Microchemical Journal 109 (2013) 122–127

Contents lists available at SciVerse ScienceDirect

Microchemical Journal

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

Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis

Carolina Santos Silva [a], Flávia de Souza Lins Borba [a], Maria Fernanda Pimentel [b,*], Marcio José Coelho Pontes [c], Ricardo Saldanha Honorato [d], Celio Pasquini [e]

[a] Departamento de Química Fundamental, Universidade Federal de Pernambuco, Recife, PE, Brazil
[b] Departamento de Engenharia Química, Universidade Federal de Pernambuco, Recife, PE, Brazil
[c] Departamento de Química, Universidade Federal da Paraíba, João Pessoa, PB, Brazil
[d] Superintendência Regional em Pernambuco, Polícia Federal, Brazil
[e] Instituto de Química, Universidade Estadual de Campinas, Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 29 November 2011
Received in revised form 27 February 2012
Accepted 24 March 2012
Available online 30 March 2012

Keywords:
Infrared reflectance spectroscopy
Blue ink pen identification
Linear discriminant analysis
Successive projections algorithm
Genetic algorithm
Stepwise selection

ABSTRACT

Attenuated total reflectance (ATR) Fourier transform infrared (FTIR) spectroscopy associated to linear discriminant analysis (LDA) was employed to perform classification of blue pen ink according to types and brands, in a nondestructive way. To build a representative data set, blue pens of 3 types, namely ballpoint (5 brands), roller ball (2 brands) and gel (3 brands) were purchased from local dealers. Ten different pens, representing the best seller of each brand, were purchased, making a total of 100 pens. Circular areas were painted five times with each pen and spectra were taken in 2 different locations, using a Universal Attenuated Total Reflectance accessory (UATR), within the range of 4000 to 650 cm^{-1}. Three types of paper were employed: two brands of A4 sulfite paper (paper 1 and paper 2) and one recycled paper (paper 3). The genetic algorithm (GA), stepwise formulation (SW) and successive projections algorithm (SPA) were employed to select spectral variables employed in LDA. LDA models were built using the blue pen ink spectra obtained from paper 1. Three test sets were employed using the blue pen ink spectra obtained from papers 1, 2 and 3, in order to evaluate the influence of the paper on the predictions. The LDA models used to classify the pens according to their type (gel, rollerball and ballpoint) achieved a correct classification rate of 100% in the test set composed of blue pen ink spectra obtained from paper 1, using GA and SPA. Using SW, the rate achieved was 99.5%. For paper 2, SPA, GA and SW provided 100%, 97.3% and 93.8% of correct classification, respectively. For paper 3, SPA, GA and SW achieved a correct prediction rate of 100%, 100% and 94.9%, respectively. LDA models for classifications of pens according to their brand were 100% correct in their classification when the test set was composed of blue pen ink spectra obtained from papers 1 and 2. For the test set composed of blue pen ink spectra obtained from paper 3, LDA-SPA, LDA-GA and LDA-SW classified them correctly at 91.3%, 100% and 100%, respectively. The method developed was able to differentiate successfully all brands of pen used on each type of paper and could be a helpful tool for detection and confirmation of counterfeits in documents of legal importance.

1. Introduction

Forensic document examination is a field in forensic science that studies the manipulation and falsification of documents, such as bank checks, medical certificates, working papers and passports. The evaluation of inks used to produce manuscripts can help in the identification of counterfeits [1]. Pen inks consist of complex systems, which require a combination of compounds to provide, not only the color, but also other characteristics associated to quality. For ink production, dyes, lubricants, resins, pigments, solvents, surfactants, emulsifiers and substances for pH control are usually employed.

The methods for ink analysis can be destructive or non-destructive. Using destructive methods, a portion of the ink has to be removed from the document using appropriate solvents. Different destructive analytical methods have been evaluated for the study of organic composition in pen inks, such as: UV–vis spectrometry [2–4]; Fourier transform infrared spectroscopy (FTIR) [4–6]; thin layer chromatography [4]; positive and negative-ion electrospray ionization mass spectrometry [7]; and high performance liquid chromatography (HPLC) [8]. Although these methods can identify some of the ink compounds, they destroy the document and cannot be applied in many forensic cases.

* Corresponding author at: Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Departamento de Engenharia Química. Rua Prof. Arthur de Sá S/N, Cidade Universitária, 50740-521, Recife, PE, Brazil. Tel.: +55 81 21268729; fax: +55 81 21267235.
E-mail addresses: mfernanda.pimentel@gmail.com, mfp@ufpe.br (M.F. Pimentel).

0026-265X/$ – see front matter
doi:10.1016/j.microc.2012.03.025

Non-destructive methods for forensic discrimination of pen inks have been reported mainly using Raman [5], UV-Visible, luminescence [9] and reflectance infrared [8] spectroscopies. Raman and reflectance infrared spectroscopy provides information about the chemical constituents of the inks, such as solvents, and resins, other than colorants.

Reflectance microspectrophotometry in the visible range is one of the standard non-destructive techniques used to discriminate among types of pen ink [10]. Forensic discrimination of ballpoint ink writing on paper is also routinely carried-out, non-destructively, using luminescence spectroscopy [3]. Zieba-Palus and Kunicki have shown that Raman spectrometry is valuable as a complementary technique to infrared spectrometry for analyzing the ink composition of 69 samples of blue and black inks originating from ballpoint pens and gel pens of various brands and manufacturers, commonly available in Poland [5]. Synchrotron-based FT-infrared micro-spectroscopy has been used by Wilkinson et al. for the comparative identification of writing inks [11]

Mass spectrometry has been also employed for non-destructive analysis of pen inks. Jones et al. reported on initial studies using mass spectrometry with an ion source called Direct Analysis in Real Time (DART) in situ for the non-destructive analysis of writing inks on paper without visible alteration [12]. Forty-three different black and blue ballpoint, black fluid, and black gel inks were examined. They have shown that mass spectrometry using DART produces spectra of sufficient quality and diversity to distinguish among all but the most similar of inks. Denman et al. [13] have shown that surface analysis by time-of-flight secondary ion mass spectrometry (TOF-SIMS) of blue ballpoint pen ink markings can non-destructively analyze organic and inorganic ink components directly with no interference from the paper substrate. Despite the fact that these techniques have been reported as non-destructive, a partial ink-extraction still occurs. Consequently, the documents are slightly modified.

Although research has been carried out to develop efficient analytical methods regarding the composition of inks, in most of the previously mentioned works, discrimination is carried-out by visual inspection to decide whether the two samples show the same spectra/chromatogram under the same experimental conditions [2]. Few papers describe the use of chemometric multivariate analysis to explore the analytical data efficiently and to enable unbiased decisions about the similarities among the ink samples.

Adam has shown that principal component analysis (PCA), followed by linear regression of the loadings, applied to a data set obtained from 10 black ballpoint pens using a standard luminescence spectrometer facilitates the separation of the ink luminescence from that of the paper and also allows a direct comparison between two ink lines drew on a document [9]. PCA has been applied to the UV–Vis spectra of inks obtained from black ballpoint pens available on the market in United Kingdom. For the complete set of 25 pens, interpretation of the loadings for the first few principal components showed that both the pen inks and the extracted (with ethanol) ink lines could be classified in an objective manner and in agreement with the results of parallel thin layer chromatography studies [3]. Thanasoulias et al. have employed cluster analysis (CA), PCA and discriminant analysis (DA) to UV–vis spectral data. Five commercially available brands of blue ball-point pen inks were used for the study. For each brand, each of 10 pens from the same batch were sampled by means of a stainless steel needle that was used to penetrate the wall of the plastic ink reservoir and transfer a small portion of the sample into the solvent (ethanol) [2]. This study showed a 100% correct classification of the training dataset between inks of different brands. The authors state that a more thorough study of the system could be made by using more ink samples, from different batches and that the models should also be tested on new samples, instead of items from the training dataset, to ensure a better assessment of the usefulness of the models. Using TOF-SIMS, organic and inorganic information taken from a total of 24 blue ballpoint pens (including replicates) were collected simultaneously and processed with PCA, identifying 41 out of 45 pairs (91%) of the pens analyzed [13]. Wang et al. developed a method for classifying blue ballpoint pen inks using FT-IR, that allowed a total of 108 samples to be divided into 35 subgroups using a pattern recognition system [6]. This method was found to be fast and reliable for bulk ink analysis, however, it has not yet been applied to identify inks on paper.

Kher et al. [8] applied PCA and linear discriminant analysis (LDA) to a data set composed of Micro-ATR infrared spectra obtained, non-destructively, from eight brands of six blue ballpoint pens each (totalizing 48 pens). HPLC analyses of the inks were also carried-out. On the IR data set, two approaches were used for data reduction before LDA analysis. The first approach consisted of using the PCA decomposed matrix while the second consisted of choosing a few discrete spectral features. The best result was obtained using the PCA scores which came up with a correct classification of only 62.5% of the pen ink samples. Applying LDA to the HPLC data set resulted in a correct classification of 97.9%.

The use of PCA scores or a few select wavenumbers, chosen after a priori consideration, as input variables for LDA, may cause loss of information relevant for classification. In order to overcome this problem, variable selection methods such as the genetic algorithm (GA) [14] and stepwise formulation (SW) [15] were employed to select the spectral variables for LDA in several classification problems. Recently, the successive projections algorithm (SPA) has been successfully applied to the classification of edible vegetable oils [16,17], diesel fuels [16], Brazilian soils [18], cigarettes [19], coffee [20] and diesel/biodiesel [21] samples.

The present paper proposes an analytical method based on infrared spectroscopy to classify blue inks from different types of pens (gel, rollerball and ballpoint pens), as well as inks from different brands for each type. For this purpose, the successive projections algorithm, genetic algorithm and stepwise formulation were employed to choose an appropriate subset of wavenumbers for a linear discriminant analysis model.

2. Materials and methods

2.1. Samples

Three different types of paper were employed: two brands of A4 print paper (paper 1 and paper 2) and one brand of recycled A4 paper (paper 3). To build a representative data set, blue pens of 3 types: ballpoint (5 brands), roller ball (2 brands) and gel (3 brands) were purchased commercially. Ten different pens, representing the best seller of each brand, were purchased: 5 from the same batch and 5 from different batches, making a total of 100 pens. Circular areas with a 1 cm radius were painted five times with each pen, on paper 1, and spectra were taken from 2 different locations of each. On paper 2 and 3, three pens of each brand were chosen from different batches and used to paint circular areas from which 5 spectra were taken.

2.2. Spectra acquisition

Spectra were acquired using *Spectrum400* (Perkin Elmer) equipment with a universal attenuated total reflectance (UATR) accessory, in the range 4000 to 650 cm^{-1}, with 4 cm^{-1} resolution and by averaging 16 scans. Spectra of each paper were also taken and subtracted from the ink pen spectrum. The integrity of the sample was completely preserved during spectra acquisition.

2.3. Classification models

LDA is a well-known technique for dimensionality reduction. In LDA, the dimensional embeddings are reduced in such a way that the