



Artificial neural network for on-site quantitative analysis of soils using laser induced breakdown spectroscopy

J. El Haddad^{a,b}, M. Villot-Kadri^c, A. Ismaël^c, G. Gallou^c, K. Michel^d, D. Bruyère^d, V. Laperche^d, L. Canioni^{a,b}, B. Bousquet^{a,b,*}

^a Univ. Bordeaux, LOMA, UMR 5798, F-33400 Talence, France

^b CNRS, LOMA, UMR 5798, F-33400 Talence, France

^c IVEA Solution, Centre Scientifique d'Orsay, Bât 503, 91400 Orsay, France

^d BRGM, Service Métrologie, Monitoring et Analyse, 3 avenue Claude Guillemin, B.P 36009, 45060 Orléans Cedex, France

ARTICLE INFO

Article history:

Received 13 September 2012

Accepted 9 November 2012

Available online 21 November 2012

Keywords:

Laser-induced breakdown spectroscopy (LIBS)

Soil

Quantitative analysis

Artificial neural network

ABSTRACT

Nowadays, due to environmental concerns, fast on-site quantitative analyses of soils are required. Laser induced breakdown spectroscopy is a serious candidate to address this challenge and is especially well suited for multi-elemental analysis of heavy metals. However, saturation and matrix effects prevent from a simple treatment of the LIBS data, namely through a regular calibration curve. This paper details the limits of this approach and consequently emphasizes the advantage of using artificial neural networks well suited for non-linear and multi-variate calibration. This advanced method of data analysis is evaluated in the case of real soil samples and on-site LIBS measurements. The selection of the LIBS data as input data of the network is particularly detailed and finally, resulting errors of prediction lower than 20% for aluminum, calcium, copper and iron demonstrate the good efficiency of the artificial neural networks for on-site quantitative LIBS of soils.

© 2012 Published by Elsevier B.V.

1. Introduction

Laser-induced breakdown spectroscopy (usually called LIBS) is a multi-elemental analytical technique based on the analysis of the atomic spectral lines present in the spectrum of the light emitted by a laser-induced plasma [1]. Nowadays, LIBS is used for both qualitative and quantitative analyses with the possibility of fast and on-site measurements in conjunction with small or in some cases even no sample preparation [2]. Thus, LIBS presents good qualities in the framework of environmental monitoring and more specifically on soil analysis. Several studies have been conducted on this topic by different groups [3] with interesting details on the LIBS instrumentation and on sample preparation. However, let's emphasize that soils are complex and heterogeneous samples and high predictive ability is not expected in the case of on-site analysis. As a consequence LIBS should be considered only as a useful tool for smart sampling of the site, but not as an alternative of the reference ICP-AES laboratory method [3,4]. Moreover, it is well established that LIBS like other analytical methods suffers from matrix effects, i.e. when the physical properties and the chemical composition of the sample affect the signal despite the fact that the analyte concentration remains constant [1] [5]. Matrix effects are evidenced in LIBS not only in the case of

clearly different matrixes (e.g. water, steel, soil, etc...), but also in the case of different matrixes of soils. For example the LIBS emission intensities of barium were found to vary as follows: carbonate > oxide > sulfate > chloride > nitrate [6] in soil. Moreover, the authors reported that no single physical property (enthalpy of formation, vaporization, fusion, heat capacity, entropy...) could explain by itself the characteristics of the plasma emission. In addition, the ablated mass being strongly correlated to the emission, it is easy to understand that many physical and chemical parameters can affect the ablation and consequently the plasma emission. Depending on the nature of soils (calcareous, argillaceous, organic...), the absorptivity can vary over a wide range. Indeed, a mixture of sand containing graphite was found to absorb four times and to provide a LIBS signal for the analyte twice the value of a cellulose sand mixture [6]. In addition, the plasma temperature T_e and the electron density N_e are ideal parameters to probe the differences between the two soil samples due to matrix effects since they are very sensitive to the soil matrix composition [7]. This problem of matrix effect can be usually overcome by the use of matrix-matched samples for the calibration step but unfortunately, this approach can't be applied to soils because of the huge diversity of natural soils. Moreover, sample preparation like dissolution before ICP analysis is efficient to remove the matrix effects but on-site measurements require minimum preparation and consequently this approach is not suitable.

Quantitative LIBS is generally based on the construction of a calibration curve namely the LIBS signal versus the concentration of the

* Corresponding author at: Univ. Bordeaux, LOMA, UMR 5798, F-33400 Talence, France. Tel.: +33 540002870; fax: +33 540006970.

E-mail address: bruno.bousquet@u-bordeaux1.fr (B. Bousquet).

analyte. This method is univariate by definition since the LIBS signal is related to the peak intensity or peak area of the selected spectral line considered as the most relevant to describe the analyte. This simple approach is obviously not suitable to deal with matrix effects and several ways of normalization have been applied to reach better predictive ability. Let's first mention the normalization by an internal standard, namely a selected spectral line related to a chemical element among the major elements of the sample. For example, iron is frequently selected to normalize LIBS spectra of steel samples [8]. In this case, the spectral line chosen for iron may correspond to a transition with an upper energy level close to the one of the transition considered for the analyte. However, the use of an internal standard is not possible in the case of soil samples because there is no chemical element in the soil matrices with constant concentration or even with known different values of concentration [9,10]. Other ways of normalization have been proposed such as normalization by the whole spectrum [7] or by the acoustic signal monitored during the plasma ablation and supposed to be proportional to the ablated mass [11]. Finally, these different solutions of normalization efficiently allow reducing the experimental fluctuations due to physical parameters such as laser energy, focusing conditions, or sample moisture but are not sufficient to remove matrix effects due to the chemical composition of the samples.

Moreover, quantitative LIBS has also been performed without calibration through the so-called calibration-free LIBS [10] algorithm, which is supposed to get rid of matrix effects. This physical model is based on specific hypotheses: conditions of local thermodynamic equilibrium satisfied, optically thin and homogeneous plasma, and stoichiometric ablation. Finally, all the chemical elements must be detected in order to retrieve all the absolute concentrations. Corrections have been applied to the basic algorithm in order to compensate for self-absorption but the question of stoichiometry was not studied in details and remains open. The calibration-free approach was applied to the analysis of soil samples in a preliminary study but the predictive ability was very low and it was decided to continue by using calibration.

With or without calibration, quantitative LIBS also suffers from spectral interferences. This is especially true for soils since many chemical elements can be present in the matrix, some of them with low number of spectral lines (e.g. Al, Si) and some other with a very high number of spectral lines (e.g. Fe, Ti, Mn, Ca) rising up the risk of spectral interference [12]. Consequently, quantitative LIBS remains very challenging and specific studies should be conducted either to take into account the matrix effect through advanced data processing in the case of the approach with calibration or to discuss the conditions of stoichiometry for applying the calibration-free algorithm.

In this context, this paper focuses on advanced quantitative LIBS analysis with calibration. Multivariate analysis was already recommended for the quantitative LIBS analysis of soil samples [9], [13]. Indeed, LIBS quantitative analysis of soils by chemometrics was first introduced by Wisbrun et al. in 1994 [14]. Later, Sirven et al. [15] used partial least square regression (PLS) to quantify chromium in artificially prepared soil samples and obtained a good improvement of prediction of concentrations compared to univariate calibration method. But at the opposite, Essington et al. [12] applied PLS to LIBS and ICP-AES data from natural soils originating from East Tennessee and obtained relative errors of prediction not smaller than the one provided by univariate approach. They also noticed that the analytically viable methods of analysis are defined by an average relative error of prediction below 20%. Moreover, artificial neural network (here after called ANN) was applied to LIBS for qualitative purposes like classification of polymers [16], or identification of soils [17] and also for quantification [18]. Indeed, Sirven et al. [15] quantified chromium in soil samples by ANN applied to LIBS data with interesting discussion on the selection of input data for the ANN, and Ferreira et al. [19] reported quantitative LIBS analysis of copper in soils by ANN.

In this paper, the principle of a 3-layer artificial neural network in the specific case of quantitative LIBS is explained, and the selection of the spectral lines in order to obtain the best predictive ability for the ANN is detailed. The advantage of using the ANN for quantitative LIBS is illustrated in this paper in the case of selected chemical elements present in natural soil samples and analyzed directly on-site with a transportable LIBS system.

2. Artificial neural networks for laser-induced breakdown spectroscopy

It is well established that the behavior of the LIBS signal as a function of the concentration of a given analyte can often be nonlinear, especially due to self-absorption effect. As a consequence, linear regression might become insufficient and a nonlinear model might become necessary to efficiently process LIBS spectra for quantitative analysis. Artificial neural network (ANN) is a network of interconnected neurons characterized by nonlinear activation functions. So, it is by definition the ideal tool to describe nonlinear behaviors [20]. In this work, a simple 3-layer perceptron architecture was chosen. Selected LIBS data were feeding the input layer while the output layer was in charge to generate the value of the predicted concentration of the analyte. Indeed, the ANN was exploited here to predict the concentration of one chemical element at a time. The hidden layer was present for interconnecting the input and output layers. The 3-layer perceptron is recognized to be very efficient for quantification purposes [15,19,21]. Fig. 1 is a general scheme of the ANN used in this study. Each neuron compares the sum of all the weighted input signals connected to it to a value of bias, namely an activation threshold and generates an output signal by applying the sigmoid function to the results of this calculation. This algorithm of forward propagation starts with random values of weights and bias. Let's emphasize that the dataset is split into two subsets: the calibration set used to train the network and the validation set used to check the predictive ability of the network. Starting with the calibration set, the first output value generated by the ANN is compared to the actual value and the error between the actual and the predicted values is used in a back-propagation algorithm to adjust the values of the weights and bias for a second iteration and so on. Then, the two – forward and backward – algorithms are consecutively repeated over all the data (from all the samples of the calibration set) in order to achieve a good training of the ANN. The well-known statistic tool called cross-validation is finally exploited to evaluate the predictive ability of the ANN. It consists in feeding the ANN with the data of both the calibration and the validation sets. Then, the concentration of each sample from these two subsets is calculated by the ANN and the predictive ability is evaluated by the calculation of the root mean square error (RMSE) values for each subset, defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (c_i - \hat{c}_i)^2} \quad (1)$$

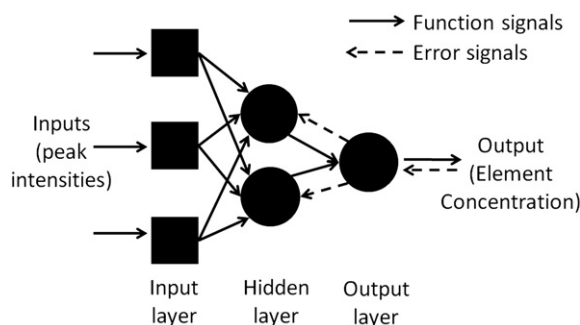


Fig. 1. Schematic of the three-layer artificial neural network used for LIBS analysis with the two basic flows of learning: forward for function signals and backward for errors signals.

where c_i and \hat{c}_i are the actual and the calculated values of concentrations, respectively, and N is the number of samples. RMSE can be calculated either for the calibration (RMSEC) or the validation (RMSEP) sets of data.

The cross-validation algorithm is applied to optimize the parameters of the ANN namely the number of neurons in the hidden layer, the speed of convergence and the memory of the network, and to stop early the training before over-fitting [20]. RMSE is frequently chosen to evaluate the predictive ability of the models of calibration; however it provides an absolute value, here in mg/kg for the concentrations. The best model corresponds to minimum values of RMSEC and RMSEP simultaneously. In this study, two other figures of merit were used instead of RMSE to evaluate the predictive ability of the ANN, namely the average relative error of calibration REC (%) resulting on a calculation over the calibration set and the average relative error of prediction REP (%) based on the validation set [15,22], and defined as follows:

$$\text{REC}(\%) = \frac{100}{N_c} \sum_{i=1}^{N_c} \frac{|c_i - \hat{c}_i|}{c_i} \quad (2)$$

$$\text{REP}(\%) = \frac{100}{N_p} \sum_{i=1}^{N_p} \frac{|c_i - \hat{c}_i|}{c_i} \quad (3)$$

where N_c and N_p are the numbers of samples in the calibration and prediction (or validation) sets, respectively. REC and REP were preferred to RMSE because they provide percentage instead of absolute values. Moreover, RMSE is strongly influenced by the highest values of concentrations, which is not the case for REC and REP. Consequently, even if the figures (REC, REP) evaluate the predictive ability of the models in a similar way as RMSE, they finally offer an easier interpretation over the whole range of concentrations. Furthermore, while the value of RMSE, here in mg/kg, doesn't allow direct conclusion, the percentage values of REC and REP can be directly compared to percentage reference values.

3. Experimental

3.1. Sample preparation

On-site quantitative LIBS analyses were conducted on an old mining site located in the South of France which potentially offered a wide range of concentrations for different chemical elements and which is especially known for presenting high concentrations of lead and zinc as well as barium and calcium. From a smart sampling assisted by the use of a portable XRF device (Niton XL3t800, Thermo Scientific), 70 samples (soils, ore concentrates and ore dumps) were collected around and surrounding the site covering a wide range of heavy metal rate.

Each sample was sifted at 2 mm maximum grain size in order to ensure the fast elimination of leaves, roots and gravels. Indeed, gravels in soil sample are not contaminated but may false the results of on-site LIBS analysis by a nugget effect. Then each sample was split into two parts, one dedicated to direct LIBS analysis and the other one to later laboratory ICP-AES analysis supposed to provide the actual values of concentration in order to evaluate the predictive ability of LIBS. The part of each sample dedicated to LIBS was completely dried with the use of a microwave oven in order to remove any bias on the LIBS signal due to moisture. Indeed, the higher is the moisture level the lower is the LIBS signal [23]. Finally the dried soils were prepared as pressed pellets of 13 mm diameter by applying 8 tons/cm² for 2 min with a manual press.

3.2. Transportable laser-induced breakdown spectroscopy setup

A commercial instrument was used for on-site LIBS measurements of soils (MobilIBS III, IVEA SAS). This complete system used a quadrupled

Nd:YAG laser at 266 nm running at 20 Hz repetition rate, with a pulse duration of 5 ns FWHM. In this study, the laser energy was optimized at the fixed value of 4.2 mJ and then the resulting irradiance was 43 GW/cm². This optimal irradiance enabled to obtain high sensitivity for the elements of interest and good repeatability for plasma-sample interaction for the soil samples. The light emitted by the plasma was collected with a patented achromatic telescope and injected in a 3-meter fused silica optical fiber of 550 μm diameter. An Echelle spectrometer was used with a resolution power $\lambda/\Delta\lambda$, close to 4000 from 200 to 900 nm, and equipped with an intensified CCD camera. The software AnaLIBS (IVEA SAS) was used to control the experimental parameters and for the data processing. The MobilIBS system was integrated in a van, as a mobile laboratory, in order to allow on-site LIBS measurements.

Each LIBS spectrum was the result of 20 laser shots accumulated at the same point of the sample, with a gate delay of 300 ns and a gate width of 3 μs. In order to estimate the macroscopic composition of each analyzed sample and to reduce the effects of heterogeneity, 64 spectra were acquired for each sample and one single average spectrum was finally calculated for each sample. This average spectrum was used for quantitative analysis. Let's point out that statistics was however calculated by running five times the ANN calculation with different initial random values of the weights but always with the same input spectra.

4. Results and discussion

Since soil samples are highly heterogeneous, a localized LIBS analysis at the microscopic scale is a non-sense when the chemical composition of the whole sample is investigated. Consequently, it's necessary to average many micro-scale LIBS spectra in order to obtain an effective spectrum closer to the reality of the sample constituents. In this work, each LIBS spectrum exploited for quantitative analysis was the result of averaging 64 LIBS spectra which are taken at different locations of the pellet surface, as detailed in the previous section. In parallel to the LIBS analysis, the same samples were analyzed by ICP-AES in order to get values of concentrations considered as the actual values. Processing these ICP-AES data through Principal Component Analysis (PCA) confirmed that the two major elements for these soils were calcium and barium. However, the samples were spread in the plane of the principal components and no small-size cluster was observed. This result demonstrated that the geological site analyzed during this campaign could not be separated into distinguishable sub-areas with clear different characteristics but at the opposite had to be considered as a whole single site, even if the concentration in major elements varied a lot from one location to the other.

The 3-layer ANN developed to process the LIBS data from the soil samples was built from the cross-validation method, as discussed above. Consequently, our data set was divided into the calibration set (3/4 of total data) and the validation set (1/4 of total data). The calibration set was used to build the quantitative model while the validation set was used to test the model. Let's emphasize that a specific ANN was built for each chemical element. As a consequence, the calibration and validation sets were also different from one chemical element to the other in order to exploit the whole range of concentrations in each case. This process has no negative consequence on the quantification of each chemical element.

4.1. Quantitative analysis by standard linear calibration

Soils are complex samples containing many chemical elements and thus related to LIBS spectra characterized by hundreds of atomic lines. In the case of aluminum which is one of the major elements for these soils (in the concentration range of % i.e. tens of g/kg), four persistent lines of Al I (308.215; 309.271; 394.400; 396.152 nm) were detected and identified thanks to the NIST atomic database [24]. For

each of these lines, the value of the correlation coefficient R^2 after a calibration with regular linear least squares regression was found to be between 0.01 and 0.2 which demonstrates that there was no linear correlation between the LIBS signal and the concentration of aluminum. Fig. 2a displays the result of the linear regression obtained with the Al I-line at 309.271 nm. Here, the 64 original spectra for each sample were used to build 10 average spectra (each new spectrum was the result of an average of 6 or 7 spectra) and the error bars correspond to the standard deviation of these ten spectra. In the case of copper which is a trace element (in the concentration range of mg/kg) two significant lines of Cu I (324.754; 327.396 nm) were exploited from the same NIST atomic database. Again for these two lines, the linear correlation was too low with $R^2 = 0.79$ for the 324.754 nm line and $R^2 = 0.81$ for the 327.396 nm line (see Fig. 2d). In order to emphasize the fact that these correlation factors were not sufficient for quantitative analysis, relative errors of calibration (REC) and prediction (REP) were calculated for the 327.396 nm line. The results were REC = 43% and REP = 37%, far above the value of 20% discussed earlier and considered by ourselves as the limit value for on-site analysis. In the case of iron which was one of the major elements of these soil samples, tens of atomic lines were identified from the NIST atomic database but again the linear correlation for each of them was poor. As an example, the Fe I line at 404.581 nm provided the results reported hereafter: $R^2 = 0.75$, REC = 48%, and REP = 46% (see Fig. 2b) while the Fe I line at 375.823 nm gave $R^2 = 0.7626$, REC = 50%, and REP = 48%. Here again, the univariate quantitative analysis of iron seems to be difficult. Our last example to illustrate quantitative LIBS through regular linear calibration method is dedicated to calcium. In these soils, calcium was one of the major elements and LIBS provided a series of spectral lines identified from the NIST atomic database. Compared to other chemical elements,

the results obtained for calcium were much better with for example for the Ca I line at 610.272 nm $R^2 = 0.88$, REC = 26%, REP = 24% (see Fig. 2c). This relatively good result might be related to the fact that calcium was one of the two most influencing elements, like barium, as revealed by the PCA results. However, despite of the better correlation observed with calcium, the relative errors were still too high for quantitative analysis with acceptable predictive ability. This preliminary study motivated us with working on chemometrics in order to achieve quantitative LIBS analyses of soil samples. Artificial neural network (ANN) was chosen as the most promising approach to overcome both matrix effects and nonlinearities in the calibration.

4.2. Quantitative analysis by artificial neural network

In this study, artificial neural network (ANN) was applied to LIBS data through the algorithm of the Igor Pro 6.11 software as already reported in Ref. [15]. The first step consisted in training the 3-layer artificial neural network starting with random values of weights and bias. Here it's worth pointing out that because the initial values were randomly chosen, the output value of the ANN was never exactly the same while the calculation was repeated, even if the input values were kept unchanged. To overcome this specific characteristic of fluctuation with the ANN, the cross-validation step was always repeated five times for each model in order to ensure that the best learning parameters were chosen, namely the number of hidden nodes, learning rate, momentum, and number of iterations. Repeating the cross-validation step five times also permitted to select the input data from the LIBS spectra offering the best quantitative predictive ability using ANN model. All the ANN parameters and the input data were chosen by minimizing both the average value of the error of prediction and its standard deviation over the five repetitions.

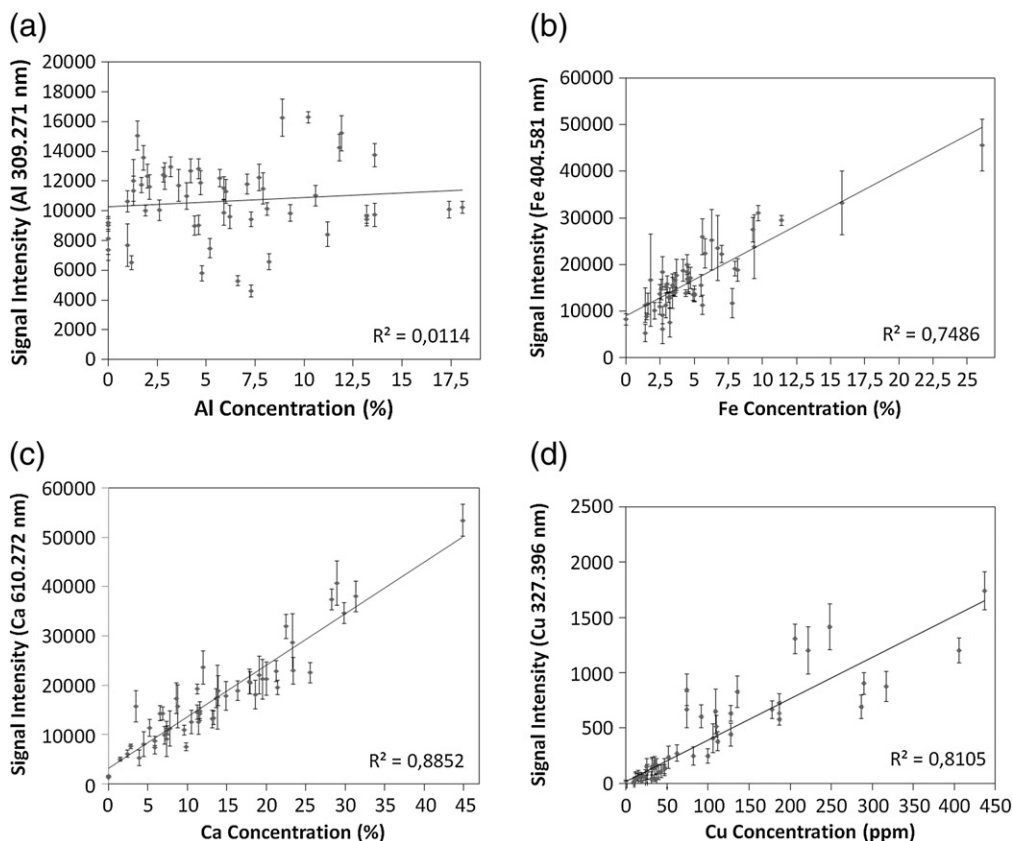


Fig. 2. Univariate calibration curves, namely LIBS intensity vs. concentration, for Al (a), Fe (b), Ca (c), and Cu (d) in the case of soil samples. Error bars represent standard deviations detailed in the text.

The input data of the ANN were the peak intensities of the LIBS spectra at the selected wavelengths. In order to normalize these input data prior to the processing through the ANN, each of them was divided by the maximum value calculated over all the spectra at the same wavelength. For aluminum, four input data were selected, corresponding to the four persistent lines already discussed in the previous section of this paper. Including these four lines at the same time as input data of the ANN, the relative errors of calibration and prediction were found to be $REC = (29.1 \pm 2.2)\%$ and $REP = (29.1 \pm 3.8)\%$, respectively. Let's emphasize here that the values of standard deviation were calculated after five repetitions with the same input data but different initial random values of weights and bias. But, even if this result was much more satisfying than the one obtained with the univariate approach described in the previous section, it was still not sufficient for on-site quantitative analysis of soils regarding the criterion of 20% already discussed.

For calcium, more than 40 lines were detected in the LIBS spectra in the range 250–800 nm. Introducing 42 lines as input data of the ANN, the relative errors were found to be $REC = (13.2 \pm 0.6)\%$ and $REP = (15.1 \pm 0.3)\%$. In this case, both the training and the prediction ability of the ANN were satisfying. However, reducing the number of input data is of major interest to avoid redundancy and to reach the best predictive ability of ANN with the minimum input data. Consequently we proposed the following approach to select the smallest number of the most relevant spectral lines: 1 – starting with the list of persistent lines provided by the NIST atomic database [24], 2 – remove from the list the lines with fundamental level as lower level of the transition in order to reduce the self-absorption effect mainly observed with these lines, 3 – remove from the list the lines characterized by low signal-to-noise ratio, namely below a threshold value that can be adjusted from case to case, and 4 – if no line is listed after the first three steps, restart the process canceling the step 2 since if only resonant lines are detected, the data processing must be based on them. According to this approach applied to our LIBS spectra of soil samples, the number of input data for the ANN in the case of calcium analysis decreased from 42 to 10 (Table 1) while the predictive ability was slightly increased with $REC = (11.2 \pm 0.4)\%$ and $REP = (13.3 \pm 0.2)\%$. This clearly demonstrated that the number of input data could be easily reduced without affecting the predictive ability of the ANN and Fig. 3 reports the results of the predicted concentrations of the validation set by ANN compared to the reference values given by ICP-AES. This result provides much more details than the REC and REP values. Indeed, it clearly demonstrates that this ANN model was well suited for the whole range of concentrations.

The same approach was then applied to iron and the number of input data was reduced to 11, reported in Table 1. In this case, the relative errors were found to be $REC = (15.3 \pm 1.1)\%$ and $REP = (20.1 \pm 1.2)\%$ that is very close to the threshold of 20%, our target for on-site analysis.

For copper, the fourth step of the approach was applied since only two resonant lines were detected, namely Cu I at 324.754 nm and Cu I at 327.396 nm. Thus, by introducing these two lines inside the ANN,

Table 1

Wavelength (nm) of the spectral lines of Al, Ba, Ca, Cu, Fe and Ti selected for LIBS analysis by ANN.

Element	Wavelength (nm) of selected spectral lines
Al	308.215, 309.271, 394.400, 396.152
Ba	652.731, 659.532, 669.384, 705.994, 728.029
Ca	442.544, 443.568, 445.478, 558.875, 610.272, 612.221, 616.217, 643.907, 646.256, 649.378
Cu	324.754, 327.396
Fe	278.81, 358.119, 373.486, 374.556, 374.826, 374.948, 375.823, 382.042, 388.628, 404.581, 438.354
Ti	365.349, 375.285, 395.633, 395.82, 399.863, 498.173, 499.106

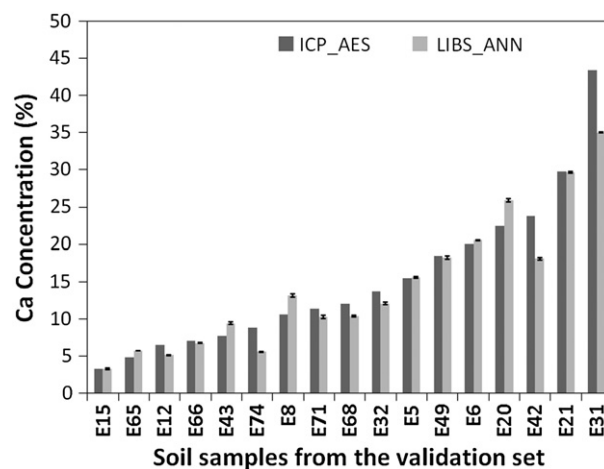


Fig. 3. Comparison chart of the Ca concentration (%) measured by ICP-AES (black) and by LIBS and ANN (gray) for the validation set of soil samples. Error bars correspond to the standard deviation of five repetitions of the ANN calculation (same input data, different initial random weights). The ANN parameters are: number of hidden nodes = 4, learning rate = 0.1, momentum = 0.1, number of iterations = 19,000.

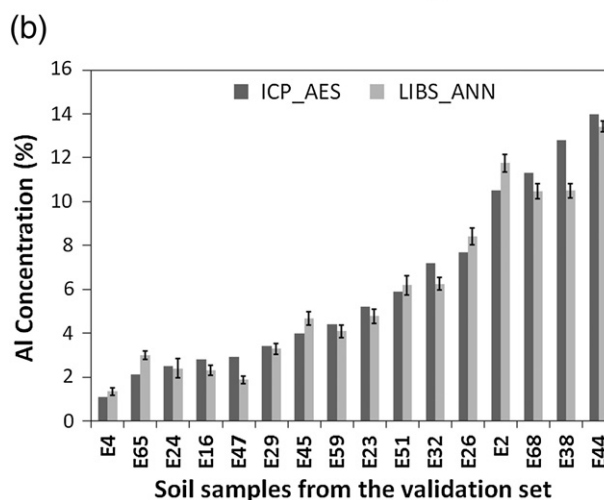
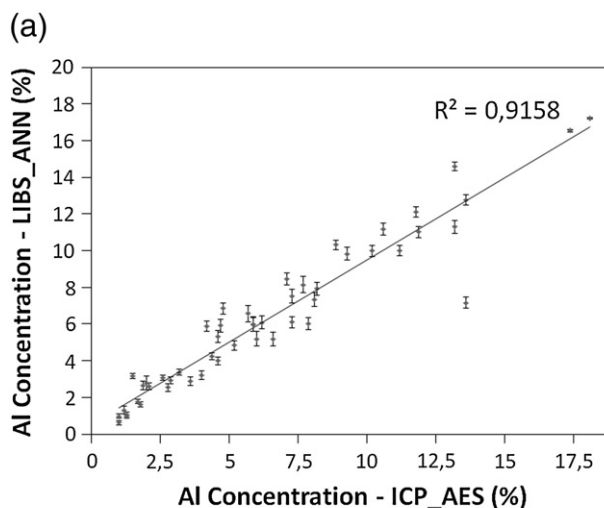


Fig. 4. Results for the Al concentrations with an ANN model based on Al and Ca spectral lines as input data. The ANN parameters are: number of hidden nodes = 8, learning rate = 0.1, momentum = 0.2, number of iterations = 7000. (a) Correlation between Al concentrations measured by LIBS and ANN and by ICP-AES for the calibration set. (b) Comparison chart of the Al concentration (%) measured by ICP-AES (black) and by LIBS and ANN (gray) for the validation set. Error bars correspond to the standard deviation of five repetitions of the ANN calculation (same input data, different initial random weights).

Table 2

Average relative errors of calibration (REC) and prediction (REP) in the case of quantitative analysis of copper in soils by LIBS for different sets of input data of the ANN. The number of nodes in the hidden layer is also given.

Output	Input element	Number of input	Number of hidden nodes	REC (%)	REP (%)
Cu	Cu	2	5	35.5 ± 3.5	31.9 ± 4.4
Cu	Cu, Ca	12	3	22.3 ± 9.7	25.8 ± 7.4
Cu	Cu, Ba	7	6	27.3 ± 2.7	27.7 ± 3.1
Cu	Cu, Fe	13	5	16.6 ± 1.9	17.4 ± 1.2
Cu	Cu, Ti	9	6	19.1 ± 3.3	20.2 ± 1.4
Cu	Cu, Ba, Ca, Fe, Ti	33	7	12.5 ± 0.8	13.7 ± 1.5

the relative errors obtained were REC = (36.2 ± 3.5)% and REP = (32.3 ± 4.4)% which was above the threshold of 20% and consequently not acceptable for on-site analysis of copper. As an intermediate conclusion, let's point out that quantitative LIBS below 20% of error was achieved using ANN for two major elements, namely iron and calcium, but not for aluminum, also a major element nor for copper which was a trace element. Consequently, instead of introducing in the ANN input data coming from only the chemical element to be quantified, namely the analyte, we studied the advantage of introducing other input data from other chemical elements in order to better take into account the well-known matrix effects.

This new advanced method was tested first on aluminum, which means that the ANN was used to quantify the concentrations of aluminum and input data were constituted not only by Al lines but also by Ca lines. More precisely, the 4 lines of aluminum discussed above were introduced in the ANN and the 10 lines of calcium (resulting of the selection described above) were also introduced. In this case, the relative errors were found to be REC = (17.1 ± 1.2)% and REP = (15.3 ± 1.6)%, which should be compared to the original values of REC = (29.1 ± 2.2)% and REP = (29.1 ± 3.8)%, with only input data from aluminum. This clearly demonstrates that the introduction of 10 spectral lines of calcium offers a significant advantage to the quantitative analysis by ANN. Fig. 4a displays the results of the predicted concentrations for samples belonging to the calibration set. This clearly shows a good correlation between the concentrations predicted by ANN and the reference concentrations of Al obtained by ICP-AES. Fig. 4b displays the concentrations predicted by ANN for samples belonging to the validation set, namely sample not used to build the ANN model, compared to the reference values of ICP-AES.

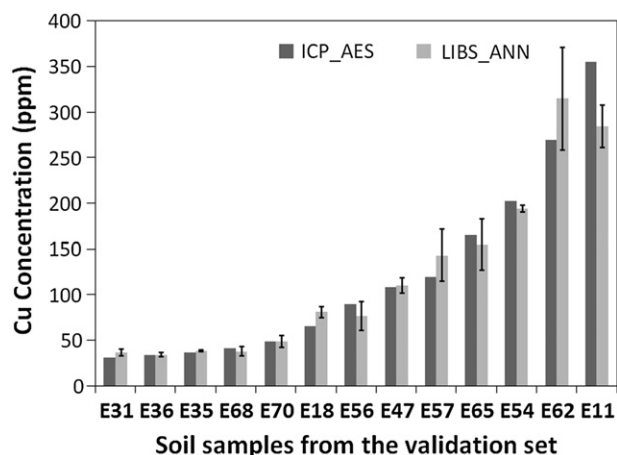


Fig. 5. Comparison chart of the Cu concentration (ppm) measured by ICP-AES (black) and by LIBS and ANN (gray) for the validation set of soil samples. The ANN model is based on selected (Cu, Ba, Ca, Fe, Ti) lines and the ANN parameters are: number of hidden nodes = 7, learning rate = 0.4, momentum = 0.1, number of iterations = 14,000. Error bars correspond to the standard deviation of five repetitions of the ANN calculation (same input data, different initial random weights).

Table 3

Average relative errors of calibration (REC) and prediction (REP) in the case of quantitative analysis of aluminum, calcium and iron in soils by LIBS after optimizing the input data of the ANN. The number of nodes in the hidden layer is also given.

Output	Input element	Number of input	Number of hidden nodes	REC (%)	REP (%)
Al	Al, Ca, Ba, Fe, Ti	37	11	18.7 ± 0.8	19.3 ± 2.1
Ca	Ca, Ba, Fe, Ti	33	6	9.4 ± 0.4	15.2 ± 0.8
Fe	Fe, Ba, Ca, Ti	33	11	15.5 ± 0.6	16.8 ± 0.9

One can conclude that the advanced selection of input data for the ANN applied here appears to be very efficient to overcome the matrix effects since the predicted values are really close to the reference ones over the whole range of concentrations.

The same protocol was repeated for the analysis of copper. In this case, the 10 lines of calcium were added to the 2 lines of copper to build the input data set of the ANN. But unfortunately, the advantage was not significant in this case, (see Table 2, line 2). Thus, it was decided to add other input data based on spectral lines of other chemical elements, namely iron, titanium and barium. For these three elements, the spectral lines were selected according to the standard approach described above for calcium. The selected lines for Fe, Ti and Ba are reported in Table 1. The addition of extra lines from different elements always resulted in better predictive ability of the ANN (see Table 2) and the lines related to iron provided the best improvement with REP = (17.1 ± 1.2)%. And finally, the addition of spectral lines from the four matrix elements (Ca, Fe, Ti, Ba) significantly improved the quantitative analysis of copper with REP = (14.2 ± 1.5)%. The results of the predicted values of the copper concentration for samples belonging to the validation set are displayed in Fig. 5. This demonstrates the existence of an additive effect of the extra input data to improve the prediction ability of the ANN.

Following the same approach, the ANN provided values of REP below 20% for Al, Ca and Fe (Table 3). However, adding extra input data was not always the best solution as one can conclude in the case of Ca and Al. Moreover, in the case of Fe, the improvement was really clear as displayed in Fig. 6. Indeed, the predicted values of concentration for samples belonging to the validation set were found to be very close to the values provided by ICP-AES. So finally, for this specific geological site, the use of selected lines of (Ca, Fe, Ti, Ba) as input data of the ANN was demonstrated to be quite efficient for the quantitative analysis of major elements such as Fe, Ca, Al as well as trace elements such as Cu.

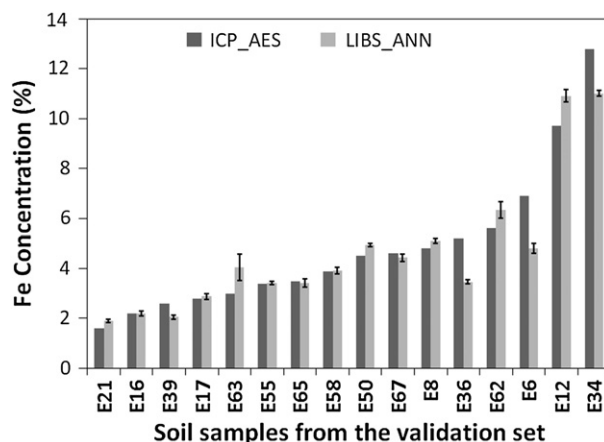


Fig. 6. Comparison chart of the Fe concentration (%) measured by ICP-AES (black) and by LIBS and ANN (gray) for the validation set of soil samples. The ANN model is based on selected (Ba, Ca, Fe, Ti) lines and the ANN parameters are: number of hidden nodes = 11, learning rate = 0.05, momentum = 0.02, number of iterations = 7000. Error bars correspond to the standard deviation of five repetitions of the ANN calculation (same input data, different initial random weights).

5. Conclusion

A transportable LIBS system was successfully exploited for on-site LIBS measurements of soil samples. Quantitative analysis was achieved by artificial neural network (ANN) to overcome the matrix effects and the nonlinear behavior of the calibration. Univariate analysis was demonstrated to be unsatisfactory for on-site analysis thanks to a series of demonstrations and ANN was successfully applied to selected LIBS data. It was demonstrated that the spectral lines from the analyte were not sufficient as input data of the ANN to reach good predictive ability and that extra lines from other chemical elements were necessary to improve the prediction ability of the ANN.

One single ANN model was built for each single chemical element to be analyzed. With the ANN, the relative error of prediction (REP) was found to be below 20% for matrix elements like Ca and Fe, for major element like Al (in the % range) and also for trace element like Cu (in the mg/kg range). This predictive ability for quantitative analysis is perfectly suitable in the framework of on-site quantitative analysis of soils. In addition, let's also emphasize that the well-known problem of overfitting was carefully controlled during this work via the average relative error of calibration (REC) value. Indeed, when the value of REC was very low compared to those of REP, overfitting was evidenced and the related ANN model was rejected.

Our future work will be organized in different steps. First, the ANN built after a first campaign will be used for quantitative LIBS analysis during a second campaign on the same site. This is to check the sensitivity of the ANN to different experimental conditions and to deal with independent data. Indeed, in the work reported here, the two data sets used for calibrating and testing the ANN were not independent because the samples were from the same site and the spectra were recorded during the same campaign. Testing the ANN calculated here on data coming from a second campaign on the same site is consequently of major interest. The other axis of our work consists in using the ANN model built on this specific site to the analysis of samples extracted from another site. This is to study the sensitivity to the matrix effects. These two steps are necessary to go ahead with on-site LIBS analysis of soils by the use of ANN.

Acknowledgments

The authors gratefully acknowledge the French Environment and Energy Management Agency (ADEME) for its financial support in the framework of the CALIPSO project dedicated to the development of a transportable LIBS instrument and software for on-site quantitative LIBS analysis of soils.

References

- [1] A.W. Miziolek, V. Palleschi, I. Schechter, *Laser-Induced Breakdown Spectroscopy (LIBS) Fundamentals and Applications*, Cambridge University Press, New York, 2006.
- [2] F.C. De Lucia Jr, J.L. Gottfried, Rapid analysis of energetic and geo-materials using LIBS, *Mater. Today* 14 (2011) 274–281.
- [3] F. Capitelli, F. Colao, M.R. Provenzano, R. Fantoni, G. Brunetti, N. Senesi, Determination of heavy metals in soils by laser induced breakdown spectroscopy, *Geoderma* 106 (2002) 45–62.
- [4] A. Ismaël, B. Bousquet, K. Michel-Le Pierrès, G. Travaillé, L. Canioni, S. Roy, In situ semi-quantitative analysis of polluted soils by laser-induced breakdown spectroscopy (LIBS), *Appl. Spectrosc.* 65 (2011) 467–473.
- [5] D.W. Hahn, N. Omenetto, Laser-induced breakdown spectroscopy (LIBS), part II: review of instrumental and methodological approaches to material analysis and applications to different fields, *Appl. Spectrosc.* 66 (2012) 347–419.
- [6] A.S. Eppler, D.A. Cremers, D.D. Hickmott, M.J. Ferris, A.C. Koskelo, Matrix effects in the detection of Pb and Ba in soils using laser-induced breakdown spectroscopy, *Appl. Spectrosc.* 50 (1996) 1175–1181.
- [7] B. Bousquet, J.B. Sirven, L. Canioni, Towards quantitative laser-induced breakdown spectroscopy analysis of soil samples, *Spectrochim. Acta Part B: At. Spectrosc.* 62 (2007) 1582–1589.
- [8] C. Aragon, J.A. Aguilera, F. Penalba, Improvements in quantitative analysis of steel composition by laser-induced breakdown spectroscopy at atmospheric pressure using an infrared Nd:YAG laser, *Appl. Spectrosc.* 53 (1999) 1259–1267.
- [9] D. DíAz, D.W. Hahn, A. Molina, Evaluation of laser-induced breakdown spectroscopy (LIBS) as a measurement technique for evaluation of total elemental concentration in soils, *Appl. Spectrosc.* 66 (2012) 99–106.
- [10] A. Ciucci, M. Corsi, V. Palleschi, S. Rastelli, A. Salvetti, E. Tognoni, New procedure for quantitative elemental analysis by laser-induced plasma spectroscopy, *Appl. Spectrosc.* 53 (1999) 960–964.
- [11] M. Pervolaraki, P.E. Dyer, P. Monk, Ablation studies using a diode-pumped Nd:YVO4 micro-laser, *Appl. Phys. A: Mater. Sci. Process.* 79 (2004) 849–854.
- [12] M.E. Essington, G.V. Melnichenko, M.A. Stewart, R.A. Hull, Soil metals analysis using laser-induced breakdown spectroscopy (LIBS), *Soil Sci. Soc. Am. J.* 73 (2009) 1469–1478.
- [13] J.B. Sirven, B. Bousquet, L. Canioni, L. Sarger, Laser-induced breakdown spectroscopy of composite samples: comparison of advanced chemometrics methods, *Anal. Chem.* 78 (2006) 1462–1469.
- [14] R. Wisbrun, I. Schechter, R. Niessner, H. Schroeder, K.L. Kompa, Detector for trace elemental analysis of solid environmental samples by laser plasma spectroscopy, *Anal. Chem.* 66 (1994) 2964–2975.
- [15] J.B. Sirven, B. Bousquet, L. Canioni, L. Sarger, S. Tellier, M. Potin-Gautier, I.L. Hecho, Qualitative and quantitative investigation of chromium-polluted soils by laser-induced breakdown spectroscopy combined with neural networks analysis, *Anal. Bioanal. Chem.* 385 (2006) 256–262.
- [16] M. Boueri, V. Motto-Ros, W. Lei, Q. Ma, L. Zheng, H. Zeng, J. Yu, Identification of polymer materials using laser-induced breakdown spectroscopy combined with artificial neural networks, *Appl. Spectrosc.* 65 (2011) 307–314.
- [17] A. Koujelev, M. Sabsabi, V. Motto-Ros, S. Laville, S.L. Lui, Laser-induced breakdown spectroscopy with artificial neural network processing for material identification, *Planet. Space Sci.* 58 (2009) 682–690.
- [18] X. Ma, Z. Zheng, H. Zhao, M. Zhang, Y. Liao, Laser induced breakdown spectroscopy algorithm using weights iteration neural network, 2011. (77532K-77532K).
- [19] E.C. Ferreira, D.b.M.B.P. Milori, E.J. Ferreira, R.M. Da Silva, L. Martin-Neto, Artificial neural network for Cu quantitative determination in soil using a portable laser induced breakdown spectroscopy system, *Spectrochim. Acta Part B: At. Spectrosc.* 63 (2008) 1216–1220.
- [20] S.O. Haykin, *Neural Networks and Learning Machines: International Version*, third ed. Pearson Education, Upper Saddle River, 2009.
- [21] V. Motto-Ros, A.S. Koujelev, G.R. Osinski, A.E. Dudelzak, Quantitative multi-elemental laser-induced breakdown spectroscopy using artificial neural networks. 08011 J. Eur. Opt. Soc. Rap. Public. 3 (2008).
- [22] J.L. Tarazona, J. Guerrero, R. Cabanzo, E. Mejía-Ospino, Construction of a predictive model for concentration of nickel and vanadium in vacuum residues of crude oils using artificial neural networks and LIBS, *Appl. Opt.* 51 (2012) B108–B114.
- [23] B. Bousquet, G. Travaillé, A. Ismaël, L. Canioni, K. Michel-Le Pierrès, E. Brasseur, S. Roy, I. le Hecho, M. Larregieu, S. Tellier, M. Potin-Gautier, T. Boriachon, P. Wazen, A. Diard, S. Belbèze, Development of a mobile system based on laser-induced breakdown spectroscopy and dedicated to in situ analysis of polluted soils, *Spectrochim. Acta Part B: At. Spectrosc.* 63 (2008) 1085–1090.
- [24] NIST Basic Atomic Spectroscopic Data, in: National Institute of Standards and Technology. <http://physics.nist.gov/PhysRefData/Handbook/periodictable.htm>, Last access date 30/08/2012.