Microchemical Journal 218 (2025) 115223

Contents lists available at ScienceDirect

Microchemical Journal

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

Mixing to unmix: Painting component identification in historical graphic documents via data fusion of DRIFTS and VNIR–SWIR reflectance spectra

Francisco Moronta-Montero [a,*], Anna Sofia Reichert [b], Ana Belén López-Baldomero [a], María Rosario Blanc [c], Carolina Cardell [d], Ana López-Montes [b], Javier Hernández-Andrés [a], Eva M. Valero [a]

[a] Department of Optics, University of Granada, Faculty of Sciences, Campus Fuentenueva, s/n, Granada, 18071, Spain
[b] Department of Painting, University of Granada, Faculty of Fine Arts, Av. Andalucía s/n, Granada, 18071, Spain
[c] Department of Analytical Chemistry, University of Granada, Faculty of Sciences, Campus Fuentenueva, s/n, Granada, 18071, Spain
[d] Department of Mineralogy and Petrology, University of Granada, Faculty of Sciences, Campus Fuentenueva, s/n, Granada, 18071, Spain

ARTICLE INFO

Keywords:
Hyperspectral imaging (HSI)
Diffuse Reflectance Fourier Transform Infrared Spectroscopy (DRIFTS)
Spectral unmixing
Data fusion
Painting mock-ups
Illuminated historical manuscripts

ABSTRACT

The identification of historical painting materials is essential for understanding artistic techniques, proposing conservation strategies, and supporting dating and provenance studies. However, the presence of pigment and dye mixtures, along with complex pigment–binder interactions, poses significant challenges for non-invasive analysis. This study presents a spectral unmixing approach based on the data fusion of hyperspectral imaging (HSI) and Diffuse Reflectance Fourier Transform Spectroscopy (DRIFTS) data to identify the painting components of mock-up samples that replicate historical objects. Results show that fusion significantly improves the identification of mixture painting components compared to single-technique analysis. The complement of the Goodness-of-Fit Coefficient (cGFC) outperformed other metrics, achieving the highest rate of correct painting component identification and lowest error for the greater presence of some elements compared to others. Preprocessing steps including Savitzky–Golay derivative, spectral cropping, and normalization proved essential for maximizing performance. The method was further validated on a set of historical manuscripts, correctly identifying painting materials in the majority of cases.

1. Introduction

The analysis of historical pigments and dyes plays a critical role in studying past cultures, planning conservation strategies [1], and aiding in the accurate dating of artworks based on the materials present [2]. Given the fragility and importance of heritage objects, non-invasive analytical methods are indispensable tools for its study and conservation.

Among these techniques, Hyperspectral Imaging (HSI) has gained relevance in cultural heritage research for its ability to map materials, offering both spectral and spatial information [3]. However, traditional classification approaches often assume that each pixel or measurement corresponds to a single pure material, which is rarely the case in painted surfaces. In these cases, a single pixel often contains signals from multiple pigments, binders, or even surface irregularity effects, which can limit the accuracy of material identification. These limitations are particularly pronounced when binders or varnishes alter the optical properties of pigments, requiring complementary methods such as Raman spectroscopy, X-ray fluorescence (XRF) or Fourier Transform Infrared Spectroscopy (FTIR) for confirmation [4].

Diffuse Reflectance Fourier Transform Spectroscopy (DRIFTS) is widely employed to identify pigments, dyes and binders [5,6]. Acquisition of DRIFTS spectra using portable systems offers non-destructive information about the molecular composition of the samples [7], and allows identifying organic and inorganic components, even in layered structures. However, DRIFTS can struggle to resolve mixtures due to overlapping absorption bands [8].

The integration of multiple techniques has become increasingly common, as in most of the cases, single techniques are insufficient for identification. For example, techniques such as Raman spectroscopy or DRIFTS can be used for the identification of molecular vibrational modes, and then complemented using HSI, which allows taking into account the spectral reflectance of the sample in the visible and infrared ranges, as well as mapping the results [9–11].

Recent advances in data fusion offer a promising solution to these challenges. In fields such as food analysis [12,13], low-level data fusion, performed through simple concatenation of DRIFTS and HSI

* Corresponding author.
E-mail address: fmoronta@ugr.es (F. Moronta-Montero).

https://doi.org/10.1016/j.microc.2025.115223
Received 10 July 2025; Received in revised form 4 September 2025; Accepted 6 September 2025
Available online 17 September 2025

reflectance spectra, has led to significant improvements in classification and component identification when compared to using the techniques independently.

Similar strategies have been proposed for pigment mixture classification with promising results [14], although such studies are often limited in scope, for instance, by focusing on a narrow range of pigments of similar color or composition, while more complex mixtures of pigments and/or dyes remain largely unexplored.

Spectral unmixing has proven to be a powerful analytical tool for this purpose. These models aim to decompose a spectrum into a linear or nonlinear [15] combination of known reference spectra, called endmembers, weighted by their relative concentration [16]. In the field of cultural heritage, this procedure has been successfully used with hyperspectral data for pigment component identification [17–20], but its efficacy in other types of spectra, such as DRIFTS, has not yet been studied.

This work aims to address the problem of pigment and dye unmixing in a multimodal manner. It seeks to apply unmixing algorithms with different spectral comparison metrics to two different types of data: VNIR–SWIR reflectance and DRIFTS spectra, both separately and merged by low-level data fusion, with the goal of getting one step closer to non-invasive material identification in real historical documents.

Through quantitative validation on a controlled mock-up dataset, and qualitative assessment on real historical samples confirmed by XRF, this study demonstrates that data fusion significantly improves component identification via unmixing over single-modality approaches.

2. Materials and methods

2.1. Samples description

To optimize and test the proposed method, a subset of a series of painting mock-ups previously presented in [21], which are part of the HYPERDOC dataset [22], was used.

The full dataset presented in [21] contains 156 painting mock-ups, of which we selected 54 samples deposited on paper, composed of a 1:1 mixture of cotton and linen fibers, and 54 deposited on parchment. The paintings components chosen as endmembers for this work are shown in Table S1 of supplementary information (including the abbreviations and chemical composition of each painting component). These are:

- 14 pigments: azurite (AZ), smalt (SM), lapislazuli (LAP), lead white (LW), calcite (CA), cinnabar (CIN), hematite (HMT), minium (MIN), orpiment (ORP), lead tin yellow (LTY), malachite (MLC), verdigris (VG), yellow ochre (OC), and burnt umber (BU) mixed with either gum arabic (GA) or egg glair (EG), with the exception of hematite, which was bound exclusively with EG, making a total of 27 samples.
- 6 dyes: carmine (CARM), gamboge (GMB), saffron (SA), indigo (ING), and imperial yellow ink (IYI, made as a mixture of buckthorn with alum), all of them bound exclusively with GA, and a buckthorn (BCKT) dye solution with no binder.

Therefore, this results in 33 endmembers, representing the possible individual painting constituents of the more complex mixtures.

In addition, 42 mixtures also described in [21] were selected to test our method. 38 of these mixtures are triphasic, (i.e. composed of 2 pigments and 1 binder, or 1 pigment, 1 dye and 1 binder), while the remaining 4 are quadriphasic, containing a mixture of 3 pigments (CIN, LTY and LW) bound with either GA or EG. All these mock-ups are combinations in equal proportions of its painting components (pigments, dyes and binder). All of the painting samples were applied on paper and on parchment supports. The full list of the triphasic and quadriphasic mixtures can be found on Table S2 of supplementary information.

The samples were applied to the support (paper or parchment) in two ways: as flat-tone over a 2x2 cm surface, and as writing strokes, in both cases applied with a brush. For this study, only the spectral information of the squares was used, since it provides a more homogeneous surface.

In addition to the mock-up samples, some points of interest from 11 different historical documents were analyzed to test the applicability of the proposed method in real cases. These historical documents are preserved in the Archive of the Royal Chancellery of Granada (Spain), and consist of 7 illuminated manuscripts with parchment support that correspond to nobleman's lawsuits dated between 1459 and 1608, and 4 maps on paper from the 18th century. All of them decorated with a significant variety of bound pigments and dyes, some of which have not yet been unequivocally identified.

Fig. 1 shows a series of examples of the pigments, dyes and mixtures containing binder measured for this study, as well as the comparison between the spectra of the endmembers and their mixture for the case of a mixture of azurite and lead white bound with GA.

2.2. Analytical techniques

To characterize the painting materials present in both mock-up samples and historical documents, a multi-modal analytical approach was adopted. For spectral measurements (HSI and DRIFTS), a representative area of each color was chosen for each document. The points of interest were chosen to completely cover an area of about 7 mm^2, which roughly corresponds with the spot area of the XRF measurement device used as external validation technique for material identification in this study (as explained in Section 2.2.3).

2.2.1. Hyperspectral imaging (HSI)

Hyperspectral imaging (HSI) captures were acquired for both mock-up samples and historical documents using two line-scan spectral cameras from Resonon: the Pika L [23] and Pika IR+ [24] systems, enabling data acquisition in the visible-near infrared (VNIR) and short-wavelength infrared (SWIR) spectral ranges, respectively. The VNIR camera covers the 380–1080 nm range with a spectral resolution of 2.1 nm and a spatial resolution of 900 pixels per line. The SWIR camera operates over the 888–1732 nm range with a 2.4 nm spectral resolution and 640 pixels per line. Due to signal degradation at the extremes of each range, the outer portions were discarded. The spectra were then interpolated to a common 5 nm sampling interval, resulting in 121 bands in the VNIR range (400–1000 nm) and 161 bands in the SWIR range (900–1700 nm). Dark current noise subtraction was performed, and flat-field correction was applied using the 90% reflectance patch of the Sphere Optics Zenith Lite Multistep reference white. Illumination was provided by four halogen lamps arranged to minimize specular reflections. Acquisition distances were set at 50 cm for VNIR and 40 cm for SWIR, resulting in fields of view of approximately 14.5 cm and spatial resolutions of 0.16 mm/pixel and 0.22 mm/pixel, respectively. Fig. 1 shows the capture configuration and an example spectra for this technique.

To ensure spatial alignment between the VNIR and SWIR datacubes, spectral images were registered using MATLAB's Registration Estimator App. A region of interest (ROI) corresponding to the painted square in each sample was extracted from the registered cubes for both spectral ranges. The spectral reflectance curves within each ROI were averaged, and the VNIR and SWIR spectra were concatenated. To ensure spectral continuity, the final ten bands of the VNIR and the initial eleven bands of the SWIR spectra were removed, producing a single reflectance spectrum for each sample.

2.2.2. Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS)

Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) analyses were performed on both mock-up samples and historical documents using a portable 4300 Handheld FTIR Spectrometer from Agilent Technologies [25] equipped with a diffuse reflectance interface and a 6 mm spot diameter. Spectra were acquired in the mid-infrared

region (4000–650 cm^{-1}) with a spectral resolution of 4 cm^{-1}, averaging 10 scans per spectrum. Background correction was carried out using a Coarse Gold Reference Cap (G8180-67560).

Spectral pre-processing was conducted using OMNIC [26] software (version 9.12.993), applying both smoothing and baseline correction procedures to minimize instrument-induced errors [27]. Measurements were performed by avoiding direct contact with the surface of the samples, aiming to establish an applicable non-invasive procedure for the analysis of real historical documents. This was achieved through the use of an adjustable tripod and a reclining bookrest, which enabled a stable positioning of the instrument. Fig. 1 shows the position of the document on the bookrest and how the spectrometer is placed for the measurement, together with an example of a DRIFTS spectrum.

As DRIFTS spectra are usually expressed in terms of wavenumbers, with units of cm^{-1}, in our case we calculate the inverse of these values in order to express them in wavelengths (nm) and make them comparable with HSI spectra.

2.2.3. X-ray Fluorescence (XRF)

To evaluate the effectiveness of the proposed method for historical document analysis, our results were compared with those obtained through portable X-ray fluorescence spectroscopy (XRF), a complementary technique that provides elemental information of the inorganic constituents present in the measured points [28].

XRF analysis was performed using a handheld NITON XL3t GOLDD+ (Thermo Fisher Scientific, Waltham, MA, USA) [29] with a silver anode (50 kV, 200 A). The analyzer was fitted with a camera and a suitably equipped Small Spot analyzer, with which analysis could be restricted to a small diameter of the camera angle (3 mm). After waiting five minutes to allow the instrument's electronics to stabilize, a system check was performed to calibrate the detector and ensure it was operating according to the specifications. Spectra were collected using the measuring mode "test all geo". The analyzer was equipped with four excitation filters (main, high, low, and light) that optimize the sensitivity of the analyzer for the various elements. Measurements of 30 s for each filter were set, meaning that each spot analysis took approximately 120 s to complete. The geometrical setup was the same for all the measurements. NITON Data Transfer (NDT) software 6.1 was used to control the instrument and for data management and transfer.

XRF analysis has been carried out exclusively for validation in historical documents, given that the mock-ups had already been thoroughly characterized in a previous study [21].

2.3. Data fusion

To integrate the complementary information provided by VNIR–SWIR reflectance and DRIFTS spectra, a low-level (feature-level) data fusion strategy was employed. Each VNIR–SWIR reflectance and DRIFTS spectrum, both in the endmembers and the problem mixtures, was concatenated into a single feature vector, preserving the full spectral resolution of both modalities while allowing the unmixing algorithm to operate on a unified dataset.

The motivation behind this strategy lies in the spectral complementarities of the two techniques: VNIR–SWIR reflectance spectra provide information about the actual color of the sample (VNIR range), and its behavior under infrared illumination (SWIR range), but they do not provide compositional information regarding the chemical bonds and vibrational modes of the molecules present within the material. On the other hand, DRIFTS does provide this information, but automatic algorithms using this spectra can confuse painting materials of similar chemical composition but different color, such as pigments with a high presence of lead, like red minium and lead white. Our main hypothesis is that, by combining both spectra, each will complement the weaknesses of the other, improving the results of using both individually.

In a preliminary test, mid-level data fusion was also explored using Principal Component Analysis (PCA) to reduce the dimensionality of each spectral technique prior to fusion. However, this approach produced poor results. This was likely caused by the loss of spectral physical meaning: this type of data fusion can be very useful for classification tasks, but in our case, the resulting fused data no longer represented an interpretable spectral function, which is essential for effective application of the unmixing algorithm. However, the PCA analysis with two Principal Components has been used as a mean of data visualization for supporting the hypothesis, with the aim of observing the difference between the information provided by each technique.

2.4. Spectral unmixing

Spectral unmixing is a procedure used to estimate the concentrations of the constituents of a spectrum resulting from the mixing of multiple known pure materials. In our work, the objective is to estimate the concentration for each of the individual painting materials to be present in the painting mixtures, which in our case are the 33 endmembers described in Section 2.1.

Although non-linear models such as the subtractive model have been shown to be more effective in unmixing problems with HSI reflectance spectra [15,17], preliminary tests with DRIFTS data and fused spectra showed that better results are obtained with the linear additive model. In this model, it is assumed that the resulting spectrum will be a linear combination of the spectra of the different endmembers:

$$x = \sum_{i=1}^{q} \rho_i \alpha_i \tag{1}$$

where x is the reflectance of the mixture, q is the total number of endmembers, ρ_i is the reflectance of endmember i and α_i represents its concentration. We will then seek the linear combination of endmembers that most closely approximates the problem spectrum of the mixture whose painting components we want to identify, according to the metric used, which will be selected as the best performing of those described in Section 2.4.2.

For this search, the MATLAB R2023a optimization function *fmincon* was used with two conditions: non-negativity ($\alpha_i > 0, \forall i$), which assumes that no component can be found in a negative proportion in the mixture, and sum to one ($\sum_{i=1}^{q} \alpha_i = 1$), since the sum of all concentrations cannot exceed or fall below 100%. This last condition means assuming that all real constituents are found within the endmembers library, which in the case of mock-ups we know is true, but cannot be certain in historical documents.

2.4.1. Unmixing evaluation

Two metrics were used for evaluating the unmixing performance:

First, the number of painting components correctly identified by the algorithm was computed. For this, only the two (for triphasic mixtures) or three (for quadriphasic mixtures) endmembers with the highest unmixing-predicted concentrations, independently of their binder, were compared with the real constituents of the mixtures. Hence, each of those highest-predicted endmembers that matched one of the actual constituents would count as a correct guess. In the case of binders, it will be taken as the one present in the endmember with the highest predicted concentration, since it is the one with the greatest contribution to the reconstructed spectrum.

The evaluation of the concentrations of the endmembers was performed by calculating the Root Mean Squared Error (RMSE = $\sqrt{\text{MSE}}$, see Eq. (3)) between the unmixing-predicted concentrations and the real proportions used for making the mixtures. As the triphasic and quadriphasic mixtures were produced by mixing in equal parts the biphasic mixtures that we take as endmembers, a concentration of 1/2 was expected for each endmember in the case of triphasic mixtures (two pigments + binder), and 1/3 for the quadriphasic ones (three pigments + binder). Those endmembers that were not present in the mixture were considered to have a concentration equal to zero.

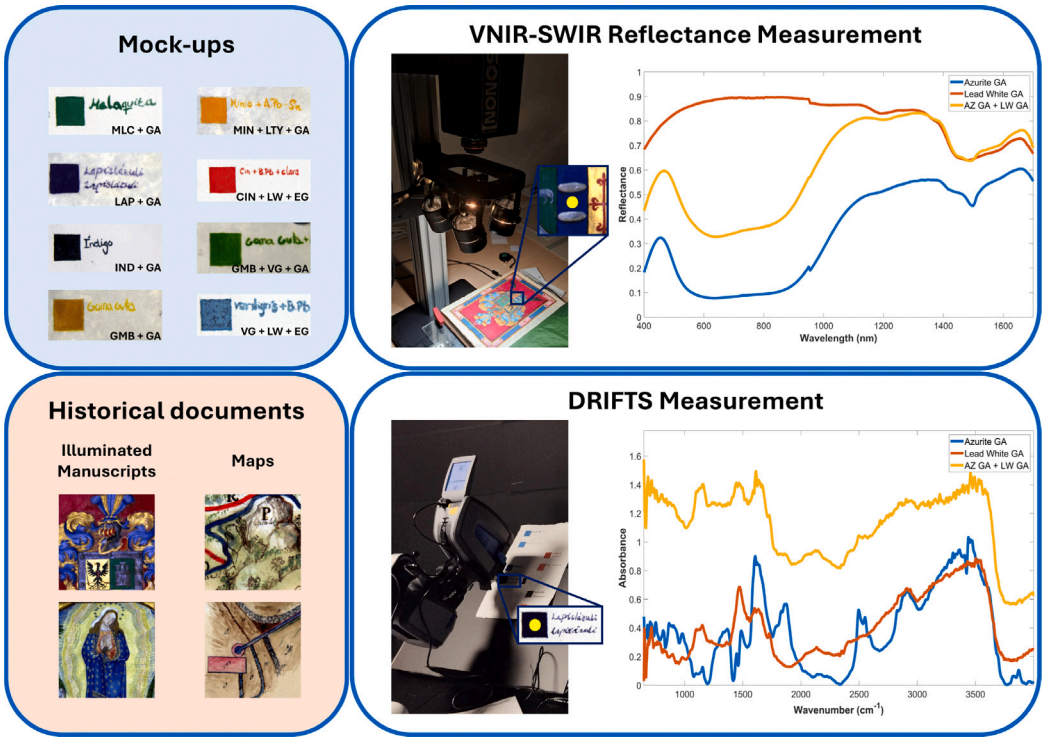

Fig. 1. Materials used for obtaining the HSI and DRIFTS spectra. Top left: examples of the endmembers (left column) and triphasic and quadriphasic mixtures (right column). Bottom left: examples of historical documents, illuminated manuscripts (left column) and maps on cotton paper (right column). Measurement configurations for HSI (top right) and DRIFTS (bottom right), both with example spectra of azurite bound with GA (blue), lead white bound with GA (orange) and their mixture (yellow).

2.4.2. Spectral metrics

Different spectral metrics can be used to compare the reconstructed spectrum with the one measured from the mixture, so five widely used spectral comparison metrics [30] were tested to verify which one was the most appropriate to work with our data. All of them were computed as measurements of difference between the reconstructed spectra x, and the reference spectra $\hat{x}$, both along N bands, of the mixture to be unmixed.

The unmixing algorithm described above will be applied to the complete set of 42 mixtures, using each of the metrics as a comparison between the reconstructed spectrum and that of the mixture. Once the results are obtained, the evaluation described in Section 2.4.1 will be carried out, and the metric with the best performance will be used for the optimization of the model.

Complement of the Goodness-of-Fit Coefficient (cGFC). The complement of the Goodness-of-Fit Coefficient (cGFC) is a similarity measure representing the cosine of the angle between the two vectors [31]:

$$\mathrm{cGFC}(\mathbf{x},\hat{\mathbf{x}}) = 1 - \frac{\sum_{i=1}^{N} x_i \hat{x}_i}{\left(\sum_{i=1}^{N} x_i^2\right)^{1/2}\left(\sum_{i=1}^{N} \hat{x}_i^2\right)^{1/2}} \tag{2}$$

with values closer to zero indicating greater similarity between the spectra.

Mean Squared Error (MSE). The Mean Squared Error (MSE) is one of the most widely used error metrics in regression and signal processing. It quantifies the average of the squared differences between corresponding elements of two vectors. It is defined as:

$$\mathrm{MSE}(\mathbf{x},\hat{\mathbf{x}}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{3}$$

where the lowest values are given for the most similar spectra. MSE is sensitive to both amplitude and shape discrepancies. This should be taken into account, as differences in signal intensity would give high MSE values even though the spectra might in fact be similar.

Spectral Information Divergence (SID). Spectral Information Divergence (SID) [32] is a statistical measure derived from the Kullback–Leibler divergence, and it treats spectra as probability distributions by converting them to normalized positive values.

$$\mathrm{SID}(\mathbf{x},\hat{\mathbf{x}}) = \sum_{i=1}^{N} p_i \log\left(\frac{p_i}{q_i}\right) + \sum_{i=1}^{N} q_i \log\left(\frac{q_i}{p_i}\right) \tag{4}$$

where $p_i = \frac{x_i}{\sum_{i=1}^{N} x_i}$ and $q_i = \frac{\hat{x}_i}{\sum_{i=1}^{N} \hat{x}_i}$. SID is advantageous when discriminating spectra with similar global shape but differing local features. However, it may be sensitive to noise or very small intensity values due to its logarithmic nature.

Combined SAM and SID (SAMSID). A fifth metric, defined as a combination of the Spectral Angle Mapper (SAM) and SID [33] was tested. SAM is an angular similarity metric that interprets spectra as vectors in a high-dimensional space and computes the angle between them [34]. It is defined as:

$$\mathrm{SAM}(\mathbf{x},\hat{\mathbf{x}}) = cos^{-1}\left(\frac{\sum_{i=1}^{N} x_i \hat{x}_i}{\left(\sum_{i=1}^{N} x_i^2\right)^{1/2}\left(\sum_{i=1}^{N} \hat{x}_i^2\right)^{1/2}}\right) \tag{5}$$

which can also be interpreted as the arccosine of GFC. SAMSID is defined as:

$$\mathrm{SAMSID} = sin(\mathrm{SAM}) \cdot \mathrm{SID} \tag{6}$$

SAMSID can be useful in cases where small deviations in spectral shape need to be captured. However, it presents a greater computational cost and is more complex to interpret quantitatively.

2.4.3. Unmixing optimization

Once the most appropriate metric has been selected, according to the evaluation procedure described in Section 2.4.1, it is important to attempt improving the results by applying a pre-processing to the spectra before running the unmixing algorithm. This pre-processing to optimize the final result consists of three main steps: derivative computation, edge cropping, and normalization. As this preprocessing considers a number of variables that can be arbitrarily chosen, an optimization phase is performed to find which combination of them gives the best results.

The first step consisted in computing the first derivative of each spectrum using the Savitzky–Golay (SG) filter, a pre-processing proposed in [12]. This derivative will highlight local spectral changes, such as peak slopes and inflection points, which are often more informative than raw intensity values when distinguishing between painting materials. The SG filter was applied using two parameters for each technique: the polynomial degree, which defines the order of the polynomial used to fit the data, and the frame size, which specifies the number of neighboring points considered in each local fit.

The edge cropping step was introduced to address the boundary artifacts introduced by the derivative calculation at the edges of the spectrum, and also to eliminate non-relevant edge bands that did not contain characteristic information of the analyzed painting materials. Two additional parameters were defined here for each technique, which determine the number of bands to crop at the beginning and at the end of each spectrum.

Finally, the cropped derivatives were normalized, dividing by the largest value in absolute value to ensure that they are between −1 and 1. This was added in order to make all the spectra comparable, so that the intensity of the signal would not be relevant, instead focusing on spectral variations.

Overall, the optimization for the data fusion spectra consisted of 8 parameters, with their minimum and maximum possible values shown in brackets:

- DRIFTS polynomial degree for SG filter (1–4).
- DRIFTS window size for SG filter (5–155).
- VNIR–SWIR reflectance polynomial degree for SG filter (1–4).
- VNIR–SWIR reflectance window size for SG filter (5–155).
- DRIFTS bands cropped at the beginning of the spectra (0–400).
- DRIFTS bands cropped at the end of the spectra (0–400).
- VNIR–SWIR reflectance bands cropped at the beginning of the spectra (0–20).
- VNIR–SWIR reflectance bands cropped at the end of the spectra (0–20).

These parameters were optimized using the surrogateopt function of MATLAB R2023a, which was chosen due to its effectiveness in low-dimensional problems, since in this problem we have a small number of variables. Additionally, unlike gradient-based solvers, surrogate optimization can explore the global landscape more effectively, avoiding getting trapped in local minima as easily.

5000 iterations were chosen for optimization, and the same seed was fixed in all cases to avoid randomness-related variations between experiments. To avoid the bias of optimizing the same spectra we are evaluating, the mixtures were randomly distributed into four approximately equal groups: sets 1 and 2 contained 10 mixtures, while sets 3 and 4 contained 11. With this division, k-fold cross-validation with $k = 4$ was performed to ensure consistency of results: on each experiment, one of the sets was used to optimize the parameters (train), while the other three were used for evaluation (test).

This optimization process was performed independently for the three types of spectra: DRIFTS (4 parameters), VNIR–SWIR reflectance (4 parameters) and data fusion of both (8 parameters). All three cases were computed in order to directly compare the best result for each of them, so it could be assessed if data fusion performed better than both techniques separately.

Fig. 2 illustrates the workflow for the unmixing procedure for a historical document spectrum, including pre-processing and data fusion. The gap between the HSI and DRIFTS spectra in the central part of Fig. 2 arises only from the visual representation of the spectra, since it has no influence on the final result. Both spectra are continuously concatenated and no wavelength information is used.

2.4.4. Method validation

Additionally, OMNIC software was used in order to assess the efficiency of the proposed method on both mock-ups and historical documents, as it remains as a widely used software for spectra processing [35,36] and pigment recognition [37–39]. For this purpose, the DRIFTS spectra of the mixtures was compared with a reference library containing the DRIFTS spectra of the endmembers, evaluating the correct guesses in the same way as unmixing: by comparing the actual painting components with the two or three most similar spectra of the library.

The best optimization was also tested in the case of historical documents. In this case, the XRF spectra obtained were analyzed to check which elements are present in the different mixtures. By using the identified elements and taking the color of the mixture into account, a hypothesis was made regarding the possible painting materials present in the measured points, and this is compared with the results of the best optimization for unmixing with data fusion.

In this case, the three endmembers with the highest predicted concentration are taken into account, and if these results match the hypothesis or offers another valid alternative (for example, by proposing the presence of a dye, undetectable by XRF, that matches the color of the sample), our model is considered to have given a reasonable result and it is counted as a success.

3. Results and discussion

3.1. Principal component analysis

Principal Component Analysis (PCA) was applied to the fused spectra as an exploratory tool to assess spectral variance across the dataset and to examine the distribution of endmembers in reduced-dimensional space.

Fig. 3 (a) shows the loadings for all wavelengths in the fused spectra. PCA loadings represent how much each original variable, wavelengths in our case, contributes to a principal component in PCA, as they indicate the weight or influence that each variable has in defining the direction of a principal component. In PC1, which accounted for 63% of the total variance, a clear separation was observed: VNIR–SWIR reflectance wavelengths were grouped on the left side of the loading plot, while DRIFTS wavelengths were on the right. This indicates that the primary axis of variance in the data arises from differences between the two spectral techniques, suggesting that each contributes with different information to the set. PC2, which explained 13% of the variance, showed both VNIR–SWIR reflectance and DRIFTS wavelengths close to each other, with HSI contributing slightly more positively and DRIFTS slightly more negatively. The fact that both wavelength curves fold back on themselves indicates that there is redundant information within each technique, but the clear separation between the two in PC1 reinforces the hypothesis that data fusion highlights the differences between painting materials even more than both techniques separately.

The scores scatter plot for all of the endmembers is shown in Fig. 3 (b). As PC1 seems to differentiate samples according to the spectral techniques, samples with large negative PC1 values, like orpiment-EG, cinnabar-EG and minium-EG, are spectrally dominated by VNIR–SWIR reflectance wavelengths. Conversely, samples with high positive PC1 values, such as burnt umber-EG and verdigris with both binders (GA or EG), show stronger influence from DRIFTS features, likely due to prominent vibrational bands in the mid-IR region.

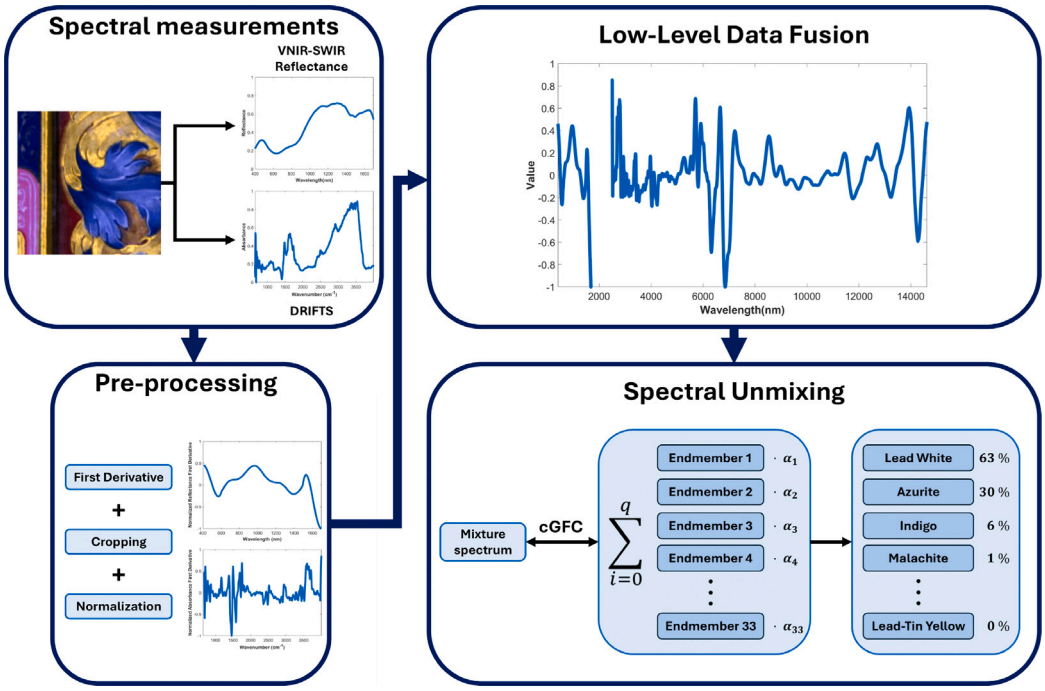

Fig. 2. Summarized workflow for the unmixing procedure using fused spectral data.

Table 1
Performance comparison using different spectral metrics for data fusion, DRIFTS and HSI VNIR–SWIR reflectance.

Metric	Fusion		DRIFTS		HSI	
	Correct endmembers	RMSE	Correct endmembers	RMSE	Correct endmembers	RMSE
cGFC	63.64%	0.0948	48.86%	0.1067	50.00%	0.1129
MSE	61.36%	0.0955	46.59%	0.1117	57.95%	0.1105
SID	40.91%	0.1119	31.82%	0.1238	57.95%	0.1159
SAMSID	54.55%	0.0969	42.05%	0.1158	48.86%	0.1173

Differences in binder are very prominent in this figure: in several cases, the sample of a pigment bound with GA separates clearly along PC1 and PC2 from the same pigment bound with EG, such as azurite, verdigris and lapislazuli, suggesting that the data fusion spectra is sensitive to binder composition.

3.2. Metric selection

The performance of five spectral metrics (cGFC, MSE, SID, and SAMSID) was evaluated using the full dataset across the three spectral modalities: HSI VNIR–SWIR reflectance, DRIFTS, and their fusion, using the full spectra without any pre-processing. Table 1 summarizes the percentage of correctly identified endmembers and proportions RMSE values for each technique and metric:

Among all metrics, the complement of the Goodness-of-Fit Coefficient (cGFC) provided the best results for the fused spectra, achieving 63.64% correct painting component identification and a minimum RMSE of 0.0948.

SID showed the weakest performance in data fusion and DRIFTS, with accuracy dropping to 31.82% for the latter, but it was one of the best performing metric with VNIR–SWIR reflectance spectra, along MSE. SAMSID demonstrated intermediate behavior for fusion and DRIFTS, outperforming SID but not reaching the accuracy of cGFC.

In general, when using the DRIFTS and VNIR–SWIR reflectance spectra individually, a better performance was observed in the latter in all metrics, with a great difference in SID. However, even in the pre-optimization step, data fusion provides better results for all metrics than DRIFTS and VNIR–SWIR reflectance separately. The only exception is SID, where fusion presents a worse result than VNIR–SWIR reflectance in terms of correctly guessed endmembers, but a better RMSE. This reinforces the hypothesis proposed in this work and demonstrates that, with the concatenation of spectra, both techniques complement each other by providing different information for painting component identification.

Although the best results have been obtained for data fusion unmixing using cGFC, a success rate of 63.64% can still be considered low, as more than one third of the painting components are not correctly identified. Therefore, the optimization phase will be of great importance to improve this result and try to achieve a more accurate model.

These results indicate that cGFC is not only more effective in selecting correct endmembers but also more accurate in predicting concentrations, especially when combining the information from both techniques. Therefore, from now on, only this metric will be used for optimization and application on historical documents.

3.3. Unmixing optimization results

Once the metric for unmixing has been chosen, the next step is to perform the cross-validation with the four previously defined training sets (see Table 2).

After optimization, data fusion newly showed a greater performance in all sets than the two techniques individually, with an average of 75.39% correct endmembers, 96.02% correct binder and an RMSE of 0.0804. DRIFTS had the second best result (59.44%, RMSE = 0.0976), followed by VNIR–SWIR reflectance (53.41%, RMSE = 0.1152).

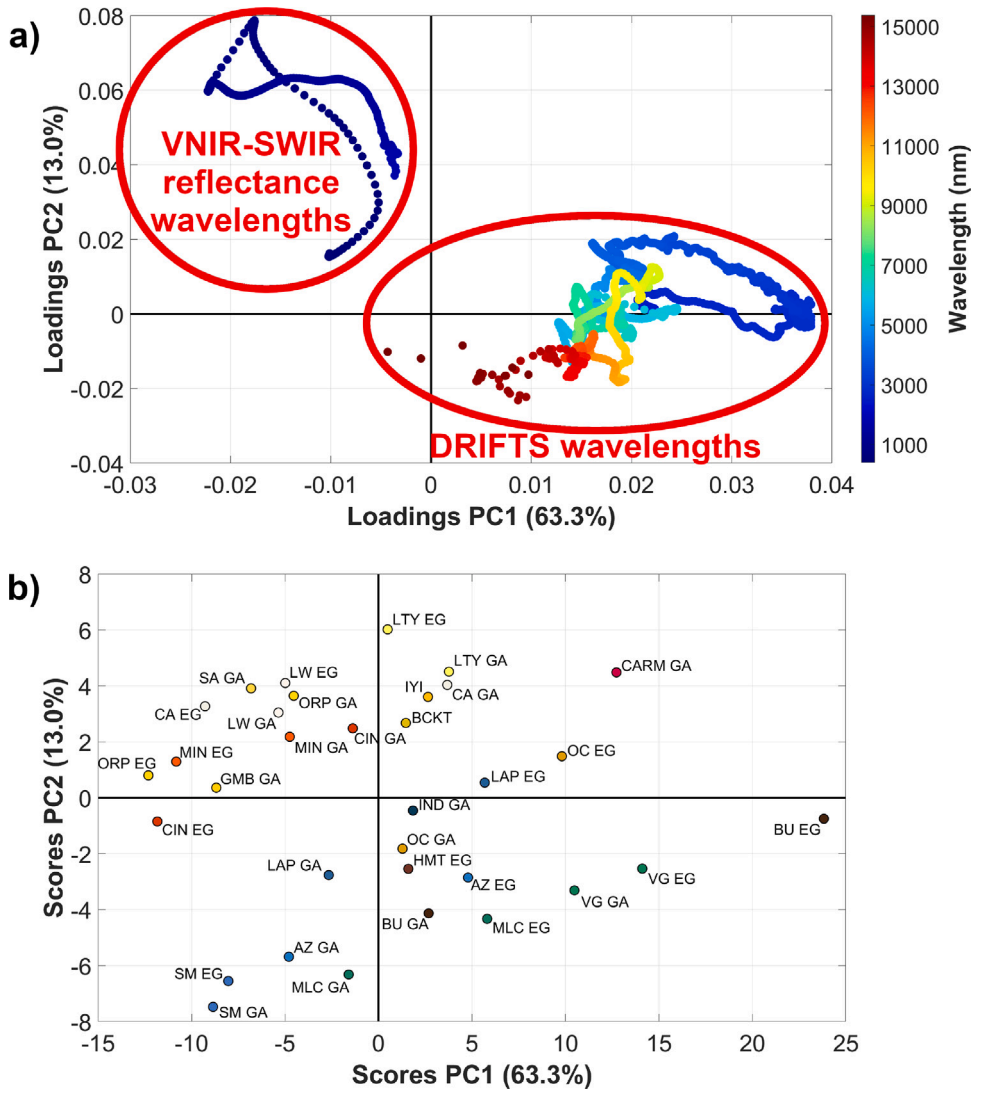

Fig. 3. PCA Loadings (a) and Scores (b) for the data fusion spectra of the endmembers (biphasic mixtures and buckthorn).

Table 2
Cross-validation for the unmixing performance after optimization for each set and technique.

Spectra	Train set	Correct endmembers	Correct binder	RMSE	Average correct endmembers	Average correct binder	Average RMSE
Fusion	1	74.63%	96.88%	0.0809	75.39%	96.02%	0.0804
	2	74.63%	96.88%	0.0775			
	3	75.76%	96.77%	0.0788			
	4	76.56%	93.55%	0.0844			
DRIFTS	1	59.70%	93.75%	0.0949	59.44%	92.87%	0.0976
	2	61.19%	90.63%	0.1009			
	3	60.61%	93.55%	0.0957			
	4	56.25%	93.55%	0.0987			
HSI	1	56.72%	81.25%	0.1074	53.41%	74.60%	0.1152
	2	52.24%	68.75%	0.1210			
	3	50.00%	83.87%	0.1141			
	4	54.69%	64.52%	0.1183			

Preprocessing of the data has also shown an improvement in the unmixing results in all cases, being less influential in the case of VNIR–SWIR reflectance-only spectra. This can be explained by the larger variations in the DRIFTS spectrum compared to the VNIR–SWIR reflectance spectrum, with a large number of peaks that provide relevant information about the painting material, so that calculating the first derivative has a greater effect in this technique. The optimization increased the proportion of correct painting components in data fusion by twelve percentage points, and caused a 15% decrease in the RMSE value of the concentrations.

Data fusion has also shown to have lower variability between sets than the two techniques separately, both in terms of painting components and binders, with a standard deviation of 0.8 versus 2.5 in the case of DRIFTS and 1.9 in the case of HSI.

Table 3
OMNIC results and data fusion unmixing results using set 3 as the training set on paper support. Green: All correct components. Red: No correct components. Yellow: One correct component (triphasic mixtures). Orange: One correct component (quadriphasic mixtures). Blue: two correct components (quadriphasic mixtures).

Mixture		Fusion		OMNIC	
Pigments/Dyes	**Binder**	**Endmembers**	**Binder**	**Endmembers**	**Binder**
Azurite Lead White	GA	LW GA AZ EG	GA	SM EG CIN GA	EG
Azurite Lead White	EG	AZ EG LW EG	EG	CARM GA CIN GA	GA
Carmine Gamboge	GA	GMB GA IYI	GA	SA GA MIN EG	GA
Carmine Lead White	GA	CARM GA LW EG	GA	IND GA LTY GA	GA
Cinnabar Lead White	GA	LW GA CIN GA	GA	LW GA LW EG	GA
Indigo Orpiment	GA	IND GA CIN GA	GA	LW GA ORP EG	GA
Malachite Lead White	GA	MLC GA LW EG	GA	MLC GA MLC EG	GA
Malachite Lead White	EG	LW EG MLC GA	EG	LW EG ORP EG	EG
Verdigris Lead White	GA	LW GA VG GA	GA	LW GA LAP EG	GA
Verdigris Lead White	EG	LW EG VG GA	EG	IND GA OC EG	GA
Cinnabar Lead Tin Yellow Lead White	GA	LW GA LW EG LTY GA	GA	CA EG ORP EG SM GA	EG
Cinnabar Lead Tin Yellow Lead White	EG	CIN EG MIN EG LW EG	EG	ORP EG LW GA MIN EG	EG

To detail the results, Tables 3 and 4 show a comparison between OMNIC and data fusion results using set 3 as the training set, as this was the case with performance closest to the average.

For samples deposited on paper support, unmixing with data fusion chooses a correct constituent of the mixture as the endmember with the highest concentration in all cases. In most of the triphasic samples, the error in the second endmember with the highest proportion predicted occurs only in the binder, so they are considered as a correct guess, as only the pigment or dye is taken into account for the second endmember. Furthermore, in all samples of this subset, the binder of the endmember with the highest predicted concentration is the correct one present in the triphasic or quadriphasic mixture for the fused spectra (100%), while OMNIC identifies the binder in two thirds of the cases. For the paper support, the proposed method using data fusion correctly identifies 84.6% of the endmembers, while OMNIC correctly identifies only 23.1%.

In the case of samples deposited on parchment, a similar behavior is observed. Again, in all cases, the endmember with the highest predicted concentration is in fact one of the pigments or dyes actually used to make the mixture. Data fusion correctly identifies 70% of endmembers, while OMNIC does so for only 20%. As for the binders, in this case the data fusion approach does well in almost all cases (94.7%), as it only fails for the carmine and lead white mixture bound with gum arabic. As for OMNIC, the binder is correctly identified in almost half of the mixtures (47.4%).

Confusion between painting materials similar in color (lead tin yellow for imperial yellow ink or gamboge, all yellow) or composition (lead tin yellow for lead white or minium, all containing lead) is commonly observed, which is an expected behavior when using two techniques where specific bands may have been masked in the case of complex mixtures.

OMNIC presents considerable problems for the automated identification of the painting components, probably because it is using the DRIFTS spectrum and does not take into account the color of the samples, although it does not perform too badly in the identification of the binder.

In the majority of quadriphasic samples, two of the endmembers are correctly identified, but again, it is common to find that confusion has occurred between lead-tin yellow and minium.

As for the proportions of endmembers, the first usually has a predicted concentration around 40%, and the second around 30%, with the remaining 30% distributed among the rest of the candidates. The case of the mixture of malachite and lead white bound with egg glair on paper stands out, in which the concentration of lead white is predicted to be above 99%, indicating the dominance of lead white in the spectrum of the mixtures in which it is present.

In light of our results it is observed that, for the mock-ups, when using data fusion for unmixing, the endmember with the highest predicted concentration corresponds to one of the actual painting components of the mixture, offering a good result for binder recognition, which is in line with the conclusions of the PCA analysis.

3.4. Unmixing validation in historical documents

A total of 64 points between the manuscripts on parchment and the maps were chosen for the validation of the proposed method on real documents. From these measurements, Fig. 4 shows four examples, chosen in such a way that two of them correspond to good results and two of them to bad results.

Fig. 4 shows four examples of XRF spectra captured in illuminated manuscripts from the Archive of the Royal Chancellery of Granada. The spot in Fig. 4(a) was captured on the hand of the drawn character. Its XRF spectrum shows a high presence of lead (peaks Pb α, Pb Lβ and Pb Lγ) and copper (peaks Cu Kα and Cu Kβ). Given the color of the hand, it can be assumed that it is a mixture of lead white with another painting material, possibly a dye, undetectable with XRF. The copper detected could be due to the green clothes painted around or beneath the hand, that may have malachite or verdigris, and is very close to the point of measurement. Unmixing with data fusion spectra identifies this painting material as a mixture of lead white (50%) and gamboge (30%), with the rest of endmembers scoring a concentration below 10%.

The spot in Fig. 4(b) was taken on a blue feather from a coat of arms. XRF analysis shows a high presence of lead (peaks Pb M α, Pb

Table 4
OMNIC results and data fusion unmixing results using set 3 as the training set on parchment support. Green: All correct components. Red: No correct components. Yellow: One correct component (triphasic mixtures). Orange: One correct component (quadriphasic mixtures). Blue: two correct components (quadriphasic mixtures).

Mixture		Fusion		OMNIC	
Pigments/Dyes	**Binder**	**Endmembers**	**Binder**	**Endmembers**	**Binder**
Azurite Gamboge	GA	AZ GA GMB GA	GA	BU EG CARM GA	EG
Azurite Lead White	GA	LW GA OC GA	GA	BU EG OC EG	EG
Azurite Lead White	EG	AZ EG LW EG	EG	AZ GA CA EG	GA
Carmine Gamboge	GA	GMB GA IYI	GA	VG GA LTY GA	GA
Carmine Lead White	GA	LW EG IYI	EG	HMT GA GMB GA	GA
Cinnabar Lead White	GA	CIN GA LW GA	GA	VG GA LW GA	GA
Cinnabar Lead White	EG	CIN EG LW EG	EG	CARM GA ORP EG	GA
Gamboge Malachite	GA	GMB GA MLC GA	GA	BU EG CIN GA	EG
Gamboge Verdigris	GA	VG GA GMB GA	GA	CIN GA LAP GA	GA
Indigo Orpiment	GA	IND GA LAP GA	GA	LAP GA LAP EG	GA
Lead Tin Yellow Lead White	GA	LW GA IYI	GA	LTY EG AZ EG	EG
Malachite Lead White	EG	LW EG MLC EG	EG	GMB GA SM GA	GA
Minium Lead Tin Yellow	GA	MIN GA GMB GA	GA	LTY EG LAP EG	EG
Minium Lead Tin Yellow	EG	MIN EG ORP EG	EG	HMT EG IYI	EG
Saffron Gamboge	GA	GMB GA IYI	GA	BU GA ORP GA	GA
Verdigris Lead White	GA	VG GA LTY EG	GA	VG GA LW GA	GA
Verdigris Lead White	EG	LW EG VG GA	EG	CIN EG HMT EG	EG
Cinnabar Lead Tin Yellow Lead White	GA	LW GA GMB GA IYI	GA	AZ EG ORP GA BU GA	EG
Cinnabar Lead Tin Yellow Lead White	EG	MIN EG LW EG CIN EG	EG	GMB GA LTY EG CIN EG	GA

L α, Pb Lβ and Pb Lγ) and copper (peaks Cu Kα and Cu Kβ), together with some peaks of gold that may come from the surrounding golden details. The copper suggests the presence of azurite, which is consistent with the blue color of the sample, which may be mixed with lead white. The proposed method identifies the painting material as a mixture of lead white (63%) and azurite (30%), which is in line with the color of the sample and the XRF results.

Fig. 4(c) shows another carnation on a hand. In this case, in the XRF spectrum, the lead peaks stand out, suggesting the presence of lead white. Some peaks from gold are observed, probably due to the surrounding golden details. However, in this case the unmixing with the data fusion spectrum does not detect the presence of lead white, but identifies it as a mixture of equal parts of saffron, imperial yellow ink and cinnabar, which does not agree with XRF results, showing the limitations of the proposed model. This could be due to the presence of a colorant in the fleshings, such as carmine, whose identification is usually complex by these methods and whose proportion compared to lead white would be minimal. In addition, it is also possible that the interaction of the lead white with the binder makes its identification more difficult.

Finally, the brown feather in Fig. 4(d) shows a high presence of calcium and iron, according to XRF spectra. The latter may be due to the iron oxide present in the brown ochre pigment, and the calcium might indicate the occurrence of calcite. Indeed, brownish and umber pigments belong to the so-called earthy pigments which are polymineral natural mixtures colored with free iron oxides, where clay minerals and calcite can be present [40]. According to unmixing, it is a mixture of equal parts of minium or cinnabar, which does not agree with the XRF spectrum, since this technique would detect the presence of lead (minium, (Pb_3O_4) and mercury (cinnabar, HgS).

Considering the complete set of measurements in historical documents, unmixing with data fusion correctly identifies at least one painting component among the three endmembers with the highest predicted concentration in 65% of the cases. In general, the proposed method easily detects the presence of azurite, lead-tin yellow, cinnabar and malachite or verdigris. Lead white is often correctly identified in mixtures with other pigments, but present difficulties when mixed with dyes. The model presents problems with brown samples, that could contain many painting components. It also struggles in cases where paint layers are very thin, where limitations are mainly caused by the detection limit of the instruments and the high influence of the support on the DRIFTS spectra, as studied in [21].

In summary, for historical illuminated manuscripts, the proposed method can be of great help for painting materials characterization, especially when considering the possible presence of organic dyes. In this way, the results of automatic unmixing assists confirming hypotheses about the presence of certain pigments and dyes, as observed in the cases of Fig. 4(a) and (b).

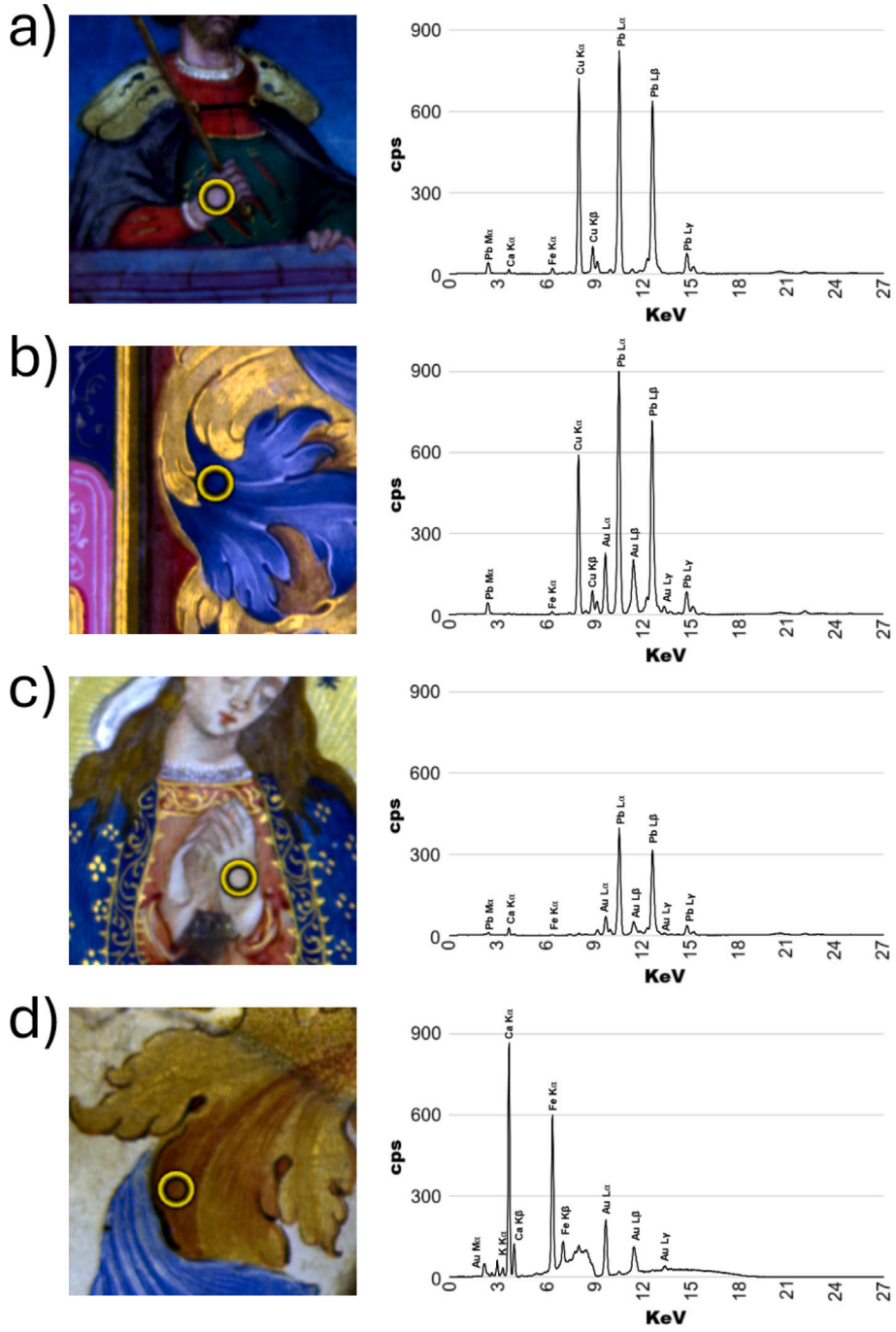

Fig. 4. Examples of XRF spectra in historical documents, all of them with parchment support.

4. Conclusions

This study presents a methodology for non-invasive identification of painting materials in historical documents and mock-ups, including pigments, dyes, binders and its mixtures using spectral unmixing and data fusion. By combining Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS) and Hyperspectral imaging (HSI), we demonstrate that low-level data fusion significantly outperforms individual techniques in both accuracy and concentration prediction.

As the first step for the proposed unmixing pipeline, the complement of the Goodness-of-Fit Coefficient (cGFC) was chosen as the comparison metric. It gave the best results for all modalities of spectra, achieving the highest percentage of correctly identified painting components and the lowest RMSE in predicted concentrations.

A pre-processing step was proposed, including the computing of Savitzky–Golay derivative, edge cropping, and normalization. The optimization of this step further enhanced the unmixing results, particularly for fused data, correctly choosing one of the real painting components of the sample as the highest concentration predicted endmember in almost all mock-up cases, which also led to correct binder identification.

In the complete set of analyzed historical documents, of which some examples were presented in Fig. 4, the proposed method was able to correctly identify at least one true painting component in 65% of the cases, often detecting key pigments and dyes such as lead white, azurite, malachite, and cinnabar. However, it is likely that these results could be improved if artificially aged samples were used to constitute

the endmember library, since the degradation of the different painting materials over time can change their spectral response [41].

However, this correct identification rate for historical documents is not only the result of pigment alteration and material aging, but also reflects the limitations of DRIFTS and HSI. Both techniques can struggle with weak signals and overlapping spectral features, which, in addition to the presence of complex mixtures, may reduce identification accuracy. Furthermore, it is also important to consider the limitations of the sensors that were used. Increasing spectral resolution and expanding the sensitivity limits to higher infrared wavelengths may improve the identification of painting components. This 65% rate should be regarded as a realistic outcome that highlights the challenges of applying these non-invasive techniques to historical documents.

It should be noted that DRIFTS and HSI are surface-sensitive techniques with shallow penetration depths, and their spectra can be influenced by the optical and chemical properties of the support material [21] in the case of thin painting layers. In contrast, XRF penetrates more deeply into the sample and provides elemental information, which is less affected by surface heterogeneity. Therefore, the three techniques should be interpreted as complementary: by combining the proposed unmixing method with fused spectra and XRF elemental analysis, hypotheses about the painting materials present in historical documents can be proposed and confirmed more reliably than by using the three techniques separately.

An unavoidable limitation of this type of study is the fact that the endmember library, although extensive, does not include all pigments, dyes and binders present in historical illuminated manuscripts. The development of new datasets with spectral information of painted objects from different time periods and places is essential to facilitate the work of conservators and restorers, providing them with more precise information on what materials may be present in the documents in a non-invasive way.

Future work should focus on expanding the endmember library to include a broader range of historical painting components, including the presence of laquers and aged pigment spectra, either by artificial aging or by applying mathematical aging models to existing endmember spectra. It is also important to explore unmixing models that incorporate non-linearities or spatial constraints, which could further improve accuracy. Integration of spatial–spectral data from hyperspectral imaging (HSI) within the fused spectra may also provide new insights, particularly in the mapping of painting components across the manuscript pages. In addition, fusion with other techniques, such as Raman spectroscopy, or a tandem between X-ray diffraction and X-ray fluorescence, can be explored to increase the amount of independent information that can provide more characteristic signatures of each painting component.

CRediT authorship contribution statement

Francisco Moronta-Montero: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Anna Sofia Reichert:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ana Belén López-Baldomero:** Writing – review & editing, Software, Methodology, Investigation, Data curation, Conceptualization. **María Rosario Blanc:** Writing – review & editing, Visualization, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Carolina Cardell:** Writing – review & editing, Supervision, Resources, Conceptualization. **Ana López-Montes:** Writing – review & editing. **Javier Hernández-Andrés:** Writing – review & editing. **Eva M. Valero:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work supported by grant PID2021-124446NB-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU; grant PRE2022-101352 funded by MICIU/AEI/10.13039/501100011033 and “ESF+”; and grant FPU2020-05532 funded by Ministry of Universities (Spain). This work was also supported by the Andalusian Research Group RNM-179.

Funding for open access charge: Universidad de Granada / CBUA.

The authors wish to thank David Torres Ibañez, archivist at the Archive of the Royal Chancellery of Granada (Spain).

Data availability

Data will be made available on request.

References

[1] L. Burgio, Pigments, dyes and inks: their analysis on manuscripts, scrolls papyri, Archaeol. Anthropol. Sci. 13 (2021) 194.
[2] M. Melo, P. Nabais, M. Guimarães da Silva, R. Araújo, R. Castro, M. Oliveira, I. Whitworth, Organic dyes in illuminated manuscripts: A unique cultural and historic record, Philos. Trans. R. Soc. Math. Phys. Eng. Sci. 374 (2016) 20160050.
[3] C. Cucci, A. Casini, Hyperspectral Imaging for Artworks Investigation, in: Data Handling in Science and Technology, 32, Elsevier, 2019, pp. 583–604.
[4] F. Daniel, A. Mounier, J. Pérez-Arantegui, C. Pardos, N. Prieto-Taboada, S. Fdez-Ortiz de Vallejuelo, K. Castro, Hyperspectral imaging applied to the analysis of Goya paintings in the Museum of Zaragoza (Spain), Microchem. J. 126 (2016) 113–120.
[5] D. Buti, F. Rosi, B. Brunetti, C. Miliani, In-situ identification of copper-based green pigments on paintings and manuscripts by reflection FTIR, Anal. Bioanal. Chem. 405 (2013) 2699–2711.
[6] E. Tomasini, G. Siracusano, M. Maier, Spectroscopic, morphological and chemical characterization of historic pigments based on carbon. Paths for the identification of an artistic pigment, Microchem. J. 102 (2012) 28–37.
[7] G. Capobianco, M.P. Bracciale, D. Sali, F. Sbardella, P. Belloni, G. Bonifazi, S. Serranti, M.L. Santarelli, M. Cestelli Guidi, Chemometrics approach to FT-IR hyperspectral imaging analysis of degradation products in artwork cross-section, Microchem. J. 132 (2017) 69–76, http://dx.doi.org/10.1016/j.microc.2017.01.007.
[8] K. Samanian, Identification of green pigment used in Persian wall paintings (ad 1501–1736) using PLM, FT–IR, SEM/EDX and GC–MS techniques, Archaeometry 57 (4) (2015) 740–758.
[9] I. Maybury, D. Howell, M. Terras, H. Viles, Comparing the effectiveness of hyperspectral imaging and Raman spectroscopy: a case study on armenian manuscripts, Herit. Sci. 6 (2018) 42, http://dx.doi.org/10.1186/s40494-018-0206-1.
[10] C. Balas, G. Epitropou, A. Tsapras, N. Hadjinicolaou, Hyperspectral imaging and spectral classification for pigment identification and mapping in paintings by El Greco and his workshop, Multimedia Tools Appl. 77 (2018) 9737–9751.
[11] A. Gestels, F. Gabrieli, T. De Kerf, F. Vanmeert, H.F. García, J. Delaney, K. Janssens, G. Steenackers, S. Vanlanduit, High-resolution compound-specific mapping in works of art via data fusion of MA-XRPD with hyperspectral data (part 1: Method evaluation), Talanta 280 (2024) 126731.
[12] C. Cevoli, E. Iaccheri, A. Fabbri, L. Ragni, Data fusion of FT-NIR spectroscopy and vis/NIR hyperspectral imaging to predict quality parameters of yellow flesh jintao kiwifruit, Biosyst. Eng. 237 (2024) 157–169.
[13] Z. Jiang, L. Zhong, J. Xue, J. Lv, F. Zhou, Y. Zhou, Y. Xu, Q. Shao, A. Zhang, Data fusion based on near-infrared spectroscopy and hyperspectral imaging technology for rapid adulteration detection of Ganoderma lucidum spore powder, Microchem. J. 193 (2023) 109190.
[14] A.L. de Queiroz Baddini, J.L.V. de Paula Santos, R.R. Tavares, L.S. de Paula, H. da Costa Araújo Filho, R.P. Freitas, PLS-DA and data fusion of visible reflectance, XRF and FTIR spectroscopy in the classification of mixed historical pigments, Spectrochim. Acta Part A: Mol. Biomol. Spectrosc. 265 (2022) 120384.
[15] F. Grillini, J.B. Thomas, S. George, Comparison of imaging models for spectral unmixing in oil painting, Sensors 21 (2021) 2471.

[16] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, J. Chanussot, Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches, IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 5 (2) (2012) 354–379.
[17] E.M. Valero, M.A. Martínez-Domingo, A.B. López-Baldomero, A. López-Montes, D. Abad-Muñoz, J.L. Vílchez-Quero, Unmixing and pigment identification using visible and short-wavelength infrared: Reflectance vs logarithm reflectance hyperspaces, J. Cult. Herit. 64 (2023) 290–300.
[18] H. Deborah, M. Ulfarsson, J. Sigurdsson, Fully constrained least squares linear spectral unmixing of the scream (verso, 1893), in: 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing, WHISPERS, 2021.
[19] N. Rohani, E. Pouyet, M. Walton, O. Cossairt, A.K. Katsaggelos, Nonlinear unmixing of hyperspectral datasets for the study of painted works of art, Angew. Chem. Int. Ed. 57 (34) (2018) 10910–10914.
[20] N. Rohani, E. Pouyet, M. Walton, O. Cossairt, A.K. Katsaggelos, Pigment unmixing of hyperspectral images of paintings using deep neural networks, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2019, pp. 3217–3221.
[21] A.S. Reichert, A.B. López-Baldomero, F. Moronta-Montero, A. López-Montes, E.M. Valero, C. Cardell, Database of diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS) and hyperspectral imaging (HSI) spectra of pigments and dyes for historical document analysis, Anal. Bioanal. Chem. (2025) 1–22.
[22] A.B. López-Baldomero, F. Moronta-Montero, M.A. Martínez-Domingo, Y. Lefier, J.L. Nieves, J. Hernández-Andrés, R. Fernández-Gualda, A.S. Reichert, A. López-Montes, T. Espejo, E.M. Valero, Exploring the HYPERDOC database: Advancing hyperspectral imaging for historical document analysis, Arch. Conf. (2025) 1–6.
[23] Resonon Inc., Resonon PikaL, https://resonon.com/Pika-L. 2025.
[24] Resonon Inc., Resonon PikaNIR, https://resonon.com/Pika-IR.
[25] Agilent Technologies Inc, 4300 handheld FTIR spectrometer, 2025, https://www.agilent.com/en/product/molecular-spectroscopy/ftir-spectroscopy/ftir-compact-portable-systems/4300-handheld-ftir.
[26] Thermo Fisher Scientific Inc., Software OMNIC™ specta, 2025, https://www.thermofisher.com/order/catalog/product/833-036200.
[27] F. Zhang, X. Tang, L. Li, Origins of baseline drift and distortion in Fourier transform spectra, Molecules 27 (2022) 4287.
[28] L. Zhang, Z. Song, S. Zuo, F. Hou, S. Chen, Precise in-situ detection of inorganic pigments in ancient architectural color paintings by HH-XRF, Herit. Sci. 11 (2023) 230.
[29] Thermo Fisher Scientific Inc., Niton™ XL3t GOLDD+ XRF analyzer, 2025, https://www.thermofisher.com/order/catalog/product/XL3TGOLDDPLUS.
[30] A.B. López-Baldomero, M.A. Martínez-Domingo, E.M. Valero, R. Fernández-Gualda, A. López-Montes, R. Blanc-García, T. Espejo, Selection of optimal spectral metrics for classification of inks in historical documents using hyperspectral imaging data, in: Optics for Arts, Architecture, and Archaeology (O3A) IX, 12620, International Society for Optics and Photonics, SPIE, 2023, 126200E.
[31] J. Romero, A. García-Beltrán, J. Hernández-Andrés, Linear bases for representation of natural and artificial illuminants, J. Opt. Soc. Amer. A 14 (5) (1997) 1007–1014.
[32] C.-I. Chang, Spectral information divergence for hyperspectral image analysis, in: IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293), vol. 1, 1999, pp. 509–511, 1.
[33] Y. Du, C.-I. Chang, H. Ren, C.-C. Chang, J. Jensen, F. D'Amico, New hyperspectral discrimination measure for spectral characterization, Opt. Eng., Bellingham 43 (2004) 1777–1786.
[34] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, A. Goetz, The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data, Remote Sens. Environ. 44 (2) (1993) 145–163.
[35] S. Vahur, A. Teearu, P. Peets, L. Joosu, I. Leito, ATR-FT-IR spectral collection of conservation materials in the extended region of 4000-80 cm–1, Anal. Bioanal. Chem. 408 (2016) 3373–3379.
[36] A. Passaretti, L. Cuvillier, C. Künzi, L. Brambilla, E. Joseph, Multi-analytical characterisation of Art Déco dinanderie : Single-point and map analysis of Jean Dunand's metal artworks, J. Raman Spectrosc. 55 (07) (2024).
[37] J. Ferrer, I. Sandu, T. Syversen, A. Cardoso, A. Candeias, C. Borca, Investigating colour changes in red and blue paints – a preliminary study of art materials and techniques in Edvard Munch's old man in Warnemünde (1907), 2019, pp. 209–218.
[38] E.G. Sandbakken, E. Chan, I.T. Flåte, H. Kutzke, L. Puskar, I.C.A. Sandu, Powdery pigments from the original materials collection of munch: Conservation treatment and analytical characterization, Int. J. Conserv. Sci. 13 (2022) 1681–1690, (Special Issue 1).
[39] M. Edreira, M. Feliu, C. Fernández-Lorenzo, J. Martín, Mérida, Spain, Talanta, Spectroscopic analysis of roman wall paintings from Casa del Mitreo in Emerita Augusta, vol. 59, (6) 2003, pp. 1117–1139.
[40] D. Hradil, T. Matys Grygar, J. Hradilová, P. Bezdicka, Clay and iron pigments in the history of painting, Appl. Clay Sci. 22 (2003) 223–236.
[41] I. Ciortan, T. Poulsson, S. George, J.Y. Hardeberg, Predicting pigment color degradation with time series models, Color. Imaging Conf. 30 (2022) 250–257.