

Encoding of Luminescent Ink Markers Using Low-Level Data Fusion and Chemometrics

Caroline R. Carneiro,^{1a} Isabela B. Serwy,^a Carolina S. Silva^{1b,c} and Ingrid T. Weber^{1*,a}

^aLaboratório de Inorgânica e Materiais (LIMA), Instituto de Química, Universidade de Brasília (UNB), P.O. Box 04478, 70904-970 Brasília-DF, Brazil

^bDepartment of Food Sciences and Nutrition, University of Malta, Msida MSD 2080, Malta

^cCentre for Biomedical Cybernetic, University of Malta, Msida MSD 2080, Malta

The identification and analysis of documentary fraud is always a challenge for forensic science. Document analysis has proven to be an important branch of forensics in elucidating the authenticity of documents. The development and incorporation of luminescent inks in authentic documents have proved to be an excellent security feature. This paper purposes the use of a possible luminescent ink marker for anti-counterfeiting applications, aiming to create a document encoding process that is simple, robust, sensitive, and non-destructive. Since luminescent inks markers provide a visual, chemical, and spectral signature, and can be easily detected by using a UV lamp, the aid of unsupervised chemometric tools makes it possible to differentiate the luminescent markers inserted in the ink. Unsupervised models of principal component analysis (PCA) and K-mean were successful in correctly associating marked inks with their respective pure markers, while a supervised classification model based on partial least squares discriminant analysis (PLS-DA) correctly classified all samples from the prediction set and the blind test samples. For comparison, a soft independent modeling of class analogy (SIMCA) model was also built, which despite showing a misclassified sample it is also a strong candidate for future applications.

Keywords: metal-organic framework, luminescent markers, lanthanides, ink, chemometrics

Introduction

In the last few decades, the use of scientific knowledge not only for supporting but often as a decisive tool in the elucidation of crimes has intensified within forensic science, with forensic chemistry having an important role. Forensic chemistry is the field of forensic science focused on the analysis of evidence of judicial interest through the application of chemical science knowledge to criminal problems.¹ Within forensic chemistry, one of the areas of great importance today is document analysis, a part of criminalistics that studies the authenticity of a document and investigates falsification, to determine the perpetrator and the means of forgery.^{2,3}

Criminal examination of documents involves a wide variety of methods, from visual inspection to the use of expensive and sophisticated equipment, which must preferably be non-destructive, simple, fast,

reliable, inexpensive and not require sample preparation.³ Non-destructive techniques are performed directly on the document surface; thus, document integrity is preserved.⁴ The preservation of documentary evidence is critical to maintaining sample integrity throughout the investigative process. For non-destructive methods, which usually do not need sample preparation, the study of variations in document composition has been combined with chemometric models to assist in the interpretation of the results obtained.³

In general, spectroscopic analysis, such as UV-Vis, infrared, Raman, X-ray fluorescence (XRF), and easy ambient sonic-spray ionization mass (EASI-MS) meet these requirements.⁵⁻⁹ These analytical techniques are usually employed in investigations where a questioned document must be identified as authentic or not. These techniques can also be particularly useful, however, for preventive approaches when security elements such as security inks are used to ensure document authenticity or to encode documents by means of security seals. Security elements are a broad class of structures (e.g., drawings, text,

*e-mail: itweber@unb.br

Editor handled this article: Ivo M. Raimundo Jr. (Associate)

inks, fibers, and holographs) incorporated in documents to provide greater guarantee of authenticity just by simply identifying its presence, hampering possible forgeries.^{2,10} Among the safety features, safety inks and pigments stand out for having many different functions that can be easily checked.^{1,11,12} Due to their versatility, fluorescent inks are a widely used and studied class.¹³⁻¹⁷ Although security features are incorporated into some documents, these remedies do not exhaust all possible kinds of forgery, making solution of these cases a difficult task. This area, then, has become a promising field of study, with a growing demand for innovations, where fluorescent inks has gained prominence in recent years.¹³⁻¹⁷

Luminescent materials can provide color encoded customized patterns for identification or visualization under appropriate external activations and a gamut of full color-encoded patterns can be produced through simultaneous use of several luminescent inks with different colors.^{18,19} Li and Hu¹⁷ have described an easy and robust strategy for manufacturing photoluminescent nanofilms for application as safety inks, successfully applied to banknote marking. High luminescent films were prepared by low-cost synthesis by incorporating fluorescent quantum dots into the cellulose nanofibers. da Luz *et al.*¹⁴ developed an easy, fast, and cost-effective strategy for synthesizing photoluminescent metal-organic networks with terbium, europium, gadolinium, and neodymium ions as markers in jet printer inks. The materials were printed on flexible substrates with a conventional inkjet printer and observed under UV light irradiation. Liang *et al.*¹⁶ synthesized luminescent inks containing LuVO_4 :Eu nanoparticles with polyacrylic acid as the surfactant. The aqueous solution of the nanoparticles was used as a marker and was added to pen inks and printer cartridges.

The contribution of the present article is not only to explore the idea of producing ink markers using luminescent metal-organic frameworks (MOF), but also of changing ink chemical composition to enable ink encoding through its spectral signature. Metal-organic frameworks modified with lanthanide ions (L-MOFs) have great potential as markers, since they are not naturally occurring. The L-MOFs have many advantages such as high chemical and thermal stability,²⁰ can be easily detected by the visualization, with a UV lamp due to the high photoluminescence²¹ and provide a visual, chemical, and spectral signature. Moreover, compared to standard fluorescent commercially used dyes, the lanthanide-based luminescent inks exhibit several advantages, such as, sharper transitions without photobleaching and both, down and up-conversion processes, in a single host lattice which makes them intrinsically more difficult to be frauded.²² Then, the use of lanthanide-based materials appears as an

interesting alternative in situations where a more complex encoding system is intended. For example, when only the emission color is not considered sufficient, the use of MOFs containing lanthanides offers a spectral signature (composed of small variations in the spectra and excitation and emission, modulated by both the ion and the ligand) as well as a chemical signature.

Despite the great advantage of using colors as an encoding system they are colorimetrically inaccurate or involve utilization of complex half-toning algorithms to decipher, due to immense spectral cross-talking between emission bands and substantial background interference.¹⁸ Eu- and Tb-based inks are, of course, easily distinguishable by their emitting colors. The use of chemometric tools become interesting when we distinguish two red emitting inks (or green emitting) by combining excitation and emission spectra. The chemometric tools used in this work are quite easy to handle and offer the possibility of fine tuning the encoding system, using a relatively large combination of inks/markers (in this case, 4). With that in mind, the aid of chemometric tools can be an excellent tool for encoding unequivocally, it is possible to classify different MOFs inserted in an ink, thus creating a document coding process that can be verified by spectroscopy, thus a simple, robust, sensitive, and non-destructive method. In a previous study, Carneiro *et al.*²⁰ demonstrated the possibility of using a system of encoding markers with MOFs, making use of different organic ligands (with europium as emitting center); and employing chemometric tools, for the use of marking of ammunition. When this idea is applied to document analysis, the difference in composition as well as differences in optical behavior make marking tasks possible, which may lead to the development of an ink encoding system.

Unsupervised learning models such as principal component analysis (PCA) and K-means can be used as simple alternative to identify similarity between samples. Although unsupervised models do not classify samples, they are simple to use and widely accepted by forensic experts. While PCA is a popular multivariate statistical model for reducing dimensionality, K-means is a simple clustering technique, providing robust association of similar samples by means of simple distance measurement. PCA performs linear combinations of the original variables so that the derived variables capture the maximal variance. After the decomposition proposed by the PCA, the samples become points in a dimension-reduced space defined by the principal components (PCs),²² and can be subjected to a clustering model to identify meaningful groups of similar samples. Those two models can be used to identify the presence of the synthesized marker and authenticate a questioned document.

It is worth clarifying that there is a conceptual difference with respect to document authentication from the forensic and the chemometric point of view. In forensics, authentication relates to the use of security elements incorporated in documents that make it possible to distinguish the original document from its copies and therefore identify counterfeits. The term authentication in the context of chemometrics, on the other hand, is related to a classification process of determining whether the identity of an object is, in fact, what it has been declared to be. This problem is often solved using supervised classification models, often requiring the use of class-modeling approaches.²³

In the present study, as a proof of concept, we proposed a synthesis and application of non-commercial luminescent MOFs containing terbium and europium ions, as possible luminescent inks markers for anti-counterfeiting applications aiming to create a document encoding process that is simple, robust, sensitive, and non-destructive. We initially aimed to demonstrate the feasibility of a low-cost synthesis of a possible marker for luminescent inks that could be used as a security device for printed documents; then, we proposed to attest the feasibility of emission/excitation spectra and chemometrics to discriminate between the different markers in the encoding system. To do this, four different MOF-based luminescent markers were produced by microwave-assisted hydrothermal synthesis and inserted in common printer inks and painted on vegetable paper. The samples were analyzed by X-ray diffraction and fluorescence emission and excitation spectroscopy, and the results analyzed with chemometric models. Blind tests were performed, and the results analyzed with chemometric models. With the results of this project, we intended to develop luminescent optical markers showing the efficiency of use as anti-fraud technology and to demonstrate an application to enable the accurate, nondestructive classification of luminescence MOFs as printer ink markers, using excitation and emission spectra.

Experimental

Samples preparation

A set of four MOFs with potential for use in ink markers was synthesized by a microwave-assisted hydrothermal method, $[\text{Ln}(\text{BTC})]_n$, and $[\text{Ln}_2(\text{BDC})_3(\text{H}_2\text{O})_2]_n$ (wherein Ln = Eu^{3+} and Tb^{3+} , BTC = trimesic acid and BDC = terephthalic acid). In the rest of this study, these will be referred to as Eu(BTC), Tb(BTC), Eu(BDC) and Tb(BDC), respectively.

For the syntheses, the oxides Tb_4O_7 (99.9%, Sigma-Aldrich, St. Louis, USA) and Eu_2O_3 (99.5%, Sigma-Aldrich, St. Louis, USA) were used to prepare the respective nitrates. The ligand BTC (98%, Sigma-Aldrich, St. Louis, USA) was used without any treatment, and the ligand BDC (98%, Alfa Aesar, Tewksbury, USA) were used to prepare the salt Na_2BDC , as described by Wanderley *et al.*²⁴ These reagents were used as received without further purification.

All the samples were hydrothermally prepared, using a microwave reactor and autogenous pressure (Monowave 300 Anton Paar, São Paulo, Brazil) with power of 400 W. For the $[\text{Eu}(\text{BTC})]_n$ marker, Eu_2O_3 (0.175 mmol), trimesic acid (H_3BTC) (0.35 mmol), and 12 mL of distilled water were mixed and placed in a 30 mL quartz reactor, under agitation. And for the $[\text{Tb}(\text{BTC})]_n$, $\text{Tb}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ (0.175 mmol) and trimesic acid (H_3BTC) (0.175 mmol) were mixed in 12 mL of distilled water and placed in a 30 mL quartz reactor, under agitation. Both reactions were performed at 150 °C for 20 min. After each reaction, the powder obtained was washed with distilled water and acetone and dried at 100 °C for 24 h.

For the $[\text{Eu}_2(\text{BDC})_3(\text{H}_2\text{O})_2]_n$, $\text{Eu}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ (0.7 mmol) and terephthalic acid disodium salt (Na_2BDC) (0.7 mmol) were solubilized in 12 mL of distilled water and placed in a 30 mL quartz reactor, under agitation. And finally, for the marker $[\text{Tb}_2(\text{BDC})_3(\text{H}_2\text{O})_2]_n$, $\text{Tb}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ (0.7 mmol), terephthalic acid disodium salt (Na_2BDC) (0.7 mmol), and 12 mL of distilled water were mixed and placed in a 30 mL quartz reactor, under agitation. The last two reactions were performed at 160 °C for 20 min. The salt of the ligand BDC (Na_2BDC) was synthesized based on a previous method as described by Wanderley *et al.*²⁴ After each reaction, the powder obtained was washed with distilled water and acetone and dried at 100 °C for 24 h.

Each synthesis was performed five times, resulting in 20 samples: five Eu(BTC), five Tb(BTC), five $[\text{Eu}_2(\text{BDC})_3(\text{H}_2\text{O})_2]_n$ and five $[\text{Tb}_2(\text{BDC})_3(\text{H}_2\text{O})_2]_n$. The samples were characterized by X-ray diffraction (Rigaku/MiniFlex 300, Austin, USA, with the following parameters: 2θ range from 5 to 50° by 0.01 step and 1.5° min⁻¹ speed) and photoluminescence spectroscopy (spectrofluorometer Fluorolog Horiba Jobin Yvon FL3-22, Kyoto, Japan, with excitation from a xenon lamp).

Excitation and emission spectra of all syntheses of the four markers were acquired in triplicate, generating a total of 60 emission and excitation spectra. The excitation spectra, for markers of europium, were obtained by monitoring the emission at 613 nm, scanning with a range between 200-400 nm. The emission spectra were acquired exciting at 293 nm for the BTC ligand and at 323 nm for de BDC ligand, ranging from 400 to 750 nm. On the other hand,

for the markers containing the ion terbium, the excitation spectra were obtained by monitoring the emission at 545 nm with range of 200-400 nm, and the emission spectra were acquired by exciting at 293 nm for the BTC ligand and 323 nm for de BDC ligand and analyzing in the range of 400-750 nm. No sample preparation was needed before the spectra acquisition.

Testing as ink marker

To evaluate the performance of these MOFs as luminescent markers for inks, each marker synthesized was individually added to blue and black common printer inks at a ratio of 10 mmol L⁻¹, in order to produce 40 marked samples, 20 containing marked blue inks and 20 containing marked black inks. The markers were added directly to the inks and were maintained under magnetic stirring, for 15 min at room temperature. Then, for each ink sample, traces were made on sheets of vegetable paper (Canson® Vegetal 90, 95 g m⁻²). When dry, the samples were visualized under UV light ($\lambda = 254$ nm), to have a visual confirmation of the lanthanide ions; then characterized by emission and excitation spectra to analyze the chemical composition.

Excitation and emission spectra of all samples of the four markers were acquired in triplicate with a spectrofluorometer Fluorolog Horiba Jobin Yvon FL3-22, with excitation from a xenon lamp. The excitation spectra were obtained by monitoring the emission at 545 nm, scanning with a range between 200-400 nm. The emission spectra were acquired by monitoring the excitation at 323 nm and analyzing in the range of 400-750 nm. Those parameters were selected that presented the best results when used for all markers.

Blind tests

For the blind tests, ten inks marked were selected by a volunteer unrelated to the study, so that the composition of the sample remained unknown, avoiding interference in the final result. To keep the composition unknown, each sample received a code from S1 to S10, and this information was kept in a sealed envelope until the end of the experiments. Then, each sample was deposited on the surface of a vegetable paper, with its code indicated on the other side of the paper.

When dried, the papers were analyzed by excitation and emission spectroscopy to evaluate the chemical composition. A manual scanning was performed, where spectra of excitation and emission were observed at random points of the sample. Each spectrum was acquired in

triplicate. The same procedure was performed on the other nine unknown samples. Based on the results, the analysts tried to identify the markers. At the end of the experiments, the envelope was opened to confirm the coded results of the evaluation of analysts and volunteers.

Chemometric treatment

For chemometric analysis, the PLS Toolbox 8.6.2 (Eigenvector Research Inc., USA) was used in MATLAB²⁵ (R2021b, MathWorks Inc., USA). The following preprocesses were used: Savitzky-Golay smoothing filter using 2-order polynomial and 21-point window width for spectral noise removal, automatic weighted least squares baseline correction, mean center and normalization by the maximum (raw and preprocessed spectral profiles from the pure markers and the inks can be seen in the Supplementary Information (SI) section, Figure S1). For PCA, a low-level data fusion approach was taken in which the data matrices containing the excitation (200 to 400 nm) and emission (400 to 750 nm) spectra were concatenated. Afterwards, K-means was applied to PCA scores using the Mahalanobis distance. Unsupervised models are of great importance for the analysis of questioned documents; among their advantages is the fact that they are simple techniques, widely accepted by forensic experts, and do not require the use of reference databases.

Supervised approaches of this application can be employed for proposing an encoding system based on the identification of a specific spectral signature. Figure 1 shows the scheme of a possible 2-step protocol to be followed. The protocol consists in a detection fraud step and an encoding step. Questioned documents without security markers do not show luminescence under the UV-light, making fraud easily identifiable as early as in the first step (fraud detection). If luminescence is observed, there are two scenarios: the first one being the case of luminescence resulting from another luminescent material; the second scenario is the case of it being a controlled security marker (encoding step). The first scenario is related to another type of fraud and is currently out of the scope of this article. The second scenario, the encoding step, was the object of the present study. As an example of application, a general QR code produced with marked inks can provide different information once parts of the code have different markers. In this scenario, a classification task is needed to expedite the readiness and since the proposed markers have a known spectral profile (with no possibility of unknown markers that have not been modelled) and can be initially identified by UV light, discriminant analysis approaches can be employed instead of class-modeling to identify the specific

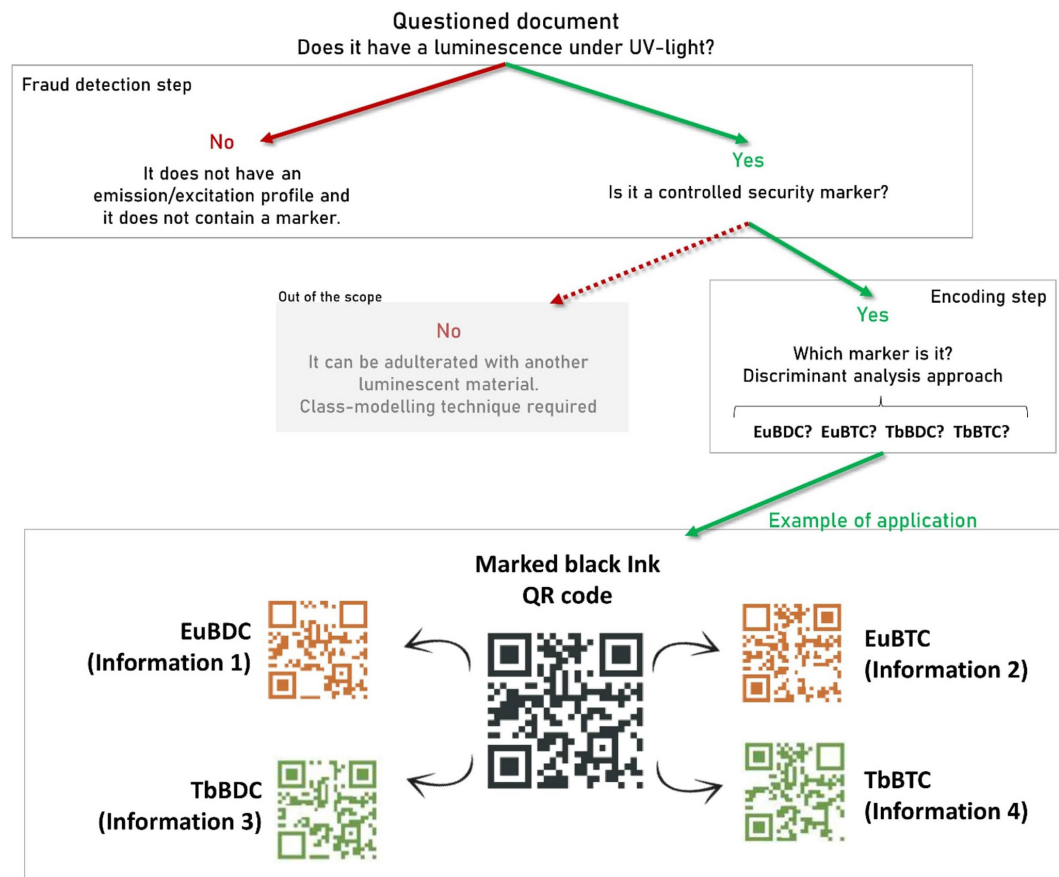


Figure 1. Scheme for a suggested protocol. As example, a QR code is produced with different markers and each one provides a different information. Possible frauds regarding the markers are not included within the scope of the present work.

chemical profile of the marker in the document. In this case, the partial least squares discriminant analysis model (PLS-DA) was proposed as a classification model. The pure marker samples (60 samples) and 12 samples containing the marker in inks (3 *per* marker) were employed as the training set while the remaining marked inks (28 samples) and the blind test samples (10 samples) were used as the prediction set. The PLS-DA model was also built using only the pure markers as training set (60 samples) and all the marked inks and the blind test samples used as prediction set (50 samples). Random subsets cross-validation (10 data splits and 5 iterations) was employed, the figures of merit rates of true positive (TPR) and true negative (TRN), as well as sensitivity (Sn) and specificity (Sp) were assessed as well as the variable importance in projection (VIP scores). For the sake of comparison, a class-modelling approach was also considered and the soft independent modelling by class analogy (SIMCA) model was build.

Results and Discussion

Figures 2a and 2b shows the average excitation and emission spectra acquired for pure Eu-based markers

and Figures 3a and 3b shows the average excitation and emission spectra acquired for pure Tb-based markers. In both cases of the excitation spectrum (200 to 400 nm), it is possible to see the ligand band $\pi \rightarrow \pi^*$, which refers to the antenna effect; for markers with BDC ligands, however, this band is shifted to longer wavelengths. In the emission spectra, which brings information about the central metal, the characteristic bands of each lanthanide are observed. In Figures 2a and 2b, it is possible to observe the Eu^{3+} characteristic bands, related to ${}^5\text{D}_0 \rightarrow {}^7\text{F}_J$ (as $J = 0-4$) transitions. With respect to the emission spectra of the markers based on Tb^{3+} , in Figures 3a and 3b, ${}^5\text{D}_4 \rightarrow {}^7\text{F}_J$ (as $J = 6-2$) transitions are noted. The markers were also analyzed by X-ray diffraction and compared to CIF to confirm the phase obtained; the diffractograms can be seen in the SI section (Figures S2, S3, S4 and S5).

The spectral differences of each marker became more prominent when the excitation and emission spectral profiles overlap and are fused in a low-level data fusion approach after normalization to provide the complete luminescence profile (since the information from the excitation and emission spectra are complementary once excitation spectrum brings information about the ligand

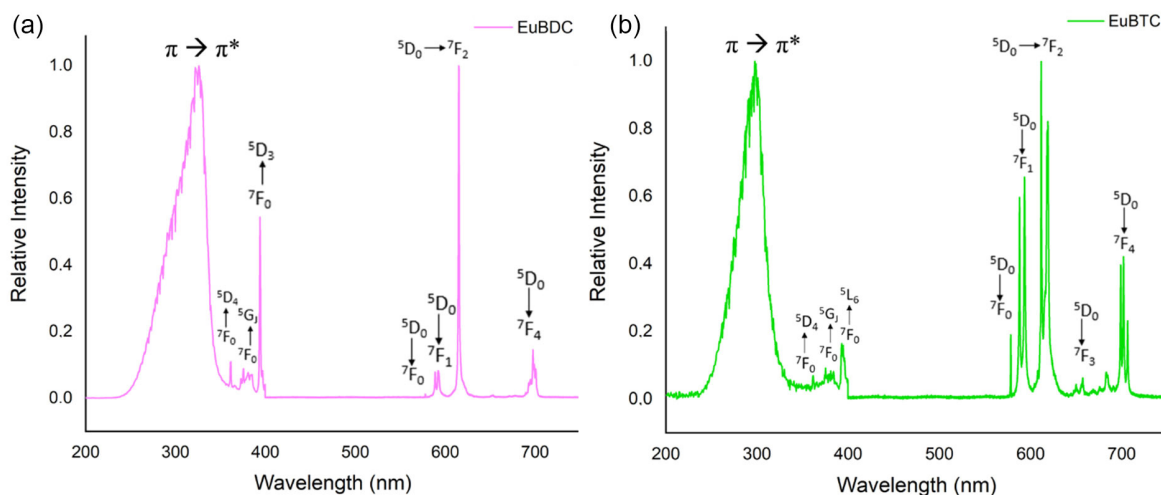


Figure 2. Average excitation (200–400 nm) and emission (400–750 nm) spectra with the correspondent transitions for (a) EuBDC marker and (b) EuBTC marker.

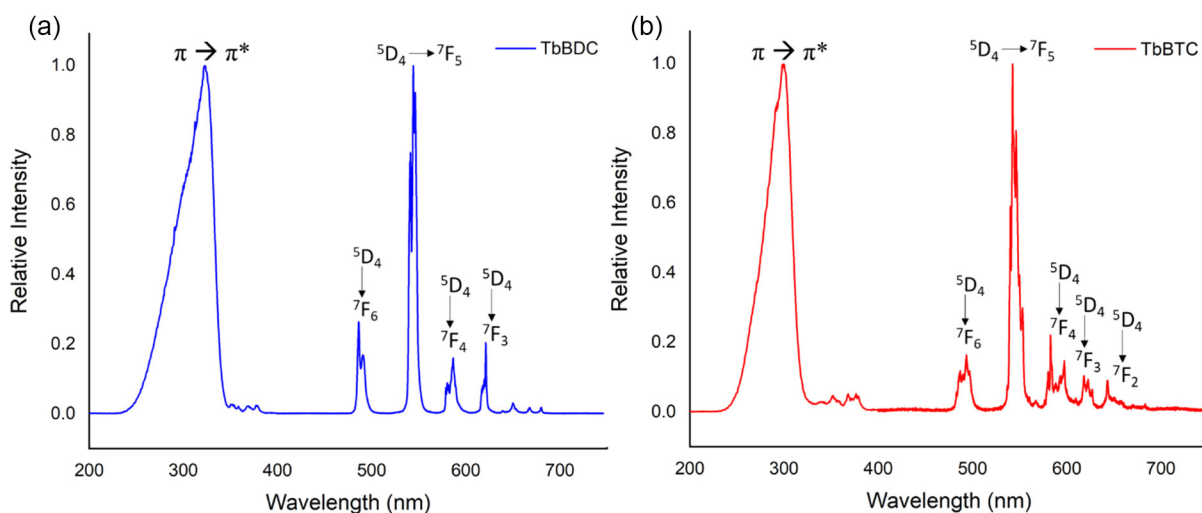


Figure 3. Average excitation (200–400 nm) and emission (400–750 nm) spectra with the correspondent transitions for (a) TbBDC marker and (b) TbBTC marker.

and the emission spectrum about lanthanide by itself); the mean spectra acquired can be seen in Figure 4.

For an exploratory analysis of the data, PCA was performed to evaluate the four chemical markers under study. It was possible to observe a 2-component model, explaining 89.02% of data variation, showing a clear difference between the four markers in the scores scatter plot using the PC1 and PC2 as illustrated in Figure 5. Note that samples are coded as follows: same color for the same Ln (red for Eu and green for Tb), same symbol for the same organic ligand (diamond for BDC, circles for BTC) and edges for the inks according to the ink color (blue edge for blue ink and black edge for black ink) and grey circles for the blind test samples. The PCA model was built using only the synthesized pure markers, while the samples of markers incorporated in the inks and the blind test samples

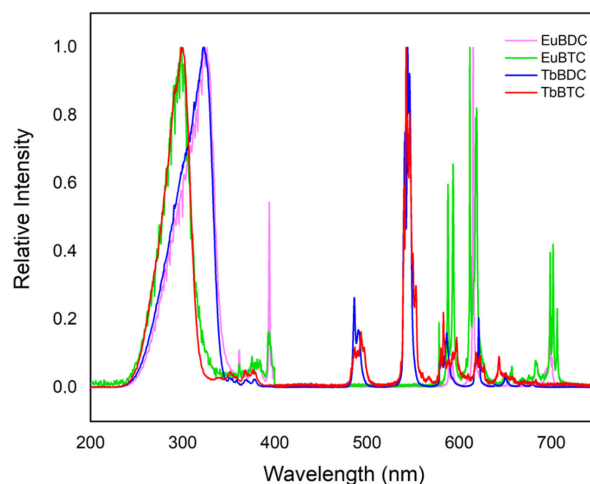


Figure 4. Mean normalized luminescence spectra concatenated of excitation and emission spectra from EuBDC, EuBTC, TbBDC and TbBTC (low-level data fusion).

were only projected in the model as prediction set. The corresponding loadings plot can be seen in Figure 6 and the residuals can be seen in Figure S6 (SI section).

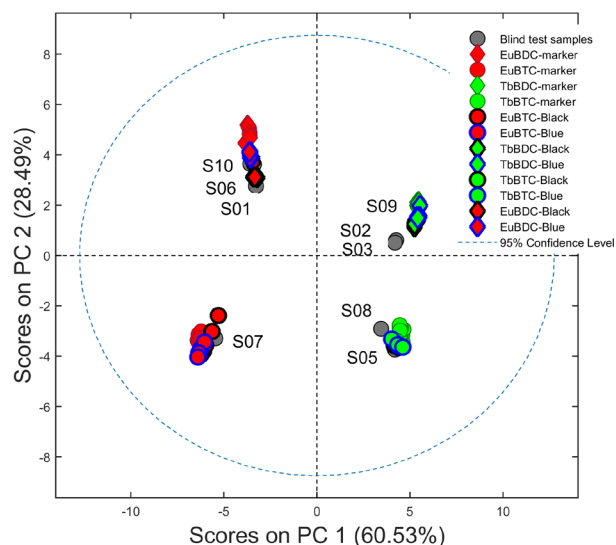


Figure 5. Scores scatter plot of PC1 and PC2 from a 2-component PCA model for the concatenated luminescence spectrum of MOFs Tb(BDC), Tb(BTC), Eu(BTC) and Eu(BDC). MOFs with the same metallic center were represented by the same color: red for Eu and green for Tb; MOFs with the same ligand are represented by the same symbol: circles for BTC and diamonds for BDC; ink color used are represented by the color in the edge of the symbol: black for black ink and blue for blue ink.

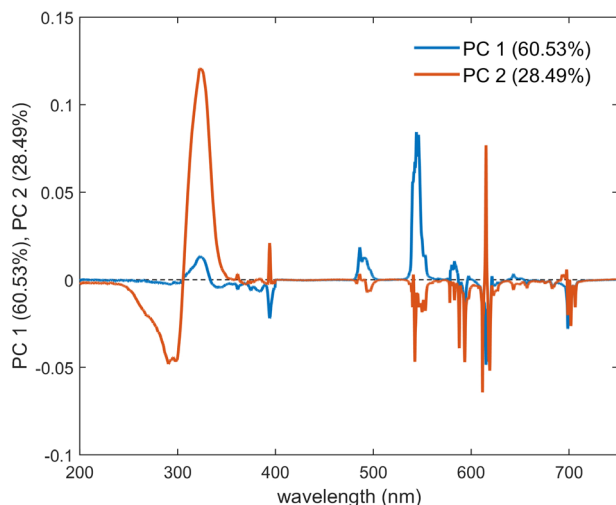


Figure 6. Loadings plot of PC1 and PC2 from the PCA model built with the concatenated luminescence spectra of excitation and emission profiles of MOFs Tb(BDC), Tb(BTC), Eu(BTC) and Eu(BDC).

It is important to bear in mind that excitation and emission profiles are being used as spectral signatures. These analytical signals have a particular behavior in this particular scenario, given that the presence of the ink chemical matrix does not affect the spectral profile of the marker. The clear presence of specific emission and excitation bands related to the Ln-based markers do not

carry information superimposed on any other constituent present in the inks. This can be seen in the scores scatter plot depicted in Figure 5, where no significant differences between the pure markers and the inks is observed. This, along with the fact that 89.02% of data variation is explained by the 2 first components, provides an important consequence for further data analysis. PCA loadings interpretation is straightforward, and it provided unambiguous chemical interpretation in this particular case.

The loading plot shows interesting features regarding the simplicity of the PCA model. It can be observed that PC1 clarifies the differences between the markers Tb(BTC), Tb(BDC) and Eu(BTC), Eu(BDC), since the first principal component differentiated markers from two distinct metal centers: Eu and Tb. This is also observed in the loadings plot, since the PC1 profile shows that relevant contributions were mainly related to the emission profile (from 400 to 750 nm). The Eu-based and Tb-based markers can be distinguished visually by emissions in different colors: orange for Eu and green for the Tb. Moreover, looking at the loadings plot, this difference is mainly due to the fact that the markers containing Tb have transitions in regions of shorter wavelengths, which does not occur for markers with Eu. Note for markers Tb(BTC) and Tb(BDC) the transitions $^5D_4 \rightarrow ^7F_6$ (at approximately 490 nm) and $^5D_4 \rightarrow ^7F_5$ (at approximately 545 nm), whereas for the Eu markers, no transition can be observed in this region, the first transitions appear only after 550 nm. The specific spectral transitions for Tb have positive loading values in PC1 and are associated with samples in the scores plot with positive scores, which are the samples containing the Tb marker. In contrast, the spectral regions associated to Eu transitions ($^7F_0 \rightarrow ^5D_3$ at approximately 400 nm; $^5D_0 \rightarrow ^7F_2$ at approximately 620 nm; and $^5D_0 \rightarrow ^7F_4$ at approximately 700 nm) have negative values for PC1 loadings and are associated with samples with negative scores in PC1, which are all markers containing Eu in their composition.

As is well known, the aggregation of lanthanides as the metallic center in MOFs produces luminescent materials, with narrow and well-defined spectral lines. As well, the relative intensities of the transitions in the luminescence spectra can be used to probe the local environment of the ion, in the case of europium.²⁶ So, by altering the lanthanide ion, it is possible to differentiate markers by the color of the emitted light. If the same ion is fixed, it is possible to distinguish the MOFs by analyzing the organic ligands. This is observed in PC2 scores and loadings. The second principal component clarifies the difference between markers regarding the organic ligand used. The scores scatter plot (Figure 5) shows that both Eu(BDC) and Tb(BDC), represented by diamonds, have

positive scores for PC2, whereas Eu(BTC) and Tb(BTC) have negative values of PC2, represented by circles in Figure 5. PC2 loadings plot shows that both excitation and emission profiles made important contributions towards that differentiation.

The main contribution observed in PC2 (orange line in Figure 6) is observed at approximately 300-330 nm, which is related to the ligand band $\pi \rightarrow \pi^*$. The loadings of PC2 shows a characteristic profile of a shift related to the spectral profiles in that region, in which BDC appears, displacing the profiles to longer wavelengths for the Eu(BDC) and Tb(BDC), at 330 and 323 nm, respectively. In contrast, Eu(BTC) and Tb(BTC) shows the $\pi \rightarrow \pi^*$ transition at smaller wavelengths, 300 and 297 nm, respectively.

Moreover, contributions to emission profiles can also be observed. It can be seen that the differentiation of the Tb(BTC) and Tb(BDC) MOFs is based on the small spectral profile differences of both markers. For Tb(BTC), in general, transitions appear with greater multiplicity than observed for Tb(BDC). For the europium-based markers, the differentiation is mainly due to the characteristic spectral profile, $^5D_0 \rightarrow ^7F_1$ and $^5D_0 \rightarrow ^7F_2$ transitions, at approximately 590 and 615 nm, respectively, and have negative values for PC2. These two transitions in the Eu(BTC) profile appear with relatively similar intensity, as this MOF experiences a high symmetry chemical environment, which is not the case for Eu(BDC), where transition $^5D_0 \rightarrow ^7F_2$ shows up with higher relative intensity.

After this exploratory analysis, the complete spectra acquired for the MOFs incorporated in the blue and black inks on the vegetal paper and the spectra were projected onto the PCA model previously built. Results can be seen in the scores scatter plot in Figure 5, where the blue and black inks are coded as the symbols with blue and black edges, respectively. As can be seen, all spectra acquired on vegetal paper for both blue and black ink show similarities with their corresponding marker. Small variations in the samples with black and blue inks were apparently caused by the reduction of the signal to noise ratio observed for these samples. Possibly the high spectral noise observed is due to interference caused by some chemical component in the inks. Despite this, no sample deviated from the respective pure MOF to the point of causing data interpretation by exploratory analysis to be vague. However, to ensure correct association of the inks with their respective pure markers, a cluster analysis was performed with the data.

A model based on K-means using the marked inks and pure markers was built and all ink samples were clustered according to their respective marker. The dendrogram in Figure 7 shows the four groups, each related to a particular marker. To simplify data visualization, the sample labels

from the inks with known markers and the pure markers were removed and only the blind test samples can be seen.

Finally, a blind test was performed, in which one volunteer randomly selected 10 samples for the acquisition of luminescence spectra and subsequent projection in the PCA and K-means models proposed in order to reaffirm the methodology. The scores scatter plot in Figure 5 shows the blind test samples in grey circles projected on the PCA model, whereas Figure 7 shows to which marker the blind test samples were associated using the K-means clustering. The results obtained using the two unsupervised models are summarized in Table 1.

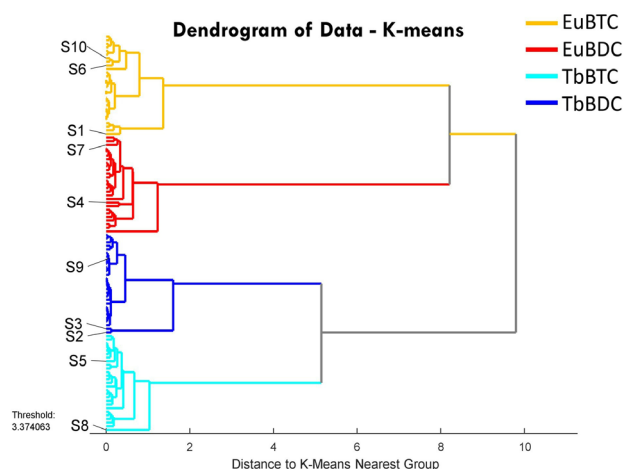


Figure 7. K-means dendrogram with 4 clusters. Blind test samples are highlighted in the insert.

Table 1. Results of blind test summarized. 10 samples and the corresponding MOF association after PCA performed

Sample	MOF association	Answer	
S1	EuBDC	EuBDC	✓
S2	TbBDC	TbBDC	✓
S3	TbBDC	TbBDC	✓
S4	EuBTC	EuBTC	✓
S5	TbBTC	TbBTC	✓
S6	EuBDC	EuBDC	✓
S7	EuBTC	EuBTC	✓
S8	TbBTC	TbBTC	✓
S9	TbBDC	TbBDC	✓
S10	EuBDC	EuBDC	✓

S1-S10: samples 1 to 10 from blind test; MOF: metal-organic frameworks; BTC: trimesic acid; BDC: terephthalic acid.

As shown, all 10 samples analyzed in this test were correctly associated to their corresponding MOF; even taking into account the sample dispersion caused by the noisy profiles obtained with the marked inks, the K-means correctly associated the samples with the

corresponding MOF. It is important to remember that PCA is an exploratory data model, therefore does not provide a classification itself.

To implement a classification model, it is important to notice another important consequence of the emission/excitation profiles. The documents without specific markers will not provide a spectral signal if there is no luminescent constituent inserted. Possible samples with forgeries (i.e., documents without the luminescent markers) that lack spectral signature for these particular analytical techniques do not need to be included in either the training or the prediction sets. Moreover, the spectral signal is exclusively related to the specific marker without any other constituent interference. This makes PLS-DA a perfect candidate as classification model for this authentication issue, since this is, in fact, an encoding problem.

In a recent study, Sharma *et al.*²⁷ also implemented a multivariate analysis (using linear discriminant analysis (LDA) model) for discrimination and classification of marker pen inks, achieving a discriminative power of 98.21% for permanent marker inks and 100% of discrimination power for whiteboard marker inks. This method has provided better results compared to the visual examination, showing the importance of this type of analysis.

A PLS-DA model was proposed and, using 3 latent variables, made it possible to achieve 100% specificity and

sensitivity for both training and prediction sets. Figure 8 shows the Y predicted values for all samples (training and prediction) while Figure 9 shows the original variables that were meaningful for the classification model proposed. It is possible to visualize that PLS-DA results for blind test samples correspond to the association made when using the simple K-means. It is important to mention that the PLS-DA model built using only the pure markers as the training set showed the same performance of the model using both pure and marked inks samples in the training set. Additionally, the VIP scores plot shows that the original variables that are significant for the classification are the same as those discussed in the loadings plot of PCA models, corresponding to important ligand bands and metal transition.

For comparison reasons between a discriminant analysis approach and a class-modelling, SIMCA model was also built for the fused data (Table 2). For all classes, 2 PCs were necessary to build each one of the independent models. SIMCA models shows advantages and disadvantages in comparison to PLS-DA model. One misclassified sample was observed in the prediction set, reflecting into a decrease of TNR and TPR of EuBDC and EuBTC classes, respectively. The residuals plots for the SIMCA models can be found in the Figure S9 (SI section). Although the results seem to be outperformed by PLS-DA, SIMCA models has the advantage of being independent from the non-target

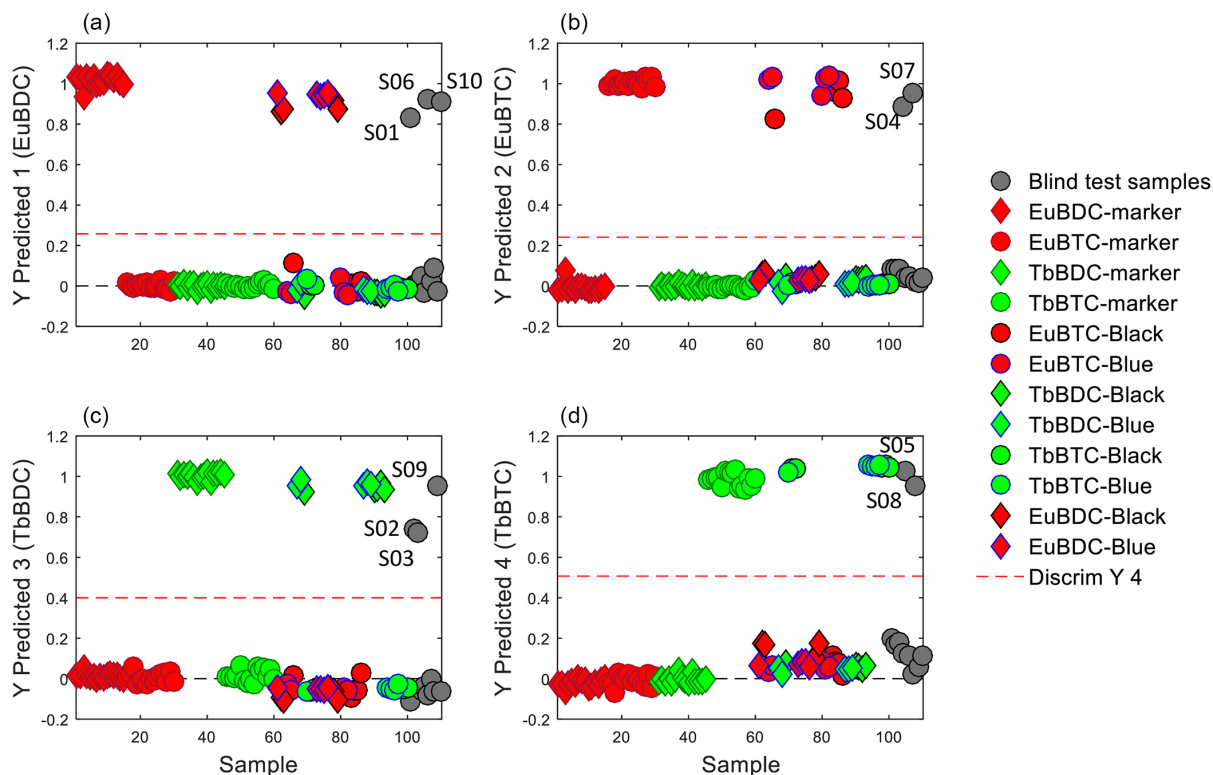


Figure 8. PLS-DA results with the Y-predicted values for all samples (training and prediction sets) for each class: (a) Eu(BDC); (b) Eu(BTC); (c) Tb(BDC); (d) Tb(BTC).

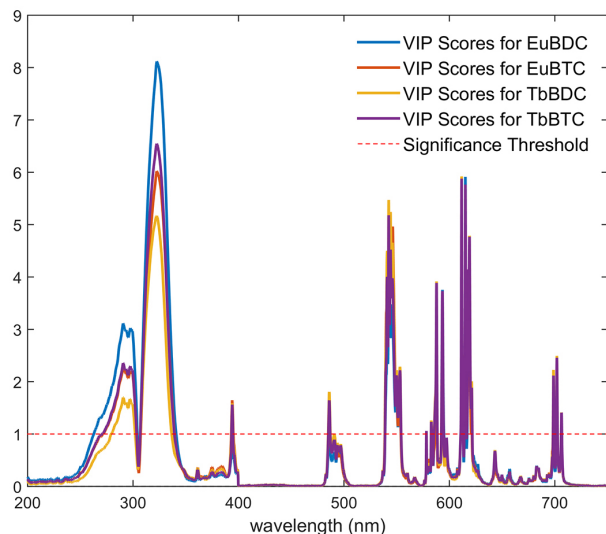


Figure 9. Variable importance in projection (VIP) scores of PLS-DA model.

samples. The models are built using only the information of the target class to define the decision boundaries, which is not affected by the presence of non-target samples, in contrast to PLS-DA. On the other hand, to build a more robust SIMCA model, more samples need to be included, especially the pure markers with inks, in the training set. Nevertheless, SIMCA also showed a correct classification for all the blind test samples.

Table 2. Results of the SIMCA model

	PCs	TPR _{training} / %	TPR _{prediction} / %	TNR _{prediction} / %
EuBDC	2	100	100	95
EuBTC	2	100	85	100
TbBDC	2	100	100	100
TbBTC	2	100	100	100

PCs: principal components; TPR: true positive; TRN: true negative; BTC: trimesic acid; BDC: terephthalic acid.

Conclusions

The purpose of the study was to propose new MOF-based markers for anti-fraud inks that could be employed to create a document encoding process using simple, robust, sensitive and non-destructive method that acts like a safety feature to assist in the authentication of documents. The chemometric tools of principal component analysis and K-means were used as unsupervised models in order to differentiate four marked luminescent inks, based on Eu and Tb. These were associated to the known pure marker profiles through their luminescence spectral profiles. Even using a spectroscopic luminescence technique that, fundamentally, provides molecular information, it was possible to obtain

information that was associated with the ligand (excitation spectrum) and with lanthanide (emission spectrum). With this complementary information, with low-level data fusion approach, it was possible to observe clear differences between the Tb(BDC), Tb(BTC), Eu(BTC) and Eu(BDC) using a simple 2-component PCA model for both pure MOFs and after their incorporation in blue and black inks and painted on the vegetal paper. The K-means dendrogram corroborated with that observed in the PCA and also made a correct association between the inks and their respective pure markers. The blind test performed was also effective, confirming the robustness of the proposed method. All 10 samples analyzed in this step were correctly associated with their correspondent marker, using only unsupervised models.

In addition, two classification approaches were employed. Using a PLS-DA model, with only pure markers as the training set, it was possible to achieve 100% of specificity and sensitivity for both training and prediction sets. PLS-DA results for blind test samples corresponded to the association made using simple K-means. Although SIMCA models showed a misclassified sample for EuBTC class, it has the advantage of class-modelling approaches that do not depend on non-target samples to define the decision boundaries, being also a strong candidate for future applications. For future work and improvement of the current methodology, the association of variable selection with classification models and hyperspectral images can make the acquisition of spectral profiles simpler, due to the selection of few spectral channels, thereby providing more accuracy to the decision to implement a confidence limit for each class.

Supplementary Information

Supplementary information, which contains the analysis and comparison of the X-ray diffraction of the synthesized samples with the CIF, is available free of charge at <http://jbcbs.sbj.org.br> as PDF file.

Acknowledgments

To Brazilian agencies CNPq, CAPES, FACEPE and FAPDF. Special thanks for NUQAAPE, NEQUIFOR, CAPES-PROFORENSE, 3509/2014. The English text of this paper has been revised by Sidney Pratt, Canadian, MAT (The Johns Hopkins University), RSAdip-TESL (Cambridge University).

References

1. de Farias, R. F.; *Introdução à Química Forense*, 3rd ed.; Editora Átomo: Campinas, Brazil, 2010.

2. D'Almeida, M. L. O.; Koga, M. E. T.; Granja, S. M.; *Documentoscopia: O Papel como Suporte de Documentos*, 1st ed.; ITP: São Paulo, Brazil, 2015.
3. da Silva, V. G.; Talhavini, M.; Peixoto, I. C. F.; Zacca, J. J.; Maldaner, A. O.; Braga, J. W. B.; *Microchem. J.* **2014**, *116*, 235. [Crossref]
4. Somma, F.; Aloe, P.; Spagnolo, G. S.; *J. Phys.: Conf. Ser.* **2010**, *249*, 12060. [Crossref]
5. Heudt, L.; Debois, D.; Zimmerman, T. A.; Köhler, L.; Bano, F.; Partouche, F.; Duwez, A. S.; Gilbert, B.; de Pauw, E.; *Forensic Sci. Int.* **2012**, *219*, 64. [Crossref]
6. Braz, A.; López-López, M.; García-Ruiz, C.; *Forensic Sci. Int.* **2013**, *232*, 206. [Crossref]
7. Raza, A.; Saha, B.; *Sci. Justice* **2013**, *53*, 332. [Crossref]
8. Rodrigues, V. C.; Diniz, M. F.; Mattos, E. C.; Dutra, R. C. L.; *Polimeros* **2016**, *26*, 68. [Crossref]
9. Council of the European Union General Secretariat, *Public Register of Authentic Travel and Identity Documents Online PRADO Glossary - Technical Terms Related to Security Features and to Security Documents in General*, <https://www.consilium.europa.eu/prado/en/prado-glossary/prado-glossary.pdf>, accessed in June 2022.
10. Anderson, R.; *Security Engineering: A Guide to Building Dependable Distributed Systems*, 2nd ed.; Wiley Publishing, Inc.: Indianapolis, USA, 2008.
11. Ezcurra, M.; Grávalos, G. R.; *Análise Forense de Documentos - Instrumentos de Escrita*, 1st ed.; Millennium: São Paulo, Brazil, 2012.
12. United Nations Office on Drugs and Crime (UNODC); *Introduction to Security Document Examination: Training Programme and Trainers' guide*, https://www.unodc.org/documents/scientific/Brochure_doc_kits.pdf accessed in June 2022.
13. Song, B.; Wang, H.; Zhong, Y.; Chu, B.; Su, Y.; He, Y.; *Nanoscale* **2018**, *10*, 1617. [Crossref]
14. da Luz, L. L.; Milani, R.; Felix, J. F.; Ribeiro, I. R. B.; Talhavini, M.; Neto, B. D.; Chojnacki, J.; Rodrigues, M. O.; Júnior, S. A.; *ACS Appl. Mater. Interfaces* **2015**, *7*, 27115. [Crossref]
15. Singh, A. K.; Singh, S.; Gupta, B. K.; *ACS Appl. Mater. Interfaces* **2018**, *10*, 44570. [Crossref]
16. Liang, L.; Chen, C.; Lv, Z.; Xie, M.; Yu, Y.; Liang, C.; Lou, Y.; Li, C.; Shi, Z.; *J. Lumin.* **2019**, *206*, 560. [Crossref]
17. Li, X.; Hu, Y.; *Carbohydr. Polym.* **2019**, *203*, 167. [Crossref]
18. Tam, M.; Li, F.; Wang, X.; Fan, R.; Chen, G.; *ACS Nano* **2020**, *14*, 6532. [Crossref]
19. Jiang, K.; Zhang, L.; Lu, J.; Xu, C.; Cai, C.; Lin, H.; *Angew. Chem., Int. Ed.* **2016**, *55*, 7231. [Crossref]
20. Carneiro, C. R.; Silva, C. S.; Pimentel, M. F.; Talhavini, M.; Weber, I. T.; *Microchem. J.* **2020**, *159*, 105362. [Crossref]
21. Zou, H.; Hastie, T.; Tibshirani, R.; *J. Comput. Graphical Stat.* **2006**, *15*, 265. [Crossref]
22. Kumar, P.; Singh, S.; Gupta, B. K.; *Nanoscale* **2016**, *8*, 14297. [Crossref]
23. Rodionova, O. Y.; Titova, A. V.; Pomerantsev, A. L.; *TrAC, Trends Anal. Chem.* **2016**, *78*, 17. [Crossref]
24. Wanderley, K. A.; Alves Jr., S.; Paiva-Santos, C. O.; *Quim. Nova* **2011**, *34*, 434. [Crossref]
25. MATLAB, (*R2021b*); The MathWorks, Inc., Natick, Massachusetts, United States, 2021.
26. Binnemans, K.; *Coord. Chem. Rev.* **2015**, *295*, 1. [Crossref]
27. Sharma, V.; Kumar, R.; Devgan, K.; Mishra, P. K.; Ekielski, A.; Kumar, V.; Kumar, V.; *Spectrosc. Lett.* **2018**, *51*, 5. [Crossref]

Submitted: December 8, 2021

Published online: July 5, 2022

