Dra. Mónica Marro Sánchez Institut de Ciències Fotòniques (ICFO)

Dra. Anna de Juan Capdevila Departament d'Enginyeria Química i Química Analítica



Treball Final de Grau

Raman spectroscopy imaging and chemometrics to investigate forensic documents involving delinquent actions such as crossed lines and obliteration.

Imatge d'espectroscòpia Raman i quimiometria per investigar documents forenses relacionats amb accions delictives com línies creuades i obliteració.

Manel Madrid Valero

January 2022





Aquesta obra esta subjecta a la llicència de: Reconeixement–NoComercial-SenseObraDerivada



http://creativecommons.org/licenses/by-nc-nd/3.0/es/

An expert is a person who has made all the mistakes that can be done in a certain field.

Niels Bohr

En especial, a Mónica y Anna por la implicación y el apoyo prestado, ha sido un placer trabajar junto a vosotras.

A toda la gente de ICFO y al equipo de Pablo Loza, por acogerme y por hacerme sentir uno más durante estos meses. A Adri y Marina, por hacerme partícipe de este gran proyecto y por implicaros tanto en mi investigación estando tan lejos de mí. Os deseo lo mejor en los caminos que habéis emprendido.

Por último, y no menos importante, a mi familia y amigos por ser mi gran apoyo y confiar siempre en mí durante estos 4 maravillosos años de aprendizaje constante y de crecimiento personal, os quiero.

Gracias.



CONTENTS

1. SUMMARY	3
2. Resum	5
3. INTRODUCTION	7
3.1. Fighting document fraud throughout history	7
3.2. Hyperspectral imaging. Present and future of document analysis	8
3.3. Document forgery types	10
3.4. Pen inks. Classification and composition	11
4. OBJECTIVES	13
5. EXPERIMENTAL	14
5.1. Samples. Ink typologies and forgery simulations	14
5.2. Raman measurements. Single point spectra and hyperspectral images	15
6. DATA TREATMENT. CHEMOMETRIC TOOLS	17
6.1. General spectral preprocessing	17
6.2. Principal Component Analysis (PCA). Ink exploratory analysis	19
6.3. Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS).	
Ink image unmixing	20
7. RESULTS AND DISCUSSION	24
7.1. Ink exploratory analysis. PCA results	24
7.2. Raman images unmixing. MCR-ALS results	27
7.2.1. Crossed lines cases	28
7.2.2. Obliteration cases	35
8. CONCLUSIONS	38
9. REFERENCES AND NOTES	39
10. ACRONYMS	40
APPENDICES	41
Appendix 1: Local PCA models	42
Appendix 2: Distribution maps of crossed lines and obliterations	44

1. SUMMARY

Document forgery is a serious crime followed by judicial authorities. The most common cases of fraud in handwritten documents are related to character modification or text deletion, known as falsifications through crossed lines and obliterations, respectively.

The use of destructive and subjective techniques in the last decades has hindered forensic investigations. This is the reason why this work proposes a non-destructive, quick, objective, and trustworthy analytical technique for forensic detection of fraud in handwritten documents, based on the combination of hyperspectral Raman imaging with the image unmixing method Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS).

The beginning of the research consisted of investigating the potential of Raman spectroscopy to distinguish commercial blue inks. Ink differentiation using Raman spectroscopy has proven to be very satisfactory for inks of different brands and typologies.

Hence, the next step was the acquisition of Raman images in simulated cases of document falsification using confocal microscopy. The distribution maps and spectral signatures obtained for every ink in these forgeries applying spectral preprocessing and MCR-ALS have been very satisfactory, proving that the used analytical techniques are appropriate and allow obtaining detailed information regarding the distribution of inks throughout the paper in any written document.

It has been observed that the nature of inks leads to different interactions in the crossed strokes and in the ink distribution on the paper. The chronological order of ink strokes application has been correctly solved in almost all crossed lines studied cases, except in a reported case of inks of a particular brand.

Obliterated texts have been satisfactorily recovered, irrespective of the ink typology applied over them. We have proven that fast Raman imaging operates correctly to detect obliterations and that an increase of spatial resolution allows recovering complicated and small characters very accurately.

Keywords: Chemometrics, Raman spectroscopy, hyperspectral images, MCR-ALS, forensic, ink, crossed lines, obliterations.

2. RESUM

La falsificació de documents és un greu delicte molt perseguit per les autoritats judicials. Els casos més comuns de frau es relacionen amb la modificació de caràcters o l'eliminació de text, denominats falsificacions per línies creuades i obliteracions, respectivament.

La utilització de tècniques analítiques destructives i subjectives en les últimes dècades ha dificultat les investigacions forenses. Per aquest motiu, aquesta memòria proposa una tècnica analítica no destructiva, ràpida, objectiva i fiable per a la detecció forense de frau documental en documents escrits, que combina la imatgeria hiperespectral de Raman amb el mètode quimiomètric de desmesclat d'imatge anomenat resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS).

L'inici del treball va consistir en estudiar el potencial de l'espectroscòpia Raman per a la diferenciació de tintes blaves comercials. La diferenciació de tintes mitjançant Raman va ser molt satisfactòria entre tintes de fabricants diferents i de natura diversa.

Com a conseqüència, la següent etapa va consistir en l'adquisició d'imatges Raman en casos simulats de frau documental emprant la tècnica de microscòpia confocal. Els mapes de distribució i les signatures espectrals obtingudes per a cadascuna de les tintes de les falsificacions aplicant les eines de preprocessat espectral i MCR-ALS van ser molt satisfactòries i van confirmar que les tècniques analítiques emprades són adients per obtenir informació detallada sobre la distribució de les tintes presents en qualsevol document escrit.

La diferent natura de les tintes ha permès observar diferents interaccions entre els traços i diferents formes de distribuir-se sobre el paper. L'ordre cronològic en l'aplicació dels traços de les tintes s'ha resolt correctament en quasi tots els casos de línies creuades estudiats, excepte en un cas descrit de tintes d'una marca concreta.

Els textos obliterats han estat recuperats satisfactòriament, independentment del tipus de tinta aplicada sobre ells. S'ha vist que la imatgeria Raman ràpida funciona correctament en la detecció d'obliteracions i que l'augment de resolució espacial permet recuperar acuradament caràcters complicats i de petites dimensions.

Paraules clau: Quimiometria, espectroscòpia Raman, imatges hiperespectrals, MCR-ALS, forense, tintes, línies creuades, obliteracions.

3. INTRODUCTION

3.1. FIGHTING DOCUMENT FRAUD THROUGHOUT HISTORY

Despite being one of the earliest fields in forensic science, forensic investigation of questioned documents has always been one of the most important issues in police and judicial investigations. Nevertheless, the most important forensic advances in the development of physicochemical methods to detect these criminal actions occurred about halfway through the 20th century¹.

The 60's and 70's were characterized by the physical examination of documents using optical microscopy^{2,3,4} and Scanning Electron Microscopy (SEM)^{5,6}. Optical microscopy was very simple but SEM was an expensive and complicated technique that damaged the document. Besides, both techniques were very dependent on visual examination from an expert.

In the 80's, the lifting technique was developed. It consisted of transferring the inks in a crossing to another paper or suitable surface by applying pressure and heat, expecting the ink applied last to be transferred in a greater quantity. Lifting combination with SEM was the most popular trend in document examination⁷. However, the results this technique offered were very affected by different factors, such as the pressure applied, the inks that were used, the subjectivity in results examination and the damaging nature of the lifting technique.

The 90's continued with the introduction of new methods such as Electrostatic Deposition Analysis (ESDA)⁸ and Laser Scanning Confocal Microscopy (LSCM)⁹.

Until this moment, great advances in document analysis were reached. Nonetheless, all the techniques continued to rely on the visual interpretation of a professional, since all of them were based on physical features. Meanwhile, alternative chemical methods, such as Thin Layer Chromatography (TLC), High-Performance Liquid Chromatography (HPLC) or Gas Chromatography (GC)¹⁰, were destructive and only provided information on the inks used. These examinations were useful in some cases, but the inks' behaviour when mixed was still unknown and generated disagreement and inconclusive results.

Since the beginning of the 21st century, after the promising initial studies of Claybourn and Ansell¹¹, who carried out a single point Raman depth analysis in questioned documents, researchers noticed that it was necessary to cover and analyse spatially all the suspicious regions of the document. In this way, a real representative analysis and spectral information of

all the dubious regions could be obtained and, consequently, an objective methodology to demonstrate the sequence of writing in crossing lines and the covered text when treating obliteration cases.

In the last decade, hyperspectral imaging emerged in document examination. This technique, combined with several chemometric methods, emerged as the most powerful option to carry out automated, precise, more informative and reliable examinations of all the questioned zone of the document. Unlike the rest of the previously mentioned techniques, it provides non-destructive and more objective examinations employing the spectral signature of the different components that are in the document, in order to identify them and sequence the order of writing or find out the covered text. The fact of using chemical information and chemometric methods to visualize and understand the obtained results improves the issue related to the uncertainty of visual examination and the different interpretations people could extract with earlier physical techniques¹.

3.2. HYPERSPECTRAL IMAGING. PRESENT AND FUTURE OF DOCUMENT ANALYSIS

The large variety of optical phenomena (absorption, emission, scattering, photoluminescence, chemiluminescence) and the possibility of studying these optical responses in different ranges of the electromagnetic spectrum opens up a huge number of spectroscopic techniques to study, identify and distinguish atoms, ions, molecules and supramolecular structures.

Often, the samples we are used to treating are homogeneous and uniform in the space because they tend to be in solution form. Therefore, we only need to obtain a single point spectrum that represents accurately the totality of the sample.

Nevertheless, sometimes we will have to deal with solid samples that have different compositions depending on the spatial zone analysed. In these cases, a single point spectrum will not represent the totality of the sample and we will need a technique able to obtain spatial and spectral information simultaneously. This is the situation when we think about document forgery, as we will face zones of our document that will consist only of one ink and areas in which we will have several inks mixed or superimposed. Moreover, we cannot forget that we will always have the spectral contribution of the paper where the document has been written. Hyperspectral images are the best option to record spectra of non-homogenous samples.

They are a chemical image of a sample surface, where every pixel is associated with a full spectrum. Thus, we will have spatial information in a plane (x and y pixel coordinates), and in addition, we will generate a spectral coordinate (λ n), creating a hyperspectral cube¹² (*Fig. 1*).



Figure 1. Scheme of a hyperspectral image. The hyperspectral cube is formed by the spatial coordinates x and y that form the image, and by the spectral coordinate (λ). Measuring the spectral response of a determined pixel at different wavelengths allows obtaining the complete spectrum at that point.

Payne et al. were the first to use absorption and photoluminescence (fluorescence) in the UV-vis-NIR regions to obtain chemical images of different document forensic samples, such as paints, tapes, adhesives, inks and firearms propellants¹³. The authors compared this technology with the traditional techniques and remarked its potential, the short analysis time, the ability to display visual results, the lack of sample preparation, the non-destructive character and the greater discriminating power compared to earlier techniques.

After this first contact with hyperspectral imaging, document investigation focused on two types of spectroscopy that gave the best results when obtaining chemical images of questioned zones:

The first kind of images were based on infrared spectroscopy. In 2008, Bojko et al. examined different cases of crossed lines using Attenuated Total Reflectance - Fourier Transform Infrared Spectral Imaging (ATR-FTIR)¹⁴. This technique resulted to be successful in determining sequences of ballpoint pen and laser printer lines but failed when gel pens, rollerball pens, felt-tip pens or crossings between two different ballpoint pens were used.

Silva et al. used hyperspectral Images in the near infrared range (HSI-NIR) in diverse added text, crossed lines and obliteration cases, together with different chemometric techniques such as Principal Component Analysis (PCA), Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) and Partial Least Squares-Discriminant Analysis (PLS-DA)⁴. The authors were able to obtain good results analysing distribution maps in some cases, but they reported difficulties discriminating infrared spectra among inks and also among inks and the paper, which showed a high spectral contribution.

On the other hand, Raman hyperspectral imaging has given very interesting results in different independent investigations and has positioned itself as the most promising and powerful spectroscopic tool, especially due to the great discriminating power Raman spectroscopy offers.

Braz et al. presented the potential of Raman imaging in combination with MCR-ALS in the study of crossed lines between different blue ink pens. They described the significant differences of ink arrangements in the intersections depending on the type of inks that were used (gel-based oil-based and liquid-based inks) as well as the consistency in the inks distribution and order of crossing when different times of application of the second ink were used. However, the study of crossed lines only at the paper surface gave rise to some inconclusive results.¹⁵

De Juan et al.¹⁶ carried out a depth profiling study of crossed lines and obliteration cases combining confocal Raman imaging and MCR-ALS. The results were very satisfactory and the analysis of the sample in-depth improved the crossed lines results interpretation and allowed deciphering the obliterated text with easiness and precision. Unfortunately, very few cases were analysed and only between oil-based pens, which did not describe the full power of the method.

We have the challenge of extending the trustworthiness of Raman imaging combined with MCR-ALS to different combinations of blue pen inks so that this method proves as the future of forensic document investigation whatever the used inks and their application order is.

3.3. DOCUMENT FORGERY TYPES

The major objectives when forging a handwritten document with another pen is to modify some relevant character (letters, numbers, signs, signatures...), to get rid of some outstanding phrase, word or number or to add false text after the document has been officially written. All of these efforts are done with the maximum possible discretion, trying to use similar pen ink colours and calligraphy. We classify these criminal actions as the following (*Fig. 2*):

<u>Crossed lines</u>: Action of modifying the meaning of a character or a group of them in a document by writing above and around it or them, causing an intersection between different ink strokes.

<u>Obliteration</u>: Action of eliminating partially or completely some text of a document crossing it out. In partial obliteration, we can obtain information of the pure ink that has been obliterated (original text ink)¹⁶.

<u>Addition of text</u>: Adding untrue text after the finalisation of a document. Forgery by adding text will not be studied directly in this research since it is an easier case, thanks to the access at both pure ink Raman spectra and the absence of ink mixing⁴.



Figure 2. Examples of crossed lines (a), added text (b) and obliteration (c) forgeries.

3.4. PEN INKS. CLASSIFICATION AND COMPOSITION

Despite the contemporary growth of digital documents and electronic signatures, lots of people and institutions still use traditional handwriting to fill their documents or to sign them. We can find a large variety of pens in the market with different colours, brightness, and stroke widths. This diversity arises from the different chemical components of inks, such as colourants, solvents and resins. There is also the possibility of adding additives to modify the physical properties of the ink. We can divide the current writing tools into two groups¹⁸:

- Ballpoint pens: Their ink solvents are oil-based and the colourants, called dyes, are soluble in them. This type of pen is the most common one.

- Rollerball pens: They are water or gel-based. Their colourants can be dyes (dissolved) or pigments (dispersed). In this group, we can find rollerball pens, which are also very usual.

Colourants are organic or organometallic compounds that contain aromatic rings and resonant structures. These substances have their own Raman signature; this characteristic, with the help of chemometric tools, will allow us to distinguish and classify inks. *Table 1* describes the most common colourants in blue pen inks^{19,20} and the Raman shift of their main Raman bands^{21,22,23}:

Table 1. Common colourants in blue inks, their structure and their most significant Raman bands.

Dye	Other names	Structure	Main Raman Bands (cm ⁻¹)
Basic violet 1	Methyl Violet 2B		205, 442, 724, 809, 916, 1175, 1372, 1533, 1584,
Basic violet 3	Crystal Violet, Methyl Violet 10B		1617 (These dyes combined constitute Methyl Violet) ²¹
Basic blue 26	Victoria Blue B		426, 459, 533, 676, 731, 759, 911, 938, 1035, 1074, 1111, 1171, 1208, 1247, 1293, 1360, 1400, 1434, 1481, 1529, 1586, 1612
Basic blue 7	Victoria Blue BO		422, 655, 681, 710, 728, 761, 804, 863, 917, 1017, 1075, 1160, 1182, 1203, 1286, 1368, 1447, 1486, 1534, 1566, 1584, 1615
Basic Violet 10	Rhodamine B		1362, 1508, 1575, 1598, 1650
Phthalocyanine Blue	-		593, 678, 746, 833, 950, 1034, 1141, 1336, 1466, 1478, 1521, 1586

4. OBJECTIVES

Forgery in handwritten documents is a very common illegal action. Given this circumstance, developing a non-destructive analytical method that allows acquiring fast, reliable and meaningful spatial information about ink distribution in a suspicious part of a document would be a great support for forensic investigations.

For this purpose, the potential of confocal Raman microscopy in combination with chemometric analysis will be investigated. This novel combination has almost no antecedents in terms of counterfeited handwritten document analysis. Therefore, several objectives must be reached to prove this method as valid for real inquiries:

- Assessing Raman spectroscopy discriminating power through an exploratory analysis of the Raman signature of different blue inks.

- Optimizing instrumental parameters of the confocal Raman microscope to obtain hyperspectral images efficiently, with adequate spectral quality and spatial resolution.

- Applying a correct data treatment, supported by an appropriate spectral preprocessing and the accurate application of Multivariate Curve Resolution – Alternating Least Squares Method, to obtain reliable pure spectral fingerprints and distribution maps of the inks in the documents.

- From the obtained distribution maps, being able to determine the correct order of ink application in a crossed line, as well as figuring out the covered text in obliterations with sufficient detail.

5. EXPERIMENTAL

5.1. SAMPLES. INK TYPOLOGIES AND FORGERY SIMULATIONS

The same white office paper has been used for all the measurements (size A4, 80g/cm³ density). Ten models of commercial blue pens from four different brands with diverse ink-based typologies have been used in this study, as shown in *Table 2*:

Brand and model	Brand and model Acronym		Ink-based type
Bic Atlantis	Bic Atlantis BicA		Oil-based
Bic Cristal Clic	BicCr	Ballpoint	Oil-based
Bic Gelocity	BicG	Rollerball	Gel-based
Bic Original	BicO	Ballpoint	Oil-based
Pentel BK77	Pentel BK77 Pent		Oil-based
Pilot BPS-GP PBPS		Ballpoint	Oil-based
Pilot Super Grip PSG		Ballpoint	Oil-based
Pilot V Ball Grip	Pilot V Ball Grip PVBG		Liquid-based
Uniball Eye	UE	Rollerball	Liquid-based
Uniball Signo	US	Rollerball	Gel-based

Table 2. Information concerning the blue pen samples used.

Crossed lines cases between two different inks have been studied intersecting two straight lines. To facilitate interpretations, in all cases we wrote the first stroke vertically from top to bottom and the second stroke was written horizontally from left to right.

Obliteration cases have been studied covering completely with blue ink a letter or a character that was written first with another ink type.

All simulated falsification actions were measured two times by means of independent experiments, applying always the second ink 24 hours after the first one.

In all cases, we have tried to write with normal and similar pressure despite showing no significant influence in other studies²⁴. We also attempted to fit the paper on the microscope

slide as flat as possible to focus correctly all zones of the subsequent image. Despite this, the roughness created on the surface of the paper after writing demanded us to effectuate manually a cautious focusing by zones, thus creating a complete image surface with the depth adapted to the sample surface profile.

5.2. RAMAN MEASUREMENTS. SINGLE POINT SPECTRA AND HYPERSPECTRAL IMAGES

Confocal Raman microscopy is one of the most common techniques to capture Raman hyperspectral images. This method connects the spectral information that Raman spectroscopy provides with the increased spatial resolution confocal microscopy offers using a spatial pinhole, which eliminates out-of-focus signal. To obtain a full image, the motorized sample stage is automatically moved in the x and y directions to cover all the sample surface and a Raman spectrum is acquired at each pixel position, following a point-scanning acquisition mode. All measurements done during this research were carried out using the Renishaw InVia[™] Raman Microscope (displayed in the simplified schema in *Figure 3*) located at Institut de Ciències Fotòniques (ICFO).



Figure 3. Confocal Raman microscope simplified configuration scheme. (Image obtained from R. Smith et al., "Raman spectroscopy: an evolving technique for live cell studies", *Analyst* 2016, 141 (12), 3590–3600).

In order to obtain single point spectra and hyperspectral images maintaining the time analysis as short as possible but showing an acceptable signal-to-noise ratio, several tests were made to adjust the different confocal microscope parameters to achieve optimal measurement conditions, which turned out to be the following:

- A 20x objective lens was used to work and visualize the samples. Higher magnifications (50x) lengthened the sample analysis time focusing too much and lower magnifications (5x) required long measurements to obtain quality spectra. At 20x, the laser beam spot focused on specific zones of the sample provided representative Raman spectra of satisfactory quality in fast measurements.
- The sample electrons were excited using a 532 nm laser. The Raman intensity depends directly on the laser power. A powerful laser releases a higher number of photons which will enhance the Raman response, but it can also damage the sample. The studies concerning the effect of laser power established that a 0.25 mW laser power (1% of total power) was appropriate to execute our measures, as it offered the largest Raman intensity without burning the sample.
- The exposure time is the time the Raman signal is collected at the detector. Larger exposure times increase the intensity and quality of the Raman spectra. Exposure time in single point spectra was fixed at 1 second, whereas each spectrum of a hyperspectral image was measured in 0.1s.
- We have the option of accumulating Raman measurements on the same point of sample to reduce instrumental noise, increasing the measurement time. One accumulation was sufficient for a good signal-to-noise ratio in all Raman spectra.
- Raman wavenumber range covered 158 cm⁻¹ to 1925 cm⁻¹, to enclose the spectral signature of all the studied inks. The Charge-Coupled Device (CCD) detector (Renishaw 1024 CCD Camera), with 1800 grooves per mm grating, led to 1015 spectral channels, giving a spectral resolution of 1,7 cm⁻¹.

In the ink exploratory study, single point spectra were acquired on different points of the same stroke, dried for 1h approximately under the climatic conditions of the laboratory. Three pens of each type were used for these measurements. 21 Raman spectra were recorded per ink class (7 per stroke of the replicate pens), leading to a total dataset of 210 spectra.

In Raman images, pixel size was $(6 \times 6 \mu m^2)$ for crossed lines cases, enough to obtain an exceptional spatial resolution. All crossed lines images present the same dimensions and number of pixels, (798 x 798 μ m²) and 134 x 134 pixels, leading to 17956 spectra per image. Each image was measured in 30 minutes approximately. In images for obliteration studies, pixel size and image dimensions were modified depending on the spatial resolution required to decipher the covered text (7.2.2. section, page 36).

6. DATA TREATMENT. CHEMOMETRIC TOOLS

6.1. GENERAL SPECTRAL PREPROCESSING

Something indispensable before using any chemometric tool is the observation of the raw data to see if there are some external sources of variation unrelated to pure chemical information of the samples, such as experimental noise, problems with baselines or differences in magnitude and range of scale when they are not desired as a source of variability.

To get rid of these unwanted instrumental and external signal contributions, improving the quality of data signal and making easier the extraction of useful information, data preprocessing is essential.

Every dataset needs a particular preprocessing, which is decided according to the features, the nature and the information we want to get from the raw data (*Fig. 4*).



Figure 4. On the left, mean raw Raman single point spectra of each ink type. On the right, raw spectra of UE-PVBG crossed lines, coming from a Raman hyperspectral image.

First of all, as we can see in *Figure 4*, raw single point Raman spectra of our 10 inks presented irregular baselines due to fluorescent contributions of ink components and especially, to the fluorescence emitted by the paper. We observe samples that almost do not present fluorescence contribution (Uniball Eye, UE, and Uniball Signo, US, left plot) and samples that show a very different baseline from the rest, as Pilot Super Grip spectra (PSG). In *Figure 4*, the short spectrum acquisition time, paper fluorescence contribution and irregular baselines avoid clear ink spectra observation in hyperspectral images (right plot). Asymmetric Least Squares (AsLs) algorithm²⁵ was applied to fit with precision irregular and different baselines for a subsequent subtraction (*Fig. 5*). Two parameters had to be adjusted to reach the desired



baseline fitting³ (p (downweight of observations showing positive residuals) = 0.001, λ (fitting smoothness) = 1.10⁵).

Figure 5. On the left, an example of baseline correction with the Asymmetric Least Squares algorithm in Bic Original (BicO) mean raw single-point spectrum. On the right, UE-PVBG crossed lines Raman spectra of the related hyperspectral image after applying Asymmetric Least Squares correction, exposing the presence of cosmic peaks.

In the second place, the presence of cosmic peaks hindered the extraction of information, especially in hyperspectral images, where a large number of consecutive spectra are measured. Cosmic peaks are very intense and narrow peaks with variable bandwidths (as we can see in *Figure 5* hyperspectral image), which are measured when cosmic rays interact with the Raman microscope detector, a very common and random process in Raman measures.

Cosmic peaks were easily eliminated by applying algorithms that emphasised the presence of these unwanted signals and allowed setting threshold values to remove these intense spikes and replacing their values by interpolation of the nearest channels. In the following images, we can observe how baseline corrections and cosmic peaks removal eliminate very intense fluorescence contributions and spikes, respectively (*Fig. 6*).



Figure 6. Mean single point spectra of 10 ink types (left) and UE-PVBG crossed lines spectra (right) after applying Asymmetric Least Squares and Cosmic Peaks correction.

As we can observe in *Figure 6*, the spectral quality of images is lower compared with singlepoint spectra. This was expected since we spent only 0.1 seconds measuring every pixel spectra, while single point spectra were measured for 1 second. Even so, this spectral quality for images is enough to explain the distribution of components across the paper when applying MCR-ALS, which allows recovering high-quality pure ink spectra and clear distribution maps.

6.2. PRINCIPAL COMPONENT ANALYSIS (PCA). INK EXPLORATORY ANALYSIS

Before starting the analysis of forensic forgeries where two inks were involved (diverse crossed lines and obliteration cases) it was very important to visualize and compare the Raman spectral signature of the blue inks that were going to be used. Since many different colourants and combinations of them can be used in the formulation of inks, different spectral shapes can be found among inks. However, there may be situations, especially when comparing inks of the same brand and alike ink-based types, where spectra can be very similar or even identical^{22,23}.

Principal Component Analysis (PCA) is a chemometric method that allows getting information about the structure of the samples and variables that are enclosed in a data table. PCA can find, in a fairly straightforward way, similarities between the studied samples, groups of similar samples and anomalous samples. Moreover, PCA gives us information about the correlation among variables and the most significant variables that define our samples.

Large datasets are increasingly common and they are often difficult to interpret. This has been our case since we had a data table with 210 different single point Raman spectra, each of them showing spectral response in 1015 different wavelengths. The similarity between some of the obtained spectra of different inks and a large number of spectral bands liable to be analysed makes the visual and objective differentiation of inks incredibly difficult. Therefore, the use of chemometrics was essential.

PCA transforms large datasets (Samples x Variables), reducing their dimensionality, increasing their interpretability but at the same time minimizing information loss²⁶. This is achieved because PCA represents the data variance more efficiently with a reduced number of uncorrelated mathematic variables called Principal Components (PC) which include only chemical relevant information. PCs are linear combinations of the original variables and they describe the data variance in decreasing order (PC1 indicates the highest direction of data variance, followed by PC2 and so forth).

We can write the original data matrix in a PCA model (X, Samples x Variables) as the following:

 $X = TP^T + E$

The sample projections in the PCs coordinate system are called Scores (\mathbf{T} , Samples x Principal Components) and the description of the original variables in the space of Principal Components are called Loadings (\mathbf{P} , Variables x Principal Components). \mathbf{E} is a matrix that includes residual data variance.

In this work, PCA allowed observing the distribution of the 210 measured Raman spectra of the ink samples (21 per ink class) in the Scores plots, which helped us to see which inks were possible to be discriminated from the rest based on their Raman signature and which inks formed indistinguishable clusters. Furthermore, the analysis of Loadings plots was helpful to figure out which spectral bands defined and differentiated the inks studied (See PCA analysis results in 7.1 section, page 24).

6.3. MULTIVARIATE CURVE RESOLUTION – ALTERNATING LEAST SQUARES (MCR-ALS). INK IMAGE UNMIXING

As we know, hyperspectral images are defined by spatial pixel coordinates (x and y) and a spectral coordinate (λ), which leads to an image cube of information.

Before working with MCR-ALS, hyperspectral cubes must be unfolded into a twodimensional matrix (**D**), where each row contains the Raman spectrum of a single pixel (*Fig.7*). With this image restructuring, we lose spatial information. Nonetheless, after applying MCR-ALS, the image can be easily reconstructed.



Figure 7. Hyperspectral cube image structure and 2D restructuring data (matrix D) for MCR-ALS.

Image measurements give us a great amount of information about our crossed lines and obliteration experiments. The image data cubes contain the associated Raman spectra for every measured pixel. However, this information would be unusable without the help of chemometrics, owing to the large datasets we have to handle and the presence of multiple unknown Raman responses due to ink mixtures.

To solve the ink mixture problem, MCR-ALS supplies a chemically meaningful bilinear model based on pure contributions of the sample components, starting with a dataset that contains mixtures of these constituents²⁷.

MCR-ALS describes each measured Raman spectrum in matrix **D** as a sum of the pure spectra of every component present in the sample weighted by their related concentration in every pixel (*Fig.8*). To achieve this, the employed analytical method must follow the Beer-Lambert law, where the instrumental response of a pure component is related linearly with its concentration. All spectroscopic methods fulfil this requirement, so MCR-ALS is a suitable chemometric method to use for the analysis of Raman hyperspectral images.



Figure 8. MCR-ALS bilinear model. D can be expressed (a) as the sum of pure spectra of each component weighted by its concentration profile. This bilinear model can be rewritten (b) as a product of matrices of pure concentration profiles (C) and spectra (S^T).

After an image measurement of a document is carried out, only **D** is known since **C** and **S**^T are unknown matrices due to the presence of unknown Raman responses caused by ink mixtures and paper contributions. Through an iterative method, MCR-ALS can decompose the mixed information of matrix **D** into pure concentration and spectra profiles (**C** and **S**^T respectively) of the components that constitute the sample (inks and paper in this case). The steps followed by this chemometric method are described below and are implemented in the MCR-ALS toolbox, which was used in a MatlabR2021a environment²⁸ (*Fig.9*).

Firstly, the number of pure constituents of the sample (1) needs to be selected, which can be chosen manually, if initial information is available. If not, the number of components is chosen based on the application of chemometric methods such as Principal Component Analysis (PCA)²⁶. Then, an initial estimate of the pure spectra of the selected components (2) is done based on the simple-to-use self-modelling analysis (SIMPLISMA)²⁹ method, which selects the most different spectra in matrix **D** as an initial **S**^T matrix.

At this point, the MCR-ALS toolbox offers the possibility of applying several constraints (3) to act in the iterative optimization of **C** and **S**^T. Constraints are mathematical conditions that express natural properties or models that profiles in **C** and **S**^T follow²⁷. Therefore, the right application of constraints reduces solution ambiguities and provides chemically meaningful concentration profiles and pure spectra of each sample constituent.

In this case, considering that pixel concentrations and Raman spectra in an image should never have negative values, the non-negativity constraint was applied to both **C** and **S**^T matrices. This constraint forces all values to be positive, replacing them with zeros or positive values obtained applying algorithms, such as fast non-negative least squares³⁰.

Then, the Alternating Least Squares iterative optimization can start (4). Beginning with **D** and the **S**^T SIMPLISMA²⁸ estimation, we obtain an initial **C**^{*} matrix that is constrained (*Eq. 1*). Then, pure spectra are recalculated (**S**^T*) using the new concentration profile matrix and **D** (*Eq. 2*). At this point, the reproduced matrix **D*** (*Eq. 3*) by the MCR-ALS model is compared with the original data matrix **D**. This process is repeated, alternating the optimization between **C** and **S**^T in each iterative cycle until the difference between **D** and **D*** is small enough, according to our requirements and imposing the selected constraints.



Figure 9. MCR-ALS toolbox steps flowchart. E is an error matrix.

Once the iterative optimization has stopped, MCR-ALS provides some parameters that indicate the quality of the model, such as the percentage of explained variance by the model (% R^2 , *eq. 4*) and the percentage of lack of fit (% LOF). In our case, convergence was achieved when the difference in (% LOF, *eq. 5*) between two consecutive iterations was below 0.1%.

$$\% R^{2} = 100 \left(1 - \frac{\sum_{i,j} e_{ij}^{2}}{\sum_{ij} d_{ij}^{2}} \right) \quad (4) \qquad \qquad \% LOF = 100 \sqrt{\frac{\sum_{i,j} e_{ij}^{2}}{\sum_{ij} d_{ij}^{2}}} \quad (5)$$

Finally, MCR-ALS results are obtained. The S^{T} matrix contains the pure spectra of the image constituents. The distribution maps of these components are obtained refolding the concentration profiles in **C**, recovering the original two-dimensional spatial image structure (*Fig. 10*).



Figure 10. Final pure spectra (ST) and distribution maps (reshaped C) of two constituents coming from a hyperspectral image, through MCR-ALS application.

In the context of this work, MCR-ALS distribution maps will allow observing how pure inks are spatially distributed through the paper and consequently, to figure out which ink has been applied first in crossed lines cases or which text has been covered up in obliterations. The profiles in the **S**^T matrix will provide the spectral fingerprints of the inks used in the document analysed.

7. RESULTS AND DISCUSSION

7.1. INK EXPLORATORY ANALYSIS. PCA RESULTS

Before extracting the information provided by PCA, it is important to make a visual examination of the Raman spectra of the ten ink types analysed to observe ink differences and to see whether the visual inspection matches the results of the exploratory analysis, made without any prior information. With a visual inspection, groups of similar inks can be seen at first sight, which respond to the principal dyes that they contain (*Fig. 11*).



Figure 11. Mean preprocessed Raman spectra of each ink sample, separated into groups by visual inspection. Ink acronyms are as in *Table 2*.

From visual inspection of the preprocessed average Raman spectra of our samples, the ten ink types can be separated into 5 clear different groups, which are the following (all the comparisons with pure dyes have been made from the main Raman bands indicated in *Table 1*, page 12):

- Group 1 (Blue group: Bic Atlantis, Bic Cristal Clic, Bic Original, Pentel BK-77): These four inks present a very similar Raman signature. Going over the pure spectra²¹ and the main Raman bands of Methyl Violet dye, we can clearly see that these inks present the characteristic bands

of this colourant. Therefore, Methyl Violet is the component that provides the highest Raman signal in these four inks.

- Group 2 (Green group: Uniball Eye, Uniball Signo): These identical brand inks have a very different Raman spectrum from the rest, which agrees with the pure spectra of Blue Phthalocyanine dye, which appears to be the sole dye that shows Raman signal in these inks.

- Group 3 (Dark red group: Pilot V Ball Grip, Bic Gelocity): These inks have a similar Raman signature as those from group 1, which indicates the presence of Methyl Violet dye. However, differences in some bands shape and some Raman bands that do not appear in group 1 spectra, especially in the 600-700 cm⁻¹ and 1000-1150cm⁻¹ range, can be observed. According to literature, these bands appear owing to the presence of Victoria Blue BO dye, a very common colourant in Pilot brand inks^{20,22,23}.

- Group 4 (Black, Pilot V Ball Grip) and Group 5 (Ochre, Pilot Super Grip): These inks are similar to Group 3 inks. Nonetheless, some Raman bands differentiate these inks between them and with Group 3. Group 4 (Pilot V Ball Grip) is distinguished principally due to the presence of bands at 200, 400, 1200 and 1400 cm⁻¹ that do not appear in group 3 spectra or that present different spectral shapes. Bands at 200, 400, 600, 780, 1400 and 1620 cm⁻¹, among others, distinguish Group 5 (Pilot Super grip) from groups 3 and 4. These could be due to the presence of other dyes apart from Victoria Blue BO, variations in colourant amounts or changes in the mode of production of inks²³.

Visual inspection of the Raman signature of inks has allowed defining better the types of inks and their major components. Even so, it is necessary a PCA analysis to confirm this first approximation and to improve differentiation between inks with the help of chemometrics.

In an exploratory analysis based on spectral signatures, it is important to give the same relevance to all spectral bands without considering their intensity. Differences in the magnitude of scale were eliminated by applying Mean Center preprocessing, which subtracts to every spectral response the mean value of its variable, leading to a mean value of all variables (Raman shifts) equal to zero.

Moreover, we can observe in *Figure 6* (page 19) that there are notable differences in spectra intensities. By means of a Euclidean normalization, all spectra were adjusted to a comparable scale, which led to a more accurate PCA model to differentiate among Raman spectral shapes.

Figure 12 shows the Scores plot of the first PCA model performed on the matrix of the 210 measured single-point spectra. This PCA model needs five Principal Components and provides a total explained variance of 96.85%, which indicates a good description of data variability. Only PC1 and PC2 were used to form the Scores diagram (these components explain 78.70% of the total explained variance) because the rest of the PCs did not provide extra information for ink differentiation.



Figure 12. Scores plot (PC1 vs PC2) of the PCA containing all ink samples.

As we can see, PC1 reflects the great difference that exists between Raman signatures of Uniball inks (UE, US) and the rest of the studied inks, something already seen by visual inspection. PC2 is employed to explain the variability between the remainder samples. This first PCA supports the visual inspection interpretation, since the formation of 3 clusters (pointed out in *Figure 12*) of inks that are indistinguishable among them agrees exactly with the groups observed by visual inspection.

Once the most different inks have been discriminated, separate local PCA models are going to be done to every cluster separately to see if it is possible to discriminate the most similar inks in the absence of the rest, so that all Principal Components of our PCA model can focus on explaining the differences on Raman signature within the same ink cluster, if they exist.

Figure 13 shows the results for the isolated PCA models, which have been useful to improve the initial visual differentiation, i.e., now the two inks in cluster C are distinguished with the local model, and Pentel ink is distinguished from the rest in the local model of cluster A, and

to confirm that there are two groups of inks that contain identical inks in terms of Raman signature (cluster B and Bic inks in cluster A). *Appendix 1* (page 42) shows the results for the local PCA models and highlights the spectral bands that allow these complex differentiations as of the examination of Loading diagrams and mean spectra overlapping.



Figure 13. Diagram of PCA analysis performed and their results in terms of ink differentiation.

It is important to notice that both indistinguishable clusters are formed by inks of the same brand (Uniball and Bic). This may point out that these inks have the same chemical composition and that they only change in terms of line width, ink additives or simply in the external appearance of the pen. Besides, it should be noted that is not correct to extrapolate these results to other investigations, since other researchers have reported differences in Raman signatures between batches of the same brand and model inks manufactured in different times²³, which could lead to different results.

7.2. RAMAN IMAGES UNMIXING. MCR-ALS RESULTS

With the use of an exploratory chemometric method such as Principal Component Analysis (PCA), we have been able to see how different our inks are and the most important spectral bands that lead to that differentiation in case it is possible to distinguish our inks.

For the image studies, crossed lines have been studied between physically similar inks, which were impossible or unlikely to be differentiated by visual inspection with a microscope, but that were distinguishable through their Raman signature according to exploratory analysis results. In the obliteration studies, Bic Original, one of the most used oil-based ballpoint pens, has been obliterated with the three varieties of inks studied in this work: oil, gel and liquid inks.

7.2.1. Crossed lines cases

Liquid-based inks crossings

In the first place, an exhaustive description of the MCR-ALS results of the first crossed lines experiment will be done, paying attention to all details in order to analyse the remainder images straightforwardly. The results correspond to the crossing of two liquid-based type inks, from Uniball Eye and Pilot V Ball Grip pens. This is a case of inks of clearly different composition (in clusters B and C, respectively, in Figure 12). In all cases, since crossed lines involve two inks, both possible orders of ink application have been prepared and measured.

Table 3. Liquid-Liquid crossing systems and MCR-ALS quality parameters table. (1) and (2) indicate the order of application of inks, **UE** and **PVBG** designate each ink colour in the following RGB images (see next page).

Crossing number	Ink system	Number of components	% Explained variance (R ²)	% Lack of fit
1	UE (1) – PVBG (2)	4	86.3 %	37.0 %
2	PVBG (1) - UE (2)	3	85.4 %	38.3 %

As observed in *Table 3*, crossing number 1 has an additional component that was not expected, since crossing lines systems are formed of two inks plus paper. Nevertheless, additional components have been needed often, owing to the presence of unwanted instrumental noise, which has not been possible to be fully eliminated with spectral preprocessing and that impedes a good description of inks and paper distributions with only the three supposed components. In this case, as in the rest, display of instrumental noise components will be avoided, since these unwanted signals are not chemically meaningful.

Furthermore, MCR-ALS quality parameters shown in *Table 3* can be shocking at first sight, since a rather low explained variance and especially, a very high lack of fit, is obtained. However, these were the expected results in all measurements, since the signal-to-noise ratio of the collected spectra is low, as we could have observed when preprocessing (see *Figure 6*, right plot). Therefore, the unexplained variance and high lack of fit come from unexplained instrumental noise. Even so, the distribution of the components is perfectly described and the recovered pure spectra display a similar quality compared to those on exploratory analysis in all cases (See *Figure 11*, page 24).



Figure 14. UE-PVBG (crossing 1) results. Top plots, MCR-ALS distribution maps of components (paper and the two inks). Bottom left, RGB image made of overlaid top maps; bottom right, MCR-ALS resolved spectra of inks and paper.

Crossing 1 (*Fig.* 14) exhibits very satisfactory results. MCR-ALS recovered spectra show the most characteristic bands of the two inks used and the expected spectral differences between them. Inspecting isolated distribution maps, the following conclusions are obtained:

- As expected, paper is present primarily on the corner zones of the images, where ink is absent. Also, it presents significant concentrations at some pixels in the ink zone, due to some free gaps that both pen mechanisms leave.

- Uniball Eye (UE) stroke shows a clear discontinuity in the crossing zone, indicating that this ink has been covered by the PVBG one.

- Pilot V Ball Grip (PVBG) stroke presents a similar concentration throughout the horizontal stroke, which indicates that it predominates at the surface of the crossing, as was expected.

The RGB image puts together all the maps information and clearly shows that the horizontal stroke (which corresponds to PVBG) is above the vertical UE stroke, as expected. Additionally, a few darker pixels in the crossing zone indicate the presence of some light ink mixing.



Figure 15. PVBG-UE (crossing 2) results. Top plots, MCR-ALS distribution maps of components. Bottom left, RGB image of overlaid top maps; bottom right, MCR-ALS resolved spectra of inks and paper.

In Crossing 2 (*Fig. 15*), MCR-ALS recovered spectra are nearly the same in both reversed crossings (in the following crossed lines cases only spectra of one of the opposite crossings will be shown since they are very similar between them in all examples).

Crossing 2 leaves further doubts about ink order application. In the isolated distribution maps, both PVBG and UE inks decrease their concentration at the crossing zone, especially the Uniball brand ink. This could indicate that UE ink has gone through PVBG stroke and has penetrated deeper on the paper, shading doubts in the expected stroke order. RGB image reinforces this observation because inks appear highly mixed in the crossing and it seems that PVBG ink (red stroke) has almost been written over UE ink (green stroke).

This particular Uniball Eye ink behaviour has been verified repeating the same experiment and corroborated with literature. Braz et.al pointed out the same outcome when UE was applied over pilot G2, another liquid-based ink¹⁵.

Oil-based inks crossings

Crossing number	Ink system	Number of components	% Explained variance (R ²)	% Lack of fit
3	BicCr (1) – PSG (2)	5	92.2 %	28.0 %
4	PSG (1) - BicCr (2)	4	89.3 %	32.7 %

Table 4. Oil-Oil crossings systems and MCR-ALS quality parameters table.



Figure 16. BicCr-PSG (crossing 3) and PSG-BicCr (crossing 4) RGB images and MCR-ALS resolved spectra of components.

Oil-Oil crossings have generated very interesting results (*Table 4*, *Fig.16*). Once again, the MCR-ALS spectra of both inks are very similar to their pure signature collected in the exploratory analysis.

Oil-based inks are distributed on the paper much more discontinuously than liquid inks (as we have seen in crossings 1 and 2) or gel inks (as we will see in the next cases). This pattern is typical of ballpoint mechanism pens, and this phenomenon has been detected under the microscope in all six oil-based inks both separately and being part of a crossing.

Nevertheless, it is quite clear that the horizontal stroke, which was written in second place, is located on top of the crossing in both reversed cases. We can confirm these results further by looking at distribution maps of single components, where we can figure out concentration decreases significantly in vertical strokes at the crossing zone (first inks applied). (Distribution maps of single components of these crossings and the followings are in *Appendix 2*, page 44)

Gel-based inks crossings

Table 5. G	sel-Gel	crossings	systems	and M	CR-ALS	quality	parameters	table.

Crossing number	Ink system	Number of components	% Explained variance (R ²)	% Lack of fit
5	US (1) – BicG (2)	3	86.1 %	37.3 %
6	BicG (1) - US (2)	3	86.3 %	37.0 %



Figure 17. US-BicG (crossing 5) and BicG-US (crossing 6) RGB images and MCR-ALS resolved spectra of components.

Gel inks are physically similar to liquid inks. This is also observed when crossed lines experiments are done (*Fig. 17*). We can see the presence of ink mixing at the crossing zones (dark shades) and nearly uniform ink distributions through the paper (unlike oil-based inks, no large discontinuities are observed).

Crossing 5 exhibits the awaited ink behaviour, with Bic Gelocity ink (horizontal stroke) clearly on top of Uniball Signo stroke. Some minor zones of the crossing show greenish colours due to ink mixing and minor discontinuities of Bic Gelocity ink, represented in red.

However, crossing 6 suggests us the inverse order of ink application, since Bic Gelocity Raman signature predominates over Uniball Signo one, with highly mixed ink zones. Again, as in Crossing 2 (page 31), when a Uniball brand ink is applied on top of another, it is highly absorbed by the paper and the erroneous order of application is noticed looking at MCR-ALS images (RGB image on the previous page and isolated distribution maps in *Appendix 2*).

Gel-Liquid crossings

This final case has been more complex in terms of spectra differentiation for MCR-ALS since the inks used (liquid-based PVBG and gel-based BicG) have very similar Raman signatures (*Appendix 1* shows that an isolated PCA was necessary to differentiate these inks).

The application of a 5x5 spatial binning has allowed increasing the signal-to-noise ratio in the image spectra by means of spectral pixel combination, reducing the spatial resolution of images at first (*Fig. 18*). This spatial preprocessing increases spectral quality and has permitted MCR-ALS to distinguish these inks.

At last, we have been able to recover the initial spatial resolution (134 x 134) pixels, considering that we can calculate a full **C** matrix from the original **D** matrix and the pure spectra matrix obtained from MCR-ALS analysis of the binned image S^{T}_{binned} (**C** = **D** S^{T+}_{binned}). Also, we can observe in *Table 6* (page 34) that MCR-ALS quality parameters are much better than in previous cases, due to the higher signal-to-noise ratio of binned Raman spectra.



Figure 18. Example of a 3x3 binning. An 81 pixels image turns into a 9 combined pixels image.

Binning was not applied in the previous cases because the preservation of original spatial information over quality parameters improvement was preferred in the MCR-ALS analysis.

Table 6. Gel-Liquid crossings systems and MCR-ALS quality parameters table.

Crossing number	Ink system	Number of components	% Explained variance (R ²)	% Lack of fit
7	BicG (1) – PVBG (2)	3	99.4 %	7.8 %
8	PVBG (1) - BicG (2)	3	99.4 %	7.6 %



Figure 19. BicG-PVBG (crossing 7) and PVBG-BicG (crossing 8) RGB images and MCR-ALS resolved spectra of components.

Even though these inks were very similar in terms of Raman signature, the same spectral bands at 920, 1075 and 1230 cm⁻¹ approximately, which allowed differentiation in Principal Component Analysis (Cluster C PCA, *Appendix 1*) are those that enable differentiation in MCR-ALS, manifesting again that both chemometric methods are capable to deal with very similar spectral signatures.

In terms of crossings, the combination of liquid and gel inks does not change their behaviour and the same interactions when they were not merged are detected, such as continuity throughout the paper and ink mixing. The red colour of PVBG Raman shift signature predominates at the crossing zone in case 7, while Bic Gelocity green colour prevails in crossing number 8 (*Fig.19*).

In this case, it is possible that an external observer could not be so sure about which Raman signature predominates in these crossings. Distribution maps in *Appendix 2* clarify the situation further since the first written strokes show an obvious concentration decrease once again.

7.2.2. Obliteration cases

Table 7 sums up the cases in which Bic Original oil-based ink has been obliterated. Quality parameters and resolved spectra acquired are almost the same as those obtained in crossed lines examples for the same kind of inks. For this reason, MCR-ALS spectra will not be shown.

Obliteration number	Obliterating ink	Number of components	% Explained variance (R ²)	% Lack of fit
1	PBPS (oil)	5 (3x3 binning)	98.5259 %	12.1414 %
2	PVBG (liquid)	6	87.0882 %	35.933 %
3 (general image)	RicC (gel)	5	87.5902 %	35.2276 %
3 (resolution rising)	Dico (gei)	5 (3x3 binning)	98.0646 %	13.9118 %

Table 7. Obliteration systems and MCR-ALS quality parameters table. As in crossed lines cases, binning application improves R² and LOF substantially.

As we can see in *Figure 20* (page 36), the deciphering of obliterated characters has been really satisfactory, regardless of the ink-based type used to cross out the initial information.

In the RGB images shown, the obliterated text is shown in red, while the obliterating ink and the paper have green and blue colours, respectively. Distribution maps of single components of obliteration cases are located in *Appendix 2*.



Figure 20. Bic Original obliteration cases results. Image dimensions and pixel area are indicated below each sample.

- Images of obliterations 1 and 2 have been measured in less than 20 minutes and the original information (M character and benzene molecule, respectively) has been recovered in detail as a result of a high spatial resolution (60 x 60 μ m² and 50 x 50 μ m² of pixel size, respectively) in the images acquired.

- Obliteration 3 shows how to proceed in real cases, where it may not even be known whether any text has been obliterated. This method is conceivable thanks to the versatility and discriminating power that Raman imaging offers in just a short period.

A fast first image measurement (A) permits obtaining a preliminary view in order to see whether there is obliterated text and, if so, where it is located and how complex the covered text is. In this case, 2 minutes of measurement (200 x 200 μ m² pixel size) were sufficient to detect the presence of a text and an approximation of the content. After locating where the covered ink is, a second exhaustive measure (B) of 30 minutes in a reduced area is performed so as to improve resolution (30 x 30 μ m² pixel size) and leave no doubt about the covered text.

8. CONCLUSIONS

After this investigation, several promising outcomes and future avenues can be drawn, in order to enhance and carry on the development of a non-destructive method that can be finally used in documentary forensic investigations reliably and rapidly.

1. Raman spectroscopy shows an excellent ink interbrand differentiating power since all differently branded inks have been distinguished through a Principal Component Analysis of their Raman signatures. Nevertheless, some inks of the same brand present equal Raman signatures, which makes it impossible to study document forgery cases among them.

2. Instrumental parameters optimization has been successful. We have been able to obtain fast hyperspectral images (from 2 to 30 minutes in all cases) with the highest reported spatial resolution in crossed lines and obliteration cases.

3. MCR – ALS method (in conjunction with adequate spectral preprocessing tools) is a great chemometric tool to obtain spatial information about ink distribution through the paper, recovering also good spectral Raman signatures, despite paper fluorescent contribution and instrumental noise, with no prior knowledge about the number of inks used, their chemical identity and their location in the document. Quality parameters can be improved by applying powerful preprocessing, such as binning, or making a larger use of constraints.

4. The correct sequence of ink application in crossed lines has been clearly determined in almost all the studied cases, thanks to the high resolution of Raman images and the study of the ink distribution maps. Due to an unusual absorption phenomenon observed in Uniball inks, two crossings can be wrongly interpreted.

5. Obliterated texts written with Bic Original have been recovered, regardless of what type of ink covered the text. We have verified that Raman imaging combined with MCR-ALS can carry out extremely rapid measurements to get initial approximations, which can be further improved with more accurate measurements that can recover complex or small characters in great detail.

Future research could focus on the combination of different imaging techniques to reinforce the reliability of the method. The nature of inks indicates that the incorporation of infrared and fluorescence imaging can further improve the discriminating power and the spatial resolution, providing a better understanding of the physical and chemical processes that ink experiences across the paper and the interactions among them.

9. REFERENCES AND NOTES

- L. R. e Brito, A. R. Martins, A. Braz et al., "Critical review and trends in forensic investigations of crossing ink lines", *Trends Anal. Chem.* 2017, 94, 54-69.
- 2. L. Godown, "Sequence of writings", J. Crim. Law Criminol. 1963, 54, 101-109.
- 3. J. H. Bradley, "Sequence of pencil strokes", J. Crim. Law Criminol. 1963, 54, 232-234.
- S.J. Strach, "Establishing the sequence of intersecting ball-point pen strokes", *Forensic. Sci* 1978, 11, 67-64.
- M. Oron, V. Tamir, "Development of SEM methods for solving forensic problems encountered in handwritten and printed documents", *Int. Crim. Police Rev.* **1979**, 324, 24-30.
- P. A. Waeschle, "Examination of line crossings by scanning electron microscopy", J. Forensic Sci. 1979, 24, 569-578.
- 7. R.D. Koons, "Sequencing of intersecting lines by combined lifting process and Scanning Electron Microscopy", *Forensic Sci. Int.* **1985**, 27, 261-276.
- 8. R. W. Radley, "Determination of sequence of writing impressions and ball pen ink strokes using the ESDA technique", *J.Forensic Sci. Soc.* **1993**, 13, 51-54.
- 9. K.-C. Cheng, C.-H. Chao, B.-S. Jeng, S.-T. Lee, "A new method of identifying writing sequence with the laser scanning confocal microscope", *J. Forensic Sci.* **1998**, 43, 348-352.
- 10. A. A. Cantu, "Detecting Fraudulent Documents", Forensic Doc. Exam. 2005, 113-131, 2007.
- 11. M. Claybourn and M. Ansell, "Using Raman spectroscopy to solve crime: Inks, questioned documents and fraud", *Sci. Justice J. Forensic Sci. Soc.* **2000**, 40 (4), 261-271.
- C. S. Silva, M. F. Pimentel, R. S. Honorato, C. Pasquini, J. M. P. Montalbán, and A. Ferrer, "Near infrared hyperspectral imaging for forensic analysis of document forgery", *Analyst* 2014, 139 (20), 5176-5184.
- G. Payne, C. Wallace, B. Reedy, C. Lennard, R. Schuler, D. Exline, and C. Roux, "Visible and near-infrared chemical imaging methods for the analysis of selected forensic samples", *Talanta* 2005, 67 (2), 334-344.
- K. Bojko, C. Roux, and B. J. Reedy, "An examination of the sequence of intersecting lines using attenuated total reflectance-Fourier transform infrared spectral imaging", *J. Forensic Sci.* 2008, 53 (6), 1458-1467.
- A. Braz, M. López-López, and C. García-Ruiz, "Raman imaging for determining the sequence of blue pen ink crossings", *Forensic Sci. Int.* 2015, 249, 92-100.
- F. D. S. L. Borba, T. Jawhari, R. Saldanha Honorato, and A. De Juan, "Confocal Raman imaging and chemometrics applied to solve forensic document examination involving crossed lines and obliteration cases by a depth profiling study", *Analyst* **2017**, 142 (7), 1106-1118.
- 17. E. Smith, G.Dent, "Modern Raman Spectroscopy: A Practical Approach", *John Wiley & Sons* **2000**, 3-5.
- M. Ezcurra, J. M. G. Góngora, I. Maguregui, and R. Alonso, "Analytical methods for dating modern writing instrument inks on paper", *Forensic Sci. Int.* 2010, 197 (1–3), 1-20.
- J. A. Denman, W. M. Skinner, K. P. Kirkbride, and I. M. Kempson, "Organic and inorganic discrimination of ballpoint pen inks by ToF-SIMS and multivariate statistics", *Appl. Surf. Sci.* 2010, 256 (7), 2155-2163.
- J. Coumbaros, K. P. Kirkbride, G. Klass, and W. Skinner, "Application of time of flight secondary ion mass spectrometry to the in situ analysis of ballpoint pen inks on paper", *Forensic Sci. Int.* 2009, 193 (1-3), 42-46.
- I. Geiman, M. Leona, and J. R. Lombardi, "Application of Raman Spectroscopy and Surface-Enhanced Raman Scattering to the Analysis of Synthetic Dyes Found in Ballpoint Pen Inks," *J. Forensic Sci.* 2009, 54 (4), 947–952.

- F. de Souza Lins Borba, R. Saldanha Honorato, and A. de Juan, "Use of Raman spectroscopy and chemometrics to distinguish blue ballpoint pen inks", *Forensic Sci. Int.* 2015, 249, 73–82.
- A. Braz, M. López-López, and C. García-Ruiz, "Studying the variability in the Raman signature of writing pen inks", *Forensic Sci. Int.* 2014, 245, 38–44.
- N. Ozbek, A. Braz, M. López-López, and C. García-Ruiz, "A study to visualize and determine the sequencing of intersecting ink lines", *Forensic Sci. Int.* 2014, 234 (1), 39-44.
- 25. P. H. C. Eilers, "Parametric Time Warping", Anal. Chem. 2004, 76 (2), 404–411.
- I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments", Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 2016, 374 (2065).
- 27 A. De Juan, J. Jaumot, and R. Tauler, "Multivariate Curve Resolution (MCR). Solving the mixture analysis problem", *Anal. Methods* 2014, 6 (14).
- J. Jaumot, A. de Juan, and R. Tauler, "MCR-ALS GUI 2.0: New features and applications", Chemom. Intell. Lab. Syst. 2015, 140, 1–12.
- W. Windig and J. Guilment, "Interactive Self-Modelling Mixture Analysis", Anal. Chem. 1991, 63 (14), 1425–1432.
- R. Bro and S. de Jong, "A Fast Non-negative Constrained Linear Least Squares Algorithm", J. Chemom. 1997, 11, 393–401.

10. ACRONYMS

MCR-ALS: Multivariate Curve Resolution - Alternating Least Squares SEM: Scanning Electron Microscopy ESDA: Electrostatic Deposition Analysis LSCM: Laser Scanning Confocal Microscopy TLC: Thin Layer Chromatography HPLC: High Performance Liquid Chromatography GC: Gas Chromatography ATR-FTIR: Attenuated Total Reflectance-Fourier Transform Infrared Spectroscopy HSI-NIR: Hyperspectral Images in the Near Infrared Range PCA: Principal Component Analysis PLS-DA: Partial Least Squares-Discriminant Analysis CCD: Charge-Coupled Device AsLs: Asymmetric Least Squares SVD: Singular Value Decomposition SIMPLISMA: Simple-To-Use Self-Modelling Analysis LOF: Lack of Fit PC: Principal Component RGB: Red Green Blue

APPENDICES

APPENDIX 1: LOCAL PCA MODELS

This appendix shows Scores diagrams, Loadings diagrams of local PCA models on spectra from clusters A, B and C in *Figure 12*. The outcomes of these local PCA models are summarized in *Figure 13*. Large loading values highlight the Raman bands that differentiate the analysed inks. In all models, superimposed preprocessed spectra of the inks analysed by local PCA models are provided.

• PCA model 4, Cluster A (Pentel, Bic Original, Bic Cristal Clic, Bic Atlantis) (4PC, Explained Variance = 74.39%, **Pentel differentiation**)



PCA model 5, Cluster A (4PC, Explained Variance = 59.76%, No differentiation)







APPENDIX 2: DISTRIBUTION MAPS

Individual distribution maps of MCR-ALS analysis of crossed lines (section 7.2.1) and obliterations (section 7.2.2) are provided.



BicG

Paper



Crossing 8: PVBG (1) - BicG (2)



Obliteration 1: PBPS obliterates BicO (letter M)



Obliteration 2: PVBG obliterates BicO (benzene)



45

1000



- Original image (200 µm)



- Resolution rising (30 µm)

