Research article

Journal of RAMAN SPECTROSCOPY

Received: 2 October 2012 Revised: 6 February 2013 Accepted: 15 February 2013 Published online in Wiley Online Library: 3 April 2013

(wileyonlinelibrary.com) DOI 10.1002/jrs.4291

Multivariate analysis of combined Raman and fibre-optic reflectance spectra for the identification of binder materials in simulated medieval paints

Anuradha Pallipurath,[a] Jonathan Skelton,[a] Paola Ricciardi,[b] Spike Bucklow[c] and Stephen Elliott[a]*

The non-invasive identification of paint materials used in works of art is essential, both for preserving and restoring them, and also for understanding and verifying the history surrounding their creation. As such, the development of suitable non-invasive techniques has received much interest in recent years. We have investigated the use of Fourier transform (FT)-Raman spectroscopy and fibre-optic reflectance spectroscopy (FORS), together with multivariate principal-component analysis (PCA) techniques, in order to identify the pigment and binding materials used in made-up samples representative of real artwork. We demonstrate that both types of spectroscopy provide complementary information which can be used to identify the pigments and binders in paint samples. We show that PCA with FT-Raman spectra can be used to assist in the identification of oil-based binders, and that the additional data provided by FORS spectra enables PCA on combined spectra to identify more complex proteinaceous and polysaccharide-based binding media. The results presented here demonstrate that multivariate analyses of lead-based paints, using data measured by FT-Raman and FORS in conjunction, have much potential for identifying individual pigments and binders in paint samples. This provides a path towards computer-assisted characterisation of paint materials on artwork. Copyright © 2013 John Wiley & Sons, Ltd.

Supporting information may be found in the online version of this article.

Keywords: FT-Raman; FORS; PCA; paint–binder materials; non-invasive analysis

Introduction

Authentication and conservation of works of art can be carried out with a precise knowledge of the pigments and binders used in their production.[1,2] Understanding the use of specific techniques for the production of paints not only helps to date works of art but also helps in identifying the master or school that was involved in their production. Apart from being useful for complementing art-historical research, analysis of pigments and binders also helps in identifying suitable conservation treatments, and in detecting forgeries by means of the identification of a non-authentic colour palette used to produce them. A major concern during the analysis of a work of art is its preservation, and thus removal of paint samples is not ideal. A great deal of work is therefore currently being carried out in identifying the pigments and dyes used in painted artworks by both invasive and non-invasive techniques.[3–8] However, comparatively little attention has been given to the binders that hold paints together, owing mainly to the often complex nature of the biomaterials involved and the tedious experimental procedures usually required to identify them.

Binders have been analysed invasively by removing paint samples from works of art and employing a variety of analytical techniques, including proteomics,[9] gas-chromatography mass-spectrometry[10–14] and liquid chromatography.[15] Bacci *et al.* took a first step towards the non-invasive analysis of binder materials using Fourier transform-infrared (FT-IR) spectroscopy.[16] Vagnini *et al.*[17] subsequently made use of short-wave IR reflectance spectroscopy to investigate the overtones and combination bands of C–H, O–H and N–H groups occurring in the near-IR region, together with the bands due to C=O stretching modes, of carboxylic acids and amide groups in protein-based binding media. More recently, Ricciardi *et al.*[18] identified the presence of egg-yolk binders non-invasively using fibre-optic reflectance spectroscopy (FORS). FORS has also been used by Rosi *et al.*[19] non-invasively to identify organic binding materials on frescos, seccos and stanco wall paintings. They found that applying multivariate analysis to IR spectra of wall paintings was rather challenging, but were successful in

* *Correspondence to: Stephen Elliott, Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, UK. E-mail: sre1@cam.ac.uk*

a *Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, UK*

b *Fitzwilliam Museum, University of Cambridge, Trumpington Street, CB2 1RB, Cambridge, UK*

c *Hamilton-Kerr Institute, University of Cambridge, Mill Lane, Whittlesford, CB22 4NE, Cambridge, UK*

Journal of RAMAN SPECTROSCOPY

identifying egg and glue components. Daher *et al.*[20] have proposed a flowchart for the identification of four families of varnishes, viz, gums, oils, glues and resins, using characteristic peaks in the Raman and IR spectra of the materials.

Nevin *et al.*[21] used dispersive Raman spectroscopy, with 785 nm wavelength excitation, to analyse binders deposited as films with a thickness of ~15 μm onto silica-disk substrates. They employed multivariate analysis to analyse the spectral data, investigating signal intensities in the C–H stretch region between 2700 and 3200 cm^{-1}. Using this method, they were able to discriminate between some of a set of eight binding media, mainly dairy-based and collagen-based. Manzano *et al.*[22] have used FT-Raman studies to analyse aged lipidic binders and have successfully employed supervised classification methods to identify the different aged lipidic binders. Nevin *et al.*[23] have also made use of time-resolved fluorescence spectroscopy as another non-invasive method to make use of the naturally fluorescing nature of organic binders in order successfully to identify them.

Raman spectroscopy is a technique that has been widely used in art-conservation science,[24–27] as it is not only non-invasive, but also provides a wealth of structural and bonding information, from insight into the nature of functional groups present at the molecular level, to information about the crystallographic structure. A drawback of conventional Raman spectroscopy, however, is that fluorescence from the organic molecules that make up binder materials can mask the weak Raman signal. This is a particular problem when visible-wavelength lasers are used for excitation.[21] To work around this problem, it is ideal to identify the laser wavelength for which the material under study has little fluorescence.[28] The use of a near-IR excitation wavelength, e.g. 1064 nm, in a FT-Raman system substantially reduces the fluorescence from organic compounds,[4] making it an excellent tool to study binders, varnishes and resins. However, a problem with this wavelength is that it is impossible then to analyse Fe-, Si- and Cu-based pigments, since Fe and Cu have electronic energy levels that give rise to fluorescence when excited at this wavelength.[25] FORS is another non-invasive technique which has been widely used for the identification of pigments.[7,24,29–32] FORS provides spectral data from the UV-visible to the short-wave IR region,[18,33] and thus yields information not only about the pigment colour, but also chemical information on the functional groups present, such as C–H, O–H, N–H etc.

Identifying the constituents of paint materials by manually interpreting spectra can be a complex and time-consuming task. This is particularly true when investigating binding materials, since these are complex mixtures of organic compounds which together give rise to a wealth of spectral features. While differentiating classes of binders, for example proteinaceious and fat-based, may be relatively straightforward, discriminating between materials in the same class, for example different kinds of oils, is likely to be more challenging. For example, linseed oil (LO) is composed of ω^3-fatty acids and triglycerides, while poppy and walnut oils (WOs) are composed mainly of ω^6-fatty acids and triglycerides. The main difference between poppy oil (PO) and WO is just a small amount of vitamin E dissolved in PO, which gives rise to spectral features that leads to a difference in their respective spectra. To aid in the classification of spectra, different forms of multivariate analysis have been employed.[21,22,34–36] The goal of such techniques is to find combinations of spectral bands which account for the variation within a large dataset, and hence to classify the data; it is then up to the user to interpret the resulting classification.

One commonly employed multivariate technique is principal-component analysis (PCA).[37] In PCA, covariance is used to find orthogonal linear combinations of bands, termed principal components (PCs), which account for as large a fraction of the variance within the data as possible. In a 2D scatter plot, on plotting the first two PCs for each data item against one another, similar items will form clusters or bands, allowing them to be easily grouped. For this method to work well, the first two PCs should account for as much of the variance as possible. The PCA analysis produces a projection matrix which can be used to project a new set of sample data onto the PC space, and therefore if a set of reference samples is used to 'train' a PCA, it can potentially be used for identifying the components in spectra of unknown materials. Many such examples can be found in the literature, and some are included in the references.[10,16,35,38–41]

In this work, we have investigated the use of FT-Raman and FORS for the non-invasive analysis of binding materials used to make lead pigment-based paints. We demonstrate that the two types of spectroscopy provide useful, and complementary, information that could allow identification of the components present in paint films. We employed PCA to differentiate binder constituents using FT-Raman spectra and also combined FT-Raman/FORS spectra. We found that a PCA analysis, constructed using a high-quality set of references, could be used to aid in the automatic classification of spectra. While oil-based binding media could be discriminated using FT-Raman alone, combined Raman and FORS data were required for reliable discrimination among proteinaceious and polysaccharide-based media. The methods investigated in this work should be a valuable tool for the non-invasive analysis of binding media in artworks.

Experimental

Sample preparation

All the pigments used in this study, viz. lead white (LW), red lead (RL) and lead-tin yellow (LTY), were purchased from Kremer Pigments Inc. and used without any further processing. Seven binder materials used for making the paint films were purchased locally, except for gum arabic resin (Kordofan Grade), which was imported from the Middle East via L. Cornelissen and Son, London. Gum arabic (GA) binder was prepared from this resin by dissolving it in water. Egg white (EW), egg yolk (EY) and whole egg (WE) were obtained from free-range eggs, while LO, PO and WO were obtained as cold-pressed products. Significant variations in the chemical composition of these binders can be expected between batches, and this depends on various factors; for example, in the case of oils, the strain of plant and soil type, as well as variations in the harvesting and processing procedures, can be important. These factors were not accounted for in this work, although these issues will potentially exist in real art-work samples.

Paint mixtures were prepared using criteria employed by artists, e.g. by mixing a pigment and a binder until a paint with appropriate mechanical and optical properties was obtained. This procedure is purely qualitative in nature. Two sets of samples were prepared for each pigment–binder combination, one rich in the binding material and the other lean in binder. These prepared paints were then painted as single layers onto microscope glass slides and allowed to dry on a bench top for 3 months before analysis. Neat samples of the binding materials

were also prepared on microscope glass slides for comparison. After the initial drying, the paint films were stored in a light-tight box.

The paint films had rough, non-homogenous surfaces, as would be expected in real samples. Representative images can be found in the supporting information. Thicknesses of a subset of these films were measured using a profilometer, and they varied between 30 and 250 microns (data not shown). The general trend observed was that EW-, GA- and WE-based paint films were thinner than the oils, and LW-based paints were thinner than those containing LTY and RL.

Spectroscopy

Raman spectra were collected using a Bruker RAM II instrument (1064 nm wavelength, ~1 mm spot size, KBr beam splitter and Ge detector, 2.37 mm penetration depth) and recorded within the range of 20 – 3600 cm^{-1} at a resolution of 4 cm^{-1}. In order to obtain the best quality spectra, we investigated the effect of altering the laser power and acquisition time on the noise level in the acquired spectra, testing laser powers of 5, 25, 50, 75, 100 and 150 mW, and acquiring 64, 128 and 256 scans per spectrum. Typically, 256 scans take 7 min to be acquired. We found that a power of 100 mW was required in order reliably to discern carbonyl stretches, particularly prominent in samples using proteinaceious binders, from the background noise. Although this laser power is higher than that usually employed,[21] we did not observe any beam damage during our experiments. As expected, we found that integrating more scans per spectrum reduced noise significantly.

Fibre-optic reflectance spectra were obtained using a FieldSpec 4 spectroradiometer (ASD Inc., USA) in the wavelength range of 350–2500 nm, at a spectral resolution of 3 nm at 700 nm, and 10 nm at 1400 nm and 2100 nm. The instrument was calibrated using a 99% reflective spectralon standard (Labsphere) before acquiring sample spectra.

Data analysis

Data processing was implemented in Python 3.1,[42] using the NumPy,[43] SciPy[44] and Matplotlib[45] packages. For performing the PCAs, we made use of the PCA functionality in Matplotlib, while all other processing was coded using the basic functionality of the three packages.

To perform PCAs, a set of data samples was loaded, and the spectral region(s) under consideration was linearly interpolated to generate a set number of data points. After pre-processing, the spectra were loaded into rows of a data matrix, which was analysed using the Matplotlib PCA routine.

In order to account for variations in intensity due to sample geometry, present in both FORS and Raman spectra, the region of the spectra being used in the analysis was vector normalised.[46] In this procedure, the mean intensity is subtracted from all bands, and each point is divided by the standard deviation of the signal. This ensures that spectra with different absolute peak heights, but whose features are similar, will be adjusted to appear similar to one another. This prevents the PCA from using absolute peak heights to differentiate the samples, which could lead to erroneous classifications.

Various things can lead to random noise in Raman spectra, and such noise can make the identification of spectral bands more difficult, particularly in computational analyses, as in the present work. There are several approaches to denoising spectra. The simplest are averaging algorithms, e.g. rectangle, triangle or Gaussian averaging, and median filtering, which work by smoothing out spurious spikes from the spectra, while retaining more prominent spectral features. A slightly more sophisticated algorithm, proposed by Rowlands and Elliott,[47] works on a similar principle, but attempts selectively to remove noise based on its expected Gaussian statistical distribution. A more advanced alternative is to fit spectral peaks to a suitable functional form (e.g. Gaussian or Lorentzian functions), and hence to extract specific spectral parameters of interest. Colomban and Tournie[48] used this to excellent effect to analyse the spectra of silica glasses, determining their composition and degree of weathering, by extracting and analysing characteristic peaks associated with specific Si_xO_y structural motifs. Although this sort of analysis facilitates detailed interpretation of the spectra, it is not well suited to the present work, since the different binders have Raman signatures which are not shared by the others.

We tested the effect of applying the basic smoothing methods to the Raman spectra, looking at how they affected the quality of the spectra and the resulting PCA. We found that rectangle averaging and median filtering both led to a flattening of peaks, which is highly undesirable. In contrast, when used with small averaging windows, triangle and Gaussian averaging both largely preserved the shape of the peaks, while substantially reducing

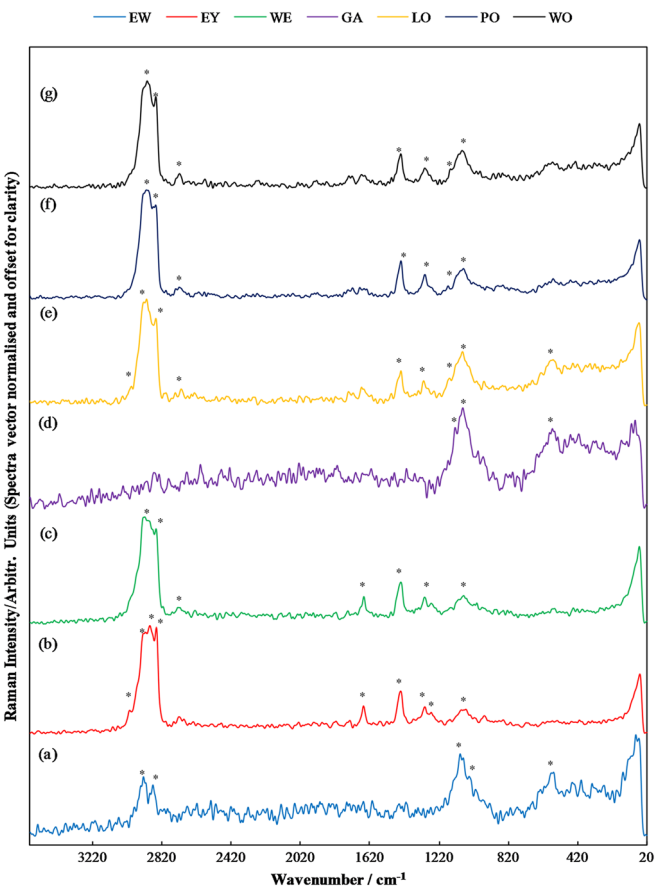

Figure 1. Representative FT-Raman spectra of pure (a) egg yolk (EY), (b) whole egg (WE), (c) linseed oil (LO), (d) poppy oil (PO) and (e) walnut oil (WO) acquired at 100 mW and with 256 scans (7 min), after one-point triangle smoothing and vector normalisation. The peaks labelled with * have been assigned in Table 1. This figure is available in colour online at wileyonlinelibrary.com./journal/jrs

Table 1. Tentative assignments of spectral peaks observed in the FT-Raman spectra of pure binder materials.*phe = phenylalanine

Gum arabic	Egg white	Egg yolk	Whole egg	Linseed oil	Poppy oil	Walnut oil	Tentative assignments[21,22,52]
967		955		955			
994							Phe* ring breathing /
	1031						C–C stretching
	1042						
			1072		1074		Phe* C–C stretching
1080	1078	1058					
1086		1076		1078		1076	C–C, C–N stretching
1122	1096						Aliphatic anti-symmetric
1131					1162	1136	stretching and aromatic rocking
			1255	1257			Amide III / =C–H symmetric
			1297	1297	1297	1295	rocking
				1303			CH_2 deformation
		1441	1436	1438	1436	1436	C–H bending
		1653	1649	1651			Amide I / C = C stretching
					1673	1657	
			1746	1735	1715	1717	C = O stretching
			1813				
	2536	2708	2710	2700	2710	2712	Aliphatic
	2617			2792	2732		C–H stretching
	2869			2853			Aromatic / unsaturated
	2914						C–H stretching
	2925	2849	2851	2905	2853		
		2885			2895	2853	
		2907					
	2976		2887		2915	2901	
		2997	2915			2979	

the noise. We found that the algorithm proposed in Ref.[47] tended to perform similarly to the best averaging methods on spectra with good signal-to-noise ratios (SNRs), but was less effective on more noisy samples. Since the triangle and Gaussian methods performed similarly, but the latter required optimisation of another parameter (the width of the Gaussian curve), we opted to use the former. To select an appropriate averaging window, we tested a range of sizes from 11 to 151 data points. We found that a 21-point window gave a good balance of retaining spectral features and reducing noise across a range of samples.

To assess the quality of spectra and PCA plots, it is desirable to have a metric by which they can be compared quantitatively. To compare spectra, we made use of the Anderson–Darling (AD) statistical test[49,50] as a crude measure of SNR. Given a set of sample data, the AD test gives the probability that the data are drawn from a given distribution. Raman spectra are known to be a convolution of Lorentzian peaks and Gaussian noise,[47] and thus if the AD test gives a high probability of a spectrum being drawn from a Gaussian distribution, it should contain more noise. A lower test statistic indicates a higher probability of the spectra intensities being drawn from a Gaussian distribution and hence a lower SNR. We found this procedure to serve as a good quantitative means for comparing the SNRs of spectra.

To compare the ability of PCAs to discriminate between samples, we found that the sum of the variance of the first two PCs could be misleading; a high explained variance did not necessarily translate into desirable tight clustering of similar spectra. Therefore, we devised an alternative metric, in which the average 'distance' of points within a cluster to the centroid is used for comparison. This provides a means of comparison which more closely resembles a qualitative 'by eye' assessment and was used as a supplement to comparing the explained variance.

Results and discussion

Raman spectroscopy

When interpreting the FT-Raman spectra, data from around 1000 to 3200 cm^{-1} were considered, as the presence of intense peaks below 1000 cm^{-1}, arising from the pigments, swamps the peaks arising from low-frequency 'accordion-type' phonon modes in macromolecules.[51] The phenylalanine ring-breathing and C–C stretching modes are quite clearly represented in the spectra of pure binder samples, but are lost in the presence of RL and LTY. With bound LW (Fig. S1 and Table S1), the phenylalanine's C–C stretching frequencies for the oils are still observed, but over a considerably narrower spectral range of around 1050–1055 cm^{-1} compared to the span of 1030–1080 cm^{-1} in the pure samples (Fig. 1). This clearly indicates the presence of pigment–binder interactions that could be valuable in discriminating between samples. The egg and gum samples, unlike the oils, only have strong C–H and C–C stretching frequencies and ring-breathing modes, with only the C–H stretching frequencies being observed in the pigment–binder samples, except in the case of WE. This is an interesting observation, given the fact that the phenylalanine's ring-breathing and C–C stretching modes

should have been more pronounced in the case of EY than in WE, drawing comparison to the other fat-based binders viz. oils. The C=O stretching frequencies observed in the pure oil–binder samples were not easily discernible against the noisy background in the pigment–binder samples. Representative spectra, together with tentative peak assignments, may be found in the supporting information (Figs. S1 – 3 and Tables S1 – 3). As discussed earlier, the S/N ratio for these spectra correlate to their sample thickness. EW, EY and GA and those paints made out of LW, particularly, have very low S/N ratio. This is also related to the fact that a 1064 nm laser has an estimated penetration depth of 2.37 mm (in polished Si), and the above mentioned films of LW-based paints have a film thickness of about 30 – 80 microns.

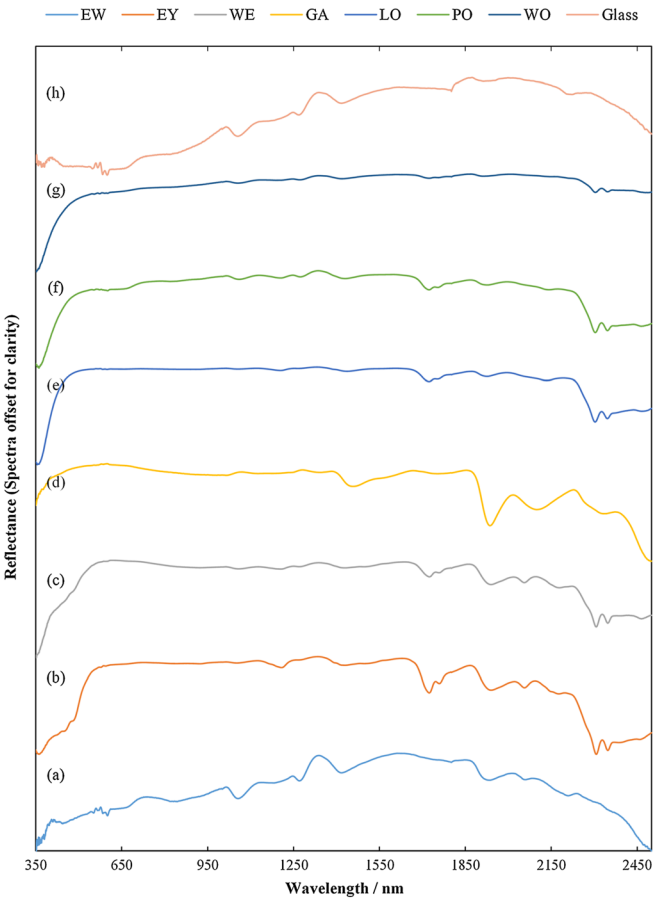

Figure 2. Representative fibre-optic reflectance spectra of samples of pure binder materials: (a) egg white (EW), (b) egg yolk (EY), (c) whole egg (WE), (d) gum arabic (GA), (e) linseed oil (LO), (f) poppy oil (PO) and (g) walnut oil (WO); (h) Plain glass. This figure is available in colour online at wileyonlinelibrary.com./journal/jrs

FORS

The most prominent difference between the FORS spectra of the pigment–binder samples was observed to be in the transition edges in the visible region, arising from the colour of the pigments. In this respect, spectra from pure binder samples yielded less information than the pigment–binder samples due to their translucent nature. The first-overtone bands of methylene (−CH_2–), methyl (−CH_3) and alkenyl (−CH=CH–) stretching modes were observed in the range of 1720–1760 nm, while the respective second-overtone bands[17] were observed between 1200 and 1215 nm (Fig. 2, Table 2 and Figs. S4 – 7 (supporting information)). Combination bands for the –CH=CH– and –CH_2–/–CH_3 stretching modes were observed in the ranges of 2100–2190 nm and 2300–2490 nm, respectively.

As previously reported,[53] the first overtone for OH stretching was observed between 1400 and 1460 nm. In the LW samples, similarly to the FT-Raman spectra, the first overtones of OH vibrations were consistently observed in the FORS spectra at 1441 nm for the fats and oils (Table S4). Corresponding peaks from egg binder components and GA fell within a 10 nm range of this. These spectral features could not only be due to possible pigment–binder interactions, but also to the OH groups present in basic LW. Interestingly, the first overtones of OH vibrations were also observed in all other bound spectra, except the egg-white and whole-egg samples in the RL and LTY samples. The OH combination bands, unlike the overtones, were observed in both pure binders and the bound-pigment samples between 1920 and 1940 nm. Finally, features in the range 2050–2060 nm were ascribed to NH combination bands[17] and were present only in egg components in the pure binder samples, and in all but, the WO-bound lead-pigment-based paints. Representative FORS spectra of the different samples and their tentative assignments are available in the supplementary information (Figs. S4 – 7 (supporting information)).

Multivariate analysis

A total of 317 Raman spectra were acquired and were subject to the pre-processing methods outlined above (see also the supplementary information). A lot of work[21,35,41] has been done

Table 2. Tentative assignments of bands observed in the fibre-optic reflectance spectroscopy (FORS) spectra of pure binder materials

Gum arabic	Egg white	Egg yolk	Whole egg	Linseed oil	Poppy oil	Walnut oil	Tentative assignments[16,17]
		1210					Second overtones of –CH_3, –CH_2–, –CH=CH– groups
1456		1414					First overtones of –OH
		1724	1725	1723		1721	First overtones of –CH_3, –CH_2–, –CH=CH– groups
		1759	1759	1755		1757	
1933		1933	1933	1922		1922	Combination bands of –OH
		2058	2058				Combination bands of –NH
2103				2131			Combination bands of –CH=CH–
2330		2306	2306	2302	2302	2302	Combination bands of –CH_3
		2348	2348	2347	2344	2345	and –CH_2–

Journal of RAMAN SPECTROSCOPY

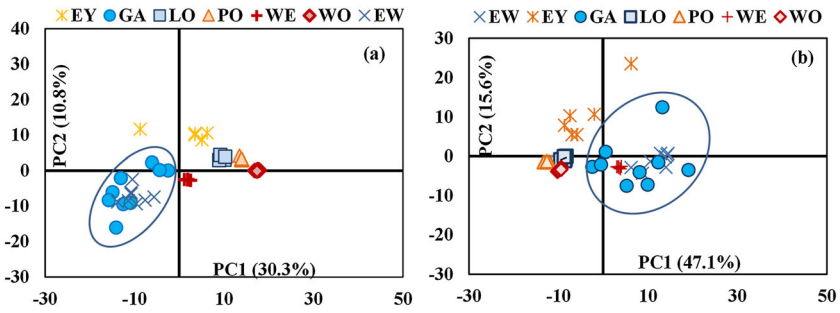

Figure 3. Principal-component analyses (PCAs) of FT-Raman spectra of red lead (RL) bound in egg yolk (EY), gum arabic (GA), linseed oil (LO), poppy oil (PO), whole egg (WE), walnut oil (WO) and egg white (EW). Triangle smoothing and vector normalisation were applied to the spectra before the analysis. The two analyses were conducted using spectral regions between (a) 1100 – 3200 cm^{-1} (full spectrum) and (b) 2700 – 3200 cm^{-1} (C–H region). In general, oils form tight clusters, while GA and EW, both consisting of polysaccharides and glycoproteins, form overlapping, widely spaced clusters. The clustering is improved when using the full spectrum instead of just the C–H region, but distinguishing between GA and EW is still difficult. EY forms tighter clusters in the full spectrum compared to the C–H region, and WE which fell on top of the EW and GA in the C–H region plot forms a separate cluster when the full spectrum is considered.

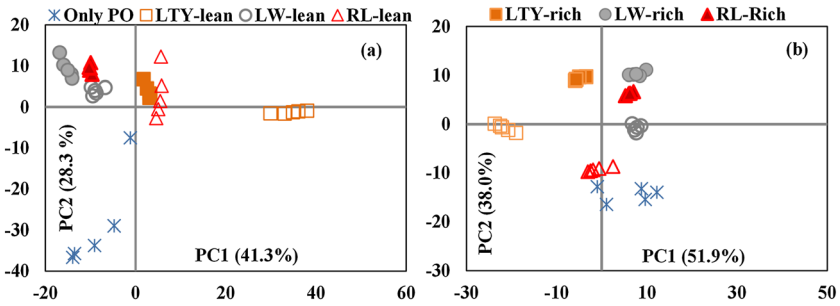

Figure 4. Principal-component analysis of Raman spectra of red lead (RL), lead-tin yellow (LTY) and lead white (LW) bound in poppy oil (PO) taken with five different laser powers on the same point. The paint films rich in binder concentration are represented by filled symbols and lean samples by hollow symbols. Triangle smoothing and vector normalisation were applied to the spectra before analysis. The two analyses were conducted using spectral regions between (a) 1100 – 3200 cm^{-1} (full spectrum) and (b) 2700 – 3200 cm^{-1} (C–H region). Rich and lean samples form different clusters, as do the different pigments.

to identify binders non-invasively using the C–H region of the acquired spectra to group binder materials. To investigate this process further, we performed separate PCAs on different spectral regions, viz. the fingerprint region (1100 – 1500 cm^{-1}), carbonyl region (1500 – 2000 cm^{-1}), C–H region (2700 – 3200 cm^{-1}), a combination of the carbonyl and C–H regions (1500 – 3200 cm^{-1}) and the full spectral region (1100 – 3200 cm^{-1}). The region between 20 and 1100 cm^{-1} was not included, since it contains prominent peaks corresponding to the pigments, which would invariably cause the PCA to assign most weight to discriminating between the pigments rather than the binders. In a similar vein, vector normalisation of the entire spectra would likely flatten peaks due to the binders, and might possibly also broaden the pigment peaks and hence have a major (undesirable) impact on the PCA analysis. We therefore performed the vector normalisation after extracting and interpolating the region of interest.

We found that using the carbonyl and fingerprint regions did not yield good clustering, but analysing the C–H region gave good results for the oils, and considering the full spectrum gave good clustering for both the oils and the proteinaceous binders (Fig. 3). Using the C–H region, the oils form tight clusters, but their polysaccharide-containing counterparts, viz. EW, GA and even WE, form rather more spaced-out clusters that overlap with each other. In contrast, a PCA analysis using the full spectral region seems to give a better cluster spacing than just the C–H region, even though the percentage variance decreases, owing mainly to the presence of peaks in the fingerprint region. However, as observed by Nevin *et al.*,[21] differentiation of GA from EW still proved difficult.

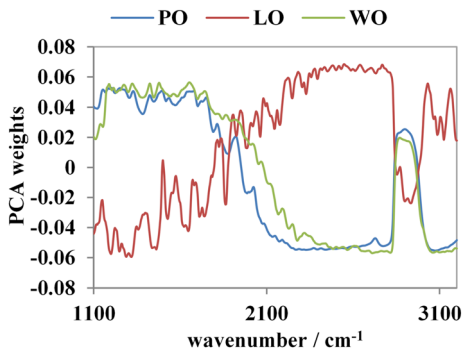

Figure 5. Loading plots of the first principal component (PC1) in principal-component analyses (PCAs) of poppy oil (PO), linseed oil (LO) and walnut oil (WO) with three different bound pigments and in two different concentrations. Higher (absolute) weights are given to parts of the spectra that contribute to the largest variations in spectral features. In this case, the weight is assigned to the width of the C–H peak and to the spectral background, perhaps due to the equalisation of peak heights through vector normalisation. This figure is available in colour online at wileyonlinelibrary.com./journal/jrs

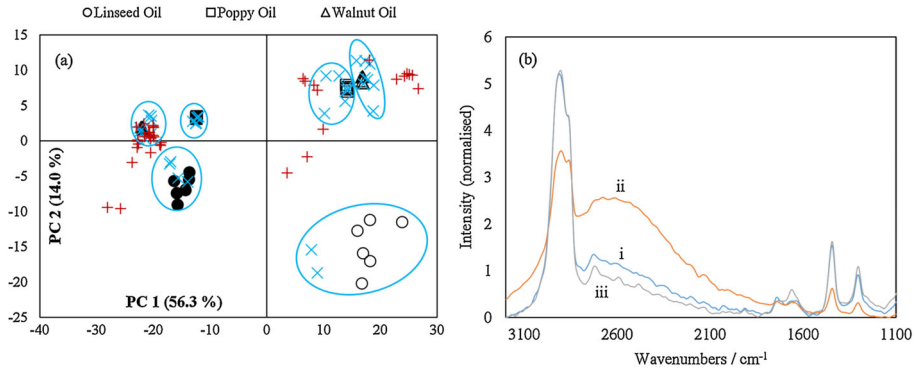

Figure 6. (a) Principal-component analysis (PCA) of the FT-Raman spectra (1100 – 3200 cm^{-1}) of lead-tin yellow (LTY) bound in poppy oil (PO), linseed oil (LO) and walnut oil (WO), used for assisting in the classification of unknown spectra. The filled symbols denote samples rich in binder, and the hollow symbols denote lean samples, with a lower binder concentration. Unknown spectra classified correctly are represented as blue 'X', and misclassified spectra are represented as red '+'. (b) Sample pre-processed spectra that were classified. (i) LTY bound in PO (rich sample) acquired at 100 mW; classified correctly (Anderson – Darling (AD) value of 58.96). (ii) LTY bound in PO (lean sample) acquired at 100 mW; classified incorrectly (AD value of 25.25). (iii) LTY bound in WO (rich sample) acquired at 100 mW; classified correctly (AD value of 73.63).

To investigate the influence of the pigments and binder concentration on the spectra, we performed PCAs on the spectra obtained from three different lead pigments bound in a single binder, in both rich (high) and lean (low) binder concentrations. Even though the pigments do not have prominent features in the spectral regions considered (between 1100 and 3200 cm^{-1}), they seem to give rise to features which are detectable by PCA. This is clearly illustrated in Fig. 4, where separate clusters are formed, not only for each binder concentration, but also for each pigment. Furthermore, this discrimination is observed regardless of whether the full spectrum or just the C–H region is used. It is also interesting to note that the samples rich in binders generally seem to form tighter clusters, presumably owing to more prominent spectral features.

To investigate the origin of the discrimination, we looked at the weights assigned to the spectral bands in the first principal component (PC1) in PCAs of each of the three oils with different pigments and binder concentrations (Fig. 5). Vector normalisation causes absolute band intensities across different spectra to be largely equalised, preventing the PCA from weighting, e.g. the height of C–H bands. This is borne out in the loading plots, where weight appears to have been assigned to the edges, suggesting that the C–H peak may be broadened to different extents by bound pigments. Perhaps unexpectedly, weight appears to have been assigned to the background from around 2100 cm^{-1}; we suspect this largely corresponds to fluorescence from LTY (see Fig. 6 b) and presumably serves as a means to differentiate samples containing this pigment from the others. We note that the relative prominence of features such as this, arising from the presence of a bound pigment, and those predominantly from the binders, may account for the observed clustering of the samples with respect to binder concentration. These findings highlight the fact that bound pigments can influence FT-Raman spectra in the region where the binder bands are found, and a technique such as that used here should take this into account when working with unknown samples.

As mentioned earlier, oils seemed to cluster much better than proteinaceous and polysaccharide-based binders. To test the utility of FT-Raman for assisting in the classification of spectra,

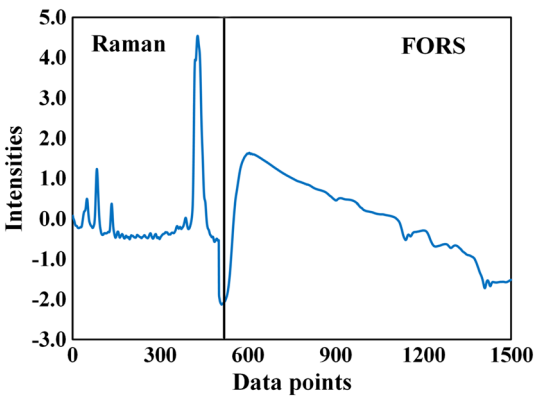

Figure 7. Representative combined Raman and FORS spectra of lead-tin yellow (LTY) bound in egg yolk (EY). The first 500 data points are made up by the Raman spectrum, and the remaining 1000 data points by FORS data.

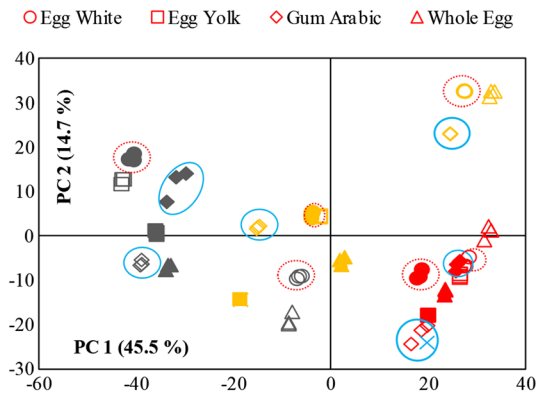

Figure 8. Principal-component analysis (PCA) of combined FT-Raman and fibre-optic reflectance (FOR) spectra of red lead (RL), lead-tin yellow (LTY) and lead white (LW) bound in egg white (EW), egg yolk (EY), gum arabic (GA) and whole egg (WE). To compare with the clustering of GA and EW in Fig. 3, GA clusters have been circled in blue and EW in dotted red. This plot was hence used for assisting in the classification of an unknown spectrum (denoted by a cross). The representation markers are coded as follows: paint films rich in binder concentration are represented by filled symbols, and lean samples by hollow symbols, and pigments are represented by colours - LW (grey), RL (red) and LTY (yellow). The use of both Raman and FORS data results in better clustering and separation of EW and GA, both of which consist mainly of polysaccharides and glycoproteins.

Journal of RAMAN SPECTROSCOPY

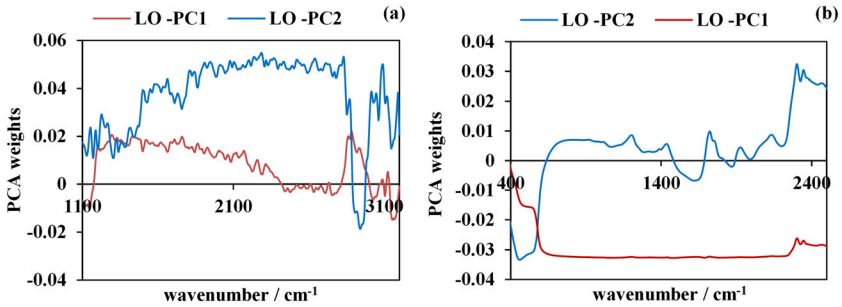

Figure 9. Loading plots for the first and second principal component (PC1 and PC2) in a PCA analysis of combined FT-Raman and fibre-optic reflectance spectroscopy (FORS) spectra of linseed oil (LO) samples. Higher (absolute) weights are given to the parts of the spectra that contribute to the largest variation in the spectra. For clarity, the loading plots for the combined PCA are represented separately as (a) Raman and (b) FORS. PC1 has more weight assigned to the FORS spectra, while PC2 assigns more weight to the Raman spectra, although there is a significant contribution to PC1 from the Raman data.

we performed a PCA with spectra of the LTY oil samples, recorded with 256 and 128 scans and used this to classify a range of 64-scan spectra treated as unknowns. These spectra were projected onto the PCA space using the projection matrix built by analysing the training set and then classified according to their Euclidean distance from each of the cluster centroids in the reference PCA (Fig. 6a). We were able to classify over 60% of the spectra using this technique, and the analysis is sensitive not only to the binder, but also its concentration. As observed earlier, the rich samples again form tighter clusters than the lean ones, and hence the rich unknowns are more easily classified. The misclassified set of data were mostly that of PO that was clustered with WO due to the presence of fluorescence from LTY (dark), which seems to have been exaggerated on vector normalisation (Fig. 6b). This hence affects the S/N ratio considerably. The fluorescence from LTY (dark) can be attributed to the presence of silicon in its composition, which absorbs at 1064 nm (glass spectrum Fig. S12 in supporting info). The AD values of sample spectra show that the misclassified spectrum has much lower AD values than the ones classified correctly. The linseed-oil spectra that were misclassified as WO were the ones acquired at low laser powers and with a small number of integrated scans, making the spectral features less prominent (data not shown).

In principle, PCA may give more accurate classifications when more information is provided for each sample under study. To test this, we created combined FT-Raman and FORS spectra by concatenating pairs of processed spectra (Fig. 7); each spectrum had 1500 data points, with the first 500 being made up using the Raman spectrum, and the remaining 1000 from the FORS spectrum. As is clearly evident in Fig. 7, vector normalisation, applied to both the Raman and FORS spectra separately before concatenation, gives spectra within a similar intensity range. This is noteworthy, since if the intensities were significantly different, the PCA would assign more weight to the data set with the higher intensities.

A PCA plot generated using the combined spectra (Fig. 8) showed very good clustering, not only of the oils, but also of EW and GA, which proved difficult to segregate using only FT-Raman data. The analysis was performed on spectra recorded for paints prepared using LW, LTY and RL bound in EW, EY, WE and GA. EW and GA, which have similar Raman spectra, have significantly different FORS spectra, owing to the presence of overtones and combination bands (Fig. 2). This can be clearly seen on a comparison of the clusters that overlap in Fig. 3 to those in Fig. 8, where they do not. The entire range in the FORS spectra (350–2500 nm) was used, since, in this case, the pigments have distinct peaks/transition edges in the visible region that do not interfere with bands from the binders and will invariably aid in the classification of samples by bound pigment. Using the projection matrix generated by the PCA analysis, we were able to identify a spectrum of RL bound in GA, regarded as unknown.

To study the origin of the enhanced performance of the combined-spectra PCAs, we looked at the loading plots used to generate them. Figure 9 shows the loading plots of a PCA of the linseed-oil samples, with the Raman and FORS spectra represented separately for clarity. The PC1 has more weight assigned to the FORS spectra and less to the Raman, while PC2 is weighted towards the Raman data. However, there is a significant Raman contribution to PC1, indicating that the PCA analysis is using both data sets in combination.

Conclusion

Knowledge of the binding media and colour palettes used by artists helps not only in complementing art-historical research, but also in the process of conservation and restoration of artwork. There is a wide interest in the non-invasive identification of historic binder materials, which are often polysaccharides, fatty acids or protein based, and whose spectroscopic identification is generally difficult due to their complex nature. In this work, we have used 1064 nm FT-Raman and FORS measurements to identify non-invasively binding media in a range of lead-based paints. We found that both types of spectra contain complementary information that could be used to distinguish between binding media as well as bound pigments, as well as to highlight pigment–binder interactions.

Using PCA, we were able to use sets of reference spectra to assist in classifying 60% of spectra of unknown samples. Using FT-Raman data, this technique worked well for classifying spectra of paints using oil-based binders. We found that using the features in a large spectral window, between 1100 and 3200 cm^{-1}, gave better results than the conventionally employed C–H region, and that the PCA assigned most weight to the spectral background and peak widths. For discrimination of the more challenging proteinaceious and polysaccharide-based media, the quality of the PCA, and hence the discrimination

between the binders, was greatly improved by using combined Raman and FORS spectra. With a reference set of such spectra, we were able successfully to classify an unknown GA sample. These results illustrate the wealth of information about paint films that can be obtained from FT-Raman and FORS measurements and demonstrate the utility of multivariate analysis techniques for assisting in the identification of paint constituents using such data. The results presented here are a first step towards computer-assisted analysis of artwork; however, more needs to be done to investigate the application of this method to real artwork, where the effects of issues such as pre-processing of products, substrate and ageing effects may become apparent.

Supporting information

Supporting information may be found in the online version of this article.

References

[1] I. R. Lewis, H. Edwards, Handbook of Raman Spectroscopy: From the Research Laboratory to the Process Line, Marcel Dekker, Inc., USA, **2001**.
[2] H. G. M. Edwards, J. M. Chalmers, Raman in art and archaeology, RSC, Cambridge, **2005**.
[3] G. Bitossi, R. Giorgi, M. Mauro, B. Salvadori, L. Dei, *Applied Spectroscopy Reviews* **2005**, *40*, 187.
[4] R. J. H. Clark, *Comptes Rendus Chimie* **2002**, *5*, 7.
[5] I. Osticioli, N. F. C. Mendes, S. Porcinai, A. Cagnini, E. Castellucci, *Analytical and Bioanalytical Chemistry* **2009**, *394*, 1033.
[6] M. Bacci, A. Casini, C. Cucci, M. Picollo, B. Radicati, M. Vervat, *Journal of Cultural Heritage* **2003**, *4*, 329.
[7] G. Dupuis, M. Elias, L. Simonot, *Applied Spectroscopy* **2002**, *56*, 1329.
[8] P. Ricciardi, J. K. Delaney, L. Glinsman, M. Thoury, M. Facini, E. R. de la Rie, in O3A - *Optics for Arts, Architecture, and Archaeology Conference II*, vol. *7391*, Munich, Germany, **2009**.
[9] S. Dallongeville, M. Koperska, N. Garnier, G. Reille-Tailefert, C. Rolando, C. Tokarski, *Analytical Chemistry* **2011**, *83*, 9431.
[10] J. Romero-Pastor, N. Navas, S. Kuckova, A. Rodriguez-Navarro, C. Cardell, *Journal of Mass Spectrometry* **2012**, *47*, 322.
[11] I. D. van der Werf, C. D. Calvano, F. Palmisano, L. Sabbatini, *Analytica Chimica Acta* **2012**, *718*, 1.
[12] J. Blasko, R. Kubinec, B. Husova, P. Prikryl, V. Pacakova, K. Stulik, J. Hradilova, *Journal of Separation Science* **2008**, *31*,1067.
[13] I. Bonaduce, H. Brecoulaki, M. P. Colombini, A. Lluveras, V. Restivo, E. Ribechini, *Journal of Chromatography A* **2007**, *1175*, 275.
[14] R. Mateo-Castro, J. V. Gimeno-Adelantado, F. Bosch-Reig, A. Domenech-Carbo, M. J. Casas-Catalan, L. Osete-Cortina, J. de la Cruz-Canizares, M. T. Domenech-Carbo, *Fresenius Journal of Analytical Chemistry* **2001**, *369*, 642.
[15] B. Witkowski, M. Biesaga, T. Gierczak, *Analytical Methods* **2012**, *4*, 1221.
[16] M. Bacci, M. Fabbri, M. Picollo, S. Porcinai, *Analytica Chimica Acta* **2001**, *446*, 15.
[17] M. Vagnini, C. Miliani, L. Cartechini, P. Rocchi, B. G. Brunetti, A. Sgamellotti, *Analytical and Bioanalytical Chemistry* **2009**, *395*, 2107.
[18] P. Ricciardi, J. K. Delaney, M. Facini, J. G. Zeibel, M. Picollo, S. Lomax, M. Loew, *Angewandte Chemie-International Edition* **2012**, *51*, 5607.
[19] F. Rosi, A. Daveri, C. Miliani, G. Verri, P. Benedetti, F. Pique, B. G. Brunetti,A. Sgamellotti, *Analytical and Bioanalytical Chemistry* **2009**, *395*, 2097.
[20] C. Daher, C. Paris, A. S. Le Ho, L. Bellot-Gurlet, J. P. Echard, *Journal of Raman Spectroscopy* **2010**, *41*, 1494.
[21] A. Nevin, I. Osticioli, D. Anglos, A. Burnstock, S. Cather, E. Castellucci, *Analytical Chemistry* **2007**, *79*, 6143.
[22] E. Manzano, J. Garcia-Atero, A. Dominguez-Vidal, M. J. Ayora-Canada, L. F. Capitan-Vallvey, N. Navas, *Journal of Raman Spectroscopy* **2012**, *43*, 781.
[23] A. Nevin, D. Comelli, G. Valentini, D. Anglos, A. Burnstock, S. Cather, R. Cubeddu, *Analytical and Bioanalytical Chemistry* **2007**, *388*, 1897.
[24] L. Appolonia, D. Vaudan, V. Chatel, M. Aceto, P. Mirti, *Analytical and Bioanalytical Chemistry* **2009**, *395*, 2005.
[25] L. Burgio, R. J. H. Clark, *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2001**, *57*, 1491.
[26] K. Castro, M. Perez-Alonso, M. D. Rodriguez-Laso, L. A. Fernandez, J. M. Madariaga, *Analytical and Bioanalytical Chemistry* **2005**, *382*, 248.
[27] T. D. Chaplin, R. J. H. Clark, M. Martinon-Torres, *Journal of Molecular Structure* **2010**, *976*, 350.
[28] M. C. Caggiani, P. Colomban, *Journal of Raman Spectroscopy* **2011**, *42*, 790.
[29] S. Acquaviva, E. Danna, M. L. de Giorgi, A. Della Patria, L. Pezzati, *Applied Physics a-Materials Science & Processing* **2008**, *92*, 223.
[30] M. Bacci, *Sensors and Actuators B-Chemical* **1995**, *29*, 190.
[31] A. Casini, M. Bacci, C. Cucci, F. Lotti, S. Porcinai, M. Picollo, B. Radicati, M. Poggesi, L. Stefani, *Proceedings of the SPIE - The International Society for Optical Engineering* **2005**, *5857*, 177.
[32] M. Picollo, M. Bacci, A. Casini, F. Lotti, S. Porcinai, B. Radicati, L. Stefani, Fiber optics reflectance spectroscopy: A non-destructive technique for the analysis of works of art, Optical Sensors and Microsystems series, *vol. 11*, Springer, New York, **2000**, pp. 259.
[33] M. Bacci, in Modern analytical methods in art and archaeology, (Eds: E. Ciliberto, G. Spoto), Chemical Analysis series, *vol. 155*, John Wiley & Sons, New York, **2000**, pp 321.
[34] M. Bacci, S. Porcinai, B. Radicati, *Applied Spectroscopy* **1997**, *51*, 700.
[35] N. Navas, J. Romero-Pastor, E. Manzano, C. Cardell, *Journal of Raman Spectroscopy* **2010**, *41*, 1486.
[36] F. Rosi, A. Burnstock, K. J. van den Berg, C. Miliani, B. G. Brunetti, A. Sgamellotti, *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2009**, *71*, 1655.
[37] L. I. Smith, www.sccg.sk/~haladova/principal_components.pdf, **2002**.
[38] M. Bacci, R. Chiari, S. Porcinai, B. Radicati, *Chemometrics and Intelligent Laboratory Systems* **1997**, *39*, 115.
[39] S. Baronti, A. Casini, F. Lotti, S. Porcinai, *Image Analysis and Processing. 9th International Conference, ICIAP '97 Proceedings* **1997**, *I*, 14.
[40] W. Fremout, S. Kuckova, M. Crhova, J. Sanyova, S. Saverwyns, R. Hynek, M. Kodicek, P. Vandenabeele, L. Moens, *Rapid Communications in Mass Spectrometry* **2011**, *25*, 1631.
[41] J. Romero-Pastor, C. Cardell, E. Manzano, A. Yebra-Rodriguez, N. Navas, *Journal of Raman Spectroscopy* **2011**, *42*, 2137.
[42] Python Software Foundation, http://www.python.org.
[43] http://www.numpy.org.
[44] Source Scientific Tools for Python bEJ, Travis Oliphant, Pearu Peterson and others, http://www.scipy.org.
[45] http://matplotlib.sourceforge.net/.
[46] P. Vandenabeele, L. Moens, *Analyst* **2003**, *128*, 187.
[47] C. J. Rowlands, S. R. Elliott, *Journal of Raman Spectroscopy* **2011**, *42*, 370.
[48] P. Colomban, A. Tournie, *Journal of Cultural Heritage* **2007**, *8*, 242.
[49] T. W. Anderson, D. A. Darling, *Annals of Mathematical Statistics* **1952**, *23*,193.
[50] M. A. Stephens, MA, *Journal of the American Statistical Association* **1974**, *69*, 730.
[51] T. G. Spiro, in Chemical and Biochemical Applications of Lasers, *vol. 2*, (Ed: C. B. Moore), Academic Press, New York, **1974**, pp. 29.
[52] H. Sadeghijorabchi, R. H. Wilson, P. S. Belton, J. D. Edwardswebb, D. T. Coxon, *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **1991**, *47*, 1449.
[53] F. Westad, A. Schmidt, M. Kermit, *Journal of near Infrared Spectroscopy* **2008**, *16*, 265.