# Laser-Induced Breakdown Spectroscopy of Composite Samples: Comparison of Advanced Chemometrics Methods

**ARTICLE** *in* ANALYTICAL CHEMISTRY · APRIL 2006

Impact Factor: 5.64 · DOI: 10.1021/ac051721p · Source: PubMed



**4 AUTHORS**, INCLUDING:

Lionel Canioni
University of Bordeaux
**195** PUBLICATIONS **1,537** CITATIONS

SEE PROFILE

Laurent Sarger
Université Bordeaux 1
**106** PUBLICATIONS **969** CITATIONS

SEE PROFILE

*Anal. Chem.* **2006,** *78,* 1462−1469

# Laser-Induced Breakdown Spectroscopy of Composite Samples: Comparison of Advanced Chemometrics Methods

**J.-B. Sirven,* B. Bousquet, L. Canioni, and L. Sarger**

*Centre de Physique Moléculaire Optique et Hertzienne (CPMOH), 351 Cours de la Libération, 33405 Talence Cedex, France*

**Laser-induced breakdown spectroscopy is used to measure chromium concentration in soil samples. A comparison is carried out between the calibration curve method and two chemometrics techniques: partial least-squares regression and neural networks. The three quantitative techniques are evaluated in terms of prediction accuracy, prediction precision, and limit of detection. The influence of several parameters specific to each method is studied in detail, as well as the effect of different pretreatments of the spectra. Neural networks are shown to correctly model nonlinear effects due to self-absorption in the plasma and to provide the best results. Subsequently, principal components analysis is used for classifying spectra from two different soils. Then simultaneous prediction of chromium concentration in the two matrixes is successfully performed through partial least-squares regression and neural networks.**

Laser-induced breakdown spectroscopy (LIBS) has been used for more than two decades for quantitative elemental analysis of solids, liquids, or gases.[1−2] Among the different techniques used to process spectra for determining the concentration of an element in a sample, one can distinguish supervised methods[3] from the calibration-free (CF) approach.[4−5] The first ones are based on a set of calibration samples of known concentrations, whereas CF-LIBS assumes local thermodynamic equilibrium in the plasma to calculate its temperature and its electron density, from which the composition of the sample is then derived, regardless of the matrix. Echelle spectrometers are particularly suited for these calculations since they allow work with single-shot, high-resolution spectra over a very large range, typically from 200 nm to 1 $\mu$m. As we do not have such a spectrometer, we rather investigated supervised methods, with the idea of comparing them in terms of prediction accuracy, prediction precision, and limit of detection.

The simplest and most widespread of these techniques is the standard calibration curve,[6] which consists of plotting the intensity of the line of interest as a function of the known concentration of a set of calibration samples. Yet LIBS may benefit from more advanced spectra treatments such as partial least-squares regression (PLSR) or neural networks, which have proved their worth and their reliability in other areas of spectroscopy, for qualitative as well as quantitative applications.[7−13] As a result, like several other atomic spectrometry techniques, LIBS is nowadays more and more associated with chemometrics methods[14] in order to improve its analytical performances with respect to the standard calibration curve.

Up to now, only linear multivariate treatments have been coupled to LIBS for quantitative purposes. Actually PLSR was mainly used, and it has been shown to be efficient when studying complex LIBS spectra,[15] to provide good selectivity and robustness to matrix effects,[16] to be able to simultaneously predict the concentrations of several elements,[17] to improve prediction accuracy,[18] and to decrease the limit of detection[19] with respect to the calibration curve. However, a limitation inherent to the linear nature of PLSR was early pointed out by Amador-Hernández et al., who found that it could not satisfactorily model the saturation effect of the signal due to self-absorption of strong lines in the plasma.[15]

* To whom correspondence should be addressed. Phone: +335 4000 2785. Fax: +335 4000 6970. E-mail" jb.sirven@cpmoh.u-bordeaux1.fr.

(1) Radziemski, L. J.; Loree, T. R.; Cremers, D. A.; Hoffman, N. M. *Anal. Chem.* **1983**, *55*, 1246−1252.
(2) Cremers, D. A.; Radziemski, L. J.; Loree, T. R. *Appl. Spectrosc.* **1984**, *38*, 721−729.
(3) Adams, M. J. *Chemometrics in analytical spectroscopy*, 2nd ed.; Royal Society of Chemistry: Cambridge, U.K., 2004.
(4) Ciucci, A.; Corsi, M.; Palleschi, V.; Rastelli, S.; Salvetti, A.; Tognoni, E. *Appl. Spectrosc.* **1999**, *53*, 960−964.
(5) Bulajic, D.; Corsi, M.; Cristoforetti, G.; Legnaioli, S.; Palleschi, V.; Salvetti, A.; Tognoni, E. *Spectrochim. Acta, Part B* **2002**, *57*, 339−353.
(6) Currie, L. A. *Anal. Chim. Acta* **1999**, *391*, 105−126.
(7) Haaland, D. M.; Thomas, E. V. *Anal. Chem.* **1988**, *60*, 1193−1202.
(8) Lalor, G. C.; Zhang, C. *Sci. Total Environ.* **2001**, *281*, 99−109.
(9) Nadal, M.; Schuhmacher, M.; Domingo, J. L. *Sci. Total Environ.* **2004**, *321*, 59−69.
(10) Cozzolino, D.; Morón, A. *Soil Tillage Res.* In press.
(11) Linker, R.; Kenny, A.; Shaviv, A.; Singher, L.; Shmulevich, I. *Appl. Spectrosc.* **2004**, *58*, 516−520.
(12) Linker, R.; Shmulevich, I.; Kenny, A.; Shaviv, A. *Chemosphere* **2005**, *61*, 652−658.
(13) Blanco, M.; Villarroya, I. *Trends Anal. Chem.* **2002**, *21*, 240−250.
(14) Fisher, A. S.; Goodall, P. S.; Hindsc, M. W.; Pennyd, D. M. *J. Anal. At. Spectrom.* **2005**, *20*, 1398−1424.
(15) Amador-Hernández, J.; García-Ayuso, L. E.; Fernández-Romero, J. M.; Luque de Castro, M. D. *J. Anal. At. Spectrom.* **2000**, *6*, 587−593.
(16) Green, R. L.; Mowery, M. D.; Good, J. A.; Higgins, J. P.; Arrivo, S. M.; McColough, K.; Mateos, A.; Reed, R. A. *Appl. Spectrosc.* **2005**, *59*, 340−347.
(17) Martin, M. Z.; Labbé, N.; Rials, T. G.; Wullschleger, S. D. *Spectrochim. Acta, Part B* **2005**, *60*, 1179−1185.
(18) Fink, H.; Panne, U.; Niessner, R. *Anal. Chem.* **2002**, *74*, 4334−4342.
(19) Wisbrun, R.; Schechter, I.; Niessner, R.; Schroeder, H.; Kompa, K. L. *Anal. Chem.* **1994**, *66*, 2964−2975.



10.1021/ac051721p CCC: $33.50 
Published on Web 01/24/2006

This observation is rather restrictive if we want to use PLSR for quantitative measurements over a wide concentration range. Keeping a linear relationship between the signal and the element concentration implies either that we limit the study to the low concentrations or that we choose in the spectrum lines free from self-absorption. As more energy is needed to excite such lines, they are generally weak, and this time we would be restricted to high concentrations.

Then we have to implement another method for treating the spectra. Since they are powerful tools to model the intrinsic nonlinearity of the signal,[20] artificial neural networks provide an efficient answer, and to our knowledge, their application to LIBS spectra processing for quantitative analysis is new.

In what follows, we perform a measurement of chromium concentration in soils samples and we propose a direct comparison—i.e., on the same spectra—between the standard calibration curve, PLSR, and neural networks, in terms of prediction accuracy, prediction precision, and limit of detection. The sensitivity of the last two techniques to matrix effects is also addressed in the last section.

## EXPERIMENTAL SECTION

**Samples and Data Sets.** In this work, we studied two different soils doped with chromium at different concentrations, of which we made pressed pellets. Samples were prepared by sorption in liquid phase and continuous agitation during 24 h, and then they were centrifuged and lyophilized.

The first one was a real agricultural soil, containing clay, sand, organic matter, several oxides, and various trace metals. Available concentrations were 66 (natural Cr concentration in this soil), 76, 116, 166, 266, and 566 ppm chromium. Ten spectra were taken for each concentration, and then our data set was composed of 60 spectra.

The second sample was kaolinite ($Al_2Si_2O_5(OH)_4$), a "model soil", and it was used only in the last section. It is a certified clay, extracted from a deposit and purified. Samples were doped with chromium at 0, 10, 50, 100, 200, 300, and 500 ppm. The data set was composed of one spectrum per concentration and 10 additional spectra at 10, 100, and 500 ppm, i.e., 37 spectra in total.

**Experimental Setup.** We used a Nd:YAG laser (Continuum Surelite) delivering 7-ns pulses at 355 nm with an incident energy on the sample of 3 mJ and a repetition rate of 3 Hz. The laser was focused onto the sample with a fused-silica lens of 50-mm focal length. The light emitted by the plasma was collimated by a parabolic mirror with a focal length of 38.1 mm and then injected by a 75-mm lens into a bundle of fibers. Each of them had a 200-$\mu$m core diameter, and they were arranged at the other end of the bundle in order to vertically match the 20-$\mu$m entrance slit of the spectrometer, an Oriel MS260i with a focal length of 250 mm and a grating of 2400 grooves/mm blazed at 400 nm. The detector was an Andor iStar intensified CCD. Spectra were accumulated over 100 laser shots, and the sample was continuously translated in its plane during each acquisition at a speed of ~0.5 mm/s, ensuring no overlap between two shots. The ICCD gate delay was fixed at 150 ns after the pulse, with a gate width of 2 $\mu$s, for an optimal signal-to-noise ratio. The gain of the intensifier was adjusted in order to fully exploit the 16-bit dynamic of the camera. Spectra were recorded over a 40-nm window centered at 360 nm containing the three chromium lines of interest at 357.869, 359.349, and 360.533 nm.

**Data Analysis.** Spectra analysis aims either at extracting qualitative information from them (for identification/classification purposes) or at calculating concentrations. Principal components analysis (PCA)[3] can be used in the first case to discriminate two classes of spectra and to detect outliers; it was successfully applied to LIBS by several authors.[17,18,21−25] PCA enables us to visualize similarities between spectra by graphically representing them in a new space of much fewer dimensions, keeping most of the information contained in the data set (see last section).

PLSR[3,7,26] was first employed for quantitative treatments. PLSR is a linear optimization technique enabling one to build a numerical model between the spectral intensities and the concentration, calculated from a set of calibrated samples. The optimization of the model is done by a "leave-one-out" cross-validation procedure, which determines the optimal number of components of the regression. Then a second set of spectra (the validation set, independent from the calibration set) allows one to evaluate the prediction ability of the regression.

PLSR was sometimes preceded by orthogonal signal correction (OSC)[27] to improve its performances. OSC is a filter that removes from the data set the information uncorrelated to the concentration. It is calculated similarly to PLSR, that is, from an optimal number of components.

To go further we also used neural networks[28] to retrieve the concentration from the spectra. Neural networks are powerful nonlinear optimization systems based on the following principle: each neuron receives a series of inputs that are dynamically weighted; it compares the weighted sum of its inputs to a given threshold value and finally applies a nonlinear function to compute the output. The neurons are arranged in layers to form the network. We used a three-layer one: the input layer receives the intensities at the different wavelengths, it is followed by a hidden layer, and in our case, the output layer has only one neuron since we are interested in only one element. The calibration phase then consists of iteratively optimizing the weights, as a function of the difference between the predicted and the true values of the concentration. Finally, the network prediction ability is evaluated from the validation set.


(20) Dieterle, F.; Busche, S.; Gauglitz, G. *Anal. Bioanal. Chem.* **2004**, *380*, 383−396.

(21) Goode, S. R.; Morgan, S. L.; Hoskins, R.; Oxsher, A. *J. Anal. At. Spectrom.* **2000**, *9*, 1133−1138.

(22) Samek, O.; Telle, H. H.; Beddows, D. C. S. *BMC Oral Health* **2001**, *1*, 1.

(23) Samuels, A. C.; De Lucia, F. C., Jr.; McNesby, K. L.; Miziolek, A. W. *Appl. Opt.* **2003**, *42*, 6205−6209.

(24) Hybl, J. D.; Lithgow, G. A.; Buckley, S. G. *Appl. Spectrosc.* **2003**, *57*, 1207−1215.

(25) Munson, C. A.; De Lucia, F. C., Jr.; Piehler, T.; McNesby, K. L.; Miziolek, A. W. *Spectrochim. Acta, Part B* **2005**, *60*, 1217−1224.

(26) Wold, S.; Sjöström, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130.

(27) Wold, S.; Antti, H.; Lindgren, F.; Ohman, J. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 175−185.

(28) Beale, R.; Jackson, T. *Neural computing: an introduction*; Adam Hilger: Bristol, 1991.

To compare the different methods, we used the average relative error of prediction (REP) to evaluate the prediction accuracy:

$$\mathrm{REP}\ (\%) = \frac{100}{N_\mathrm{v}} \sum_{i=1}^{N_\mathrm{v}} \left| \frac{\hat{c}_i - c_i}{c_i} \right|$$

where $N_\mathrm{v}$ is the number of validation spectra, $c_i$ is the true concentration, and $\hat{c}_i$ is the estimated concentration. The prediction precision is given by the average relative standard deviation (RSD):

$$\mathrm{RSD}\ (\%) = \frac{100}{N_\mathrm{conc}} \sum_{k=1}^{N_\mathrm{conc}} \frac{\sigma_{Ck}}{C_k} \quad \text{with} \quad {\sigma_{Ck}}^2 = \sum_{i=1}^{p} \frac{(\hat{c}_{ik} - c_k)^2}{p - 1}$$

where $N_\mathrm{conc}$ is the number of different concentrations in the validation set, $p$ is the number of spectra per concentration, and $\sigma_{Ck}$ is the standard deviation obtained at concentration $c_k$.

Finally, the limit of detection (LOD) is calculated from the IUPAC definition:[6]

$$\mathrm{LOD}\ (\mathrm{ppm}) = 3\sigma_a / b$$

where the signal $s$ is fitted by a line with respect to the concentration $c$, $s = a + bc$, $\sigma_a$ being the standard deviation of $a$ given by the regression.

Ortiz et al.[29] have shown that this expression could be used in the case of multivariate treatments, from a simple linear regression between the predicted and the true concentrations (applied to the validation set), so we employed it to calculate PLSR and neural networks detection limits.

## RESULTS AND DISCUSSION

**Calibration Curve.** We first present the quantitative results that we obtained by the standard calibration curve method on the measurement of chromium concentration in our agricultural soil. For this purpose, the area of the 359.349-nm line of chromium was calculated after a Gaussian fit of the peak (±0.2 nm). Various normalizations were applied to the spectra, to take into account several causes of signal fluctuations, and the effect of spectra smoothing was also investigated. In all cases, the residual baseline due to instrumental noise was first subtracted from the spectra. Three types of normalization, already described in the literature, were then used prior to plotting the calibration curve:

(1) Normalization with respect to the matrix by a line of a matrix species.[30−33] In our case, the iron line at 373.486 nm was chosen.

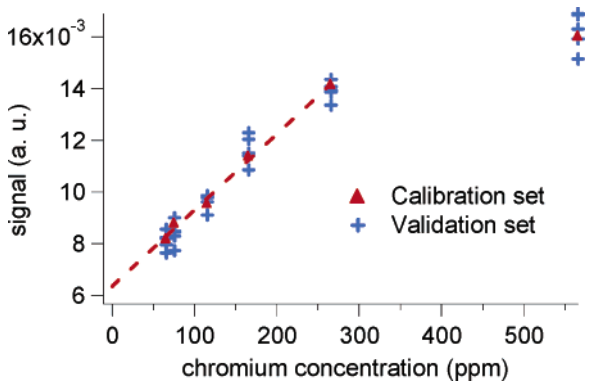


**Figure 1.** LIBS signal as a function of chromium concentration in agricultural soil samples. Calibration spectra (triangles) are fitted by a line in the low concentrations region (dash line). The regression coefficient is $r^2 = 0.9946$. Validation spectra are also reported (crosses).

(2) Normalization by the background.[34−38] In our case, we normalized by the average of the spectrum calculated between 345.6 and 346.4 nm, where no atomic spectral line is present. In this way, we aim at compensating for the fluctuations of plasma formation conditions, which can be due to laser energy, spatial mode variations, or irregular surface condition of the sample. However, working with soils can induce specific problems—spatial heterogeneity, matrix effects, etc.—not corrected by this normalization.

(3) Normalization by the area under the whole spectrum.[39] This approach can be understood as a combination of the two previous ones, since it is matrix-dependent (but in a broader way than the normalization by a single line) and it also takes into account plasma conditions, for the area of the spectrum represents the energy released by the plasma.

Additionally, smoothing was performed using a box smoothing over nine points.

We remind the reader that the data set is composed of 60 spectra, 10 per chromium concentrations among which 5 of them were randomly chosen and averaged to plot the calibration curve. This one was next fitted by an exponential function—not represented on Figure 1—as described in ref 40 to take into account the self-absorption of the strong chromium line at high concentrations. The predicted concentration was finally calculated from the fitted curve for the five other spectra. The LOD was determined from a linear fit in the low-concentration part of the curve. As we work with a real soil having a natural chromium concentration of 66 ppm, we do not have experimental measurements below this value. Yet the fit between 66 and 266 ppm provided a very good regression coefficient, then showing good linearity of the signal below ~300 ppm. Results are summarized in Table 1 and the best calibration curve is represented in Figure 1.


(29) Ortiz, M. C.; Sarabia, L. A.; Herrero, A.; Sánchez, M. S.; Sanz, M. B.; Rueda, M. E.; Giménez, D.; Meléndez, M. E. *Chemom. Intell. Lab. Syst.* **2003**, *69*, 21−33.
(30) Barbini, R.; Colao, F.; Fantoni, R.; Palucci, A.; Ribezzo, S.; Van Der Steen, H. J. L.; Angelone, M. *Appl. Phys. B* **1997**, *65*, 101−107.
(31) Galbacs, G.; Gornushkin, I. B.; Smith, B. W.; Winefordner, J. D. *Spectrochim. Acta, Part B* **2001**, *56*, 1159−1173.
(32) Bustamante, M. F.; Rinaldi, C. A.; Ferrero, J. C. *Spectrochim. Acta, Part B* **2002**, *57*, 303−309.
(33) Martin, M. Z.; Wullschleger, S. D.; Garten, C. T.; Palumbo, A. V. *Appl. Opt.* **2003**, *42*, 6174−6178.
(34) Xu, L.; Bulatov, V.; Gridin, V. V.; Schechter, I. *Anal. Chem.* **1997**, *69*, 2103−2108.
(35) Bulatov, V.; Krasniker, R.; Schechter, I. *Anal. Chem.* **1998**, *70*, 5302−5311.
(36) Carranza, J. E.; Hahn, D. W. *Spectrochim. Acta, Part B* **2002**, *57*, 779−790.
(37) Lazic, V.; Fantoni, R.; Colao, F.; Santagata, A.; Morone, A.; Spizzichino, V. *J. Anal. At. Spectrom.* **2004**, *19*, 429−436.
(38) Fornarini, L.; Colao, F.; Fantoni, R.; Lazic, V.; Spizzicchino, V. *Spectrochim. Acta, Part B* **2005**, *60*, 1186−1201.
(39) Mosier-Boss, P. A.; Lieberman, S. H.; Theriault, G. A. *Environ. Sci. Technol.* **2002**, *36*, 3968−3976.
(40) Lazic, V.; Barbini, R.; Colao, F.; Fantoni, R.; Palucci, A. *Spectrochim. Acta, Part B* **2001**, *56*, 807−820.

**Table 1. Influence of Spectra Normalization, Smoothing, and Derivation on the Prediction Ability of the Calibration Curve Method**

| type of normalization | REP (%) | RSD (%) | LOD (ppm) |
|---|---|---|---|
| Fe line | 13.3 | 19.6 | 47 |
| *after smoothing* | *20.7* | *22.7* | *46* |
| background | 16.7 | 13.6 | 52 |
| spectrum area | 28.7 | 15.9 | 47 |
| Fe line then background | 16.7 | 13.6 | 52 |
| background then Fe line | 18.2 | 20.6 | 45 |
| Fe line then spectrum area | 10.3 | 12.4 | 33 |
| *after smoothing* | *15.0* | *22.7* | *33* |
| Spectrum area then Fe line | 18.2 | 20.6 | 45 |

**Table 2. Influence of the Number of Input Variables and of OSC on the Prediction Ability of PLSR**

| number of points | REP (%) | RSD (%) | LOD (ppm) |
|---|---|---|---|
| 1024 (whole spectrum) | 41.6 | 42.4 | 84 |
| *with OSC ($N_c = 2$)* | *32.6* | *23.0* | *74* |
| 248 (10 nm bandwidth) | 34.9 | 32.5 | 63 |
| *with OSC ($N_c = 1$)* | *23.1* | *15.2* | *54* |
| 100 (4-nm bandwidth) | 29.6 | 26.8 | 63 |
| *with OSC ($N_c = 2$)* | *25.8* | *28.7* | *67* |
| 38 (3 concatenated lines) | 29.7 | 24.0 | 73 |
| *with OSC ($N_c = 2$)* | *29.4* | *22.6* | *95* |

As we can see in Table 1, simply normalizing the spectra by the iron line intensity provides good results in terms of REP that is, this normalization predicts values of the Cr concentration close to the true ones. This is what we expected since iron belongs to the matrix and its concentration in each sample remains constant. On the other hand, normalizing by the background or by the spectrum area gives a better RSD; that is, this normalization compensates well for the fluctuations of the signal from one spectrum to the other, mainly due to variations in plasma conditions. Regarding the LOD, it is similar in these first three cases. Then we combined the Fe normalization with the other ones in order to find a better compromise between REP and RSD. The best results were obtained by normalizing first by the Fe line and then by the spectrum area—see Table 1. The LOD is also improved with respect to the simple Fe normalization, but this difference may not be significant due to the uncertainty associated to the calculation of the LOD.

Regarding spectra smoothing, we tested it in two cases. As spectra are averaged over 100 laser shots, they exhibit little noise, and Table 1 clearly shows that smoothing does not improve the prediction.

**Partial Least-Squares Regression.** The calibration curve leads to good results, but which are likely to be improved by using a full-spectrum multivariate method instead. In this section, we present the results we obtained with PLSR applied to the same data set as in the previous section. As for the calibration curve method, we compared the performances of the different normalizations. For the sake of conciseness, this part will be omitted here. We will use the normalization giving the best results, that is, spectra normalized by the background, and we present a step-by-step optimization of the PLSR procedure. Calculations were performed using SIMCA-P 9.0 software (Umetrics).

As PLSR is a supervised method, considering our data set, we have to select five calibration spectra per Cr concentration that best represent it, the five others being the validation spectra. This is done through a principal components analysis, which helps to discriminate outliers. The PLSR model is then calculated and characterized.

We first investigated the effect of the number of input variables in the regression performances, that is, the influence of the spectra bandwidth. Calculations were made in four cases: with the whole spectra, with a 10- or 4-nm window centered at 360 nm, and by concatenating the three intense chromium lines at 357.869, 359.349, and 360.533 nm. The number of components was optimized in each case, and each time we obtained the best model with three components. The effect of an orthogonal signal correction was investigated at the same time. Table 2 recapitulates the results.

Let us first look at the performances of a mere PLSR as a function of the spectra bandwidth (normal characters in Table 2). We notice a general downward trend in the different parameters when the number of points decreases. In particular, the RSD for validation spectra is improved by more than 40% between the top and the bottom of the table.

OSC was next applied to the different data sets prior to the PLSR (in italics in Table 2). The optimal number of components $N_c$ from which the OSC was performed is reported in parentheses; this number should not be confused with the number of components of the following PLSR—always one in our case after the OSC, instead of three with a simple PLSR. Generally speaking, this pretreatment enhances the accuracy and the precision of prediction, although there is a risk of overfitting. With respect to the mere PLSR results, when the number of points is important, the improvement in the RSD is especially noteworthy. The best results are obtained with an OSC applied to the 10-nm window; therefore, we shall keep this one for quantitative measurements.

Smoothing and taking the derivative of the spectra was also tested to see if it improved the predictions of the PLSR, since it is often stated that this enables one to enhance spectral information and to suppress the residual baseline.[3,16] None of these two pretreatments was shown to increase the accuracy, the precision, or the LOD of the prediction, even if a preliminary OSC is used. This is due to the fact that spectra are averaged over 100 laser shots and exhibit little noise, hence this insensitivity to smoothing. Working with single-shot, noisy spectra would probably lead to a significantly different conclusion. Moreover, chromium lines are well defined; therefore, the gain in the enhancement of the signal by taking the derivative is negligible.

Finally, we investigated the effect of the linearity of the input data with respect to Cr concentration. Indeed, PLSR performances may be limited by self-absorption of intense lines, as illustrated by Figure 1, which shows that the calibration curve is clearly nonlinear after a Cr concentration lying between 200 and 300 ppm, due to the self-absorption of the intense 359.348-nm Cr line. Then we expect that reducing the concentration range should increase the accuracy and the precision of the method. To verify this, we simply performed the same data processing, limiting to 266 ppm the Cr concentration, that is, by removing from our data set the 10 spectra of the most concentrated sample. To compare the results to the calibration curve, concentrations were recalculated in the linear region from the calibration line shown on Figure 1

**Table 3. Influence of Chromium Concentration Range on the Prediction Ability of PLSR**

| Cr concn range (ppm) | data processing | REP (%) | RSD (%) | LOD (ppm) |
|---|---|---|---|---|
| 66−566 | OSC ($N_c = 1$)/PLSR (1 component) background normalization 10-nm window | 23.1 | 15.2 | 54 |
| 66−266 | OSC ($N_c = 1$)/PLSR (1 component) background normalization 10-nm window | 15.0 | 11.4 | 34 |
| 66−266 | PLSR (1 component) Fe line/spectrum area normalization 3 concatenated lines | 11.0 | 11.0 | 24 |

instead of the exponential fit. No improvement was obtained since we found a REP of 10.2%, RSD of 12.7%, with the LOD being unchanged.

The first line of Table 3 repeats the previous results we obtained between 66 and 566 ppm. The second line shows the performances of PLSR in the linear range, in the same conditions: background normalization, 10-nm window, OSC with $N_c = 1$ and PLSR with 1 component.

Additionally, the whole PLSR optimization procedure was carried out again in order to get the best results as possible between 66 and 266 ppm—last line of Table 3. This was obtained by normalizing by the Fe line and then by the area of the whole spectrum, concatenating lines, and calculating the PLSR with a single component. No OSC was performed since it did not improve the prediction.

The last two lines of Table 3 show that the regression gains from being reoptimized. Then, if we compare the first and the last lines, we see that the benefit obtained by improving the linearity of the data with respect to Cr concentration especially lies in the average REP and in the LOD, which are both divided by 2—yet the precision is also significantly better. The prediction accuracy and precision become comparable to the values of the calibration curve method—see Table 1. We can therefore conclude that in the linear range PLSR is more efficient than over the entire range and that it has performances at least as good as with the calibration curve.

This last observation shows that PLSR is not suited for LIBS measurements over a wide dynamic, since only intense lines enable us to have enough signal with trace elements, and they are self-absorbed. Therefore, as a well-known nonlinear modeling technique, we explored then the use of artificial neural networks for predicting Cr concentration over a broad range.

**Artificial Neural Networks.** For this study, we treat the same data set, composed of 10 spectra of our agricultural soil per Cr concentration, with the three-layer neural network described previously. For each concentration five calibration spectra were used, the five other ones being the validation spectra. We used the neural network functions of Igor Pro 5.0 (Wavemetrics) for calibration and for prediction. As for PLSR, due to the large number of parameters, we propose a step-by-step optimization of the method. For that purpose, we defined an ordered list of several parameters and we made them vary in the order we specified. At each step, the already optimized parameters kept their best value while the others had their default value.

**Table 4. Influence of the Number of Input Variables on the Prediction Ability of Neural Networks**

| | REP (%) | RSD (%) | LOD (ppm) |
|---|---|---|---|
| whole spectrum (939 points) | 14.1 | 21.0 | 41 |
| 10-nm bandwidth (248 points) | 6.4 | 6.7 | 15 |
| 4-nm bandwidth (100 points) | 5.4 | 5.9 | 12 |
| 3 lines concatenated (38 points) | 5.8 | 8.5 | 35 |

**Table 5. Optimization of the Neural Network (Upper Part) and Influence of Spectra Smoothing and Derivation on Its Prediction Ability (Lower Part)**

| optimized parameter | REP (%) | RSD (%) | LOD (ppm) |
|---|---|---|---|
| number of hidden nodes | 8.5 | 7.8 | 30 |
| learning speed | 6.0 | 6.3 | 13 |
| number of iterations | 5.4 | 5.9 | 12 |
| smoothing | 6.5 | 6.8 | 12 |
| smoothing/derivation | *4.5* | *5.1* | *16* |

The first point we examined was, once more, the normalization of the spectra, and we reached the conclusion that the neural network was rather insensitive to this pretreatment. For all normalizations used in Table 1, the REP was found within 2−3%, as well as the RSD. Yet the best performances were again obtained by normalizing by the Fe line at 373.486 nm and then by the area of the spectrum, so we kept this pretreatment.

Next we studied the influence of the spectral bandwidth, and we tested the same spectral windows as in the previous section, i.e., the whole spectra, a 10-n or 4-nm window, and the three concatenated Cr lines. Table 4 summarizes the results. As for PLSR, there is a clear optimum in terms of accuracy, precision and LOD with a reduced bandwidth, here with a 4-nm window. This means that there is an optimal *proportion* of points correlated to chromium in the spectra; that is, there must be "useless" wavelengths in the spectra for an optimal operation of the network.

Then the number of nodes in the hidden layer was optimized by making it vary between 1 and 10. Four hidden nodes lead to the best performances. Next, specific parameters of the network were optimized in order to refine its learning speed. After that, the number of iterations during the calibration of the network was increased from 20 000 to 100 000 with a step of 10 000. The optimum was obtained for 50 000 iterations. Beyond this value a slight overfitting was observed. Table 5 shows the prediction performances of the network at each step of the procedure for the Fe line/spectrum area normalization.

Finally, we tested the effect of smoothing and of taking the derivative of the spectra. The whole optimization procedure was repeated for smoothed and smoothed/derived spectra. Table 5 shows that smoothing did not improve the results, for the same reason as for PLSR. Taking the derivative afterward induced a better accuracy and precision than raw spectra, but a slightly higher LOD. Yet the difference in the results might be specific to our data set and we shall not draw a general conclusion about the effect of derivation.

Compared to the calibration curve method—cf. Table 1—the REP, the RSD, and the LOD are improved by a factor of 2 at least. With respect to PLSR—cf. Table 3—they are divided by more than

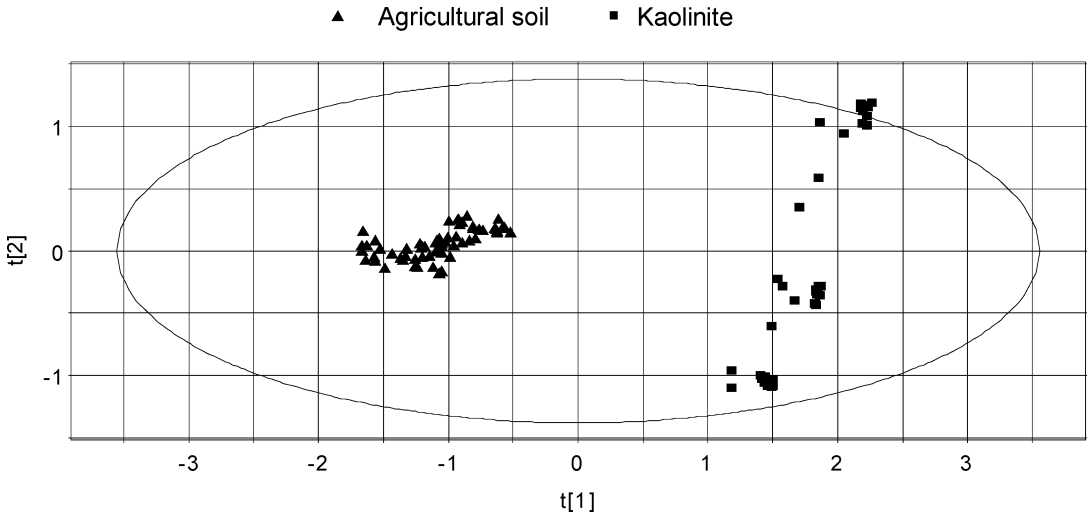


**Figure 2.** Data set projection onto the plane of the first two components obtained by PCA. The data set consists of agricultural soil spectra (triangles) and kaolinite spectra (squares).

3. Furthermore, we insist on the fact that these results were obtained over the whole Cr concentration range. This illustrates the powerful ability of neural networks in taking into account nonlinear relationships between the variables in order to extract quantitative information from the spectra. Additionally, we removed the 116 ppm sample spectra from the calibration set to test the aptitude of the network to predict a concentration not included in the calibration set. We verified that the results were as good as previously. We also checked that we obtained identical performances when the concentration range was reduced. In conclusion, we can state that neural networks constitute the most powerful tool we tested for the quantitative measurement of chromium in soil samples.

**Simultaneous Analysis of Two Soils.** An interesting study about the influence of matrix effects on chemometrics data treatments was recently published by Green et al.[16] These authors used LIBS followed by univariate least-squares regression to measure magnesium in pharmaceutical powders. They found that a single model could be built for four slightly different matrixes, with a similar prediction accuracy compared to the one obtained with an individual calibration for each matrix. These results encouraged us to investigate the sensitivity of PLSR and of neural networks to matrix effects in order to determine whether these advanced techniques can increase the versatility of LIBS and to estimate the accuracy and precision that we can expect for quantitative measurements.

Therefore, we gathered the spectra obtained with the agricultural soil and with kaolinite samples in a single treatment. All spectra were acquired with the same laser energy, same focusing conditions, and the same gate delay and gate width of the camera. The calibration set included five agricultural soil spectra per concentration between 66 and 566 ppm and one kaolinite spectrum per concentration between 0 and 500 ppm, that is, 37 spectra in total. The validation set was composed of 60 spectra, the 5 other agricultural spectra per concentration, and 30 kaolinite spectra, 10 at 10 ppm, 10 at 100 ppm, and 10 at 500 ppm. Concerning the normalization, we could not use the spectrum area since it is specific to the matrix, nor the iron line since its concentration is different in the two soils. Then we normalized the spectra by the background as described previously. Additionally, we kept the 10-nm spectral window centered at 360 nm, which we have already used.

**Table 6. Simultaneous Prediction of Chromium Concentration in Agricultural Soil and Kaolinite through PLSR and Neural Networks**

| | | REP (%) | RSD (%) |
|---|---|---|---|
| | Without Outliers (50−566 ppm) | | |
| PLSR | total data set | 26.4 | 18.4 |
| | *agricultural soil spectra* | *19.5* | *21.8* |
| | *kaolinite spectra* | *34.5* | *9.9* |
| neural networks | total data set | 10.5 | 7.0 |
| | *agricultural soil spectra* | *12.0* | *6.9* |
| | *kaolinite spectra* | *8.8* | *7.2* |
| | All Spectra (0−566 ppm) | | |
| neural networks | total data set | 9.9 | 9.8 |
| | *agricultural soil spectra* | *8.5* | *7.1* |
| | *kaolinite spectra* | *11.3* | *15.2* |

Our objective is to do quantitative measurements over a Cr concentration range as large as possible; therefore, even if PLSR was shown to be more efficient when working in the linear part of the calibration curve, we first tested it on the whole data set. We obtained unexploitable results, since negative values were found for low-concentration kaolinite samples and for the 76 ppm agricultural sample. In this latter case, we had an abnormally weak overall signal—which had not been critical when doing PLSR with only the agricultural soil, presumably due to a water content unintentionally higher than in the other samples. Therefore, the 10 spectra of this sample and all spectra from kaolinite samples with a Cr concentration lower than 40 ppm were removed from the data set.

In this case, PLSR gave poor results, which could be improved by still optimizing the data set. Then a thorough detection of outliers was subsequently performed using a principal components analysis. When treating the whole data set, spectra from the two soils were perfectly separated by a PCA in the plane of the first two components—see Figure 2. Yet to be more accurate in the identification of outliers, two additional PCA were done, one for

**Table 7. Simultaneous Prediction of Chromium Concentrations by Neural Networks Applied to Agricultural Soil Spectra (Left) and to Kaolinite Spectra (Right), between 0 and 566 ppm**

| agricultural soil samples | [Cr] (ppm) true | [Cr] (ppm) predicted | REP (%) | RSD (%) | kaolinite samples | [Cr] (ppm) true | [Cr] (ppm) predicted | REP (%) | RSD (%) |
|---|---|---|---|---|---|---|---|---|---|
| AS1 | 66 | 62.75 | 4.56 | 0.46 | K1 | 10 | 10.87 | 8.71 | 9.81 |
| AS2 | | 63.13 | 3.99 | | K2 | | 10.14 | 1.39 | |
| AS3 | | 63.32 | 3.69 | | K3 | | 9.37 | 6.26 | |
| AS4 | | 67.51 | 2.68 | | K4 | | 10.02 | 0.19 | |
| AS5 | | 66.36 | 0.93 | | K5 | | 9.23 | 7.66 | |
| AS6 | 76 | 91.29 | 20.52 | 18.75 | K6 | | 9.58 | 4.17 | |
| AS7 | | 68.26 | 9.89 | | K7 | | 7.38 | 26.17 | |
| AS8 | | 74.18 | 2.08 | | K8 | | 9.74 | 2.57 | |
| AS9 | | 74.86 | 1.17 | | K9 | | 10.01 | 0.13 | |
| AS10 | | 84.21 | 11.17 | | K10 | | 8.87 | 11.32 | |
| AS11 | 116 | 123.91 | 7.05 | 5.01 | K11 | 100 | 86.58 | 13.42 | 32.77 |
| AS12 | | 94.18 | 18.64 | | K12 | | 36.95 | 63.05 | |
| AS13 | | 124.17 | 7.27 | | K13 | | 97.30 | 2.70 | |
| AS14 | | 117.08 | 1.15 | | K14 | | 130.77 | 30.77 | |
| AS20 | | 123.89 | 7.03 | | K15 | | 114.19 | 14.19 | |
| AS16 | 166 | 171.50 | 3.47 | 20.50 | K16 | | 46.73 | 53.27 | |
| AS17 | | 184.62 | 11.39 | | K17 | | 93.73 | 6.27 | |
| AS18 | | 180.63 | 8.98 | | K18 | | 117.39 | 17.39 | |
| AS19 | | 209.47 | 26.38 | | K19 | | 120.88 | 20.88 | |
| AS20 | | 177.95 | 7.36 | | K20 | | 99.36 | 0.64 | |
| AS21 | 266 | 316.64 | 19.15 | 3.30 | K21 | 500 | 484.64 | 3.07 | 3.00 |
| AS22 | | 295.62 | 11.24 | | K22 | | 474.63 | 5.07 | |
| AS23 | | 295.06 | 11.03 | | K23 | | 476.70 | 4.66 | |
| AS24 | | 312.07 | 17.43 | | K24 | | 490.68 | 1.86 | |
| AS25 | | 327.66 | 23.30 | | K25 | | 471.44 | 5.71 | |
| AS26 | 566 | 560.77 | 0.88 | 3.76 | K26 | | 486.26 | 2.75 | |
| AS27 | | 512.33 | 9.44 | | K27 | | 439.56 | 12.09 | |
| AS28 | | 552.14 | 2.41 | | K28 | | 478.10 | 4.38 | |
| AS29 | | 563.73 | 0.36 | | K29 | | 472.90 | 5.42 | |
| AS30 | | 559.27 | 1.15 | | K30 | | 484.62 | 3.08 | |

each matrix. Two outliers were found in kaolinite 100 ppm spectra and four of them were discovered in the agricultural soil set, two at 66 ppm and two at 116 ppm. Thus, these six spectra were removed from the data set. The calibration set was finally composed of 30 spectra, the validation set of 39 spectra, between 50 and 566 ppm.

Table 6 summarizes in its upper part the results obtained with a PLSR calculated from six components. No OSC was used since it did not improve the performances. Calibration accuracy, prediction accuracy, and prediction precision are given for the whole data set. These figures were also calculated for each soil, that is, the "total data set" value is the average of the two soils values weighted by their respective number of spectra. The overall performances are quite satisfactory, with an accuracy of 26.4% and a precision of 18.4%. The results show that the REP and the RSD of the agricultural soil samples are comparable to the ones we calculated for a single soil in Table 2 for a 10-nm window. This is probably due to the fact that agricultural soil spectra form the major part of the calibration set, so the prediction ability should not be very different. For the same reason, we notice that the relative error of prediction is high for kaolinite spectra, since they are not well represented in the calibration set. However, the precision obtained is rather good, of the order of 10%. In conclusion, PLSR needs a detailed optimization of the data set in order to function properly, and even in this case, the simultaneous prediction of two soils is not uniform and gives only approximate results.

An analogous study was carried out using the neural network. Spectra were normalized by the background, as previously, and a 4-nm window was used. Table 6 (lower part) presents the results obtained after optimizing the calibration of the network, as described in the previous section, on one hand with the same data set as for PLSR—that is, after removing the low concentration spectra and the outliers, middle of the table—and on the other hand with the full data set (37 calibration spectra and 60 validation ones, bottom of the table). In the case when the data set is the same as in the previous paragraph, the overall accuracy and precision of the prediction are more than twice as better than with PLSR. Furthermore, it is noticeable that despite the imbalance of the calibration set the neural network produces much more homogeneous results for the two soils, in terms of accuracy as well as precision. When all spectra are included, overall accuracy and precision are similar to the ones we had with the previous data set—of the order of 10%—which means that the neural network is rather insensitive to the presence of outliers in the data set and predicts well the lowest concentrations. We find better results with the agricultural soil with respect to kaolinite, but the performances remain acceptable for both.

Finally, Table 7 exhibits the predicted concentrations of validation samples when all spectra are included in the model. We notice that 10 ppm samples are well predicted, with a concentration found within 10% accuracy in most cases and a RSD below 10%. RSD for 100 ppm kaolinite samples is high due to the presence of the two previously identified outliers—K12 and K16 in Table 7. Considering the agricultural soil, the 76 ppm sample, not surprisingly, shows a high RSD but the prediction is accurate within 10% for three spectra out of five. The RSD of the 166 ppm sample was also found quite high, but the reason is unknown.

Tables 6 and 7 illustrate again the predictive power of the neural network. We emphasize that these results are very encouraging since they demonstrate that, when a suitable spectra treatment is used, a simultaneous prediction of two different matrixes is possible with a good accuracy and a good precision over a wide concentration range.

## CONCLUSIONS

An intensive comparative study was performed between the standard calibration curve method, partial least-squares regression, and neural networks for the prediction of chromium concentration in soils samples. Data set consists of spectra obtained in a standard LIBS arrangement on various controlled specimens. These three data treatments were optimized in detail, the influence of spectra normalization, smoothing, derivation, and spectral bandwidth was particularly investigated. Neural networks clearly gave the best results, with accuracy and precision at least twice as good as the calibration curve. PLSR had the worst performances, but this was mainly due to the nonlinearity between the spectral intensities and the concentration induced by self-absorption in the plasma. We also studied the sensitivity of the two multivariate treatments to matrix effects by simultaneously predicting chromium concentration in two different soils. PLSR had acceptable performances provided there wasa reduction of the concentration range and a thorough search for outliers in the initial data set. Neural networks gave better results even over the whole concentration range, with both accuracy and precision of the order of 10%. At last, PCA enabled us to efficiently classify samples from the two soils. We have demonstrated that advanced chemometrics methods can widely improve LIBS analytical performances with respect to the traditional calibration curve and, thus, open the way to simultaneous quantitative analysis of several matrixes.

## ACKNOWLEDGMENT

We gratefully thank Martine Potin-Gautier and Sylvaine Tellier (LCABIE, Pau, France) for sample preparation and Jean-Pierre Dubost (University Bordeaux 2, France) for fruitful discussions on chemometrics. This research is financially supported by CNRS, ADEME, and Conseil Régional d'Aquitaine.

Received for review September 26, 2005. Accepted December 2, 2005.

AC051721P