. 46 (2013) 421–426.
[10] S. Maurice, R.C. Wiens, M. Saccoccio, et al., The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: science objectives and mast unit description, Space Sci. Rev. 170 (2012) 95–166.
[11] R.C. Wiens, S. Maurice, B. Barraclough, et al., The ChemCam instrument suite on the Mars Science Laboratory (MSL) rover: body unit and combined system tests, Space Sci. Rev. 170 (2012) 167–227.
[12] C. Sun, Y. Tian, L. Gao, Y. Niu, T. Zhang, H. Li, Y. Zhang, Z. Yue, N. Delepine-Gilon, J. Yu, Machine learning allows calibration models to predict trace element concentration in soils with generalized LIBS spectra, Sci. Rep. 9 (2019) 11363.
[13] X. Zeng, X. Li, S. Wang, S. Li, N. Spring, H. Tang, Y. Li, J. Feng, JMSS-1: a new Martian soil simulant, Earth Planets Space 67 (2015) 1–10.
[14] https://physics.nist.gov/PhysRefData/ASD/lines_form.html.
[15] C.M. Galloway, E.C. Le Ru, P.G. Etchegoin, An iterative algorithm for background removal in spectroscopy by wavelet transforms, Appl. Spectrosc. 63 (2009) 1370–1376.
[16] R.B. Anderson, O. Forni, A. Cousin, et al., Post-landing major element quantification using SuperCam laser induced breakdown spectroscopy, Spectrochim. Acta B 188 (2022), 106347.
[17] J. Guezenoc, A. Gallet-Budynek, B. Bousquet, Critical review and advices on spectral-based normalization methods for LIBS quantitative analysis, Spectrochim. Acta B 160 (2019), 105688.
[18] Y. Zhang, C. Sun, L. Gao, Z. Yue, S. Shabbir, W. Xu, M. Wu, J. Yu, Determination of minor metal elements in steel using laser-induced breakdown spectroscopy combined with machine learning algorithms, Spectrochim. Acta B 166 (2020), 105802.
[19] L. Zou, S. Chen, M. Wu, Y. Zhang, Z. Yue, W. Xu, S. Shabbir, F. Chen, B. Liu, W. Liu, J. Yu, Online simultaneous determination of H2O and KCl in potash with LIBS coupled to convolutional and back-propagation neural networks, J. Anal. At. Spectrom. 36 (2021) 303–313.
[20] W. Xu, C. Sun, Y. Zhang, Z. Yue, S. Shabbir, L. Zou, F. Chen, L. Wang, J. Yu, Accurate determination of structural H2O in rocks using LIBS coupled with machine learning algorithms extensively exploring the characteristics of the Ha line, J. Anal. At. Spectrom. 37 (2021) 317–329.
[21] S. Shabbir, W. Xu, Y. Zhang, C. Sun, Z. Yue, L. Zou, F. Chen, J. Yu, Machine learning and transfer learning for correction of the chemical and physical matrix effects in the determination of alkali and alkaline earth metals with LIBS in rocks, Spectrochim. Acta B 194 (2022), 106478.
[22] T.H. Cormen, C.E. Leiserson, R.L. Rivest, et al., Introduction to Algorithms, Second edition, MIT Press and McGraw-Hill, 2001.
[23] R.A. Back, J.R. Cao, The photochemistry of 1,2-cyclobutanedione in the gas phase, J. Photochem. 33 (1986) 161–171.
[24] M.A. Lindon, E.E. Scime, CO_2 dissociation using the versatile atmospheric dielectric barrier discharge experiment (VADER), Front. Phys. 2 (2014) 55.
[25] B.M. Devassy, S. George, Dimensionality reduction and visualisation of hyperspectral ink data using t-SNE, Forensic Sci. Int. 311 (2020), 110194.
[26] C. Sun, W. Xu, Y. Tan, Y. Zhang, Z. Yue, L. Zou, S. Shabbir, M. Wu, F. Chen, J. Yu, From machine learning to transfer learning in laser-induced breakdown spectroscopy analysis of rocks for Mars exploration, Sci. Rep. 11 (2021) 21379.