Spectrochimica Acta Part B 117 (2016) 16–22

Contents lists available at ScienceDirect

Spectrochimica Acta Part B

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

Evaluation and optimization of the robustness of a multivariate analysis methodology for identification of alloys by laser induced breakdown spectroscopy

Evelyne Vors *, Krikor Tchepidjian, Jean-Baptiste Sirven

CEA Saclay, DEN, DANS, DPC, SEARS, Laboratoire de Développement Analytique Nucléaire Isotopique et Elémentaire, 91191 Gif-sur-Yvette, France

ARTICLE INFO

Article history:
Received 20 April 2015
Accepted 13 December 2015
Available online 6 January 2016

Keywords:
Laser induced breakdown spectroscopy
Chemometrics
Model robustness
Alloy identification
Experimental design

ABSTRACT

In this study, laser induced breakdown spectroscopy (LIBS) with chemometrics was used for classification and identification of alloys, with a particular focus on the issue of the model robustness. A supervised classification model, Soft Independent Modeling of Class Analogy (SIMCA) was calculated with calibration spectra of 13 representative materials. These measurements were reproduced, with the same samples and using the same LIBS instrument, on two different dates (seven and eight months after the calibration measurements): during this period, instrumental variations occurred and the robustness of sample classification was assessed by the prediction error rate. Then, the optimization of SIMCA model parameters, including spectral preprocessing and wavelength selection, was performed using a full factorial experimental design, and a prediction error rate of 0% with a robustness of 100% was achieved for this period extending until eight months after the model calibration. The study was completed two and a half years later by a test of the robustness of the previously optimized model, carried out with an additional series of measurements on test samples with the same LIBS instrument. The predictive ability of the model on spectra acquired more than two years after validation remained good.

1. Introduction

Laser induced breakdown spectroscopy (LIBS) is a fully optical, multi-elementary and fast analytical technique, requiring no or little sample preparation. These features make the LIBS technique particularly suited for portable, in situ or online measurements, and have been applied for process control in industry [1,2,3], geological and environmental measurements [4] and classification and identification of materials [5]. Multivariate data processing is often used in those applications in order to overcome fluctuations of the LIBS signal in a weakly controlled environment. Hence, Soft Independent Modeling of Class Analogy (SIMCA) [5,6], Partial Least Squares Discriminant Analysis (PLS-DA) [5,6,7], Artificial Neural Network [8,9,10], Support Vector Machine [11,12] or Independent Component Analysis [13] have been tested by several authors for that purpose.

However, in the field of in situ or online analysis, the measurement robustness is a crucial question. Indeed, analytical instruments are designed to be continuously used over a long period, with no human intervention for maintenance, recalibration, etc. Therefore, when using a chemometric model to measure a property, its predictability has to be assessed over a representative period of time. Surprisingly, most papers dealing with chemometric models applied to LIBS data do not address this issue. Instead, it is a common practice to perform a single set of experiments, and to split in two parts the dataset obtained under the same experimental conditions. The first part is used to calibrate the model, and the second one to validate it. In this case, the model efficiency with data acquired over a longer timescale cannot be assessed.

Infrared spectroscopy is the field of analytical spectroscopy in which most of the robustness studies have been made. Zeaiter et al. have reviewed robustness studies for on-line monitoring by NIR spectroscopy in industrial applications [14,15]. The different definitions of robustness found in the literature and the different ways to assess the robustness are reported in [14]. The definitions used in the literature may be shared into two main categories [16,17]. The first one deals with transferability problems and is related to the transfer of the method between laboratories, operators or instruments. In the second one, robustness is related to the influence of environmental and instrumental changes on the prediction performances of a multivariate model. Several methods have been employed to improve the prediction performances of a model [15]: spectral preprocessing methods (different spectral normalization methods, smoothing and differentiation) and reduction of the data dimensions (selection of informative variables by statistical methods [18] or by wavelength selection [19]). Statistical treatments have also been tested to improve the on-line robustness [20].

In LIBS, the issue of the robustness of predictions based on multivariate analysis was addressed in only a few studies [5,12,10]. Yet in [5,12],

* Corresponding author.
E-mail address: evelyne.vors@cea.fr (E. Vors).

http://dx.doi.org/10.1016/j.sab.2015.12.004
0584-8547/

robustness refers to the model ability to cope with unknown samples, i.e. not included in the learning set of spectra. In [10], the authors deal with the problem of assessing the robustness of an Artificial Neural Network approach against variations of the laser energy (on simulated spectra). In the present work, we adopt the definition of Ref. [14]: "the sensitivity of the model's predictions to changes in external factors, such as variations in environmental, instrumental, and sample conditions, under which the spectra were measured during the calibration phase". We focused our study on the sensitivity to changes in instrumental conditions over time. The aim of our study was the assessment and the optimization of the robustness when the experimental conditions undergo variations due to the aging of the LIBS device (we had no prior knowledge of the phenomena involved in the aging of our system). For that purpose, classification of alloys was used as an example to test our approach. A SIMCA model was built with calibration samples and predictions were made with the same samples and the same instrument after a period of 7 to 8 months of instrumental aging. The selection of spectral data, the model parameters and the spectral preprocessing were optimized using a full factorial design of experiments in order to maximize the model robustness. After optimization of the model, a 100% rate of correct classification was achieved for spectra obtained several months after the calibration phase. The study was completed by a test of robustness over a longer timescale. Two and a half years later an additional series of test measurements was performed, with the same LIBS instrument, on three test samples and the prediction performances of the previously optimized model were evaluated for these later experiments.

2. Materials and methods

2.1. Experimental setup

A commercial instrument (IVEA SAS Mobilibs) was used for LIBS measurements. The laser was a quadrupled Nd:YAG laser (Quantel Brio) emitting at 266 nm pulses of duration 5 ns (FWHM) at 20 Hz repetition rate. On the target surface, the energy was about 6 mJ per pulse and the $1/e^2$ spot diameter is 50 μm: the incident irradiance was about 60 GW/cm^2. Plasma light was collected with an achromatic telescope and injected into a fused silica optical fiber of 200 μm diameter connected to the spectrometer. An echelle spectrometer is used (ESA3000, LLA Instruments). The spectral bandwidth of the spectrometer extends from 200 to 780 nm with a resolving power :/Δ: of about 10,000 (each spectral line corresponds to about ten points of the spectrum). The spectrometer is equipped with an intensified CCD camera (KAF 1001, Kodak). Plasma light is detected within a temporal gate whose delay (with respect to the laser pulse) and width are controlled.

LIBS measurements were carried out on 13 alloys (various steels, inconels and some other alloys). The samples, numbered from 1 to 13, can be split into several material categories (Table 1). Each alloy (sample) is considered as a class in the calculation of the models. The samples present a wide range of concentrations for different elements. The compositions of some of these elements enable the differentiation of the samples: Fe, Cr, Ni, Mo, Cu, Al and Ti (the concentrations of Al and Ti are low). Table 1 gives the available dataset on concentrations of these elements in the samples. In some cases only concentration ranges can be indicated and the symbol — marks the unavailable values of concentration. Samples 1, 2 and 5 are certified reference materials (Techlab) and correspond to the concentrations with the weakest uncertainties.

For each of the thirteen samples, 25 spectra were acquired. For each spectrum 10 laser shots were accumulated on the same location of the sample, with a gate delay of 1 μs and a gate width of 1 μs. As an example, Fig. 1 shows a spectrum of a stainless steel sample.

Table 1
Categories of the samples and concentrations (wt.%) of the discriminating elements.

Category	Class	Fe	Cr	Ni	Mo	Cu	Al	Ti
Stainless steels	9	~69	17–19	9–11	–	–	–	–
	10	~66	16–18	11–13	2–2.5	–	–	–
	11	~68	17–19	10–12	–	–	–	–
Other steels	1	68.6	0.008	18.44	4.88	0.008	0.058	0.41
	2	66.5	0.034	18.51	4.97	0.047	0.12	0.69
	5	62.9	0.053	18.4	4.71	0.023	0.11	1.47
	12	~63	<0.5	18–19	~5	<0.5	0.1	1.3–1.6
Inconels	6	6–10	14–17	72	–	<0.5	–	–
	7	8–17	21–25	58–63	–	<1	–	–
	8	40	19–23	30–35	–	–	0.3–0.6	0.3–0.6
Monel	13	2.5	–	63	–	28–34	–	–
Other alloys	3	33	25	37	2.5	–	0.1	–
	4	3	22	55	13	–	–	–

2.2. Dataset measurement and data treatment

Three sets of experimental data were acquired at three different periods of time with the same LIBS instrument and the same spectrometer. Each dataset was built with measurements on the same 13 samples (25 spectra per sample). The rows of the matrix of the dataset correspond to the spectra. The variables are the intensities corresponding to the wavelengths (about 50,000 variables per spectrum) and the matrix contains 325 rows. The measurements were made in September 2011 (dataset 1), in April and May 2012 (dataset 2), and in December 2014 (dataset 3). The wavelength calibration of the spectrometer was made before each series of measurements. The incident energy on the sample surface undergoes variations due to a drift of the laser: 6 mJ for set 1 and 4.5 mJ for set 2. Multivariate analysis was performed using PLS_Toolbox 7.5 (Eigenvector Research) running under MATLAB R2014 (MathWorks).

A supervised classification method, SIMCA (Soft Independent Modeling of Class Analogy) was used. SIMCA calculates an independent Principal Component Analysis (PCA) model for each class [21]. The prediction of the class of an unknown spectrum is made in the following way: the distance between the spectrum and the centroid of each PCA model is calculated and a probability of class membership is calculated using this distance. The decision criterion for allocating the spectrum is based on this probability of membership.

To optimize the model robustness, we proceeded as follows. The initial model was calculated using randomly selected 80% of spectra of dataset 1, and its predictive performances were determined using cross-validation on the 20% remaining spectra. Predictions of dataset 2 were used to evaluate the model robustness. Then a design of experiments was implemented in order to optimize it, using dataset 2 as a validation set. Four factors were taken into account and are detailed below: input variables, preprocessing, distance definition and classification

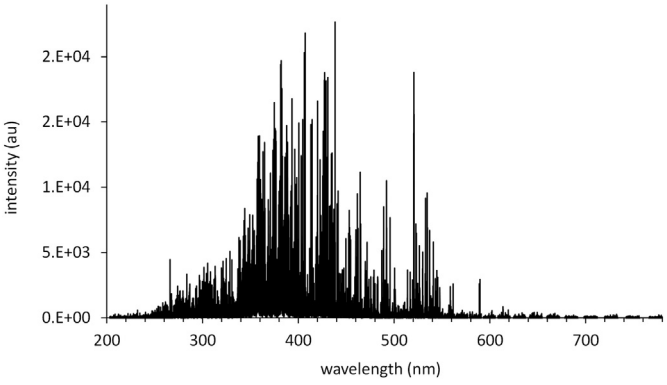

Fig. 1. Spectrum of a stainless steel.