

RESEARCH

Open Access



# NIR spectroscopy in conjunction with multivariate analysis for non-destructive characterization of Xuan paper

Yujia Luo<sup>1,2\*</sup>, Yun Liu<sup>2,3</sup>, Quan Wei<sup>4</sup> and Matija Strlič<sup>2,5</sup>

## Abstract

In the process of conservation mounting, starch paste made from wheat flour is the glue of choice to paste reinforcing strips and backing papers, often Xuan paper, for mending and decorating aged and damaged paper-based Chinese artworks. To keep objects intact and select appropriate conservation materials, this research explored the applicability of near-infrared (NIR) spectroscopy coupled with multivariate analysis for the characterization of contemporary unsized Xuan paper with and without starch. In particular, partial least squares (PLS) regression was used to predict the degree of polymerization (DP) of the paper, i.e., one of the most important properties of paper materials, and principal component analysis (PCA) was used to detect starch and to distinguish between papers with different amounts of starch. Using 12 contemporary unsized Xuan papers, the NIR-PLS method for DP prediction was validated, and the best-performed model was generated using the logarithmic transformation of  $DP^{-1}$  as a response variable, with root mean square error of prediction (RMSEP) of DP 128. The NIR-PCA method was also found to be applicable to separate starch-free papers and papers treated with starch, and the results indicate that the higher the content of starch, the better PCA classification performs. This research provides supporting data for the non-destructive characterization of Xuan paper-based objects and differentiates between Xuan paper before and after treatment with starch.

**Keywords** Xuan paper, Wheat starch, Conservation mounting, Near-infrared spectroscopy, Heritage science

## Introduction

Conservation mounting is an important intervention treatment for mending aged and damaged paper-based Chinese artworks. Xuan paper and starch paste are routinely applied in many heritage institutions [1–3]. The process has been changed little from that used during the Ming Dynasty (1368–1644 CE) [1]. With centuries of practice, this treatment is able to preserve the appearance of painting and calligraphy [1, 4], where wheat starch paste (WSP) serves as a versatile adhesive to paste reinforcing or lining strips, backing papers, and decorative silks, given its thickening and gel-forming properties [4, 5]. Gelatinization of wheat starch starts when mixed with water and heated at ~70 °C, during which the highly ordered structure of starch granules is disrupted by

\*Correspondence:

Yujia Luo  
yujialuo@snnu.edu.cn

<sup>1</sup> School of Materials Science and Engineering, Engineering Research Center of Historical Cultural Heritage Conservation, Ministry of Education, Shaanxi Normal University, Xi'an 710119, Shaanxi Province, China

<sup>2</sup> Institute for Sustainable Heritage, University College London, Gower Street, London WC1E 6BT, UK

<sup>3</sup> Leuven Core Facility for Heritage Science and Digitisation Technologies, KU Leuven, Charles DeBériotstraat 26, 3000 Leuven, Belgium

<sup>4</sup> Sichuan Museum, Wuhou District, Chengdu 610072, Sichuan Province, China

<sup>5</sup> Heritage Science Laboratory, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Kongresni Trg 12, 1000 Ljubljana, Slovenia



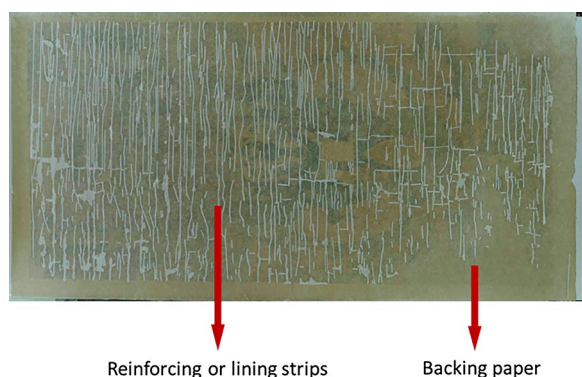
© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, 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/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

breaking hydrogen bonds, leading to irreversible phase transition to a highly viscous liquid state [6].

For the preparation of WSP, the common ratio of mixing water and starch is 5:1 as used in most contemporary Chinese heritage institutions [7, 8]. The mixture is stirred constantly over high heat to obtain a thick paste, after which it is diluted to a desired thickness, depending on the purpose of conservation mounting [1, 8]. Either pure wheat starch or wheat flour can be used; the former is cooked directly, while in practice the latter is used by the conservators to cut down the expenses [7, 8]. The wheat flour is mixed with water to make dough with hand kneading, and then the wheat starch suspension is washed off from the dough. This process disposes of insoluble gluten [1, 9].

WSP, associated with Xuan papers, is used for two steps in the process of conservation mounting [10]. WSP in lower concentrations is used to paste backing and framing papers on the verso of aged or damaged paintings in order to provide structural support [7, 8], as shown in Fig. 1. Slightly thicker paste is used for reinforcing strips (~3–5 mm wide) on the backing papers of the paintings to fill holes or mend cracks [7, 8]. A higher amount of WSP makes the paper stiffer, making it vulnerable to further mechanical damage, thus it is only applied to small areas, as discussed by Catcher et al. [1]. The concentrations of WSP are often decided based on subjective judgments of experienced conservators [1, 7, 8]. Sometimes, pure wheat starch can also be used to prepare starch paste in practice.

Given that conventional destructive characterization methods cannot be applied to valuable paper artifacts, the feasibility of near-infrared (NIR) spectroscopy coupled with multivariate analysis was investigated, as



**Fig. 1** The verso of an aged Chinese painting (Shaanxi Normal University), where the white slim paper strips are used to fill holes and mend cracks (reinforcing or lining strips), while the yellowish paper attached to the verso of the painting is the backing paper, made of dyed Xuan paper

previously explored for other cellulosic materials [11–14]. NIR spectra are dominated by combination and overtone vibrations of bands like C–H, N–H, O–H, S–H etc. [15], therefore, it is of great significance in the characterization of organic materials. However, due to the overlapping peaks, multivariate data analytical techniques are usually used to extract useful information for data interpretation [16].

Partial least squares (PLS) regression is of particular interest for quantitative analysis of NIR spectra [17]. It is known that the degree of polymerization (DP) is one of the most important parameters to evaluate the fitness-of-use of paper materials. For DP determination, viscometry is the most frequently used technique, while other methods, e.g. size exclusion chromatography, require specialized instrumentation and skills [18–20]. Successful application of NIR-PLS methods for the estimation of paper DP has been developed for European rag paper [21, 22] and nineteenth–twentieth-century Chinese paper [13]. The model performance can be affected if multiple analytes are present, leading to less satisfactory predictions for historic samples because of their higher complexity in terms of both chemical and physical properties, including, e.g., non-cellulosic organic substances such as rosin, protein, or mold [23, 24]. It was reported that models developed using the NIR-PLS regression method show better performances for ground, powder, and pulp fibers. NIR spectra can also be used for qualitative classification of paper properties, such as detecting the presence of lignin [13, 23] or classifying fiber groups [24], where principal component analysis (PCA) is one of the most frequently used methods.

In this study, the material properties of contemporary unsized Xuan paper with and without the application of starch are explored. Given the NIR-PLS method presents good performance for homogeneous samples. This method was used for DP prediction of the Xuan paper without additional treatments, and therefore in practice, developed models can be used to evaluate paper durability in a non-destructive and efficient way by curators. On the other hand, the NIR-PCA method may help curators to detect the presence of starch when visual assessment is uncertain, or spot tests are not possible.

## Methodology

### Sample preparation

In this research, 12 Xuan papers, denoted as XP (of three grades) were used, as well as 4 imitation Xuan papers (3 papers made of pure paper mulberry or mixed with straw fibers, denoted as GP, and 1 paper made of bamboo fiber mixed with straw fibers, denoted as No. 12). These were provided by conservators from the Cultural Heritage Conservation Centre, Sichuan

Museum, Chengdu, China. The descriptions and the basic paper properties were made available by Luo et al. [10]. In NIR-PLS, the four imitation Xuan papers were regarded as outliers and were also removed from the sample set for PCA analysis.

The conservators manually prepared 12 wheat starch-treated Xuan papers following the routinely used WSP recipe for paste backing. The samples (12×12 cm) with and without WSP treatment were cut and weighed in triplicate. The mass difference between each paper pair (with and without WSP), was regarded as the starch content, with an average of  $11.3 \pm 2.4\%$  and a coefficient of variation (CV) of 21.6%. These samples were used as reference starch-treated samples. In practice, as suggested by the conservators, ~10% starch content is used for pasting the backing paper, and starch content at ~20% is preferred for strengthening, which guided sample preparation. This treatment aims to minimize the effect of pH on NIR-PCA model performance.

Prior to the preparation of starch-treated model samples, all Xuan paper samples were deacidified to have similar pH of ~8 using a  $\text{Ca}(\text{HCO}_3)_2$  solution prepared by transferring 0.01 mol of  $\text{Ca}(\text{OH})_2$  to a soda bottle, followed by the addition of 1 L of deionized water and  $\text{CO}_2$  pressurization. The bottle was shaken until the hydroxide dissolved and the resulting solution was used for deacidification. Upon immersion of a sheet in a plastic tray, the sheet was kept in the bath for 10 min and then air dried. This treatment aims to exclude the effect of paper pH on model performance.

Model papers were then treated with pure wheat starch (Sigma Aldrich, Merck). 5 g of pure wheat starch (PWS) (Sigma-Aldrich, Merck) was mixed in 50 mL deionized water (Millipore, Molsheim), then the suspension was heated for 30 min with continuous stirring at 80 °C, until a creamy thick paste was obtained. When the thick WSP turned translucent and began to thicken, it was diluted to 100 mL. The samples (2×2 cm) taken from the 12 Xuan papers were immersed in the diluted paste for ~1 min, then taken out and sandwiched between two sheets of Whatman No.1 filter paper to remove excess solution and achieve a homogeneous coating. This step was repeated several times until  $10 \pm 2\%$  and  $20 \pm 2\%$  (w/w%) concentrations of starch were obtained. This was determined by weighing the samples before and after coating and drying. To make wheat flour starch (WFS) paste, commercially available wheat flour was mixed with water to make dough by kneading by hand, and then the starch suspension was washed away, leaving the insoluble gluten. The starch suspension was kept overnight in the refrigerator (4 °C), the supernatant was then poured off and ~10 mL of the suspension was used to prepare the starch paste following the above procedure. 12 Xuan papers treated

with 10% PWS and WFS and 12 Xuan papers treated with 20% PWS and WFS were prepared.

#### pH

The standard cold extraction procedure [25] was applied to evaluate pH homogeneity of the deacidified samples, modified to reduce sample consumption [26].  $1.0 \pm 0.1$  mg of a sample was extracted in 100  $\mu\text{L}$  of deionized water (Millipore, Molsheim) and soaked overnight. A Mettler Toledo SevenGo pro™ pH/Ion Meter (Columbus) with a micro-combined glass electrode (Mettler Toledo Inlab® Micro 51343160) was used to perform measurements. pH homogeneity was evaluated by triplicate measurements for each paper sheet, with the average uncertainty at 0.1 pH unit.

#### Degree of polymerization

On the basis of the standard [27], the intrinsic viscosity of samples was determined using a modified viscometer. ~20 mg of a sample was dispersed in 10 mL of deionized water and mixed with 10 mL of CED solution (1 mol  $\text{L}^{-1}$ , Merck). The weight of cellulose was corrected by subtraction of the contents of water and ash, as reported in the study of Luo et al. [10]. The limiting viscosity number was calculated using Wetzel-Elliott-Martin's equation. DP was then calculated using the Mark-Houwink-Sakurada equation:  $\text{DP}^{0.85} = 1.1[\eta]$ . Duplicate measurements were performed to obtain average efflux times for non-degraded papers. Given the good homogeneity of this paper set (CV < 2%), one measurement was carried out for papers after degradation under two sets of environmental conditions: 90 °C and 30% relative humidity (RH) for 10, 20, 30, and 40 days; and 60 °C and 70% RH for 4, 8, 12 and 16 weeks for XP3, XP5, and XP10, where 24 degraded samples were obtained. The three types of papers were selected as representative samples for the three grades of Xuan paper, i.e., cotton-like, pure, and extra-pure grades.

#### Near-infrared (NIR) spectroscopy

NIR spectra were recorded for 12 non-degraded and 24 degraded Xuan papers for DP prediction using the PLS method, and NIR spectra were collected to detect starch presence using the PCA method for 12 Xuan papers without starch treatment, 12 Xuan papers treated with 10% PWS and WFS, and 12 Xuan papers treated with 20% PWS and WFS. A UV–VIS–NIR Labspec 5000 spectrometer (Analytical Spectral Device, USA) was used to collect NIR reflectance spectra. The spectrometer is configured with three separate detectors: a 512-element silicon photo-diode array detector for the spectral region 350–1000 nm (resolution: 3 nm), and two TE-cooled extended range InGas photodiodes for spectral regions

1000–1830 nm and 1830–2500 nm (resolution: 10 nm). This instrument was calibrated with Spectralon® Multi-Component Wavelength Calibration Standard (WCS-MC-010, Labsphere, US).

Spectra were taken at 0° angle with 100 scans using a fiber-optic probe (diameter 1 mm) to interface with the spectrometer from three spots on a paper sheet, and the averages were used for further analyses. A Spectralon® White Diffuse Reflectance Standard (99%, Labsphere, US) was applied to calibrate the instrument before each spectrum was taken, and it was also used as a sample background for spectra collection.

### Multivariate analysis

NIR spectral range in 1000–2300 nm was used for further analysis, where wavelength range 2300–2500 nm was removed due to low signal-to-noise ratio. The cropped spectra were treated by the 1st derivative Savitzky-Golay (SG) algorithm [28] and standard normal variate (SNV) or multiplicative scatter correction (MCS) to optimize the quality of the data before further analysis was conducted.

After pre-processing, PCA was performed using MATLAB (R2018b) to reduce variances. For PLS, each sample spectrum was randomly allocated into calibration and validation sets, in which the calibration set comprised 2/3 of the pre-processed spectra, and the spectra of the remaining samples were used to perform independent validation. PLS models were optimized by performing ten-fold cross-validation on the calibration sets and selecting wavelength ranges that led to the best performance in validation. PLS factors were selected according to the number that presents the minimum of the root mean square error of cross-validation (RMSECV).

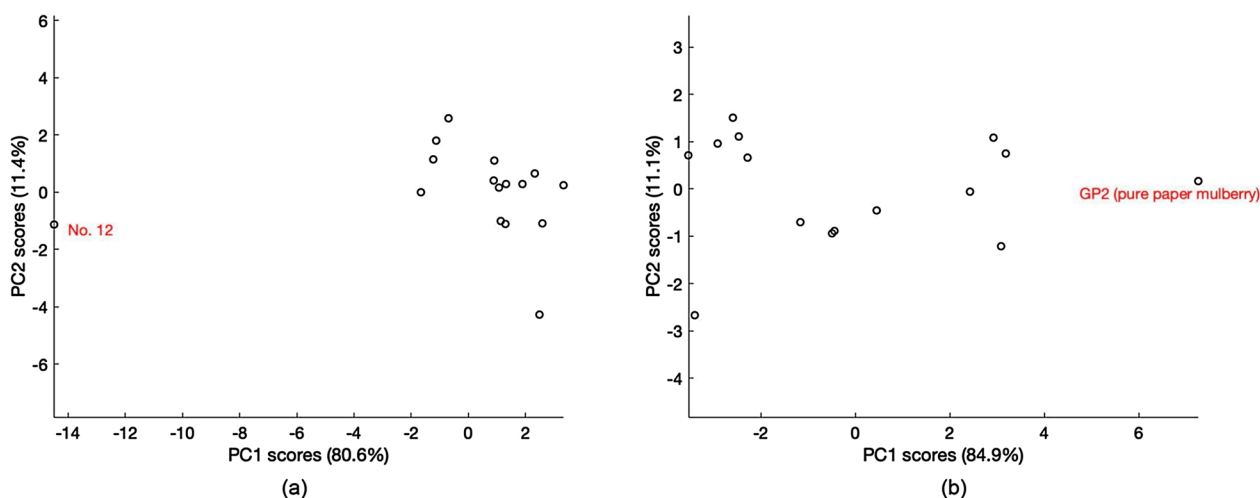
### Characterization

The Fourier transform infrared (FTIR) spectra were collected in the frequency region between 4000 and 400  $\text{cm}^{-1}$  with an attenuated total reflection (ATR) accessory equipped with a diamond cell, where 64 scans were obtained and averaged using a spectral resolution of 4  $\text{cm}^{-1}$  (Nicolet IS10, USA). A SU3500 scanning electron microscopy (SEM, Hitachi, Tokyo) operated at 2 k eV was used to capture images for paper samples, prior to which the samples were coated with platinum with an MC1000 ion sputter coater (Hitachi, Tokyo) for 90 s at 5 kV, in order to improve sample conductivity.

## Results and discussion

### NIR-PLS analysis

Prior to PLS, PCA was conducted to detect outliers using the entire sample set ( $n=16$ ), using pre-processed NIR spectral data as commonly used in previous studies [13, 29]. As shown in Fig. 2a, No. 12 was well separated from the rest of the samples and the first two PCs explained 92% of the variability. This might reflect the fiber composition of the outlier sample, therefore, this sample was removed from the sample set for PLS modeling. Further PCA performed on the rest of the samples ( $n=15$ ) indicated that GP2 (made of pure paper mulberry) differed from the rest (Fig. 2b). To ensure the quality of modeling, GP2 and the other two GP papers (GP1 and GP3) were also removed from the sample set. The remaining 12 Xuan papers, non-degraded as well as degraded, were used for further PLS analysis. Although the size of the sample sets is small, this still conformed to the American Society for Testing Materials (ASTM) standard [30], which indicates that  $\sim 30$  samples can be used to develop feasibility models.



**Fig. 2** Scores plot of the first two PCs for **a** the entire paper set ( $n=16$ ), and **b** for the paper set excluding No. 12 ( $n=15$ )

According to Lambert–Beer’s law, NIR spectral response should correlate with the concentration of bonds [31, 32], therefore, the relationship between DP, used as the response variable, and the concentration of chemical bonds was explored. The amount of three characteristic bonds in cellulose, related to DP, are as follows:

$$n_r = N \cdot DP^{-1} \tag{1}$$

$$n_g = N - N \cdot DP^{-1} \tag{2}$$

$$n_{OH} = 3N + 2N \cdot DP^{-1} \tag{3}$$

**Table 1** Comparison of cross-validation ( $R^2_{CV}$ ), root mean square error of cross-validation (RMSECV),  $R^2$  of cross-validation ( $R^2_p$ ), root mean square error of prediction (RMSEP), and numbers of PLS factors used for models with different transformations of DP for Xuan papers before and after thermal degradation

	DP	DP <sup>-1</sup>	ln(DP <sup>-1</sup> )
$R^2_{CV}$	0.72	0.83	0.85
RMSECV	136	128	114
$R^2_p$	0.67	0.81	0.82
RMSEP	168	144	128
Number of factors	8	8	8
Variance in DP explained by the 1st PLS factor (%)	52.7	54.3	55.0

All models were developed using the same wavelength range (1276–2065 nm), window width at 51 nm, as well as the same pre-processing to the spectra (SNV and 1st derivative Savitzky-Golay filter)

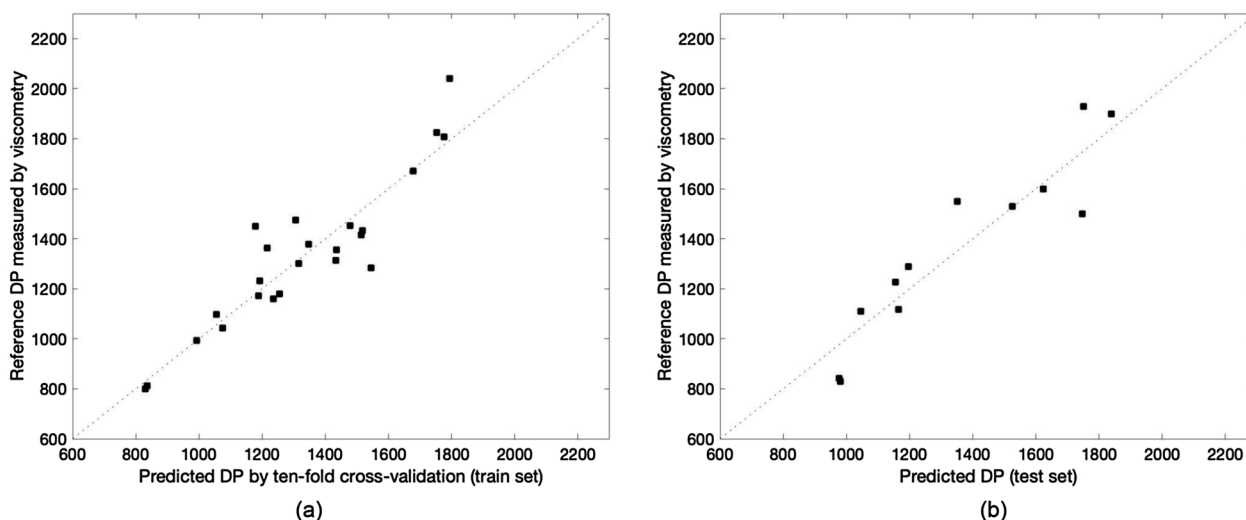
where  $n_r$  is the concentration of reducing end groups in cellulose (mol/g),  $n_g$  is the concentration of  $\beta$ -1,4 glycosidic bonds between the monomers in cellulose (mol/g),  $n_{OH}$  is the concentration of –OH of cellulose (mol/g),  $N$  is the concentration of monomers (mol/g), and DP is the average number of monomers in a cellulose chain.

In these relationships, reciprocal DP is correlated with concentrations of chemical bonds in cellulose and is then used to correlate the spectroscopic response. Also, the logarithmic form of DP<sup>-1</sup> can be attempted for modeling, since it is a common transformation to correct for non-linearity [14]. Our study is consistent with this (Table 1 and Fig. 3), as the logarithmic transformation of DP<sup>-1</sup> was found to be the best-performing model, with RMSEP of DP 128 (CV at ~8%), which is higher than the typical viscometry measurement error but still in consistent with the reported RMSEPs for paper DP prediction, which are DP 69–690 [14]. Further evaluation of the errors is discussed later. A good correlation between the reference and modeled DPs in both training and test sets indicated the potential of PLS regression to model the DP for samples before and after thermal degradation, across the DP range of 500–2200.

To gain insights into errors of DP predicted by the NIR-PLS regression method, the root mean square error of prediction (RMSEP) is taken as the general error to assess the quality of PLS models for prediction of paper DP, where

$$MSEP = RMSEP^2 \tag{4}$$

MSEP stands for the mean squared prediction error, which is defined as [33]:



**Fig. 3** Correlations between the predicted DP obtained using NIR-PLS regression and reference DP measured viscometrically for **a** training and **b** test data sets of Xuan paper set before and after thermal degradation. The dashed line ( $y = x$ ) was used as a reference



$$\text{MSEP} = E[(f(x) - y)^2] \quad (5)$$

where  $f(x)$  is the fitted model developed on the basis of reference DP data in the training set, and  $y$  is DP data measured using a viscometer for test set samples. Based on the bias-variance decomposition, this equation is derived as [33]:

$$\text{MSEP} = \sigma^2 + \text{Bias}^2[f(x)] + \text{Var}[f(x)] \quad (6)$$

in which  $\sigma^2$  is the variance of DP for reference samples in the test set,  $\text{Bias}^2[f(x)]$  is the residual model error between the best-fitted PLS model and the true model for DP, and  $\text{Var}[f(x)]$  is the variance in predicted DP of the test sample set [34, 35].

Given the good homogeneity of this Xuan paper set, DP was measured twice for Xuan papers ( $n=12$ ) before degradation and measured once for papers after degradation ( $n=24$ ). Therefore,  $\sigma^2$  was represented by the average standard deviation of DP (measured in duplicate) for Xuan papers before degradation at DP 26. The variance of the predicted DP was estimated using the average standard deviation of 12 duplicates randomly selected from the test sample sets, giving DP at 124. Given that it is difficult to directly calculate the bias of the PLS regression model to the true model, the percentage contribution of  $\text{Bias}^2[f(x)]$  was inferred by subtracting the contributions of  $\sigma^2$  (4.14%) and  $\text{Var}[f(x)]$  (93.24%) from 100%, which is 2.63%.

For unsized Xuan paper set before and after degradation used in this research, ~97% of the total variance comes from the samples and ~3% comes from the squared bias of the PLS regression model, indicating that the NIR-PLS modeling is capable of predicting paper DP with good quality. Variance in prediction, contributing the most to the total variance estimated by MSEP, can be caused by various sources [12, 14]. It is reported that

instrumentation and measuring procedures might be the main cause of this variance [14]. The results indicate that the NIR-PLS regression method has the potential to predict the DP of Xuan paper, while further research is needed to validate this method and improve the accuracy and precision of the models.

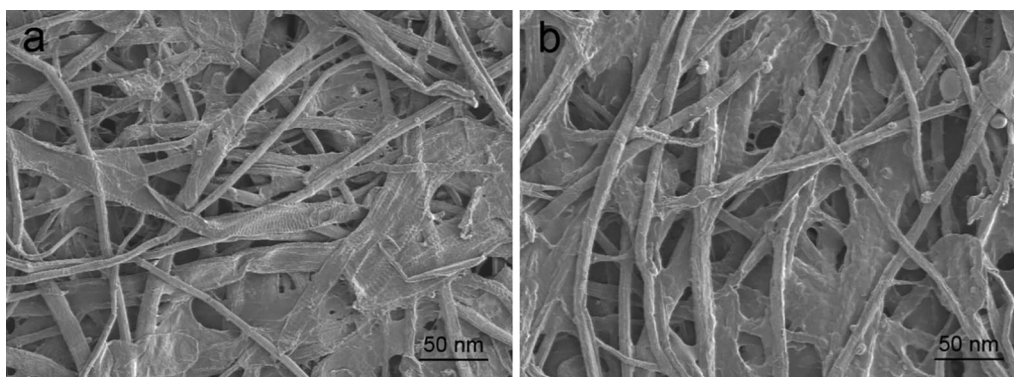
#### NIR-PCA analysis

##### Characterization of samples containing starch

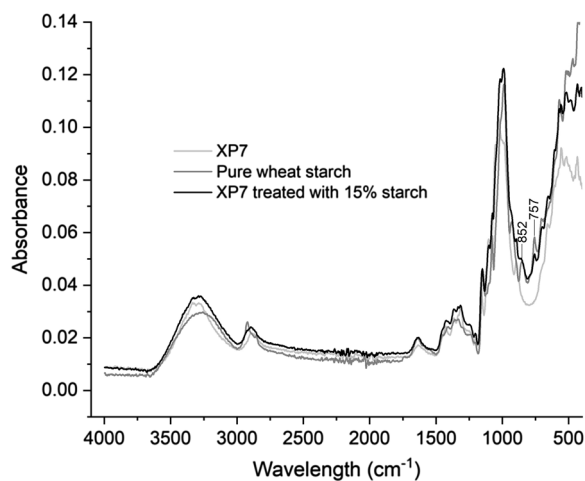
Prior to PCA analysis on separate starch-free and starch-treated contemporary unsized Xuan paper, preliminary characterization was performed. SEM images of XP7, a cotton-like Xuan paper [14], before and after starch treatment were captured and are presented in Fig. 4. Figure 4a shows the shape and size distribution of fibers before treatment with wheat starch paste, while after treatment, the starch granules can be seen Fig. 4b.

The FTIR spectra ( $4000\text{--}400\text{ cm}^{-1}$ ) of XP7, pure wheat starch from Sigma, and XP7 treated with 20% pure wheat starch are presented in Fig. 5. The bands between  $3500$  and  $3100\text{ cm}^{-1}$  are attributed to stretching vibrations of O–H, related to intra- and inter-molecular hydrogen bonds in both cellulose and starch [36, 37]. The band in the region  $3000\text{--}2800\text{ cm}^{-1}$  belongs to methyl C–H stretching vibrations [36, 38]. The peak at  $\sim 1640\text{ cm}^{-1}$  is assigned to the moisture content, indicating that all samples are hygroscopic [38, 39].

Slight differences were found in the fingerprint region at  $1500\text{--}1000\text{ cm}^{-1}$  for the three samples (XP7, starch, and a mixture of XP7 and starch). The band at  $\sim 1000\text{ cm}^{-1}$  is associated with the single helix crystalline structure, related to hydrogen bonds of anhydroglucose unit in all samples [38, 40]. It is worth noting that obvious differences were observed in the vibrational region from  $950$  to  $500\text{ cm}^{-1}$  in XP7 treated with starch, where new peaks at  $852$  and  $757\text{ cm}^{-1}$  can be observed in samples of starch and XP7 treated with starch, as the



**Fig. 4** SEM images of XP7 **a** before and **b** after impregnation with 20% pure wheat starch



**Fig. 5** FTIR spectra of XP7, pure wheat starch from Sigma, and XP7 treated with 20% pure wheat starch from Sigma

bands are attributed to the ‘secondary starch structure’ that primarily originates from vibrations of the skeletal mode of  $\alpha$ -1,4 glycosidic linkage, the skeletal modes of the C–C stretch and the skeletal modes of the pyranose ring [41].

#### NIR-PCA analysis

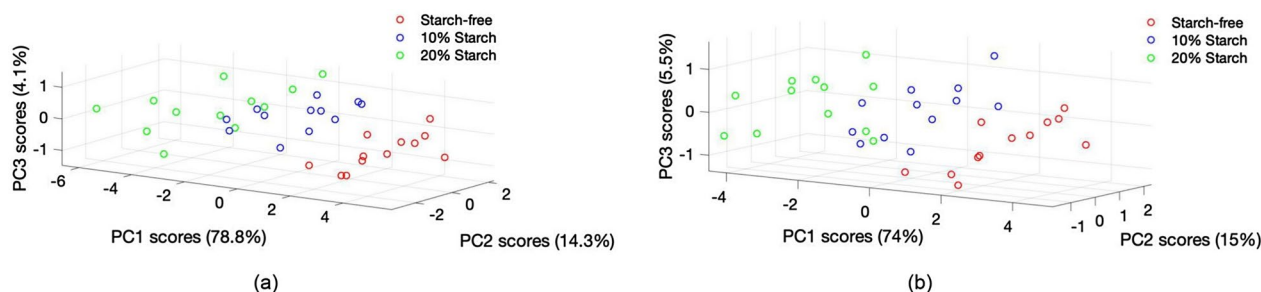
As discussed in the Introduction section, it is assumed that starch presence may affect the IR spectra of paper. To test whether this could be used to develop a qualitative method for the identification of starch in paper, PCA was used as a method of highlighting differences between the samples [42], since it is widely used to reduce the dimensionality of large data sets, transforming the original data into principal components (PCs), which summarize linear combinations of original data and contain the greatest proportion of the total variance for better interpretation [43, 44], as previously used to examine the presence of lignin [13, 23] and different fibers [24] in paper.

As a preliminary test,  $10 \pm 2\%$  and  $20 \pm 2\%$  WSP (made of pure wheat starch and wheat flour starch) were applied to the 12 Xuan paper samples, which were then used to perform PCA. The selected NIR spectral region used for analysis was 1300–2000 nm as this is the region featuring molecular vibrations associated with C=C and C=O double bonds. The spectra were smoothed and normalized using 1st derivative Savitzky-Golay filter and standard normal variate (SNV). The scores plots, as shown in Fig. 6, reveal clustering of samples with common features, i.e., with similar concentrations of starch, showing negative values for samples treated with 20% starch and positive PC1 scores for starch-free samples. The scores plots suggest that PC1 reflects starch content, regardless of the type of starch (PWS and WFS).

#### Conclusions

This paper investigated the development of a NIR spectroscopic method coupled with PLS to characterize contemporary unsized Xuan paper, including papers made of fibers of blue sandalwood and straw with different ratios. Also, the feasibility of PCA was investigated to differentiate papers treated with different concentrations of starches.

It was found that it is possible to predict paper DP for Xuan paper using log-transformed  $DP^{-1}$  as the response variable in PLS analysis, across the DP range of 500–2200. The best-performed model shows that RMSEP was DP 128 ( $\sim 8\%$ ), higher than the measurement errors of viscometry but still in the range of reported RMSEPs for similar studies. The error might be due to diverse fiber furnishes. More samples should be used to verify model applicability. NIR spectra in conjunction with PCA can be used to separate starch-free papers and papers treated with 10% and 20% starch. The scores plots show that PC1 is mainly responsible for this separation, and no obvious difference has been found among the two types of starches (pure wheat starch and wheat flour starch) in the PCA analyses. This enables us to use NIR-PCA to



**Fig. 6** Scores plots for the 12 Xuan papers **a** starch-free, treated with 10% PWS and treated with 20% PWS, **b** starch-free, treated with 10% WFS and with 20% WFS

identify the presence of starch when visual assessment is uncertain, or spot testing is not possible, which may help curators and conservators to detect starch in a fast and non-invasive way.

#### Acknowledgements

The authors are grateful to the Sichuan Museum, Sichuan Province, China for having provided the samples used in this study. Special thanks go to the Fundamental Innovation Project in the School of Materials Science and Engineering (SNNU).

#### Author contributions

YL (Yujia Luo), YL (Yun Liu), WQ, and MS jointly designed the experiment. YL (Yujia Luo) carried out all the experiments and data analysis. All the authors discussed the results and contributed to the manuscript. All authors have read and approved the final manuscript.

#### Funding

This research was financially supported by the EPSRC Centre for Doctoral Training in Science and Engineering in Arts, Heritage and Archaeology (SEAHA) (EP/L016036/1), the Fundamental Research Funds for the Central Universities (GK202304013) and Shaanxi Key Research and Development Program of China (2024GX-YBXM-560). Further funding by the Slovenian Research and Innovation Agency is acknowledged (N1-0271, J4-3085, J7-50226, P1-0447, IO-E012).

#### Availability of data and materials

The datasets used and/or analyzed in the current study are available from the corresponding author on reasonable request.

#### Declarations

#### Competing interests

The authors declare that they have no competing interests.

Received: 19 March 2024 Accepted: 21 May 2024

Published online: 30 May 2024

#### References

- Catcher S, Chang G, Zhu Q. The problem of Chinese paper reinforcement strip repairs on a set of four hanging calligraphic scrolls. *J Inst Conserv*. 2017;40:49–63. <https://doi.org/10.1080/19455224.2016.1267655>.
- Xu K, Wang J, He Q. The influence of alum in alum gelatin solution on cellulose, calcium carbonate and gelatin in Xuan paper. *Spectrosc Anal*. 2018;38:1829–33.
- Sugiyama K, Qiu J, Hakamata H. Paper lining: techniques based on knowledge and experience. *Stud Conserv*. 2014;59:5145–8. <https://doi.org/10.1179/204705814X13975704318830>.
- Gregorová E, Pabst W, Boháčenko I. Characterization of different starch types for their application in ceramic processing. *J Eur Ceram Soc*. 2006;26:1301–9. <https://doi.org/10.1016/j.jeurceramsoc.2005.02.015>.
- Bertoft E. Analysing starch structure. In: starch in food: structure, function and applications. Amsterdam: Elsevier; 2004. p. 57–96.
- Atwell WA, Hood LF, Lineback DR, Varriano-Marston E, Zobel HF. The terminology and methodology associated with basic starch phenomena. *Cereal foods world*. 1988;33:306.
- Li H. Preparation and application of traditional wheat starch paste. In: [online]. 2011. [http://blog.sina.com.cn/s/blog\\_5b07938f0100xadp.html](http://blog.sina.com.cn/s/blog_5b07938f0100xadp.html).
- Li R. How to make wheat starch paste? In: Conserv. mounting. 2011. [http://blog.sina.com.cn/s/blog\\_5b07938f0100xadp.html](http://blog.sina.com.cn/s/blog_5b07938f0100xadp.html).
- Belard R, Higuchi H, Perry J. Furunori (aged wheat starch paste): challenges of production in non-traditional settings. *J Inst Conserv*. 2009;32:31–51. <https://doi.org/10.1080/19455220802630735>.
- Luo Y, Cigić IK, Wei Q, Strlič M. Characterisation and durability of contemporary unsized Xuan paper. *Cellulose*. 2021;28:1011–23. <https://doi.org/10.1007/s10570-020-03554-3>.
- Roggo Y, Chalou P, Maurer L, Lema-Martinez C, Edmond A, Jent N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J Pharm Biomed Anal*. 2007;44:683–700. <https://doi.org/10.1016/j.jpba.2007.03.023>.
- Ozaki Y, McClure WF, Christy AA. Near-infrared spectroscopy in food science and technology. Hoboken, N.J.: Wiley-Interscience; 2007.
- Brown N, Lichtblau D, Fearn T, Strlič M. Characterisation of 19th and 20th century Chinese paper. *Herit Sci*. 2017;5:47. <https://doi.org/10.1186/s40494-017-0158-x>.
- Liu Y, Fearn T, Strlič M. Quantitative NIR spectroscopy for determination of degree of polymerisation of historical paper. *Chemom Intell Lab Syst*. 2021. <https://doi.org/10.1016/j.chemolab.2021.104337>.
- Kramer R. Chemometric techniques for quantitative analysis. Boca Raton: CRC Press; 1998.
- Siesler HW. Near-infrared spectroscopy principles, instruments, applications. Weinheim: Wiley-VCH; 2002.
- Naes T, Isakson T, Fearn T, Davies T. A user-friendly guide to multivariate calibration and classification. Chichester: NIR Publications; 2002.
- Zou X, Uesaka T, Gurnagul N. Prediction of paper permanence by accelerated aging I. Kinetic analysis of the aging process. *Cellulose*. 1996;3:243–67. <https://doi.org/10.1007/BF02228805>.
- Zou X, Uesaka T, Gurnagul N. Prediction of paper permanence by accelerated aging II. Comparison of the predictions with natural aging results. *Cellulose*. 1996;3:269–79. <https://doi.org/10.1007/BF02228806>.
- Strlič M, Kolar J. Ageing and stabilisation of paper. Ljubljana, Slovenia: National and University Library; 2005.
- Strlič M, Cséfalvayová L, Kolar J, Menart E, Kosek J, Barry C, Higgitt C, Cassar M. Non-destructive characterisation of iron gall ink drawings: not such a galling problem. *Talanta*. 2010;81:412–7. <https://doi.org/10.1016/j.talanta.2009.12.017>.
- Trafela T, Strlič M, Kolar J, Lichtblau DA, Anders M, Mencigar DP, Pihlar B. Nondestructive analysis and dating of historical paper based on IR spectroscopy and chemometric data evaluation. *Anal Chem*. 2007;79:6319–23.
- Mahgoub H, Bardon T, Lichtblau D, Fearn T, Strlič M. Material properties of Islamic paper. *Herit Sci*. 2016;34:1–14.
- Oriola M, Možir A, Garside P, Campo G, Nualart-Torroja A, Civil I, Odlyha M, Cassar M, Strlič M. Looking beneath Dalí's paint: non-destructive canvas analysis. *Anal Methods*. 2014;6:86–96. <https://doi.org/10.1039/c3ay41094c>.
- TAPPI T 509 om-11. Hydrogen ion concentration (pH) of paper extracts (cold extraction method). 2011.
- Strlič M, Kolar J, Kočar D, Drnovsek T, Šelih V-S, Susič R, Pihlar B. What is the pH of alkaline paper. *e-Preservation Sci*. 2004;1:35–47.
- ISO 5351. Pulps—Determination of limiting viscosity number in cupri-ethylenediamine (CED) solution. 2010.
- Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem*. 1964;36:1627–39. <https://doi.org/10.1021/ac60214a047>.
- Henniges U, Schwanninger M, Potthast A. Non-destructive determination of cellulose functional groups and molecular weight in pulp hand sheets and historic papers by NIR-PLS-R. *Carbohydr Polym*. 2009;76:374–80. <https://doi.org/10.1016/j.carbpol.2008.10.028>.
- ASTM E1655–05. Standard practices for infrared multivariate quantitative analysis. *ASTM Int*. 2012;05:1–29. <https://doi.org/10.1520/E1655-17>.
- Swinehart DF. The Beer-Lambert law. *J Chem Educ*. 1962;39:333–5. <https://doi.org/10.1021/ed039p333>.
- Mirabella FM. Modern techniques in applied molecular spectroscopy. Chichester: Wiley Interscience; 1998.
- Hastie T, Tibshirani R, Friedman J. Elements of statistical learning: data mining, inference, and prediction. New York: Springer; 2009.
- Marsili-Libelli S. Environmental systems analysis with MATLAB®. New York: Taylor & Francis; 2016.
- Bibby J, Toutenburg H. Prediction and improved estimation in linear models. Chichester: Wiley; 1977.
- Kizil R, Irudayaraj J, Seetharaman K. Characterization of irradiated starches by using FT-Raman and FTIR spectroscopy. *J Agric Food Chem*. 2002;50:3912–8. <https://doi.org/10.1021/jf011652p>.



37. Van Soest JGG, De Wit D, Tournois H, Vliegthart JFG. Retrogradation of potato starch as studied by fourier transform infrared spectroscopy. *Starch - Stärke*. 1994;46:453–7. <https://doi.org/10.1002/star.19940461202>.
38. Liang CY, Marchessault RH. Infrared spectra of crystalline polysaccharides. I. Hydrogen bonds in native celluloses. *J Polym Sci*. 1959;37:385–95.
39. Ferreira-Villadiego J, García-Echeverri J, Vidal MV, Pasqualino J, Meza-Castellar P, Lambis-Miranda HA. Chemical modification and characterization of starch derived from plantain (*Musa paradisiaca*) peel waste, as a source of biodegradable material. *Chem Eng Trans*. 2018;65:763–8. <https://doi.org/10.3303/CET1865128>.
40. Sevenou O, Hill SE, Farhat IA, Mitchell JR. Organisation of the external region of the starch granule as determined by infrared spectroscopy. *Int J Biol Macromol*. 2002;31:79–85. [https://doi.org/10.1016/S0141-8130\(02\)00067-3](https://doi.org/10.1016/S0141-8130(02)00067-3).
41. Fan D, Ma W, Wang L, Huang J, Zhao J, Zhang H, Chen W. Determination of structural changes in microwaved rice starch using Fourier transform infrared and Raman spectroscopy. *Starch - Stärke*. 2012;64:598–606. <https://doi.org/10.1002/star.201100200>.
42. Smith MJ, Holmes-Smith AS, Lennard F. Development of non-destructive methodology using ATR-FTIR with PCA to differentiate between historical Pacific barkcloth. *J Cult Herit*. 2019;39:32–41. <https://doi.org/10.1016/j.culher.2019.03.006>.
43. Marchioro G, Daffara C. PCA-based method for managing and analyzing single-spot analysis referenced to spectral imaging for artworks diagnostics. *MethodsX*. 2020;7:100799–807. <https://doi.org/10.1016/j.mex.2020.100799>.
44. Ortiz R, Ortiz P, Colao F, Fantoni R, Gómez-Morón MA, Vázquez MA. Laser spectroscopy and imaging applications for the study of cultural heritage murals. *Constr Build Mater*. 2015;98:35–43. <https://doi.org/10.1016/j.conbuildmat.2015.08.067>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.