Analytica Chimica Acta xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Analytica Chimica Acta

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

Discriminant analysis in the presence of interferences: Combined application of target factor analysis and a Bayesian soft-classifier

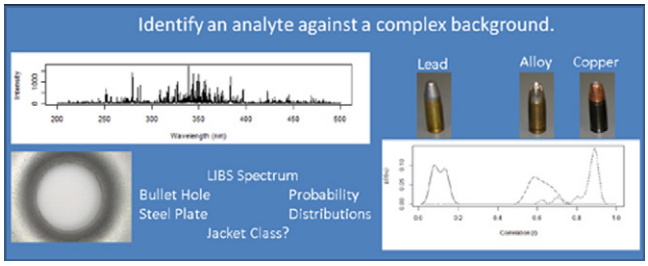
Caitlin N. Rinke[a], Mary R. Williams[a], Christopher Brown[b], Matthieu Baudelet[b], Martin Richardson[b], Michael E. Sigman[a,b,*]

[a] National Center for Forensic Science and Department of Chemistry, College of Science, University of Central Florida, PO Box 162367, Orlando, FL 32816-2367, United States
[b] Townes Laser Institute, College of Optics and Photonics, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816-2367, United States

HIGHLIGHTS

- A novel background-independent soft classification method is described.
- Target factor analysis is combined with Bayesian decision theory.
- Target factors are taken from a library of target analytes.
- Trace metal transfers to complex backgrounds are analyzed by LIBS.
- The method is shown to be both conservative and accurate.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 14 June 2012
Received in revised form
25 September 2012
Accepted 26 September 2012
Available online xxx

Keywords:
Target factor analysis
Discriminant analysis
Bayesian classifier
Laser-induced breakdown spectroscopy

ABSTRACT

A method is described for performing discriminant analysis in the presence of interfering background signal. The method is based on performing target factor analysis on a data set comprised of contributions from analyte(s) and interfering components. A library of data from representative analyte classes is tested for possible contributing factors by performing oblique rotations of the principal factors to obtain the best match, in a least-squares sense, between test and predicted vectors. The degree of match between the test and predicted vectors is measured by the Pearson correlation coefficient, *r*, and the distribution of *r* for each class is determined. A Bayesian soft classifier is used to calculate the posterior probability based on the distributions of *r* for each class, which assist the analyst in assessing the presence of one or more analytes. The method is demonstrated by analyses performed on spectra obtained by laser induced breakdown spectroscopy (LIBS). Single and multiple bullet jacketing transfers to steel and porcelain substrates were analyzed to identify the jacketing materials. Additionally, the metal surrounding bullet holes was analyzed to identify the class of bullet jacketing that passed through a stainless steel plate. Of 36 single sample transfers, the copper jacketed (CJ) and non-jacketed (NJ) class on porcelain had an average posterior probability of the metal deposited on the substrate of 1.0. Metal jacketed (MJ) bullet transfers to steel and porcelain were not detected as successfully. Multiple transfers of CJ/NJ and CJ/MJ on the two substrates resulted in posterior probabilities that reflected the presence of both jacketing materials. The MJ/NJ transfers gave posterior probabilities that reflected the presence of the NJ material, but the MJ component was mistaken for CJ on steel, while non-zero probabilities were obtained for both CJ and MJ on porcelain. Jacketing transfer from a bullet to steel as the projectile passed through the steel also proved difficult to analyze. Over 50% of the samples left insufficient transfer to be identified. Transfer from NJ and CJ jacketing was successfully identified by posterior probabilities greater than 0.8.

* Corresponding author at: National Center for Forensic Science and Department of Chemistry, College of Science, United States. Tel.: +1 407 823 3420; fax: +1 407 823 3162.
E-mail address: msigman@mail.ucf.edu (M.E. Sigman).

0003-2670/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.aca.2012.09.042

1. Introduction

An important problem in chemical analysis is the assignment of object membership among a set of defined classes. A detailed discussion of modern classification methods has been given in several texts [1,2]. Current literature provides examples of the use of discriminant analysis in forensic evidence evaluation [3–6], homeland security applications [7,8], and disease diagnosis [9–11], among other applications.

Most classification methods are not particularly well adapted to classifying an unknown in the presence of a significant background signal that is independent of the analyte class and highly variable from sample to sample. Classification procedures that have been optimized using class examples without including interferences in the examples, may fail to perform well on samples that contain impurities. However, recent results have shown some success in performing discriminant analysis in the presence of interfering factors. For example, >97% correct classification of the maximum temperature reached by burning soils has been accomplished by partial least squares discriminant analysis (PLS-DA) in the presence of up to 20% ash interference [12]. It is often the case that the number of analyte classes is small in comparison to the number of potential interfering species and background signatures, thereby complicating the task of securing training sets that contain a representative set of possible interferences.

This paper examines an approach to analyte classification in the presence of a complex and undetermined interference. The method explored here makes use of target factor analysis (TFA), as described by Malinowski [13], to determine the presence of an analyte in a mixture without requiring that every component of the mixture be identified. The method introduced here differs from a previously reported use of TFA for discriminant analysis (also referred to as procrustes discriminant analysis), which has been shown to be mathematically equivalent to PLS-DA [14,15]. In contrast, the method discussed here relies on testing a library of analyte spectra from representative classes to identify the class or classes that contribute to the data set. The data sets examined in this paper are taken from spectra obtained by laser-induced breakdown spectroscopy (LIBS) from steel plates that had been penetrated by 9 mm bullets with either copper jackets (CJ), metal alloy jackets (MJ) or non-jacketed lead (NJ) projectiles. The same bullets were applied in trace quantities to porcelain tiles and metal plates as one and two analyte (mixed) samples. The bullet jacketing materials constitute the analyte classes. The objective is to test the TFA method to identify the class of bullet that penetrated a steel plate or was applied to porcelain or steel based on the posterior probabilities calculated from the LIBS spectra collected at the edges of the holes and in the application areas where residues of jacket materials were deposited. The LIBS spectra reflect contributions from the steel or porcelain substrates and the bullet's jacket material. The class of bullet with the highest probability of contributing to the spectral set is identified by the posterior probabilities. Similarly, significant posterior probabilities for the more than one class of analyte can indicate the presence of multiple classes. The method also defines criteria that must be met before a posterior probability is calculated for a given class. Consequently, if none of the analyte classes meet the criteria, the sample may be evaluated as not containing any of the analyte classes, and the method may truly serve as a soft classifier.

2. Materials and methods

2.1. Experimental

2.1.1. Instrumentation

The LIBS instrument used in this research (LIBS2000+, Ocean Optics, Dunedin, FL, USA) was equipped with a Q-switched Nd:YAG nanosecond pulsed laser (CFR200, Big Sky Lasers, Bozeman, MT, USA). Fundamental (1064 nm) laser output (pulse width of 9 ns and 63 mJ pulse^{-1}) was used to generate laser-induced plasmas. The spectrometer delay for spectral collection was optimized at 2.5 μs, giving maximum signal intensity and signal-to-noise ratio and an integration time of 1 ms. The plasma emission between 200 and 900 nm was collected by a fiber optic bundle connected to seven sequential CCD spectrometers, providing a spectral resolution of 0.06 nm per channel. The sample was placed in a chamber consisting of a plastic box equipped with an *X*–*Y*-adjustable sample stage, as previously reported [16,17]. Data acquisitions were performed using the Ocean Optics OOILIBS software.

2.1.2. Samples

The six different manufacturer and type combinations of 9 mm cartridge used in the tests comprised three classes of bullets based on the jacketing. The bullet classes consisted of copper jacketed (CJ), metal alloy jacketed (MJ) and non-jacketed (NJ). The CJ class was comprised of four manufacturer/type cartridges, including Remington/UMC (4 cartridges), CCI/Blazer (5 cartridges), Independence (5 cartridges), and Hornady/'Tap' (5 cartridges). The MJ class was comprised of Winchester/Silver-Tip (5 cartridges) the NJ class consisted of hand-loaded rounds (5 cartridges). The Hornady/'TAP' and the Winchester/Silver-Tip were hollow-point bullets. The bullets of each manufacturer/type were obtained courtesy of East Orange Shooting Sports, Orlando, FL.

The cartridges, one of each manufacturer/type, were taken to the Orlando office of the Florida Department of Law Enforcement where the bullets were safely removed with a "bullet puller hammer" to allow for safe LIBS analysis of the outside of the jacketing materials (see Section 2.1.3). Each of the pulled bullet manufacturer/type was used to establish the LIBS spectral library, as described below. The pulled bullets were also used to generate two sets of samples involving trace quantities of the bullet jacketing transferred to steel and porcelain substrates. The first set of samples was prepared by manually transferring trace quantities of each bullet onto the steel and porcelain substrates. The transfers were made by pulling the projectile along the substrate in a straight line with lightly applied pressure. The transfers were intermittently visible along the line. Three transfer lines were made with each manufacturer/type bullet to produce a total of 18 transfer lines on steel and 18 lines on porcelain. The second set of samples were prepared in the same way as the first set; however, the jacketing from two different bullet types were overlain on a single substrate. The combinations CJ/MJ, CJ/NJ and MJ/NJ were generated by preparing three transfer lines for each combination to give a total of nine transfer lines on steel and nine lines on porcelain.

The remaining live cartridges were fired through steel plates to produce bullet holes and tested for trace quantities of the jacket material deposited on the edge of the hole. The total numbers of bullet holes from each class were 15 CJ, 4 MJ and 4 NJ. The plates were 1 mm steel sheeting obtained from Home Depot (Orlando, FL). The firing of the bullets through the steel plates was performed at the Seminole County Shooting Range by an officer from the University of Central Florida Police Department. The rounds were fired through the sheet metal positioned approximately 7–10 feet from the shooter. A total of three plates were used in the experiments, with each plate being used for two bullet manufacturer/types. The first two plates were shot with the four of the CJ class and the last plate was shot with bullets from the MJ and then NJ classes. Two clusters of bullet holes were positioned on opposite ends of each plate and labeled at the time of the shooting. Plates were wrapped individually in plastic bags to prevent possible contamination on transportation back to the laboratory for analysis.

Please cite this article in press as: C.N. Rinke, et al., Discriminant analysis in the presence of interferences: Combined application of target factor analysis and a Bayesian soft-classifier, Anal. Chim. Acta (2012), http://dx.doi.org/10.1016/j.aca.2012.09.042

2.1.3. Sample analysis

A library of spectra corresponding to each class was generated by collecting LIBS spectra from the unused projectiles. To account for the high variability often observed in replicate LIBS spectra, the library was made up of six spectra from each manufacturer/type, with each spectrum comprised of an average of ten single-shot spectra. Each bullet was placed on the sampling stage and rotated until six average spectra were collected. Similarly, spectra of the steel plates and porcelain tiles were also collected; however, these spectra were not included in the library since they constitute the background interference against which the projectile class is to be determined.

Each bullet transfer line on the steel and porcelain substrates was sampled by collection of LIBS spectra along the line. A total of 12 single shot spectra were collected along each line. This procedure was used for each of the lines corresponding to one-bullet or two-bullet transfers. The plates containing the bullet holes were cut to obtain individual bullet holes and surrounding plate from cluster of holes. Each bullet hole was positioned on the LIBS sample stage so the laser would hit the edge of the bullet hole in an area where material transfer from physical contact would be expected to occur. Twelve single shot spectra were collected from around each bullet hole.

2.2. Data analysis

Spectral data analyzed by the TFA tests was restricted to 5636 points for each spectrum in the range of 200–500 nm. The emissions comprising each spectrum were normalized to a summed intensity of one. Twelve spectra were collected from each bullet hole and along each transfer line on porcelain or steel in order to account for variations arising from laser coupling to the matrix. Each set of spectra, corresponding to a single transfer line or bullet hole, were concatenated into a single data matrix and each matrix was subjected to target factor analysis (TFA), as described below. The data manipulation and analysis was performed using Microsoft Excel (Microsoft Corp., Seattle, WA, USA) to compile the data, MatLab (MathWorks, Natick, MA, USA) to perform the TFA and Bayesian soft classification (code written in-house), and R (The R Project for Statistical Computing, http://www.r-project.org) to graph the data.

The mathematical basis for the combination of target factor analysis (TFA) with a Bayesian posterior probability calculation to provide a multiple class Bayesian soft classifier is discussed below. Application of the combined TFA and Bayesian decision theory to the detection of ignitable liquid residue in fire debris has been reported elsewhere [18].

2.2.1. Target factor analysis (TFA)

A data matrix comprised of a set of spectra, each composed of different contributions from a finite set of components, was analyzed by abstract factor analysis to first reduce the data dimensionality and identify an orthonormal set of abstract factors. The abstract factors may be divided into the principal factors which comprise the signal and a set of secondary factors which contribute to noise in the data. The total number of abstract factors corresponds to the rank of the matrix. The rank of a matrix may be defined as the maximum number of linearly independent rows or columns and for a matrix with r rows and c columns, the maximum possible rank is the smaller of r and c. The matrices analyzed in this application are comprised of relatively small number of rows, each corresponding to a single spectrum, and a large number of columns, each corresponding to the spectral variables (wavelength). The number of spectra comprising the matrix determines the maximum possible rank, and therefore the maximum number of abstract factors, which must exceed the number of principal factors contributing to the data. The matrices $[R^{\ddagger}]$ and $[C^{\ddagger}]$ in Eq. (1) contain the scores and loadings for only the principal factors. In the work presented here, determination of the number of principal factors to retain is accomplished using the determination of rank by median absolute deviation (DRMAD) and $[D]$ is factored by singular value decomposition [19]. The DRMAD method works to identify the secondary factors (the error) by the residual standard deviation (RSD) of the set of associated error eigenvalues. When a principal eigenvalue is added to the set of error eigenvalues, the RSD will increase and the principal eigenvalue will be an outlier in the set of error eigenvalues. The outlier status of a value x_0 is determined if $||x_0 - \text{median}(x_i)||/\text{MAD}$ exceeds a value of 5, where MAD is the median absolute deviation of the set of eigenvalues. The error matrix $[E]$ in Eq. (1) accounts for the error removed from the data after accounting for the principal factors. Each row of $[D]$ corresponds to a sample spectrum and each column corresponds to a variable (i.e., wavelength).

$$[D] = [R^{\ddagger}][C^{\ddagger}] + [E] \tag{1}$$

The primary factors in $[C^{\ddagger}]$ constitute a set of eigenvectors that do not represent physically meaningful spectra. The spectra comprising $[D]$ may contain contributions from analytes of interest as well as contributions from background/substrate materials. Target factor rotations are oblique rotations used to identify physically meaningful spectra contributing to $[D]$. The vector transformations are expressed in Eqs. (2)–(4), where $[T]$ is the transformation matrix that brings about the oblique rotations and the error contribution $[E]$ has been dropped in Eq. (2).

$$[D] = [R^{\ddagger}][T][T]^{-1}[C^{\ddagger}] \tag{2}$$

$$T_l' = \overline{\overline{C}}_l[C^{\ddagger}]^T \tag{3}$$

$$\overline{C}_l = T_l'[C^{\ddagger}] \tag{4}$$

Individual vectors contributing to $[T]$ or $[T]^{-1}$ may be identified without the need to identify the entire matrix. The transformation vector T_l' corresponding to the lth row transformation vector of $[T]^{-1}$ is found by Eq. (3), where $\overline{\overline{C}}_l$ is the test spectrum (vector). The predicted spectrum (vector) $\bar{C}_l$, obtained by Eq. (4), represents the maximum-likelihood prediction of a spectrum that best matches the test spectrum [13]. The test spectra are taken from a library of potential analytes. Eq. (3) may be thought of as the projection of the test vector into the abstract vector space represented in Eq. (1). Similarly, Eq. (4) represents the recovery of the projected vector and the comparison between $\overline{\overline{C}}_l$ and $\overline{C}_l$ reveals the ability to accurately represent the test spectrum in factor space generated in Eq. (1). Individual analytes in a library may be ranked based on the comparison of the test and predicted spectra by a similarity metric, such as the Pearson product moment correlation coefficient, r. In a case where a set of analytes possess class characteristics that are represented in their spectra, target factor analysis may be combined with a Bayesian approach to produce a classifier, as described below. This approach may be fruitful if identification of a class of analyte can be achieved whereas individualization may not be possible (i.e. identification of a bullet casing by class, as opposed to identifying the exact manufacturer and bullet).

2.2.2. Assigning class membership

Bayesian decision theory is applied to a multiclass classification problem using Eq. (5), where $P(\omega_i|r)$ is the posterior probability, $p(r|\omega_i)$ is the class-conditional probability, and $P(\omega_i)$ is the prior probability that an object belongs to class ω_i [2]. In this work, the prior probabilities for all classes will be set equal and cancel out of Eq. (5). Each class-conditional distribution is defined by a set of correlations coefficients which are distributed on the interval $[-1,1]$. The class-conditional probabilities in Eq. (5) are approximated by the kernel function given in Eq. (6), where there are n_i

Please cite this article in press as: C.N. Rinke, et al., Discriminant analysis in the presence of interferences: Combined application of target factor analysis and a Bayesian soft-classifier, Anal. Chim. Acta (2012), http://dx.doi.org/10.1016/j.aca.2012.09.042

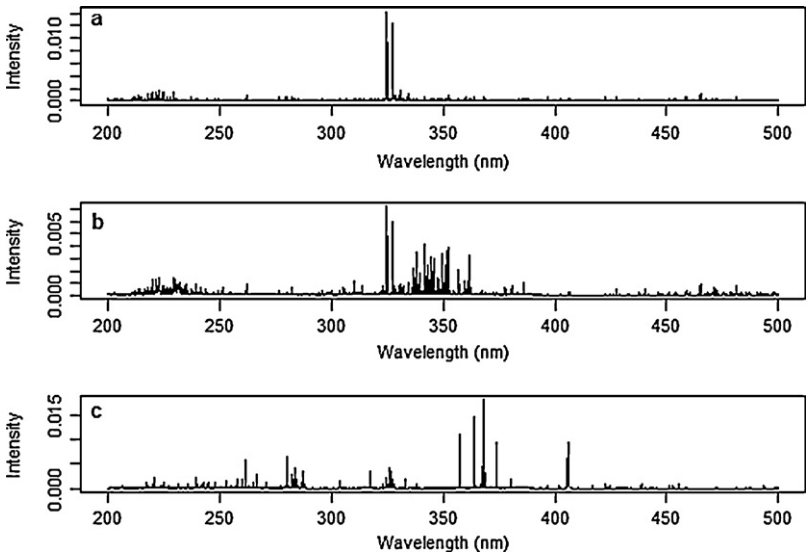


Fig. 1. LIBS spectra for bullet jacketing material (a) CJ, (b) MJ, (c) NJ. See Table 1 for composition information.

correlation values for each class ω_i [20]. The number of correlations, n_i, correspond to the number of samples within the library class. The adaptive bandwidth h_i in Eq. (6) is calculated by Eq. (7), wherein A_i is an adaptive estimate of the spread in the correlations for class ω_i [20]. The standard deviation of the Pearson correlations for class ω_i is given by s_i in Eq. (8), and the inter-quartile range of the correlations for class ω_i is given by R_i. The operation "min" in Eq. (8) indicates that the minimum of the two values (s_i and $R_i/1.34$) will be assigned to A_i. Adaptive bandwidth methods are known to perform well in density estimates, including for multimode distributions and those with long tails. A fixed bandwidth method can result in over-smoothing or under-smoothing as the density of points varies throughout a distribution. The method of Silverman, used here, has been reported to give a mean integrated squared error (MISE) within 10% of the optimum for some common non-normal distributions [20].

$$P(\omega_i|r) = \frac{p(r|\omega_i)P(\omega_i)}{\sum_i p(r|\omega_i)P(\omega_i)} \tag{5}$$

$$p(r|\omega_i) = \frac{1}{n_i}\sum_{j=1}^{n_i}\frac{1}{h_i\sqrt{2\pi}}\exp\left[\frac{1}{2h_i^2}(r-r_j)^2\right] \tag{6}$$

$$h_i = 0.9A_i n_i^{1/5} \tag{7}$$

$$A_i = \min\left(s_i, \frac{R_i}{1.34}\right) \tag{8}$$

The class-conditional probabilities, $p(r|\omega_i)$ change with every set of TFA results and it is therefore necessary to calculate a set of $p(r|\omega_i)$ values following every TFA analysis. Eq. (5) allows the calculation of $P(\omega_i|r)$ at a specified correlation, r; however, $P(\omega_i|r)$ based on Eq. (5) can vary significantly over a small range of r when the class-conditional probability distributions are narrow. An argument can be made for calculating $P(\omega_i|r)$ at $r = 1$, where the test and predicted spectra are perfectly correlated; however, this places a very stringent requirement for the identification of real factors from data with significant background contributions. A reasonable alternative is to assess classification by identifying the class with the highest probability of showing correlations that exceed a specified lower limit, r_{LL}. The integrated area under $p(r|\omega_i)$ for correlations exceeding r_{LL} yields this probability. The correlation, r, is bound on the interval $[-1,1]$; however, the integrated area under the class conditional-probability given by Eqs. (6)–(8) over the interval $[-1,1]$ may be less than one. Failure to integrate to an area of one can result from $p(r|\omega_i)$ not going to zero at $r = 1$ due to the finite bandwidth, h. This behavior is especially evident when the values of r for a class begin to cluster close to one. The boundary problem is addressed by approximating $I_i[r_{LL},1]$, the integrated area under $p(r|\omega_i)$ at $r > r_{LL}$, by the expression $1 - I_i[-1, r_{LL}]$. With this modification, the posterior probability can be calculated by Eq. (9), rather than by Eq. (5).

$$P(\omega_i|I_i[r_{LL}, 1]) = \frac{(1 - I_i[-1, r_{LL}])P(\omega_i)}{\sum_i(1 - I_i[-1, r_{LL}])P(\omega_i)} \tag{9}$$

When assignment of the sample to the class with the highest posterior probability is required, the classifier is considered a "hard classifier". Classifiers that allow a sample to be assigned to multiple classes, or not assigned to a class, are typically considered to be "soft classifiers". The posterior probabilities may also be utilized to estimate the degree of class membership, resulting in a Bayesian soft-classifier [21]. One additional consideration is important, the case where all of the class-conditional distributions reflect poor agreement between the test and predicted spectra. In this case it may be prudent to not classify the sample into any of the classes represented in the library. To address this issue, and simultaneously create a truly soft classifier, the value of $I_i[r_{LL},1]$ will be required to exceed some value α before $P(\omega_i|I_i[r_{LL},1])$ will be calculated for class ω_i. The value of α corresponds to a minimal probability of observing $r > r_{LL}$ and represents a statistical significance level for assigning a posterior probability to a given class.

3. Results and discussion

Fig. 1 shows LIBS spectra from each class (CJ, MJ and NJ) in the library, and Fig. 2 shows LIBS spectra of the steel and porcelain substrates. The two strong peaks at 324.75 nm and 327.40 nm in the CJ spectra (Fig. 1a) are attributed to Cu I transitions. The spectrum of the MJ projectile (Fig. 1b) also contains emissions attributed to the Cu I transitions, along with emissions at 361.94, 341.48 and 352.45 nm that are attributed to Ni. The representative spectrum from the NJ class projectile (Fig. 1c) contains multiple lines attributed to Pb, including the 363.96 nm and 368.35 nm Pb I transitions [21]. The spectrum from the metal plate (Fig. 2a) does not contain excessively strong lines attributable to the Cu I or Pb I transitions but does contain approximately 15 lines in common with the three projectile classes. Similarly, the porcelain substrate spectrum

Please cite this article in press as: C.N. Rinke, et al., Discriminant analysis in the presence of interferences: Combined application of target factor analysis and a Bayesian soft-classifier, Anal. Chim. Acta (2012), http://dx.doi.org/10.1016/j.aca.2012.09.042

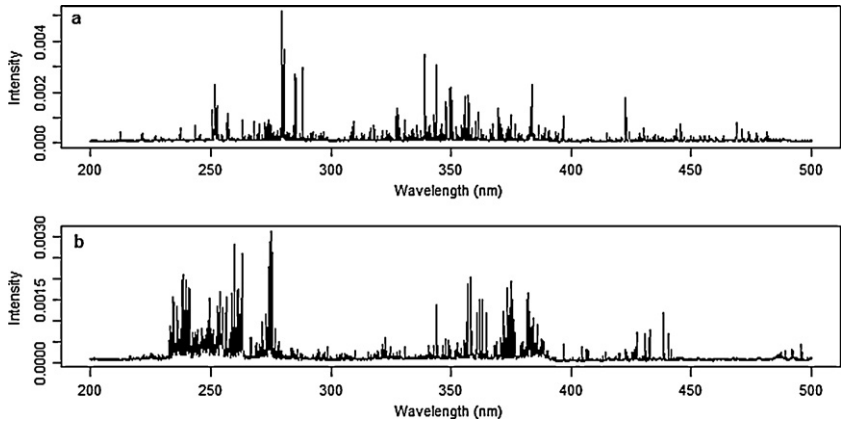


Fig. 2. LIBS spectra for substrate materials (a) steel, (b) porcelain. See Table 1 for composition information.

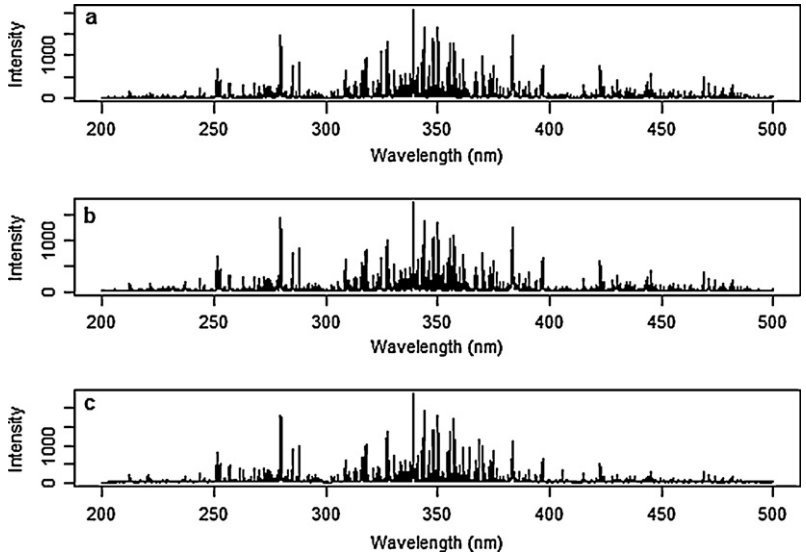


Fig. 3. LIBS spectra of (a) CJ, (b) MJ and (c) NJ transfer on steel.

(Fig. 2b) contains approximately 18 lines in common with the three projectile classes.

Fig. 3 shows representative spectra from the lines of each bullet class (CJ, MJ and NJ) deposited on the steel substrate. The complexity of the steel spectra enhances the difficulty in selecting the bullet jacketing that contributes to the spectra. When the substrate identity and associated spectral complexity vary, the problem of identifying a contribution from one of the bullet classes can become even more difficult. The spectra taken on one of the CJ bullet transfer lines on steel were processed as described above and the class-conditional distributions for CJ, MJ and NJ classes are shown in Fig. 4.

The class-conditional distribution functions calculated from the correlation of test and predicted spectra from CJ, MJ and NJ bullet classes, Fig. 4a, clearly show higher overall correlations for library spectra from the the CJ class. Fig. 4b shows $P(\omega_i|I_i[r_{LL},1])$ as r_{LL} is varied and $\alpha = 0.05$. The calculated $P(\mathrm{CJ}|I_{\mathrm{CJ}}[r_{LL},1]) = 0$ for r_{LL} greater than approximately 0.92 where $I_{\mathrm{CJ}}[r_{LL},1] < \alpha$. At r_{LL} equal to approximately 0.92, $P(\mathrm{CJ}|I_{\mathrm{CJ}}[r_{LL},1])$ jumps to a value of 1.0 and remains high until r_{LL} decreases to approximately 0.75, where $P(\mathrm{MJ}|I_{\mathrm{MJ}}[r_{LL},1])$ begins to increase as $I_{\mathrm{MJ}}[r_{LL},1]$ exceeds α. As r_{LL} continues to decrease, $P(\mathrm{CJ}|I_{\mathrm{CJ}}[r_{LL},1])$ and $P(\mathrm{MJ}|I_{\mathrm{MJ}}[r_{LL},1])$ converge on 0.5, and eventually $P(\mathrm{NJ}|I_{\mathrm{NJ}}[r_{LL},1])$ begins to contribute and all three posterior probability values converge on 0.33.

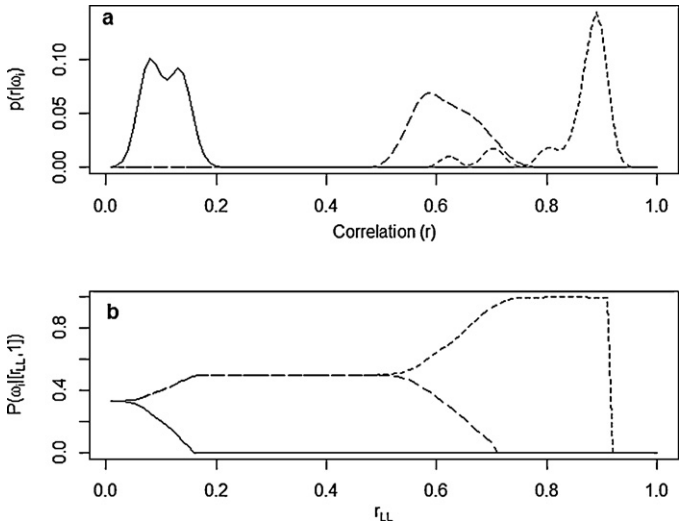


Fig. 4. (a) Class-conditional distribution functions for correlation of test and predicted spectra from CJ (short dashes), MJ (medium dashes) and NJ (solid line) bullet jacket classes. (b) Posterior probabilities with $\alpha = 0.05$ calculated as a function or r_{LL}. Data derived from analysis of a set of spectra taken from a CJ transfer line on steel. The number of library spectra (target factors) for each class conditional curve were: 24 CJ, 6 MJ and 6 NJ.

Please cite this article in press as: C.N. Rinke, et al., Discriminant analysis in the presence of interferences: Combined application of target factor analysis and a Bayesian soft-classifier, Anal. Chim. Acta (2012), http://dx.doi.org/10.1016/j.aca.2012.09.042

Table 1
Posterior probabilities calculated at $r_{LL} = 0.8$, $\alpha = 0.05$.

Substrate	Sample	# of samples	Average $P(\omega_i\|[r_{LL},1])$			Fraction $P(\omega_i\|[r_{LL},1])$ not calculated
			CJ	MJ	NJ	
Steel	CJ	12	1.00	0.00	0.00	0.00
	MJ	3	0.61	0.39	0.00	0.67
	NJ	3	0.00	0.00	1.00	0.00
Porcelain	CJ	12	1.00	0.00	0.00	0.00
	MJ	3	0.00	1.00	0.00	0.33
	NJ	3	0.00	0.00	1.00	0.00
Steel	CJ/MJ	3	0.80	0.20	0.00	0.00
	CJ/NJ	3	0.49	0.00	0.51	0.00
	MJ/NJ	3	0.28	0.00	0.72	0.00
Porcelain	CJ/MJ	3	0.73	0.27	0.00	0.33
	CJ/NJ	3	0.83	0.00	0.17	0.00
	MJ/NJ	3	0.24	0.26	0.50	0.33
Bullet hole (steel)	CJ	15	0.91	0.00	0.09	0.60
	MJ	4	0.00	0.00	0.00	1.00
	NJ	4	0.00	0.00	1.00	0.00

When the analysis described for the data in Fig. 4 is applied to the spectra from each of the transfer lines on porcelain and steel, and the bullet-hole data, the results are given in Tables 1 and 2. Table 1 gives the results for the analysis with the lower integration cutoff, $r_{LL} = 0.8$ and $\alpha = 0.05$, while Table 2 gives the results for $r_{LL} = 0.9$ and $\alpha = 0.05$. The first column gives the substrate to which the sample was applied and the second column gives the classes of bullet jacket applied to the substrate. The third column gives the number of samples analyzed and the following three columns give the average $P(\omega_i|I_i[r_{LL},1])$ for each of the three classes (CJ, MJ and NJ). The last column gives the fraction of samples for which $I_i[r_{LL},1] < \alpha$ for all three classes (i.e., the fraction of samples that were not classified).

The results for single class transfer to steel and porcelain substrates are given in the first two entries under the heading 'Substrate' in both Tables 1 and 2. All of the CJ samples on steel and porcelain met the requirements for posterior probability calculation ($\alpha = 0.05$, $r_{LL} = 0.8$, Table 1) for at least one class and the average posterior probability for the CJ class was 1.0. When the requirements for posterior probability calculation are changed to $\alpha = 0.05$ at $r_{LL} = 0.9$, see Table 2, CJ on steel and porcelain gave average $P(\omega_i|I_i[r_{LL},1]) = 1$; however, none of the classes met the requirements for posterior probability calculation in 33% of the CJ transfer samples on porcelain.

For NJ class transfers on steel and porcelain, an average $P(\omega_i|I_i[r_{LL},1]) = 1$ was calculated for $\alpha = 0.05$ at $r_{LL} = 0.8$ (Table 1) and the same performance was observed for $\alpha = 0.05$ at $r_{LL} = 0.9$ (Table 2). Two of three MJ bullet transfers on steel failed to meet the requirements for posterior probability calculation at $\alpha = 0.05$ and $r_{LL} = 0.8$ for every test class, and therefore the two samples were not classified. Average posterior probabilities of 0.61, 0.39 and 0.0 were calculated for the CJ, MJ and NJ classes respectively for the one remaining MJ dataset. On the porcelain substrate, one sample did not classify and the other two MJ transfer samples had $P(\omega_i|I_i[r_{LL},1]) = 1$, Table 1. The MJ transfer data has a high fraction of samples where $P(\omega_i|I_i[r_{LL},1])$ is not calculated and in one instance (MJ on steel) the posterior probability for the CJ class exceeds the probability for the MJ class. The poorer performance by the MJ transfer data may be related to a smaller amount of material transferred to the substrate. The MJ jacketing material is harder than copper or lead, which comprise the major chemical components in the CJ and NJ classes.

Results for the two-class transfer are given in Tables 1 and 2 in the lines under 'Sample' giving two classes (i.e., CJ/MJ), and the class-conditional probabilities and posterior probabilities for CJ/MJ on steel are shown in Fig. 5. In these tests, two classes of bullet jacketing material are present and TFA should result in higher correlations between the test and predicted spectra for members of the two classes. The calculated $P(\omega_i|I_i[r_{LL},1])$ values for each of the two classes would ideally approach 0.5; however, if the correlations are higher for one of the two classes, then the $P(\omega_i|I_i[r_{LL},1])$ values will deviate from the ideal value. The $P(\omega_i|I_i[r_{LL},1])$ values may deviate from ideal behavior, for example, if one analyte is present in higher concentration, leading to better correlations with library

Table 2
Posterior probabilities calculated at $r_{LL} = 0.9$, $\alpha = 0.05$.

Substrate	Sample	# of samples	Average $P(\omega_i\|[r_{LL},1])$			Fraction $P(\omega_i\|[r_{LL},1])$ not calculated
			CJ	MJ	NJ	
Steel	CJ	12	1.00	0.00	0.00	0.00
	MJ	3	0.00	0.00	0.00	1.00
	NJ	3	0.00	0.00	1.00	0.00
Porcelain	CJ	12	1.00	0.00	0.00	0.33
	MJ	3	0.00	0.00	0.00	1.00
	NJ	3	0.00	0.00	1.00	0.00
Steel	CJ/MJ	3	0.67	0.33	0.00	0.00
	CJ/NJ	3	0.62	0.00	0.38	0.00
	MJ/NJ	3	0.00	0.00	1.00	0.33
Porcelain	CJ/MJ	3	0.50	0.50	0.00	0.33
	CJ/NJ	3	0.87	0.00	0.13	0.00
	MJ/NJ	3	0.50	0.00	0.50	0.33
Bullet hole (steel)	CJ	15	0.37	0.00	0.63	0.93
	MJ	4	0.00	0.00	0.00	1.00
	NJ	4	0.00	0.00	1.00	0.25

Please cite this article in press as: C.N. Rinke, et al., Discriminant analysis in the presence of interferences: Combined application of target factor analysis and a Bayesian soft-classifier, Anal. Chim. Acta (2012), http://dx.doi.org/10.1016/j.aca.2012.09.042

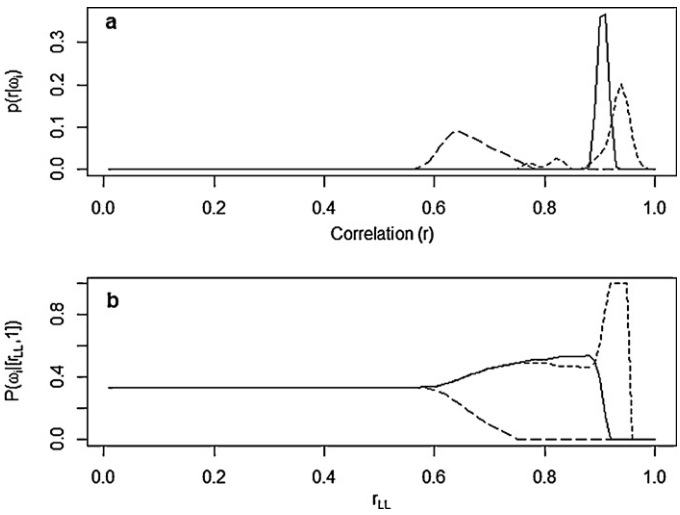


Fig. 5. (a) Class-conditional distribution functions for correlation of test and predicted spectra from CJ (short dashes), MJ (medium dashes) and NJ (solid line) bullet jacket classes. (b) Posterior probabilities with $\alpha = 0.05$ calculated as a function or r_{LL}. Data derived from analysis of a set of spectra taken from a CJ and NJ transfer line on steel.

spectra. The data in Table 1 show that most of the two-class transfer samples give $P(\omega_i|I_i[r_{LL},1]) \neq 0$ for the classes transferred; however, the $P(\omega_i|I_i[r_{LL},1])$ values differ from the ideal values of 0.5. The $P(\omega_i|I_i[r_{LL},1])$ values for the MJ class tend to be smaller than the co-transferred material, e.g. $P(\omega_i|I_i[r_{LL},1])$ of 0.8 and 0.2 for CJ and MJ classes for the CJ/MJ transfer on steel. In the MJ/NJ transfer to porcelain, non-zero posterior probabilities were obtained for all three classes, and the MJ/NJ transfer to steel, posterior probability of zero was obtained for the MJ class, while the CJ class was incorrectly identified as present with a posterior probability of 0.28. When the posterior probability calculation criteria were changed to $\alpha = 0.05$ at $r_{LL} = 0.9$ (Table 2 data), the general trends in posterior probabilities are similar to those in Table 1. The most difficult sample to correctly identify was the MJ/NJ sample on both steel and porcelain. This sample corresponds to transfer to the hardest and softest jacket materials and would likely result in the largest difference in the amount of material transferred.

The last entries in the 'Substrate' column in Tables 1 and 2 are the data from analysis around the holes produced when the bullets were fired through steel plates. One significant result from these tests is a large jump in the fraction of CJ and MJ samples which did not meet the criteria for posterior probability calculation for any of the bullet jacketing classes. None of the MJ bullets were classified. The CJ bullets that met the criteria for $P(\omega_i|I_i[r_{LL},1])$ calculation had posterior probabilities for CJ and NJ or 0.91 and 0.09, respectively at $r_{LL} = 0.8$ and $\alpha = 0.05$, whereas values of 0.37 and 0.63 were calculated at $r_{LL} = 0.9$, $\alpha = 0.05$. The posterior probability values calculated for the CJ bullets included a significant contribution from the NJ bullets. It should be noted that the CJ bullets are comprised of a copper jacket around lead. If lead is deposited on the steel plate as the CJ bullet passes through the plate, a non-zero posterior probability for the NJ class can be expected. The NJ bullet holes had the lowest fraction that failed to meet the criteria for posterior probability calculation and an average $P(\omega_i|I_i[r_{LL},1]) = 1.0$ was obtained for these samples. The NJ bullets are composed of lead, the softest material and are likely to give the highest amount of transfer.

All of the data discussed above demonstrate application of the data analysis method to samples from different and complex, substrates. The bullet-hole data provided an additional interesting observation. The spectra of the samples all contained two strong emissions at 455.40 nm and 493.40 nm that were not present in the library. These transitions are tentatively assigned Ba II and likely arise from gunshot residue deposited as a result of the proximity of the firearm to the steel plate. The gunshot residue in these samples contributes to the background interference.

4. Conclusions

The combined TFA and Bayesian soft classifier are intended to produce posterior probabilities that a sample contains a representative analyte from one of the classes in a library. The method may also indicate that a system does not contain any of the analytes represented in the library. The resulting posterior probabilities are intended to aid the analyst in classifying the sample in the presence of an unknown background signal. The method is designed to work in cases where spectra of multiple analytes from the classes in question are known. The method gives reasonable results for single analyte transfers investigated in this study when the analyte is readily transferred to the substrate (i.e., the CJ and NJ bullet jacketing). In cases where the correlations between test and predicted spectra are low and a sensitivity α is incorporated, the method becomes conservative. Generalization of the results presented here requires further testing of the method on larger data sets and larger libraries of target factors. In related work, results have been published elsewhere for application of the method, utilizing a library of 358 target factors, for the determination of ignitable liquid residues in fire debris containing interferences from pyrolysis products [18].

Acknowledgements

This work was supported under award number W911NF0610446 from the U.S. Army Research Office and award number 2004-IJ-CX-K031 from the Office of Justice Programs, National Institute of Justice, Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Army or U.S. Department of Justice. The work was done jointly at the Center for Research and Education in Optics and Lasers and the National Center for Forensic Science at the University of Central Florida. The authors acknowledge assistance by Christine Murphy (Florida Department of Law Enforcement firearms analyst), Lt. Jerry Emert and Sgt. Eric Walton (University of Central Florida Police Department).

References

[1] K. Varmuza, P. Filzmoser, Introduction to Multivariate Statistical Analysis in Chemometrics, CRC Press, Boca Raton, FL, U.S.A., 2009.
[2] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley Interscience, New York, NY, U.S.A., 2001.
[3] M.M. Barton, G.M. Miskelly, Chem. New Zealand 70 (2006) 3–6.
[4] V. Sikirzhytski, K. Virkler, I.K. Lednev, Sensors 10 (2010) 2869–2884.
[5] G.P. Campbell, J.M. Curran, G.M. Miskelly, S. Coulson, G.M. Yaxley, E.C. Grunsky, S.C. Cox, Forensic Sci. Int. 188 (2009) 81–90.
[6] B. Knauthe, S. Steffen, M. Barth, L. Niewoehner, M. Otto, J. Chemometr. 22 (2008) 252–258.
[7] L. Gottfried, F.C. De Lucia, C.A. Munson, A.W. Miziolek, J. Anal. At. Spectrom. 23 (2008) 205–216.
[8] O.M. Primera-Pedrozo, L.C. Pacheco-Londono, L.F. De la Torre-Quintana, S.P. Hernandez-Rivera, R.T. Chamberlain, R.T. Lareau, Proc. SPIE 5403 (Pt. 1, Sensors, and Command, Control Communications, and Intelligence (C31) Technologies for Homeland Security and Homeland Defense III) (2004) 237–245.
[9] J.B. Nikas, C.D. Keene, W.C. Low, J. Comp. Neurol. 518 (2010) 4091–4112.
[10] C. Sarbu, H.F. Pop, R.-S. Elekes, G. Covaci, Rev. Chim-Bucharest 59 (2008) 1237–1241.
[11] K. Murayama, M. Tomida, Y. Ootake, T. Mizuno, J.-I. Ishimaru, ITE Lett. Batt. New Technol. Med. 7 (2) (2006) 182–185.
[12] V. Arcenegui, C. Guerrero, J. Mataix-Solera, J. Mataix-Beneyto, R. Zornoza, J. Morales, A.M. Mayoral, Catena 74 (2008) 177–184.
[13] E.R. Malinowski, D.G. Howery, Factor Analysis in Chemistry, John Wiley & Sons, New York, U.S.A., 1980.
[14] D. González-Arjona, G. López-Pérez, A.G. González, Talanta 49 (1999) 189–197.

[15] D. González-Arjona, G. López-Pérez, A.G. González, Chemometr. Intell. Lab. Syst. 57 (2001) 133–137.
[16] C.M. Bridge, J. Powell, K.L. Steele, M. Williams, J.M. MacInnis, M.E. Sigman, Appl. Spectrosc. 60 (2006) 1181–1187.
[17] C.M. Bridge, J. Powell, K.L. Steele, M.E. Sigman, Spectrochim. Acta B 62 (2007) 1419–1425.
[18] M.R. Williams, M.E. Sigman, J. Lewis, K. McHugh-Pitan, Forensic Sci. Int. 222 (2012) 373–386.
[19] E.R. Malinowski, Determination of rank by median absolute deviation (DRMAD): a simple method for determining the number of principal factors fresponsible for a data matrix, J. Chemometr. 29 (2009) 1–6.
[20] B.W. Silverman, Density Estimation for Statistics and Data Analysis, Chapman and Hall, New York, 1986.
[21] J.R. Eastman, R.M. Laney, Bayesian Soft Classification for Sub-Pixel Analysis: A Critical Evaluation, Photogramm. Eng. Remote Sens. 68 (2002) 1149–1154.

Please cite this article in press as: C.N. Rinke, et al., Discriminant analysis in the presence of interferences: Combined application of target factor analysis and a Bayesian soft-classifier, Anal. Chim. Acta (2012), http://dx.doi.org/10.1016/j.aca.2012.09.042