



Non-invasive identification of textile fibres using near-infrared fibre optics reflectance spectroscopy and multivariate classification techniques

Diego Quintero Balbas^{1,a} , Giancarlo Lanterna², Claudia Cirrincione³,
Raffaella Fontana¹, Jana Striova^{1,b} 

¹ CNR-INO National Institute of Optics, National Research Council, Largo E. Fermi 6, 50125 Florence, Italy

² Laboratorio Scientifico, Opificio delle Pietre Dure – MiC, viale F. Strozzi 1, 50129 Florence, Italy

³ Laboratorio Di Arazzi E Tappeti, Opificio delle Pietre Dure – MiC, Palazzo Vecchio, Sala delle Bandiere, Piazza della Signoria, 50122 Florence, Italy

Received: 30 June 2021 / Accepted: 6 December 2021

© The Author(s) 2021

Abstract The identification of textile fibres from cultural property provides information about the object's technology. Today, microscopic examination remains the preferred method, and molecular spectroscopies (e.g. Fourier transform infrared (FTIR) and Raman spectroscopies) can complement it but may present some limitations. To avoid sampling, non-invasive fibre optics reflectance spectroscopy (FORS) in the near-infrared (NIR) range showed promising results for identifying textile fibres; but examining and interpreting numerous spectra with features that are not well defined is highly time-consuming. Multivariate classification techniques may overcome this problem and have already shown promising results for classifying textile fibres for the textile industry but have been seldom used in the heritage science field. In this work, we compare the performance of two classification techniques, principal component analysis–linear discrimination analysis (PCA-LDA) and soft independent modelling of class analogy (SIMCA), to identify cotton, wool, and silk fibres, and their mixtures in historical textiles using FORS in the NIR range (1000–1700 nm). We built our models analysing reference samples of single fibres and their mixtures, and after the model calculation and evaluation, we studied four historical textiles: three Persian carpets from the nineteenth and twentieth centuries and an Italian seventeenth-century tapestry. We cross-checked the results with Raman spectroscopy. The results highlight the advantages and disadvantages of both techniques for the non-invasive identification of the three fibre types in historical textiles and the influence their vicinity can have in the classification.

1 Introduction

During historical and archaeological textiles examination, fibre identification is fundamental to understand the constitutive materials, and technology. Historical textiles were generally obtained using yarns of a single type of fibre, with some exceptions of blended yarns, e.g. the wool blends (more than one type of wool) from medieval textiles [1] or the hop and hemp

^a e-mail: diegoivan.quinterobalbas@ino.cnr.it

^b e-mail: jana.striova@cnr.it (corresponding author)

blend identified in a nineteenth century garment [2]. However, textiles containing more than one type of yarns for the weft and warp, such as silk–wool satin [3] or the complex structures of some tapestries made of wool, silk, and metal threads [4], are more frequent and require a detailed identification of the constitutive materials.

Despite the advances and availability of several analytical techniques, microscopic examination of fibre samples, both in longitudinal and cross-sectional views, remains the most widespread method [5–7]. The chemical and microstructural information (*i.e.* polymeric orientation and crystallinity) obtained with micro Fourier transform infrared (μ -FTIR, especially in attenuated total reflectance (ATR) mode) and micro Raman spectroscopies (μ -Raman) complement the microscopic observations and report about the condition of the fibres [8–11]. However, the successful application of μ -FTIR in reflection mode may be limited due to the beam scattering producing data complex to interpret, poorly reproducible, and the superficial deposits on the fibres may produce interference. Additionally, the pressure caused by the ATR crystal can damage fragile samples [12]. FTIR in transmission mode for fibres thicker than 30 μ m in diameter may be burdensome due to the absorption of the infrared radiation resulting in spectra complex to interpret [13]. When examining dyed textiles with Raman spectroscopy, fluorescence emission may hinder the characteristic bands of the fibre [14]. Despite the broad absorption bands and their overlapping, benchtop near-infrared (NIR) spectrophotometry combined with multivariate analysis provides qualitative and quantitative information about textile fibres and their blends and is widely used in the textile industry [8, 15, 16]. Being X–H bonds (*e.g.* X=O, N, C) vibrationally active in the NIR region, they may be efficiently exploited for the textile fibres classification [8].

Microscopic techniques generally require fibre samples (few millimetres), not always available for historical or archaeological textiles, and provide local information. On the other hand, non-invasive methods such as external reflection FTIR and handheld portable NIR spectrometry—based on miniaturized technology that may impact on the instrumental performance (*e.g.* noisier detectors and lower spectral resolution) [17]—proved suitable for the identification of fibres without sampling [18–21] as well as fibre optics reflectance spectroscopy (FORS) in the NIR region [22, 23]. The latter relies on the use of optic fibres to carry the reflected light (UV, Vis, and NIR), and generally employ better performance components (*e.g.* InGaAs detectors) [24]. Exploiting the characteristic vibrational features in the NIR region, our research group classified wool and silk fibres in a seventeenth-century tapestry with the ratios of the marker bands (1507/1536 nm, 1507/1575 nm, and 1536/1575 nm) [25], Delaney et al. [26] mapped, with reflectance imaging spectroscopy (RIS), the wool and the silk fibres in tapestries. Depending on the spectral range, NIR spectroscopy may offer some information on the degradation of the fibres but only when reaching to the 2200–2300 nm spectral region [4, 27].

When dealing with big data, the examination of every single spectrum represents a highly time-consuming task. Additionally, the NIR spectra broad features hardly offer a straightforward and unambiguous interpretation. Multivariate classification techniques may overcome this problem by building mathematical models, based on the spectral information, and identifying classes or groups by coding their similarities [28, 29]. A consistent number of publications report on NIR spectroscopy and classification techniques for identifying textile fibres, both natural and synthetic, and their blends, particularly for industrial applications [15, 30–33]. For example, Howell and Davis [34] exploited Mahalanobis distances to classify four different synthetic fibres while Jasper and Kovacs [35] neural networks for classifying 17 fibre types. They point out the difficulties to correctly differentiate between cotton and linen. Soft independent modelling of class analogy (SIMCA) was employed to classify non-blended fibres, both natural and synthetic [16], and compared with other methods—partial

least squares discriminant analysis (PLS-DA) [31] and linear discriminant analysis (LDA) [30]. The fibre blends remain the main classification challenge. In this respect, multilinear regressions (MLR), extreme learning machine (ELM), and partial least square (PLS) predicted efficiently the fibre content in blended yarns with both natural and synthetic fibres [32, 36, 37].

The NIR spectroscopy combined with classification techniques is an established approach in industrial domain, whereas in the heritage science field a lower number of publications is available [4, 8, 20, 38]. For this reason, our scope was to compare the performance of two classification techniques, principal component analysis–linear discrimination analysis (PCA-LDA) and soft independent modelling of class analogy (SIMCA), to identify cotton, wool, and silk and their blends in historical textiles in a transportable, fast, and non-invasive way using FORS, in the 1000–1700 nm range, a technique easily accessible to the heritage professionals. We used cotton, wool, and silk samples, unaged and naturally and artificially aged (for 37 years and with Xenon test, respectively), and their respective blends for the training and test datasets. After the model calculation and evaluation, we analysed four historical textiles: three Persian carpets from the nineteenth and twentieth centuries and an Italian seventeenth-century tapestry. We discuss the PCA-LDA and SIMCA methods performance in terms of accuracy, sensitivity, and precision, cross-checking the outcomes with μ -Raman spectroscopy. The results highlight the advantages and limitations of each technique and the complexity of their application to real objects where there are several factors that influence the outcome of the classification.

2 Material and methods

2.1 Sample preparation and artificial ageing

Our training and test datasets included spectra from four silk, four wool, and three cotton types of fibres, the most common yarns in tapestries and carpets production since the ancient times, (Table 1) available in the sample database of the Laboratory of Conservation of Tapestries and Carpets from the Opificio delle Pietre Dure. We subdivided the fibres into three groups according to their characteristics: (a) unaged commercial fibres (UA), (b) fibres undyed (NA-UD) and dyed (NA-D) with natural dyes (yellow, red, blue, purple, and green) following historical recipes (Table S11) naturally aged for 37 years, and c) fibres dyed, with natural and synthetic dyes, and artificially aged (Xenon test) (AA-D). Only for the cotton fibres, undyed samples artificially aged were available (AA-UD). The fibre blends simulate roughly a 50–50% proportion between the two fibre types (silk–wool, cotton–wool, cotton–silk).

The artificial ageing of some of the samples was performed in 1984 using a Xenon test machine model 150 following the standard UNI 7639/1989. The light source produced irradiation in the visible (400–700 nm) and UV (300–400 nm) range (temp. 5500–6500 K), and seven IR filters were used to imitate solar radiation filtered through a window (DIN 54 004). The exposure time was 100 h and the temperature of the black panel was 50 °C; the relative humidity was 40%.

2.2 FORS measurements

Fibre optics reflectance spectroscopy (FORS) measurements were performed with a Zeiss multi-channel spectrometer (MCS) including Vis–NIR (MCS521) and NIR 17 (MCS511) modules, operating in the spectral range 304–1700 nm with a resolution of 3 nm in the

Table 1 Samples used for spectral measurements

Fibre	Aging	Dyeing	Description
Wool	Unaged	Undyed	Commercial wool. 3 threads, S-twist, Nm: 3/175
	Naturally aged	Undyed and dyed	“Venezia” weft. 2 threads, S-twist, Nm: 2/26.22
	Naturally aged	Undyed	“Siena” weft. 2 threads, S-twist, Nm: 2/10000
	Naturally aged	Undyed	Warp “Grosso.” 3 threads, S-twist, Nm: 3/8000
	Artificially aged	Dyed	“Venezia” weft. 2 threads, S-twist, Nm: 2/26.22
Silk	Unaged	Undyed	Commercial silk
	Naturally aged	Undyed	<i>Bourette</i> . Fibre obtained from silk scraps, 2 threads, S-twist, Nm: 2/36
	Naturally aged	Undyed	<i>Shappe</i> . Fibre obtained from silk scraps, 2 threads
	Naturally aged	Undyed and dyed	<i>Ritorta</i> . 6 filaments × 2 threads, S-twist
	Artificially aged	Dyed	<i>Ritorta</i> . 6 filaments × 2 threads, S-twist
Cotton	Unaged	Undyed	Commercial cotton. 3 threads, S-twist
	Naturally aged	Undyed and dyed	<i>Cordonnet</i> threat. Nm 3/54.27 (tex 18.42 × 3)
	Artificially aged	Undyed and dyed	<i>Cordonnet</i> threat. Nm 3/54.27 (tex 18.42 × 3)

visible and 6 nm in the near-infrared (NIR) range. Eleven acquisitions on a slightly different spot were recorded for each sample, using a 45°/0° lighting/detection geometry, a spot size of 3 mm in diameter, and a 100% reflecting reference Spectralon®. Each measurement is a mean of 3 acquisitions on the same spot. Data were processed through an in-house software, providing CIE-L*a*b* coordinates with standard D65 illuminant and 2° observer and a mean spectrum (out of 11 different acquisitions) for each type of sample; the reflectance intensity standard deviation in the range 1000–1700 nm was around 2–5% for cotton, 5% for wool, and 8–14% for silk. It should be pointed out that the rugosity of the textile surface influences the reflectance variability.

2.3 Classification techniques

The performance of PCA-LDA and SIMCA classification techniques was evaluated on two datasets, one containing the single fibres (S) and the other also their mixtures (B) (~50/50 ratio). For modelling, the measurements were divided into the training and test sets (Table 2). The type of sample (*i.e.* dyed or aging type) is not indicated here since it does not interfere in the classification for the given spectral range (1000–1700 nm). The matrix dimensionality is represented in Eq. 1:

$$X = \begin{pmatrix} x_{1,1} & \cdots & x_{1,i} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,i} \end{pmatrix}, \quad (1)$$

where N represents the total number of reflectance spectra (Table 2) in the matrix (X) and *i* corresponds to the 1–117 reflectance values measured in the range 1000–1700 nm.

Table 2 Number of spectra (N) in training and test sets

Data set	Abbreviation	Fibre	Training set	Test set
Single fibres	S	Wool	230	230
		Silk	168	152
		Cotton	87	76
		Total	485	458
Single fibres and their blends	B	Wool	230	236
		Silk	168	160
		Cotton	87	76
		Silk–wool	20	20
		Cotton–wool	18	22
		Cotton–silk	24	21
		Total	547	535

2.3.1 Principal component analysis–linear discrimination analysis (PCA-LDA)

To calculate LDA model with the Chemometric Agile Tool (CAT) [39], we first reduced the datasets size (*i.e.* training, test, and real samples) with PCA. We calculated the principal components (PCs) using the 1000–1700 nm reflectance spectra, and we selected the number of PCs based on the minimum number required to obtain the minimum classification error in the LDA model to avoid over-fitting. We employed 10 and 20 PCs, respectively, for the dataset with single fibres and with single and mixed fibres. We applied standard normal variate (SNV) to all the spectra to correct baseline shifts, intensity variations and reduce systematic error. We employed a column centring correction before the PCA model calculation.

We used the scores obtained from the PCA analysis as input for the LDA model calculation, the latter using five segments for cross-validation and 100 randomizations.

2.3.2 Soft independent modelling of class analogy (SIMCA)

For the SIMCA classification, we exploited the classification toolbox for MATLAB [40]. We calculated the model for each class (cotton, wool, or silk and their mixtures) using the number of PCs required to achieve the lower error rate; autoscaling was performed before the model calculation and validated using a Venetian blind cross-validation method with five segments. SIMCA classification offered four possible results: (a) correctly classified in one group, (b) classified in more than one group with one correct classification, (c) classified in more than one group with no correct classification, (d) and classified as unknown samples (not belonging to any of the groups) [16].

2.3.3 Model performance estimations

The results of the classification are summarised in a confusion matrix where the rows indicate the class to which the samples belong while the columns the class assigned by the model [28]. To estimate and compare the performance of both the classification techniques, we calculate the accuracy, sensitivity, efficiency, precision, and specificity (Eq. S11-5) [28, 41]. Specificity was calculated only for the test set predictions. For the SIMCA model, some of the values were calculated automatically by the toolbox.

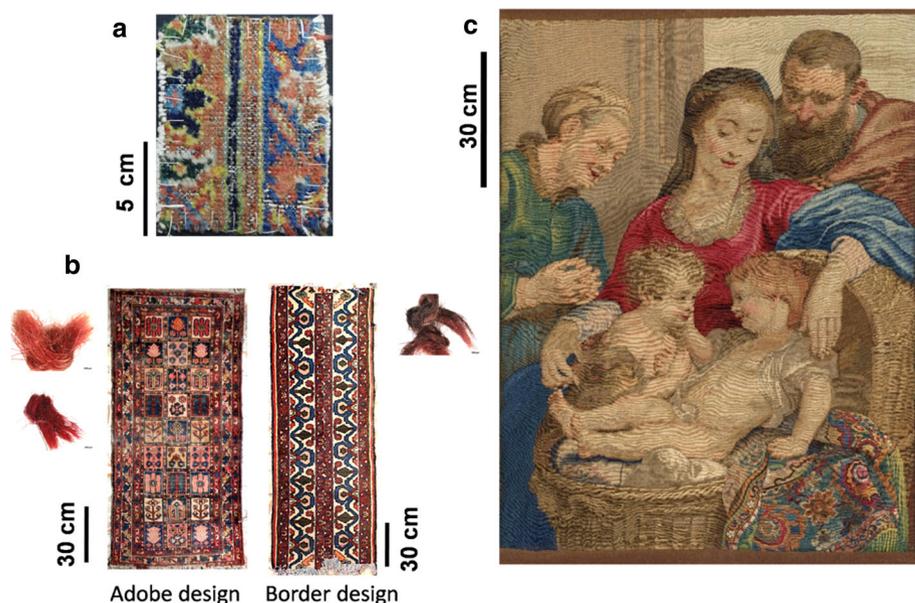


Fig. 1 Historical textiles used for the FORS multivariate fibre classification: **a** nineteenth century carpet fragment, **b** samples coming from the two twentieth century Persian carpets, **c** Italian seventeenth-century tapestry by Pierre Lefèbvre

2.4 Historical samples

After model calculation and evaluation, we classified FORS spectra from three historical Persian carpets and one historical tapestry (Fig. 1) using both models. We studied a fragment of carpet used as a patch for a Ferahan Persian carpet (Farāhān region, Iran, beginning of nineteenth century) from the Museum-Gallery Mozzi Bardini in Florence. Spectra were collected from both sides of the carpet fragment and in different colour areas. Additionally, we studied knot samples from two Persian carpets, an Adobe design carpet from the Bakhtiari-Esfahan region (Iran), dating from 1950 to 1960, and a Border design carpet from the Esfahan region (Iran), dating from 1920. The samples were kindly donated by Dr Ina Vanden Berghe and Zohreh Chahardoli and previously studied at the KIK-IRPA [42]. Finally, we used the spectra obtained from the tapestry (private collection, 1652) made by the French weaver settled in Florence Pierre Lefèbvre (also known as Pietro Fèvère). The tapestry is based on Rubens' painting *Madonna of the Basket* today at the Galleria Palatina in Florence [25].

We divided the data from the historical cases into two matrices, one for the carpets and the other for the tapestry (Table 3). We added some spectra of the reference fibres to achieve a similar distribution to the training set in the multivariate space.

2.5 μ -Raman spectroscopy

The recognition of the fibres in the historical textiles was performed using a μ -Raman Renishaw in Via Raman confocal microscope equipped with a Leica DM2700 optical microscope and a solid-state 785 nm excitation source. After the observation of the sample under the microscope, we focused the laser on a single fibre to perform the measurements in the

Table 3 Number of spectra contained in each dataset from historical samples

Dataset	Spectra	Composition	Additional reference spectra	Total
Tapestry	40	Silk	12 wool	52
Carpet	15	Wool (7) and cotton–wool mixture (8)	24 wool	39

100–3200 cm^{-1} spectral range, using a 1200 l/mm grating, and a thermoelectrically cooled CCD detector (spectral range 400–1060 nm) with a spectral resolution of 1 cm^{-1} per CCD pixel (functional resolution of 3 cm^{-1}). The laser power was 9 mW, with typically 10 s integration times and 20 accumulations. Data were collected with a $50 \times$ long distance magnification (NA Plan = 0.5; theoretical spot size at 785 nm = 0.95 μm) and processed with Wire5.1 and OriginPro software.

3 Results

3.1 NIR spectra

Figure 2 shows the reflectance spectra of cotton, wool, and silk in the range 800–1700 nm (12,500–5882.35 cm^{-1}). The absorption bands and their proposed assignments [22, 26, 30, 31] are summarised in Table 4. Apart from some minor variation in reflectance intensity, the spectra are reproducible and dye and ageing does not influence significantly the spectra in the range studied. The bands 1505 nm, 1537 nm, and 1576 nm allowed discriminating between the two proteinaceous fibres, particularly in the region 1500–1600 nm (7143–6250 cm^{-1}). The band at 1486 nm (6729 cm^{-1}) from the first overtone of the O–H stretching from semi-crystalline cellulose characterises cotton fibres.

On the other hand, the measurements obtained from the mixed samples exhibit contribution of particular features of each fibre (Fig. 2d). Silk–wool blend shows an intensity increase of the band at 1395 nm (7168 cm^{-1}) related to the adsorbed water present on both fibres (Fig. 2b, c and Table 4) [8]. The characteristic bands of silk (1540 nm and 1580 nm) are evident in both the silk–wool and cotton–silk blends, while the band at 1497 nm from cotton is present in its two blends.

3.2 Principal component analysis

We exploited the PCA to explore the data by recording the variance and by reducing their dimension for the LDA model calculation. We calculated 10 principal components (PCs) that explained 99.4% of the variance (Table SI2) of the training set and 99.5% of the test set containing single fibres. The dataset containing both the single fibres and their blends required the calculation of 20 PCs that explained 99.6% of the variance of both the training and test sets. The score plot of the first three PCs (Fig. 3), explaining, respectively, 57.4%, 20.8%, and 10.9% variance of the dataset with the three single fibres, showed an evident clustering phenomenon. PC1 described the difference between the proteinaceous (wool and silk) and the cellulosic (cotton) fibres, while PC2 and PC3 the one between wool and silk (Fig. 3). The dispersion in the multivariate space of some points of silk and cotton is related to the variability in the intensity of the adsorbed water band (1395 nm) in silk and to the

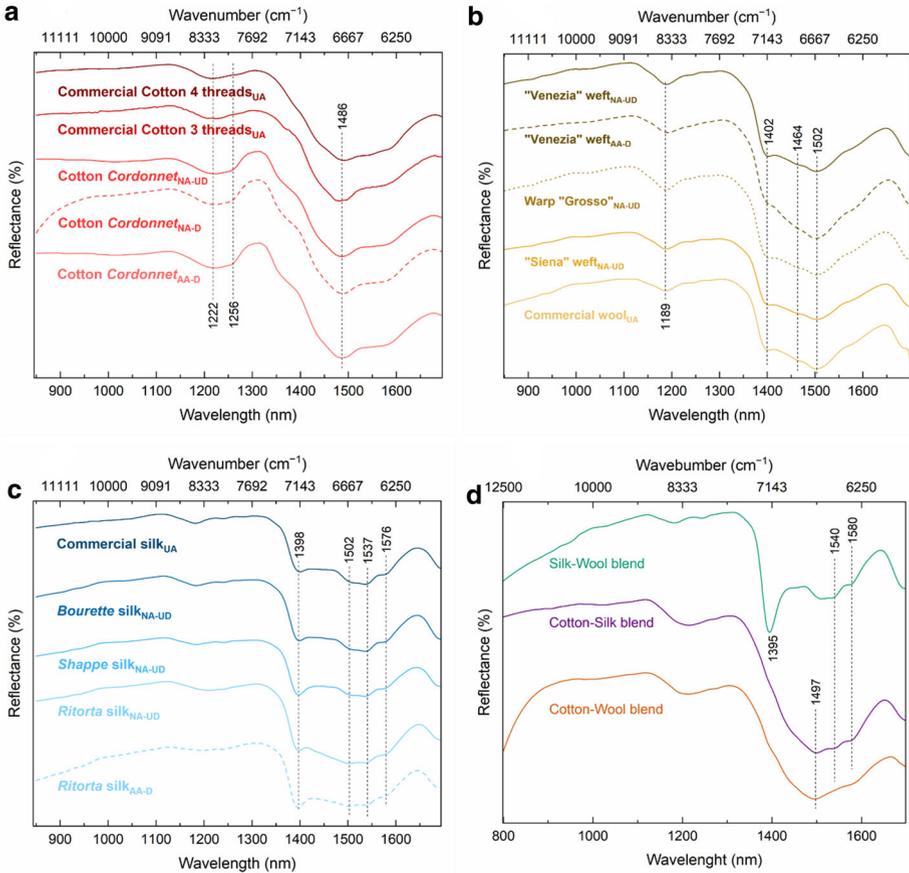


Fig. 2 NIR reflectance spectra, in the 800–1700 nm range, of the different types of **a** cotton, **b** wool, and **c** silk samples, and the spectra measured from their blends. UA = unaged, NA-UD = undyed fibres naturally aged, NA-D = dyed fibres naturally aged, AA-UD = undyed fibres artificially aged, AA-D = dyed fibres artificially aged

spectrum slope at around 1300 nm in cotton measures (Figs. 2a, c) rather than differences in the chemical composition.

When the mixtures of fibres are added, the score plot (Fig. 3b) showed a less defined separation; however, the distribution of the single fibres remained almost unaltered and the different mixtures distributed along the main two fibres (Fig. 3b). There is some overlapping between the mixtures and the single fibres as a result of a similar composition and the effect of a dominant contribution of one fibre in the mixture. Particularly, the measurements obtained from the cotton–silk mixture exhibited a trend towards the wool cluster. These particular issues may complicate their classification.

3.3 Evaluation of classification techniques

Based on error rate values in cross-validation (Fig. S11), we used 10 PCs for both PCA-LDA (error rate 0%) model of the single fibres dataset and 20 PCs for the model including the

Table 4 Band assignment of NIR spectra from silk and wool fibres

Fibre	Bands		Assignment	Reference
	cm ⁻¹	nm		
Wool	6361	1572	4th overtone of tertiary amide	[43]
	6658	1502	O–H from adsorbed water and aminoacids, 1st overtone of N–H	[30, 44]
	6831	1464	1st overtone of N–H, 3rd overtone of C=O	[30]
	7133	1402	1st overtone of O–H stretching vibration of water	[45]
	8410	1189	3rd overtone C–H	[45]
Silk	6345	1576	Amide groups in β -sheet, O–H and N–H overtones and combination bands	[8, 26]
	6506	1537	1st overtone amide A stretching N–H amorphous	[23, 27]
	6658	1502	N–H stretching from Amide A crystalline and 1st overtone stretching Amide II. Amide groups in β -sheet	[26, 27]
	7153	1398	1st overtone of O–H stretching vibration of water	[45]
Cotton	8292	1206	3rd overtone C–H	[45]
	8190	1221	2nd overtone C–H stretching CH and CH ₂ from cellulose	[46, 47]
	7962	1256	2nd overtone C–H stretching CH and CH ₂ from cellulose	[46, 47]
	6729	1486	1st overtone OH stretching from semi-crystalline cellulose	[46, 47]

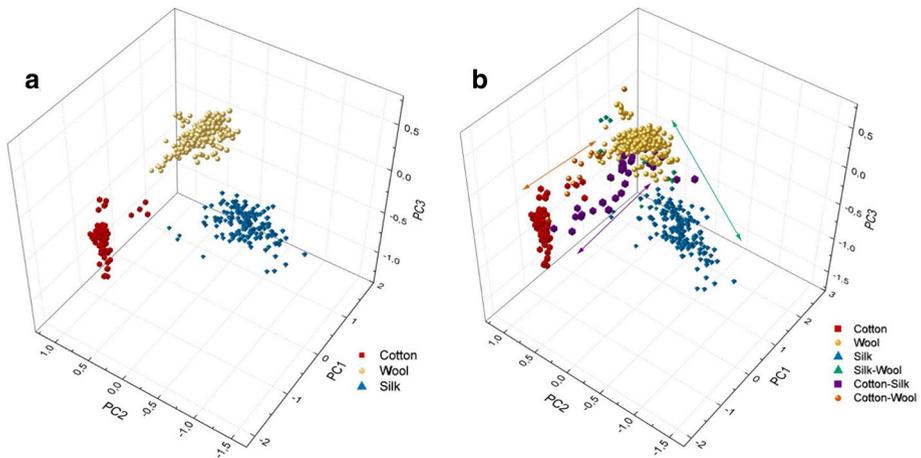


Fig. 3 Score plots PC 1, 2, and 3: **a** single fibres and **b** single fibres and their blends training sets. The first three PCs allowed separating the three fibres while the blends show a less defined separation; however, the distribution of the single fibres remains unvaried while the blend measurements distribute between the two main fibre clusters (arrows)

fibre blends (error rate 5.29%). The SIMCA model showed an average lower error rate for both datasets; however, the advantage of SIMCA technique is the possibility to calculate the model based on a target class, and therefore, the error rate varied according to the target and the number of PCs selected (Table S13). The class with the higher error rate was cotton–wool (6.95%, 11 PCs), while cotton, wool, and silk showed the lower error rates (<1%) for the two datasets.

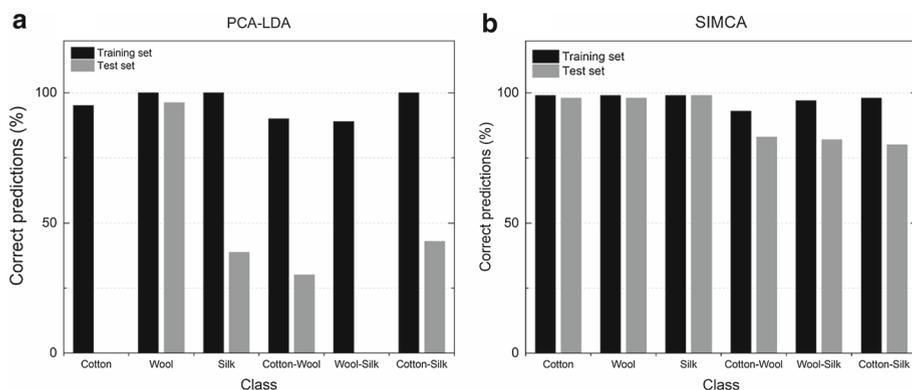


Fig. 4 Correct prediction percentage from **a** PCA-LDA and **b** SIMCA models calculated for the dataset containing single fibres and their mixtures

The PCA-LDA percentage of correct predictions (Fig. 4a) for the training set (confusion matrix in Fig. SI4) decreases from 100 to 95.7% in presence of the mixed samples. In the test set, this drop is even more evident (from 100% to 34.6%). The model works the best for the wool (96.2%) while cotton and wool-silk were the less correctly predicted classes (0%). As shown in Table SI5, the incorrect predictions are related to the contribution of the single fibres in the mixtures and the overlapping of the classes. For example, 52 silk samples were classified as cotton–silk and 1 as wool-silk or 9 wool-silk spectra as wool and 7 as silk. Silk and cotton are the most problematic to differentiate as previously reported by Davis and colleagues [16].

For SIMCA model (Figs. SI6 and SI7), the average correct prediction is comparable for the training set with the single fibres (99.8%) with that containing also mixtures (97.5%) (Fig. 4b). Similar results were obtained from the test sets, the correct predictions dropped minimally from 98% for the single fibre to 91.6% for the mixtures. These results show that the tailored selection of the number of PCs for the SIMCA model calculation based on the target class clearly allows for its better performance.

The differences in terms of accuracy, sensitivity, precision, specificity, and efficiency between the two techniques are negligible in the training set (Fig. SI2), while in the test set SIMCA showed a better performance (Fig. SI2b), especially in case of the mixtures (SIMCA scores 91.71% efficiency while PCA-LDA 56.4%). This is as a result of the detailed selection of adequate number of PCs for the model calculation for each class. SIMCA is also more sensitive (87.16% vs 34.96% of PCA-LDA), meaning it performs better in correctly assigning samples to a target class and excluding samples that do not make part of it. As mentioned before, the dye or aging did not influence any of the classifications.

3.3.1 Classification of historical samples

The spectra from the four historical textiles were divided in two datasets: one for the carpets and another for the tapestry, as input for developed models. Their performance varied considerably according to the matrix analysed. PCA-LDA model of single fibres correctly classified all the samples from the seventeenth century tapestry, which was the simplest case of study since the points analysed were made exclusively using silk fibres. Figure 5 shows the Mahalanobis distances between each tapestry spectra and the mean distribution of each

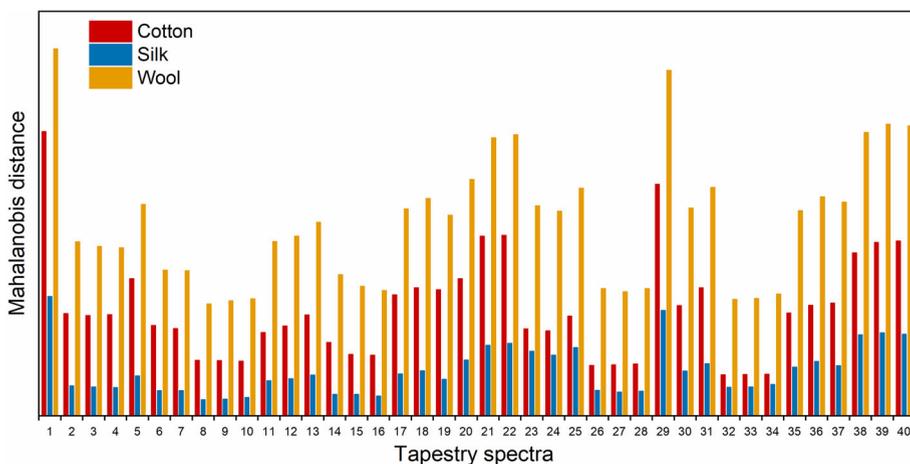


Fig. 5 Mahalanobis distances from the tapestry spectra obtained from the PCA-LDA model of single fibres

class. Silk is the closest class to all of the spectra. On the contrary, the SIMCA model showed a lower performance (Fig. SI8), it classified correctly only 3 of 40 spectra as silk, while 26 were classified as wool and 11 were not classified in any class.

The carpet dataset was more complex. The detailed examination of the nineteenth century Persian carpet fragment using μ -Raman indicated that the carpet structure was made with cotton (Fig. 6a), as suggested by the bands at 1377 cm^{-1} , 1339 cm^{-1} from $\delta(\text{CH}_2)$, 1150 cm^{-1} from $\nu(\text{CC})$ ring breathing, 1122 cm^{-1} , 1097 cm^{-1} from $\nu(\text{COC})$ glycosidic link, and 379 cm^{-1} from $\delta(\text{CCC})$ ring deformation [48]; the knots, on the other hand, are made of wool (Fig. 6b), the characteristic bands at 1652 cm^{-1} from the $\nu(\text{CO})$ from the amide I, 1447 cm^{-1} , 1315 cm^{-1} from the $\delta(\text{CH}_2)$, 1003 cm^{-1} from $\nu(\text{CC})$ aromatic, and 938 cm^{-1} assigned to $\rho(\text{CH}_3)$ terminal and $\nu(\text{CC})$ [14]. The NIR spectra indeed showed the contribution of both cotton and wool (Fig. 6c). Indeed, cotton was used in the structure of Persian carpets (see Fig. 6d) from the nineteenth century and it is still used today for their production [49, 50].

Because of the complexity of the sample, we used the models calculated considering the single fibres and their mixtures. The PCA-LDA model (Fig. 7) showed promising results since seven of the spectra from the 19th-century carpet fragment were classified as cotton and one spectrum was classified as cotton–silk. Regarding the knot samples, two spectra were correctly classified as wool, two were classified as wool–cotton, and three as silk.

The SIMCA showed better results (Fig. 7) when using the model calculated only with single fibres. It classified as wool seven of the 15 samples. Particularly, the spectra obtained from the knot samples. The other 8 samples were not assigned to a class. On the contrary, the model based on the spectra containing single fibres and their mixtures failed to classified all the spectra from the carpet, which were not assigned to any class.

4 Discussion

Both techniques have similar performance (100% of correct prediction for PCA-LDA and 99.8% for SIMCA) when the dataset is relatively simple and contains well-defined classes.

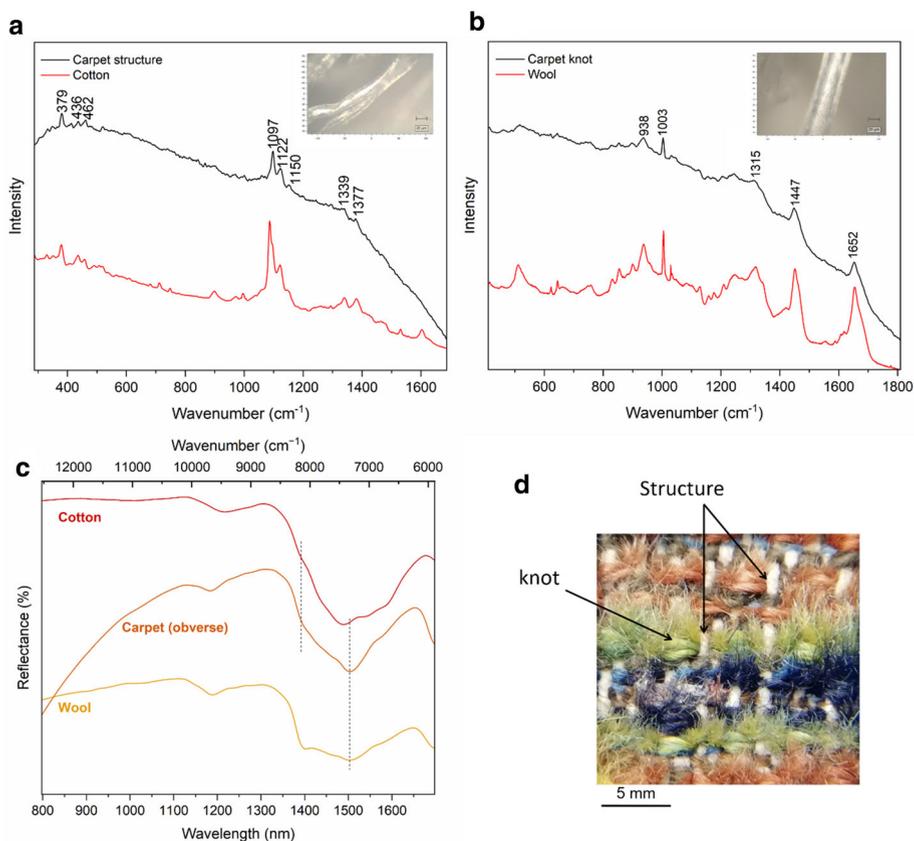


Fig. 6 μ -Raman spectra of the nineteenth century Persian carpet fragment: **a** from the fibres of the carpet structure compared to the cotton reference and cotton references, **b** from the knots fibres compared to wool fibre reference. **c** NIR spectrum from a red area of the nineteenth century Persian compared to wool. **d** Carpet structure

However, for more complex systems, in which the delimiters between classes are not well defined in the multivariate space, PCA-LDA showed a lower performance. The training and test sets results show that the main advantage of SIMCA, in comparison with PCA-LDA, is the possibility to calculate a tailored model for a specific target class, by selecting the number of PCs require to achieve the minimum error rate, thanks to the model calculation based only on sample that belong to the target class. As the training set results suggest, SIMCA may be very useful when the classes have a complex distribution in the multivariate space (e.g. mixed samples).

The application of both techniques to classify mixed fibres in real cases showed some limitations. For a single fibre type, e.g. tapestry, PCA-LDA performed better because the simple definition of delimiters between classes allowed to achieve a higher sensitivity. On the contrary, when the delimiters between classes are not well defined, *i.e.* dataset containing also the fibre mixtures, the performance of PCA-LDA is reduced considerably (Figs. 4 and SI2).

Our results pointed out the problems in the classification in presence of more fibre types. Even though we built models also considering mixtures, the contribution of each fibre varies

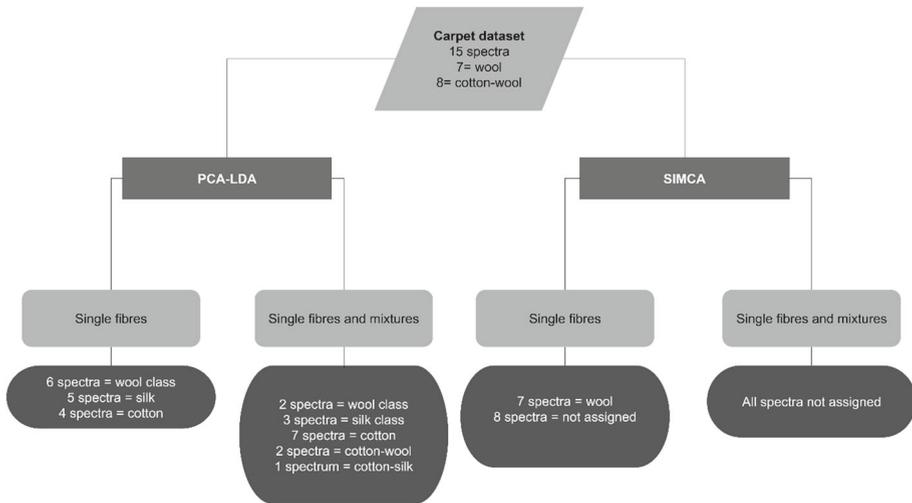


Fig. 7 Scheme of the classification results obtained from the Carpets dataset

according to the position of the spectrometer and the construction of the textile. Therefore, it is very difficult to account for this variability and to control the exact percentage of contribution of each fibre. However, the achieved results are promising as PCA-LDA allowed not only to correctly classify single fibres and also cotton–wool and cotton–silk mixtures (Fig. 7). In contrast, in all the historical samples, SIMCA was unable to correctly classify the sample, probably as a result of a high specificity (Fig. 7). Our future work will thus focus on improving the applicability to real cases by introducing a higher variability of samples to calculate the model.

5 Conclusions

This work has highlighted the advantages and disadvantages of two classification techniques (PCA-LDA and SIMCA) for the non-invasive classification of cotton, wool, and silk fibres and their mixtures in historical textiles using the FORS in the NIR (1000–1700 nm) range. Despite the well-known applicability of these techniques in the industry, a limited number of publications report on this methodology for the identification (e.g. MANOVA or PLS techniques) of textile fibres in heritage science, while PCA, which is not a classification technique, has been used more often to explore the data.

Our approach showed a good performance for both techniques when single fibres are analysed; however, when mixtures are included SIMCA showed the best performance. Despite this good performance on model samples, the study of historical samples revealed complex and showed that it is necessary to calculate a model with a high sensitivity and a lower specificity in order to manage to include the variations present in the real case not considered in the initial model.

The results suggest that PCA-LDA is a powerful tool when studying textiles with a relatively simple structure, while objects with a more complex composition, for example, embroidery or tapestries, require models built considering a higher variability and the influence of

the different material vicinity, thus a SIMCA model with a low specificity but high sensitivity can be the most adequate.

Our results showed that FORS—nowadays accessible to several cultural heritage institutions for various applications (e.g. colour measurements and material identification)—in NIR range, even considering only a reduced number of wavelengths (1000–1700 nm), is a promising tool for the characterization of textile fibres. The non-invasiveness of the technique allows extending the number of measurements and increasing the statistics and representativeness of the results compared to other spectroscopic techniques that require samples. It is a useful exploration tool to understand the homogeneity of the textiles, identified possible modification or unnoticed differences and select and reduce the sampling areas.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1140/epjp/s13360-021-02267-1>.

Acknowledgements We are thankful to Dr. Ina Vanden Berghe and Ms. Zohreh Chahardoli for the carpet samples. This work was supported by Tuscany Region as part of the Intervention program called “CNR4C”, project RS4Art, co-financed with resources of POR FSE 2014-2020 – Axis A Employment, as part of “GiovaniSi.”

Data Availability Statement This manuscript has associated data in a data repository. [Authors’ comment: All data included in this manuscript are available upon request by contacting the corresponding author.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. M.L. Ryder, T. Gabra-Sanders, *Text. Hist.* **123**, 123 (1985)
2. G. Skoglund, B. Holst, and H. Lukešová, *Archaeol. Anthropol. Sci.* 214 (2020).
3. C. Mathias, E. Moffatt, A. Murray, *J. Can. Assoc. Conserv.* **29**, 26 (2005)
4. N. Kissi, K. Curran, C. Vlachou-Mogire, T. Fearn, L. McCullough, *Herit. Sci.* **5**, 13 (2017)
5. R. Nayak, S. Houshyar, A. Khandual, R. Padhye, and S. Fergusson, in *Handb. Nat. Fibres Second Ed.*, edited by R. M. Kozłowski and M. Mackiewicz-Talarczyk, Second Edition (Woodhead Publishing, 2020), pp. 503–534.
6. M. Gleba, S. Harris, *Archaeol. Anthropol. Sci.* **11**, 2329 (2019)
7. J.A. Suomela, K. Vajanto, R. Räisänen, *Stud. Conserv.* **63**, 412 (2018)
8. E. Richardson, G. Martin, P. Wyeth, X. Zhang, *Microchimica Acta* **162**, 303 (2008)
9. J. Liu, D. Guo, Y. Zhou, W. Ziyang, W. Li, F. Zhao, X. Zheng, *J. Archaeol. Sci.* **38**(7), 1763–1770 (2011)
10. Z. Zhu, N. Tse, P. Nel, M. Tobin, *MRS Adv.* **2**, 3939–3949 (2017)
11. D.Q. Balbas, G. Lanterna, C. Cirrincione, M. Ricci, M. Becucci, R. Fontana, J. Striova, *J. Raman Spectrosc. n/a*, 1 (2021)
12. C. Margariti, *Herit. Sci.* **7**(1), 1–14 (2019)
13. L.-L. Cho, *Forensic Sci. J.* **6**, 55 (2007)
14. B. Doherty, C. Miliani, I.V. Berghe, A. Sgamellotti, B.G. Brunetti, *J. Raman Spectrosc.* **39**, 638 (2008)
15. E. Cleve, E. Bach, E. Schollmeyer, *Anal. Chim. Acta* **420**, 163 (2000)
16. C.B. Davis, K.W. Busch, D.H. Rabbe, M.A. Busch, J.R. Lusk, *J. Mod. Phys.* **6**, 711 (2015)
17. K.B. Beć, J. Grabska, H.W. Siesler, C.W. Huck, *NIR News* **31**, 28 (2020)
18. P. Peets, K. Kaupmees, S. Vahur, I. Leito, *Herit. Sci.* **7**, 1–10 (2019)
19. H.S. Rashed, P. Mishra, A. Nordon, D.S. Palmer, M.J. Baker, *Vib. Spectrosc.* **113**, 103205 (2021)
20. H. Zhao, Y. Wang, S. Liu, K. Li, W. Gao, *J. Archaeol. Sci.* **111**, 105026 (2019)

21. Analytical Methods Committee AMCTB No 75, Anal. Methods 5894 (2016).
22. M.A. Maynez-Rojas, E. Casanova-González, J.L. Ruvalcaba-Sil, Spectrochim. Acta Part Mol. Biomol. Spectrosc. **178**, 239 (2017)
23. L. Ding, T. Gong, B. Wang, Q. Yang, W. Liu, R. Pemo, T. Metok, J. Cult. Herit. **47**, 68 (2021)
24. M. Bacci, Archeometriai Műhely **1**, 46 (2006)
25. A. Dal Fovo, J. Striova, E. Pampaloni, A. Fedele, M.M. Morita, D. Amaya, F. Grazzi, M. Cimò, C. Cirrincione, R. Fontana, Microchem. J. **153**, 104472 (2020)
26. J.K. Delaney, P. Ricciardi, L. Glinsman, M. Palmer, J. Burke, Anal. Methods **8**, 7886 (2016)
27. X. Zhang, P. Wyeth, Appl. Spectrosc. **61**, 218 (2007)
28. P. Oliveri, C. Malegori, and M. Casale, in *Encycl. Anal. Sci. Third Ed.*, edited by P. Worsfold, C. Poole, A. Townshend, and M. Miró, Third Edition (Academic Press, Oxford, 2019), pp. 481–486.
29. M. Bevilacqua, R. Nescatelli, R. Bucci, A.D. Magri, A.L. Magri, F. Marini, J. AOAC Int. **97**, 19 (2014)
30. J. Zhou, L. Yu, Q. Ding, R. Wang, Autex Res. J. **19**, 201 (2019)
31. H. Chen, Z. Lin, C. Tan, Microchem. J. **144**, 489 (2019)
32. H. Chen, C. Tan, Z. Lin, Anal. Lett. **53**, 844 (2020)
33. M. Zoccola, N. Lu, R. Mossotti, R. Innocenti, A. Montarsolo, Fibers Polym. **14**, 1283 (2013)
34. H.E. Howell, J.R. Davis, Text. Chem. Color. **23**, 69–73 (1991)
35. W.J. Jasper, E.T. Kovacs, Text. Res. J. **64**, 444 (1994)
36. J. Zhou, R. Wang, X. Wu, B. Xu, Appl. Spectrosc. **7**, 2367 (2017)
37. H. Chen, Z. Lin, C. Tan, Int. J. Chem. Eng. **8**(43), 22593–22600 (2019)
38. I. Cazenobe, M. Bacci, M. Picollo, B. Radicati, G. Bacci, S. Conti, G. Lanterna, and S. Porcinai, in *13th Trienn. Meet. Rio Jan.* (ICOM Committee for Conservation, Paris, France, Rio de Janeiro, 2002).
39. R. Leardi, M. Camillo, and P. Gianmarco, *CAT (Chemometric Agile Tool)* (Società Chimica Italiana, 2021).
40. D. Ballabio, V. Consonni, Anal. Methods **5**, 3790 (2013)
41. J. Xia, J. Zhang, Y. Zhao, Y. Huang, Y. Xiong, S. Min, Spectrochim Acta Part Mol Biomol Spectrosc **219**, 8 (2019)
42. Z. Chahardoli, I.V. Berghe, R. Mazzeo, Herit. Sci. **7**(1), 1–17 (2019)
43. J.J. Workman Jr., Appl. Spectrosc. Rev. **31**, 251 (1996)
44. L.G. Angelini, S. Tozzi, S. Bracci, F. Quercioli, B. Radicati, M. Picollo, Stud. Conserv. **55**, 184 (2010)
45. X. Sun, H. Yuan, C. Song, X. Li, A. Hu, and S. Yu, Microchemical J. **149**, 104018 (2019).
46. M. Schwanninger, J.C. Rodrigues, K. Fackler, J. Infrared Spectrosc. **19**(5), 287–308 (2011)
47. Y. Liu, C. Delhom, B.T. Campbell, V. Martin, Inf. Process. Agric. **30**(1), 30–35 (2016)
48. H.G.M. Edwards, D.W. Farwell, A.C. Williams, Spectrochim. Acta Part Mol. Spectrosc. **50**, 807 (1994)
49. M.L. Eiland, *Oriental Rugs: A Complete Guide*, 2nd edn. (New York Graphic Society, New York, 1976)
50. M.K. Dolatabadi, M. Montazer, M. Latifi, J. Text. Inst. **96**, 1 (2005)