



Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Quantitative NIR spectroscopy for determination of degree of polymerisation of historical paper

Yun Liu^{a,b,*}, Tom Fearn^c, Matija Strlič^{a,b}^a Institute for Sustainable Heritage, University College London, UK^b Museum Conservation Institute, Smithsonian Institution, Washington D.C., USA^c Department of Statistical Science, University College London, London, UK

ARTICLE INFO

Keywords:

Cellulose
Degree of polymerisation (DP)
Error evaluation
Multivariate analysis
Near-infrared (NIR) spectroscopy
Heritage science

ABSTRACT

This paper discusses the development of a near infrared (NIR) spectroscopic method coupled with multivariate analysis to characterise historical paper. Specifically, partial least squares (PLS) regression was used to predict one of the most important properties of paper as a condition indicator – degree of polymerisation (DP). Supported by a set of model cellulose samples, the NIR-PLS method for DP prediction was validated and the modelling approach that led to the best prediction of DP of paper was established. The coefficient of variation of the NIR-PLS models were found to be approximately 8% and 20% of the DP of model cellulose and historical paper, respectively. The variance of the reference DP, the variance of the predicted DP, and the model bias were identified as the main sources of the total expected generalisation error of prediction. For both model cellulose and historical paper, the variance of the predicted DP by the NIR-PLS models contributed the most to the total error of prediction. This suggests that improving the instrumentation and the operation procedure is essential to improve model performance. Furthermore, the effect of water content of the samples on model performance was investigated. The model for historical paper was proven to be robust to relative humidity fluctuations between 30% and 70%, indicating the applicability of the model for collection surveys in a range of environments.

1. Introduction

The determination of degree of polymerisation (DP) is of great significance to assess the condition of polymeric materials in cultural heritage [1]. It is one of the most important molecular properties that correlate with the mechanical strength of polymers [2]. But DP is difficult to measure directly, specifically of papers of historical importance. The techniques typically used in chemistry, such as membrane osmometry, size exclusion chromatography, viscometry, and mass spectrometry, can be time consuming, inaccurate for the DP ranges involved, or require specialised instrumentation and skills [3]. On the other hand, given that significant value, including aesthetic, scientific, social, and economic, is always associated with heritage objects, substantial sampling required for destructive analytical methods is rarely an option. In heritage science, a comparably accurate non-destructive method for DP determination is highly desirable.

Quantitative near-infrared (NIR) spectroscopy provides a non-

destructive alternative to chemical analysis. Given the complexity of NIR spectra, multivariate analysis is often used to provide a correlation-based quantitative interpretation. In multivariate analysis, the spectral responses to chemical and physical properties of a sample set are modelled based on the measurement of small absorbance changes occurring at multiple wavelengths. Among several linear multivariate methods, partial least squares (PLS) regression has been the most important one for quantitative NIR analyses [4]. PLS constructs factors that capture spectral variability as well as correlating with the reference data, and is usually effective in achieving high accuracy of predictions [5]. In recent years, the NIR-PLS method has gained progresses in the analysis of complex multicomponent mixtures where the accuracy is comparable or even better than the conventional wet chemistry methods [6–9], which is especially promising for heritage materials.

A number of authors have attempted the application of NIR-PLS to analyse DP and DP-related properties, such as molecular weight (Mw) and viscosity of polymers for a range of historical and model cellulosic

Abbreviations: CoV, Coefficient of variation; DP, Degree of polymerisation; MSEP, Mean square error of prediction; NRMSE, Normalised root mean square error; PLS, Partial least squares; RMSECV, Root mean square error of cross validation; RMSEP, Root mean square error of prediction; RMSE, Root mean square error.

* Corresponding author. Museum Conservation Institute, Smithsonian Institution, Washington D.C., USA.

E-mail address: LiuY@si.edu (Y. Liu).

<https://doi.org/10.1016/j.chemolab.2021.104337>

Received 19 January 2021; Received in revised form 10 April 2021; Accepted 5 May 2021

Available online 9 May 2021

0169-7439/© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

materials. The performances of the models are summarised in Table 1. Normalised root mean square error (NRMSE) for each model was calculated as:

$$\text{NRMSE} = \text{RMSE} \cdot (y_{\max} - y_{\min})^{-1}, \quad (1.1)$$

where RMSE is root mean square error, y_{\max} and y_{\min} are the maximum and minimum reference values in the property of interest.

With the available data as shown in Table 1, a clear line can hardly be drawn between the model performance for historical samples and model samples, although historical samples generally exhibit far more complexity in both chemical and physical characteristics. It is noticeable that the results obtained using model paper in ground, powder and pulp differ significantly from those obtained using model paper sheets. This indicates that the performance of NIR-PLS models can be affected significantly by physical properties. It signifies the fact that NIR spectra represent a combination of molecular vibrations, optical properties of the instrument, and the instrument-sample interactions. Material complexity can in theory complicate the development of successful NIR-PLS models.

For historical paper, the model performances vary widely. This can be partially explained by the wide range of raw materials and manufacturing processes of the investigated materials, covering Western paper, Chinese paper and Islamic paper. It is worth noting the data of large variability reported specifically for Western paper [11,15,16]. On one hand, exceptional models were developed that outperform all the reported models for both historic and model cellulosic polymers [15]. On the other hand, difficulties in developing acceptable models were reported for similar materials [16]. Despite this divergence, an instrument using NIR-PLS models to predict DP of paper has been developed and implemented in practice with a relatively high RMSE [11]. These contrasting results for historical western paper may signify the complexity in developing NIR-PLS models for historical materials. Since it is difficult to gain insights based on the limited published data, further research is needed to shed light on the contradictions.

Lack of a clear cause-effect relationship between DP and NIR measurements also complicates model development. Successful quantitative NIR analyses usually require an underlying cause-effect relationship between the analytes and spectral data. In the literature, most successful models based on NIR-PLS have been developed for compositional analysis [6–9]. These models are mostly based on the Beer-Lambert Law, where changes in absorbance are proportional to changes in the concentration of a chemical component. In contrast, DP is not a property that can be clearly correlated to the concentration of a vibrating bond type.

Table 1

Summary of the reported data in the literature on the use of NIR spectroscopy to analyse DP, Mw, and viscosity of cellulosic polymers by PLS regression. LV represents latent variables of the models. RMSE represents root mean square error. Normalised RMSE (NRMSE) is represented using RMSE of prediction (RMSEP) where available, otherwise using RMSE of calibration (RMSEC) or RMSE. R^2 represents the coefficient of determination of the regression used to obtain the RMSEP or RMSEC (if RMSEP is not available).

| Property | Sample type | Data range | Number of LVs | RMSE | | NRMSE ^a | R^2 |
|----------------------------|---------------------------------|------------|---------------|------------------|-------|--------------------|-------|
| | | | | RMSEC | RMSEP | | |
| DP | Model pulp [10] | 943–1757 | 2 | 78 | 96 | 0.11 | NA |
| | Historical Western paper [11] | NA | NA | 400 ^b | | NA | NA |
| | Historical Chinese paper [12] | 475–4475 | 8 | 370 | 690 | 0.17 | 0.52 |
| | Historical Islamic paper [13] | 200–1800 | 3 | 298 | 177 | 0.11 | 0.78 |
| | Historical canvas [14] | 350–2300 | 14 | 270 | 275 | 0.14 | 0.77 |
| Mw (kg·mol ⁻¹) | Historical Western paper [15] | 500–4000 | NA | 79 | 119 | 0.03 | 0.92 |
| | Model paper [16] | 50–475 | 10 | 37 | 23 | 0.06 | 0.93 |
| | Historical Western paper [16] | 75–400 | NA | NA ^c | | NA | NA |
| Viscosity | Historical Western paper [15] | 150–1000 | NA | 24 | 49 | 0.06 | 0.95 |
| | Model cellulose derivative [17] | 350–2250 | 12 | 96 | NA | 0.05 | 0.97 |
| | Model paper [18] | NA | 6 | NA | 131 | NA | 0.60 |
| | Model paper ground [18] | 750–2000 | 3 | NA | 195 | 0.16 | 0.62 |
| | Model paper powder [18] | NA | 6 | NA | 479 | NA | 0.62 |
| | Model pulp [18] | NA | 4 | NA | 96 | NA | 0.54 |

^a NRMSE was calculated using RMSEP when available. Otherwise RMSEC or RMSE was used. The same goes for R^2 .

^b It is not clear whether the reported RMSE is RMSEC or RMSEP.

^c The authors reported difficulties in developing accepted models using individual samples.

This imposes an additional challenge to modelling of DP using NIR-PLS.

The issues of material complexity and the implicit DP-NIR relationship may be partially overcome by large data sets, however, the amount of available historical samples for the development of NIR-PLS models is limited due to resource constraints. In this case, to improve model performance, in-depth understanding of the prediction errors becomes essential. However, discussions of error analysis are rarely found in the literature.

To bridge the gaps in the literature and to lay a solid foundation for future research, this paper addresses the identified challenges and difficulties in the development of NIR-PLS models for historical Western paper. Experiments were designed to investigate the approaches for model development and evaluation. Through a comparison of model paper and historical paper, a plausible underlying relationship between DP and NIR response was explored, sources of prediction errors were analysed, and the robustness of the models to environmental fluctuations was assessed. These analyses not only deepen the understanding of the model development and performance, but also ensure the applicability of the NIR-PLS method to collections in practice.

2. Materials and methods

2.1. Sample sets and reference DP

The sample sets used for multivariate modelling of DP using NIR spectroscopy are summarised in Table 2. A set of model papers was prepared for a controlled feasibility study. Samples from the same sheet of Whatman filter paper No. 1 were degraded in a VWR VENTI-Line® oven (Radnor, US) at 90 °C for up to 5 months. All the samples were hung freely during the degradation and no extra humidity was added to the environment inside the oven. Intrinsic viscosity ($[\eta]$) of each sample was determined based on BS ISO 5351 [19] to calculate DP using the Mark-Houwink-Sakurada equation [20]:

Table 2

Summary of sample sets used for the calibration and evaluation of the multivariate analysis.

| Samples | Number of samples | Max DP | Min DP | CoV |
|------------------|-------------------|--------|--------|-------|
| Model paper | 68 | 2405 | 468 | 1.41% |
| Historical paper | 127 | 4071 | 425 | 2.33% |

where $[\eta]$ is the intrinsic viscosity value in $\text{ml}\cdot\text{g}^{-1}$. Given the homogeneity of the model samples [21], a single determination of DP was carried out. Duplicate determinations of DP were carried out on five randomly selected degraded samples for uncertainty estimation. The uncertainty of DP determination was 55, assessed using the pooled standard deviation. The uncertainty was also expressed as coefficient of variation (CoV), which is 1.41% estimated by the average of the CoV of the five pairs of measurements.

The same method of DP determination was used for the set of historical paper by a group of different researchers. This sample set contained mostly European rag papers from various sources, spanning from 14th to 20th century. Each sample was measured twice, in adjacent areas. The uncertainty of the mean of two DP determinations was assessed using the pooled standard deviation of pairs of DP determinations divided by $2^{1/2}$, which gave DP 29. The average CoV was found to be 2.33%.

2.2. NIR spectroscopy

Spectral responses were collected from the samples in both sample sets using a portable UV-VIS-NIR LabSpec 5000 Spectrometer (Analytical Spectral Devices, USA). The spectrometer is equipped with three separate detectors: a 512-element silicon photo-diode array detector for the spectral region 350–1000 nm (spectral resolution: 3 nm) and two TE-cooled extended range InGaAs photo-diode arrays for spectral regions 1000–1830 nm and 1830–2500 nm (spectral resolution: 10 nm). It was operated in a diffuse reflection mode and was calibrated against Spectralon® Calibrated Multi-Component Wavelength Calibration Standard (WCS-MC-010, Labsphere, US).

The NIR spectroscopy was performed in a well-controlled laboratory environment (23 ± 0.5 °C and $50 \pm 5\%$ RH). A Florilon™ Standard (EFWS-99-02c, Avian Technologies LLC, US) was used for instrument

calibration before each set of measurements was taken. Using a Florilon™ Standard as the sample background during spectra acquisition, spectral measurements of the samples were taken in circular areas of 1 mm in diameter using a fibre-optic probe in close contact with the samples at 0° angle. Full range spectra of 200 scans were obtained from three spots on a single sheet of each sample and the averages were stored and used for further analyses.

Since a Florilon™ was used as the background, it was assumed that little transmission occurred during spectra collection. Although the spectrometer was used in a diffuse reflectance mode, the spectra collected were considered to represent both the surface properties and the bulk properties due to the penetration of NIR radiation through the paper sheets. Fig. 1 shows the spectra of the model paper samples (Fig. 1 (a)) and the historical paper samples (Fig. 1. (b)). For model paper, major variability in reflectance was observed in the visible range and relative homogeneity was observed in the NIR range. For historical paper, variability in reflectance across the whole range of the wavelength was observed, with major differences in spectral shapes observed in the visible range. This is an indication of the large variation in material composition and physical properties of the historical samples. A few historical samples showed reflectance greater than 1, which is likely caused by the fluorescence of the additive materials (fillers, sizing materials, etc.) used in the production process [22].

2.3. Multivariate analysis

Given that NIR was of the main interest, the spectral range from 1000 nm to 2500 nm was used for statistical analysis for both sample sets. 2400 nm–2500 nm were further removed due to low signal to noise ratios. After the truncations, the spectra were treated by 1st derivative algorithms developed by Savitzky-Golay (SG) [23] only for outlier

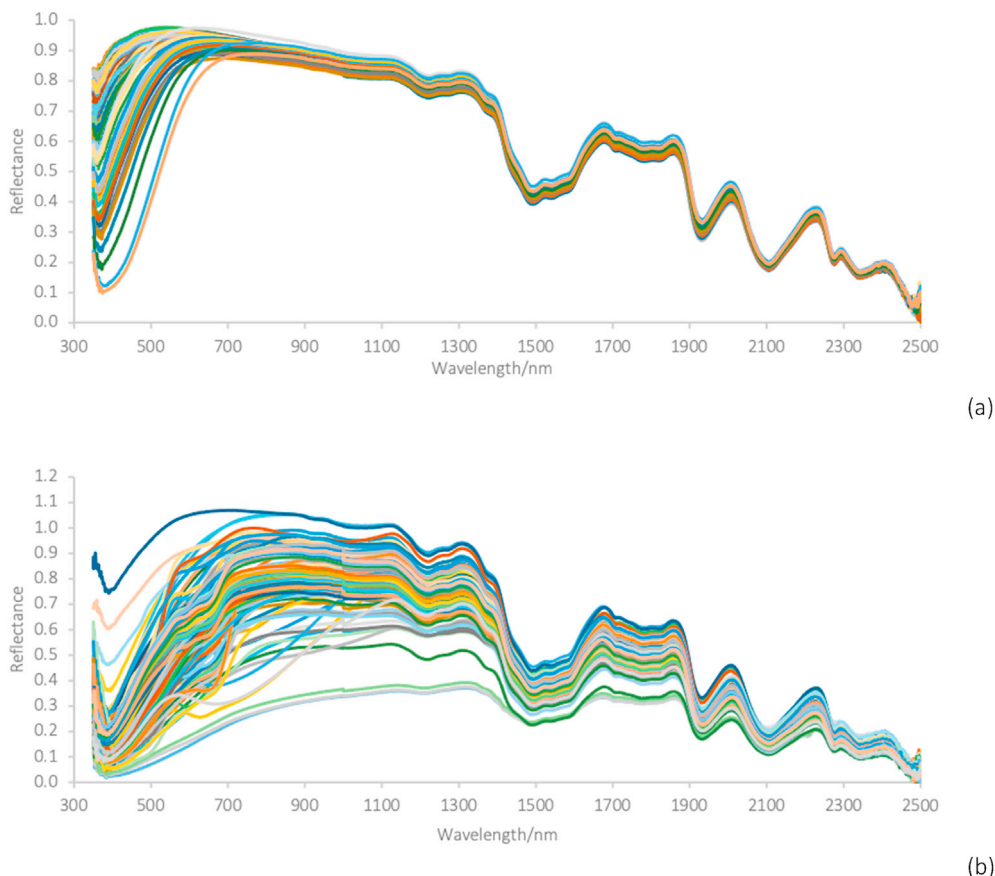


Fig. 1. Reflectance spectra of (a) model paper samples and (b) historical paper samples collected by UV-VIS-NIR LabSpec 5000 Spectrometer.

detection. Spectra showing distinctive visual features were identified as potential outliers and the corresponding samples were examined for confirmation. No outliers were found in the model paper set, whereas nine were identified in the historical paper set and excluded from further analyses. These nine samples displayed unusual material characteristics, included four thick and severely degraded rag paper, one mouldy thick rag paper, one unusually thin rag paper, one exceptionally thick rag paper, one thermal paper and one contemporary paper. Each sample set was then randomly split into two subsets: 2/3 for cross-validated training and 1/3 for independent test.

For the training sets, the spectra were pre-treated using SG derivation and standard normal variate (SNV) [24] in sequence before partial least squares (PLS) regression was carried out to model the relationship between NIR reflectance and DP. Fig. 2 presents the reflectance spectra of the model paper samples (Fig. 2 (a)) and the historical paper samples (Fig. 2 (b)) over 1000 nm–2400 nm pre-processed by 2nd order SG derivation with a window width of 49 and 51 respectively and SNV. The PLS regression models were developed, optimised, and selected based on ten-fold cross-validation on the training sets. The number of PLS factors was determined by choosing the number that gave the first local minimum of the root mean square error of cross-validation (RMSECV). The spectral ranges were estimated based on the ranges reported in the literature for model paper [16] and the overtones of –OH stretching [25]

and C=O [26]. The ranges were optimised by trial and error based on the performance in cross-validation.

The same spectral treatments and PLS model coefficients obtained from the training sets were applied to the test sets. The performance of the selected models was evaluated by the root mean square error of prediction (RMSEP) of the test sets, which was taken as the expected generalisation error in this research. NRMSE of prediction (NRMSEP) was calculated based on Equation (1.1). In cases where transformation was applied to the response variable DP, RMSECV and R^2 of cross-validation (R_{CV}^2) were calculated and evaluated in the transformed scale whereas the RMSEP and R^2 of prediction (R_p^2) were in the original scale. All the analyses were carried out in MATLAB® 2017a with Statistics and Machine Learning Toolbox™; code can be shared upon request.

3. Results and discussion

3.1. Model development

In principle, the cause-effect relationship ensures the true predictive validity of the quantitative NIR spectroscopy as an analytical method. For the quantification of analytes, this relationship is usually based on vibrations of chemical bonds. Given that DP itself is not directly

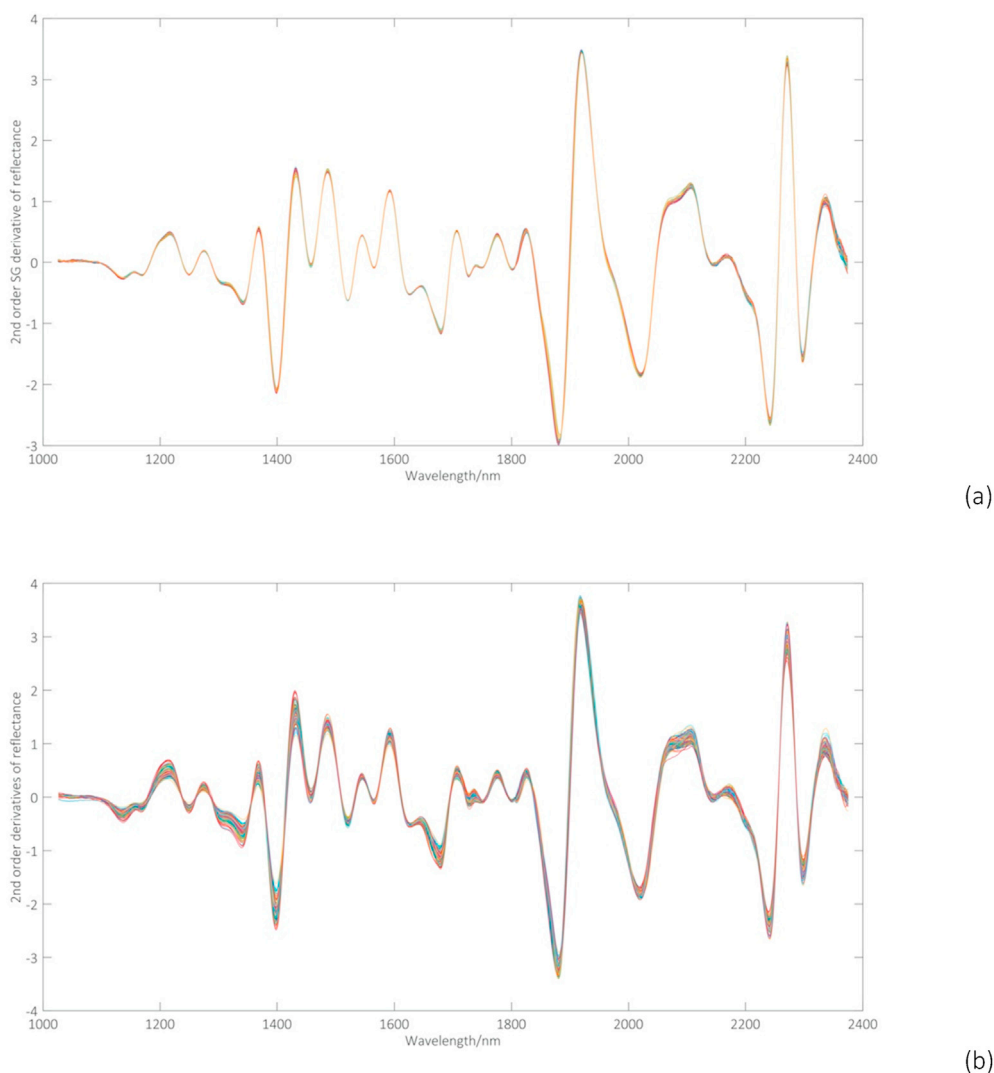


Fig. 2. Reflectance spectra of (a) model paper samples and (b) historical paper samples pre-processed by 2nd order SG derivation with a window width of 49 and 51 respectively and SNV.

represented by a concentration of chemical bonds, the relationship between DP and chemical bonds in cellulose was explored. Under the assumption that the content of oxidised and transformed groups in cellulose fragments are negligible, the amount of three characteristic bonds in cellulose in relation to DP can be estimated as the following:

$$n_g = N - N \cdot DP^{-1}, \quad (3.2)$$

$$n_r = N \cdot DP^{-1}, \quad (3.3)$$

$$n_h = 4N + 2N \cdot DP^{-1}, \quad (3.4)$$

where n_g is the concentration of β -1,4 glycosidic bonds between the monomers ($\text{mol} \cdot \text{g}^{-1}$), n_r is the concentration of reducing end groups ($\text{mol} \cdot \text{g}^{-1}$), n_h is the concentration of $-\text{OH}$ of cellulose ($\text{mol} \cdot \text{g}^{-1}$), N is the concentration of monomers ($\text{mol} \cdot \text{g}^{-1}$), and DP is the average number of monomers in a cellulose chain. In these relationships, the concentrations of bonds are all reciprocally related to DP. This reciprocal relationship still holds when oxidation takes place in the degradation process. Therefore, it was hypothesised that a correlation is more likely to be found between the reciprocal of DP and the measured spectroscopic response.

This hypothesis that the reciprocal transformation leads to better model performance was validated using the model paper set. In addition to DP and DP^{-1} , $\ln(DP)$ was also used as an option for response variable because logarithm is a common transformation to correct nonlinearity. Three different PLS calibration models were built and the results are presented in Table 3. PLS scores, which are linear combinations of the mean-centred reflectance at each wavelength multiplied by the coefficients for each PLS factor, were calculated. In all three cases, the scores of the first PLS factor explained the majority of the variance in the spectral response. The relationship between the score of the first PLS factor and different transformations of DP of the training set was plotted across the model paper samples in Fig. 3. The score of the first PLS factor clearly shows a curved dependency on DP (Fig. 3. (a)). However, this curvature was straightened by a reciprocal transformation of DP (Fig. 3 (b)), which suggests that a reciprocally transformed fit better approximates the linear relationship between DP and PLS factors.

Furthermore, for comparison, predictions obtained from all the models were converted to DP to obtain RMSEPs in the original scale, i.e. DP. As shown in Table 3, a logarithmic transformation of DP decreased the RMSEP by 76% and 44% compared to DP and DP^{-1} , respectively. Therefore, $\ln(DP^{-1})$ was considered as the most appropriate transformation of DP for model calibration, which was simplified as $\ln(DP)$.

Using a logarithmic transformation of DP, PLS regression analysis was carried out for the data sets of model paper and historical paper. Table 4 summarises the pre-processing algorithms, the spectral range, the number of PLS factors, and the performance of the models that were found to have the best predictive capacity based on cross-validation on the training sets. No outliers were excluded during model calibration and validation. Wide window widths were used for the pre-processing by SG derivatives suggesting that the spectra might be noisy. RMSECV are in logarithmic scales whereas RMSEP is in DP for comparison with

Table 3

Comparison of RMSEP, R_p^2 , and R_{CV}^2 for models using different transformations of the responsible variable DP_v of the model paper set. For comparison, all the models were developed using the same spectral range 1000–2400 nm. RMSEPs are all in the original scale.

| | DP | DP^{-1} | $\ln(DP)$ |
|--|------|-----------|-----------|
| RMSEP | 351 | 288 | 200 |
| R_p^2 | 0.55 | 0.96 | 0.86 |
| R_{CV}^2 | 0.70 | 0.89 | 0.88 |
| % variance in the spectroscopic response explained by the 1st PLS factor | 75% | 93% | 84% |
| Number of PLS factors | 1 | 6 | 4 |

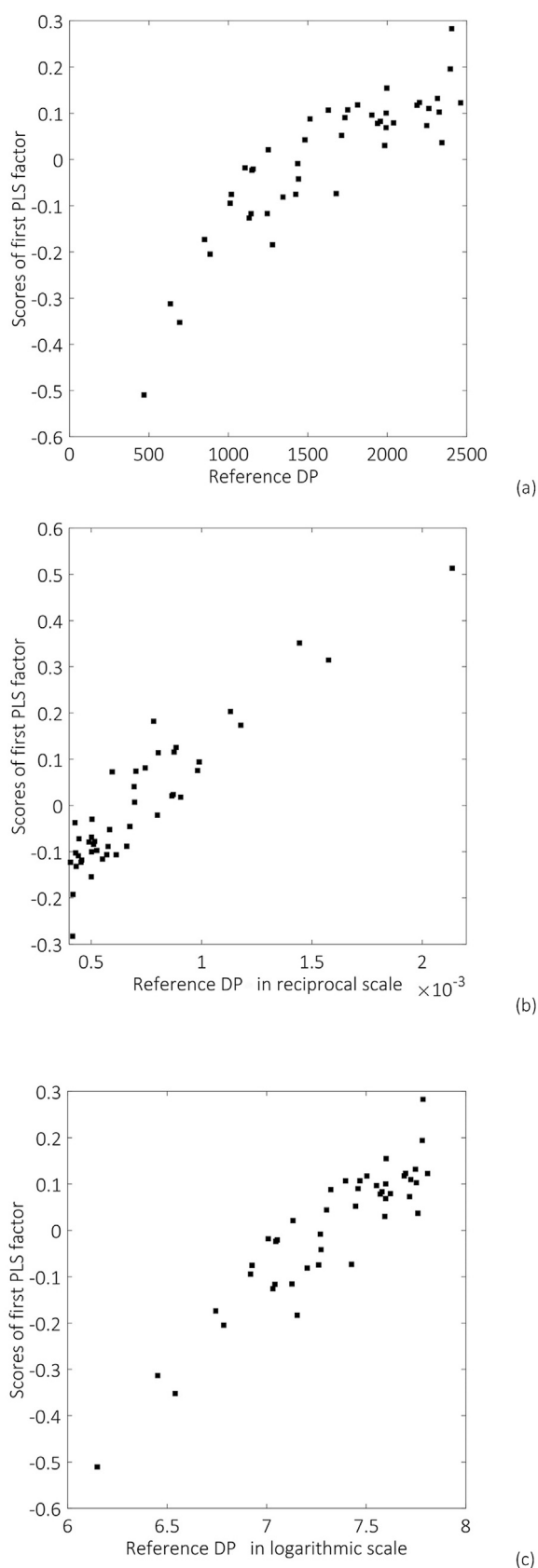


Fig. 3. Comparison of the dependence of the scores of the first PLS factors on (a) DP, (b) DP^{-1} , and (c) $\ln(DP)$ as response variables modelled by PLS regression using the training set of model paper.

Table 4

Summary of the parameters and the results of the PLS regression for DP. RMSECVs were assessed on logarithmic scale whereas individual predictions were converted to DP to obtain RMSEPs on the original scale, i.e. DP. CoV for the training and test sets are presented in the parentheses.

| Data set | Pre-processing algorithms | Spectral range (nm) | Number of factors | RMSECV | R_{CV}^2 | RMSEP | R_p^2 |
|------------------|---|---------------------------------|-------------------|---------------|------------|--------------|---------|
| Model paper | SNV, SG 2nd derivatives (window width 49) | 1130–1620 | 4 | 0.08 (8.81%) | 0.95 | 112 (7.66%) | 0.96 |
| Historical paper | SNV, SG 2nd derivatives (window width 51) | 1290–1720, 1870–1920, 2100–2160 | 8 | 0.20 (22.15%) | 0.82 | 185 (18.38%) | 0.93 |

viscometry and similar research.

The model paper set provided a controlled study of the NIR spectroscopy as an analytical method for DP prediction. Fig. 4 presents the PLS regression results for the model paper samples. Although these samples were chemically degraded, DP was considered as the dominant varying property. The changes in these samples are likely to be mainly related to the scission of cellulose chains with little effects originating from naturally occurring paper components and degradation products, which could potentially complicate the NIR modelling and increase the prediction error. As shown in Fig. 4, the good linearity between the reference values and the modelled values in both training and test sets demonstrates that a linear relationship between the spectral response and DP could be well established by PLS regression. The RMSEP of DP for model paper set was 112, ~8% across DP range 468–2462, consistent with the performance of the training set. This can be considered as a baseline of the expected generalisation error for DP prediction of cellulosic materials.

Although wavelength assignment of NIR spectra is complicated due to many broad and overlapping bands corresponding to overtones and combinations of fundamental vibrations in the mid- and far-IR region [4], the PLS regression analysis for model paper gave an idea of the most critical spectral range for DP prediction. As shown in Table 4 for model paper, the spectral range used for the PLS model is roughly from 1130 nm to 1620 nm, including bands of the 1st overtone assigned to –OH stretching (1428–1591 nm) [25] and 3rd overtone assigned to aldehyde and ketone (1436–1478 nm) and conjugated aldehyde (1461–1495 nm) [26]. This provides additional evidence that the success of the NIR-PLS method is likely to be based on the concentration of glycosidic bonds, reducing end groups, and –OH groups in cellulose.

The historical paper set required more complicated calibration models with a wider spectral range and more PLS factors than the model

paper set (Table 4). The results of the PLS regression analysis for the historical paper samples are presented in Fig. 5. Good linearity was also obtained between the reference DP and the modelled DP, however, larger errors were obtained compared to model paper. This is likely a reflection of the variability of historical paper regarding the inhomogeneity of physical and chemical properties, including surface texture, thickness, fibres, sizing, and the degradation products accumulated over the centuries. Consistent performance of the training set and the test set was observed across the DP range of 427–4071.

3.2. Error evaluation

Errors in the spectral and the reference data as well as the bias of the statistical model affect the accuracy and precision of the predictions by the NIR-PLS method [5]. To gain insights into the errors of DP prediction by NIR-PLS method, RMSEP was used to represent the expected generalisation error. Mean square error of prediction (MSEP) was derived based on the bias-variance decomposition [27]:

$$\text{MSEP} = \sigma^2 + \text{Bias}^2[f(x)] + \text{Var}[f(x)], \quad (3.5)$$

where MSEP is the square of RMSEP, σ^2 is the sample variance in reference DP, $\text{Bias}^2[f(x)]$ represents the residual model error between the best fitting PLS approximation and the true model for DP, and $\text{Var}[f(x)]$ is the variance in the predicted DP of the test sample set.

The pooled standard deviation of the reference DP specified in Methodology was used to represent the σ for model paper and historical paper samples. Given that each spectral measurement represents the average of three spectra, the variance of the predicted DP was estimated as one third of the pooled variance of 20 duplicates randomly selected from the test sample sets. Measurements were taken from two different spots on each sample which gave standard deviation of DP 98 for model

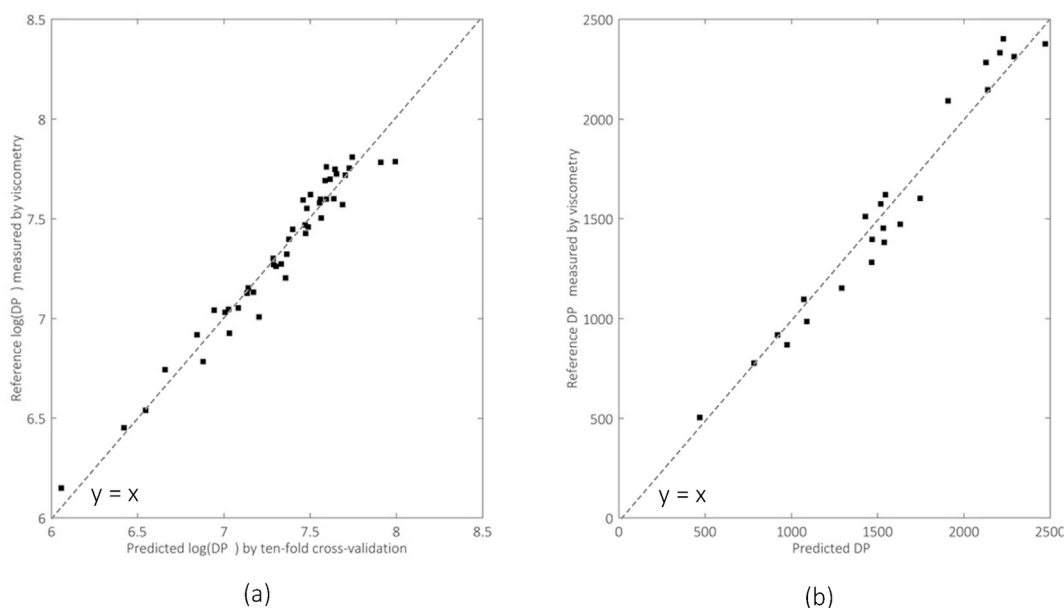


Fig. 4. Correlations between the modelled DP by PLS regression and reference DP measured chemically for (a) training and (b) test data sets of model paper samples consist of Whatman filter paper No. 1. Parameters presented in Table 4 were used for the PLS regression analysis. Note that data of the training set are presented on logarithmic scale whereas data of the test set are presented in the original scales, i.e. DP.

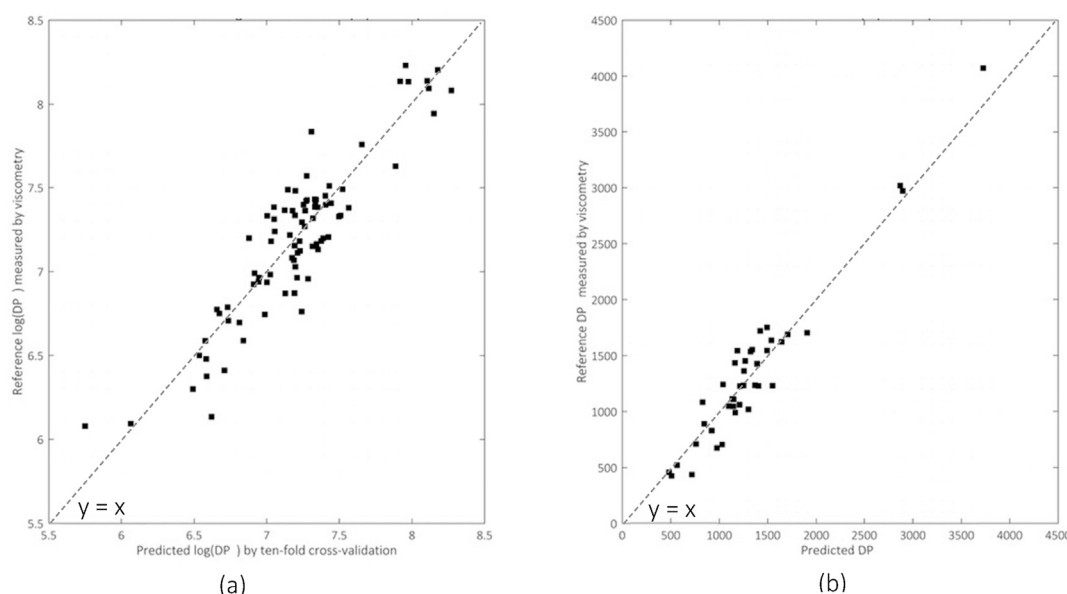


Fig. 5. Correlations between the modelled DP and reference DP measured chemically for (a) training and (b) test data sets of historical paper samples. Parameters presented in Table 4 were used for the PLS regression analysis. Note that data of the training set are presented in logarithmic scales whereas data of the test set are presented in the original scales, i.e. DP.

paper and DP 129 for historical paper. The bias of the PLS regression model from the true model was difficult to quantify directly. Therefore, the percentage contribution of bias² was inferred by subtracting the contributions of the variance of reference DP and predicted DP from 100%.

Table 5 summarises the contributions of the variance of reference DP, the variance of the predicted DP, and the square bias of the model. For both model paper and historical paper, the variance of predicted DP was found to be the largest contributor to the total variance estimated by MSEP. This is consistent with what Lu et al. [5] reported that the performance of the models for natural materials were sensitive to random noise in the spectra. Various sources can cause the variance in prediction, including the instrumentation, the chemical and physical inhomogeneity of the samples, and the measuring procedures. Given that the percentage contribution of Var [f(x)] of model paper samples (77%) was found to be higher than historical paper samples (49%), and model paper samples were inherently more homogeneous in chemical and physical properties than the historical samples, it is likely that the instrumentation and the measuring procedures contribute the most to the variances. Therefore, optimising these two factors can be crucial to the minimisation of the prediction errors.

For model paper, ~90% of the total variance coming from the samples and ~10% from the square bias of the PLS regression model. This indicates that the NIR-PLS method is capable of modelling DP of paper with good accuracy and precision. But for historical paper, the estimated square bias of the model was evidently high, ~50%. Since historical paper is much more complex in chemical and physical properties than model paper, it is likely that the NIR-PLS method was largely affected by these complexities. However, it is worth noting that this estimation was

based on the estimation of σ^2 . Unexpectedly, σ^2 of the historical paper samples was evidently smaller compared to that of the model paper samples. This could be the result of underestimation due to adjacent sampling for DP measurements by viscometry.

It was difficult to quantify the possible underestimation of σ^2 of historical paper based on available data. However, according to recent research, the CoV of historical rag paper samples was estimated 8.6% on average [21]. Given that more than 90% of the reference samples were historical rag paper, it was hypothesised that applying a correction factor of 3.78 to the CoV of the historical paper sample set could lead to more reasonable estimations. Under this hypothesis, σ^2 was estimated to contribute 33% of the total error and the square bias of the model contributed 18%. This estimation looks consistent with the estimation for model paper, where the variance of predicted DP contributed the most to the model error and the model bias contributed the least. But further research is needed to verify the validity of this approach.

3.3. Effect of moisture content

Water content is one of the most important factors that cause uncertainty in quantitative NIR spectroscopy [28]. This is because the NIR spectral range covers multiple O–H vibrations of H₂O, including a strong absorption at 1930 nm, and the O–H bond is particularly active and intense due to the large mass difference between the atoms [28]. However, the effect of moisture content of samples on quantitative NIR prediction of DP has not been studied. Since the NIR models are usually developed using samples conditioned in controlled laboratory environments, typically 23 ± 0.5 °C and 45% ± 5% RH [29], it is important to understand the effect of moisture content for applications of the models in practice, where the moisture content of the samples may vary greatly.

Given that the complexity of the NIR spectra and the chemical composition of historical paper, the sensitivity of DP prediction on changing moisture content is difficult to be derived from direct interpretation of the spectra. Therefore, an experiment was designed to assess the effect of moisture content on DP predicted by NIR-PLS models. Triplicate samples of model paper and historical paper were conditioned at different moisture contents by equilibrating the samples at the same *T* (23 ± 0.5 °C) but varying RH (20%–80%) in a climate chamber. For each condition, triplicate NIR spectra were collected from three spots on each

Table 5

Summary of the percentage contributions of the three sources to MSEP. Historical paper' represent the results where the CoV of historic paper was corrected by a factor of 4 for the possible underestimation of variance in reference DP.

| Samples | Variance of reference DP (σ^2) | Variance of predicted DP (Var [f(x)]) | Square bias of model (Bias ² [f(x)]) |
|------------------|---|---------------------------------------|---|
| Model paper | 12% | 77% | 11% |
| Historical paper | 2% | 49% | 49% |

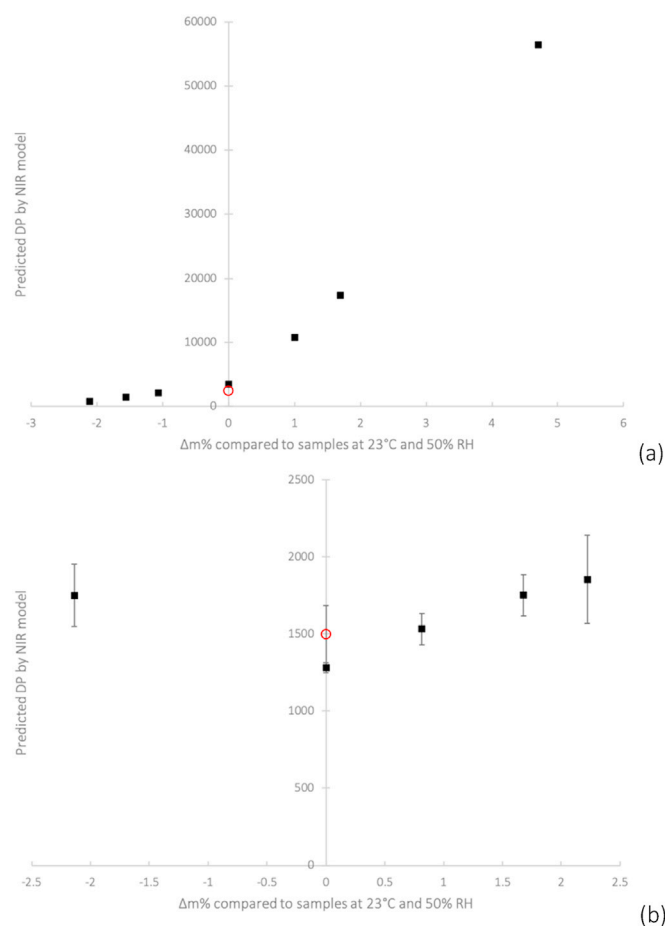


Fig. 6. The dependency of DP predicted by NIR spectroscopy on the moisture content of (a) model paper and (b) historical paper. The empty circle represents the reference DP values measured by viscometry and the error bars represent the standard deviation of the triplicate DP values predicted using the NIR-PLS model. For the model paper samples, the error bars are invisible compared to the magnitude of the change in DP caused by the change of moisture content.

sample. The average of the measurements was used to calculate the predicted DP of each sample using the PLS regression models specified in Table 4. The average mass of each sample before and after each set of NIR measurements was used to represent the mass of the sample during the measurements. Relative moisture content was expressed as the percentage mass change in the sample compared to the mass at 23 °C and 50% RH, at which the NIR-PLS model was developed.

Fig. 6 presents the scatter plots of the predicted DP in relation to the relative moisture content for model paper (Fig. 6 (a)) and historical paper (Fig. 6 (b)). Substantial change in the predicted DP of a model paper was observed when moisture content of the samples fluctuated between -3% and $+5\%$. This observation suggests that the NIR-PLS model for DP of model paper is not robust to varying moisture contents, although the spectral range including the most intense $-OH$ vibration of H_2O was excluded from the model. In this case, samples of a range of moisture content need to be used for the development of the NIR-PLS models to achieve improved robustness.

However, a closer look at the dependency of predicted DP of model paper on the relative moisture content reveals that the relationship can be approximated exponentially across the range investigated. Transformed on logarithmic scale with base 10, this dependency can be approximated with a good linear fit over the range $\pm 3\%$ (Residual Sum of Squares = 0.02, $R^2 = 0.99$). Therefore, in the case of model paper, the NIR-PLS model developed at a single moisture content can be extrapolated for predictions at other moisture contents. This extrapolation is

valid for a conditioning environment of 30%–70% RH at 23 °C, estimated by the equation developed by PaltaKari and Karlsson [30].

In contrast, when the moisture content of samples fluctuated within $\pm 3\%$ (30%–70% RH at 23 °C), the variation of the predicted DP of historical paper was found to be within the range of model uncertainty (Fig. 6 (b)). This observation suggests that the NIR-PLS model for historical paper is reasonably robust to the change of moisture content of the samples, thus no correction may be necessary for the application of the NIR-PLS model to the samples with different moisture contents. The reason for this observed stability in prediction can be complex and is likely associated with the complexity of historical paper. Part of the complexity may come from the natural ageing processes, during which the void structures in cellulose are changed and the structural resistance to the disruption caused by water absorption is increased [31]. It is worth noting that there is a tendency of increasing uncertainty as the change in moisture content increases, which may limit the accuracy of the predictions at extreme moisture contents.

4. Conclusions

To clarify the inconsistency in the literature and lay a foundation for future research, this paper systematically investigated the development of a NIR spectroscopic method coupled with PLS regression to non-destructively predict the DP of historical paper. The feasibility of the NIR-PLS method was studied using model paper samples composed of almost pure cellulose. Using log-transformed DP as response variable, satisfactory NIR-PLS models were established for model paper and historical paper. RMSEP was found to be DP 112 ($\sim 8\%$) for model paper and DP 185 ($\sim 20\%$) for historical paper. The larger error for historical paper is likely to be a result of the inhomogeneity of the samples, which is caused by the differing chemical and physical properties due to the original materials, manufacturing processes, and accumulation of degradation products over time.

RMSEP was taken as the expected generalisation error of the NIR-PLS models for error analysis. The variance of the reference DP, the variance of the predicted DP, and the bias of the NIR-PLS model from the true model for DP prediction were identified as the three sources of error. For both model paper and historical paper, the variance of the predicted DP contributed the most – 77% and 49% for model paper and historical paper respectively. This suggests that the performance of the current NIR-PLS models is mainly limited by the repeatability of the NIR measurements, which may be improved by enhanced precision and accuracy of the instrument and the operational procedures. The NIR-PLS method was found adequate to model DP of model paper but was likely to be affected by the complexity of the historical paper.

To assess the practicality of the NIR-PLS models, the effect of moisture content of the samples on DP prediction was investigated. With the fluctuation of moisture content being $\pm 3\%$, substantial variations in the predicted DP were observed for model paper, whereas the variations contained within the model uncertainty for historical paper. This suggests that the NIR-PLS models developed using model samples tend to be overly ideal for real conditions and should be avoided in real applications. Since the NIR-PLS model for historical paper was found robust to environmental conditions equivalent to 30%–70% RH at 23 °C, the quantitative NIR spectroscopic method can be applied to historical paper with confidence in a range of environments to obtain plausible predictions of DP.

Author statement

Y. L.: conceptualisation, formal analysis, investigation, methodology, software, validation, writing - original draft preparation. T. F.: conceptualisation, methodology, supervision, writing - review & editing. M. S.: conceptualisation, methodology, resources, supervision, writing - review & editing.

Declaration of competing interest

The authors declare that they have no competing financial interests or personal relationships that have influenced the research presented in this paper.

Acknowledgements

This research was generously supported by the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Science and Engineering in Arts, Heritage and Archaeology (SEAHA), UK and the Smithsonian's Museum Conservation Institute Trust Funds, USA.

References

- [1] M. Strlič, et al., Damage function for historic paper. Part III: isochrones and demography of collections, *Heritage Science* 3 (40) (2015) 1–12, <https://doi.org/10.1186/s40494-015-0069-7>. Springer International Publishing.
- [2] X. Zou, T. Uesaka, N. Gurnagul, Prediction of paper permanence by accelerated aging I. Kinetic analysis of the aging process, *Cellulose* 3 (1996) 243–267, <https://doi.org/10.1007/BF02228806>.
- [3] J.T. Oberlerchner, T. Rosenau, A. Potthast, Overview of methods for the direct molar mass determination of cellulose, *Molecules* 20 (2015) 10313–10341, <https://doi.org/10.3390/molecules200610313>.
- [4] T. Naes, T. Isaksson, T. Fearn, T. Davies, *A User-friendly Guide to Multivariate Calibration and Classification*, NIR Publications, Chichester, 2002.
- [5] J. Lu, et al., Effect of random noise on the performance of NIR calibrations, *J. Near Infrared Spectrosc.* 6 (1–4) (1998) 77–87, <https://doi.org/10.1255/jnirs.124>.
- [6] P. Baptista, et al., Multivariate near infrared spectroscopy models for predicting the iodine value, CFPP, kinematic viscosity at 40 °C and density at 15 °C of biodiesel, *Talanta* 77 (1) (2008) 144–151, <https://doi.org/10.1016/j.talanta.2008.06.001>.
- [7] R.M. Balabin, S.V. Smirnov, 'Analytica Chimica Acta Variable selection in near-infrared spectroscopy : benchmarking of feature selection methods on biodiesel data', *Anal. Chim. Acta* 692 (1–2) (2011) 63–72, <https://doi.org/10.1016/j.aca.2011.03.006>. Elsevier B.V.
- [8] M. Schwanninger, et al., Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression analysis. Part 1: wavenumber selection and evaluation of the selected range, *J. Near Infrared Spectrosc.* 19 (5) (2011a) 319–329, <https://doi.org/10.1255/jnirs.945>.
- [9] M. Schwanninger, et al., Determination of lignin content in Norway spruce wood by Fourier transformed near infrared spectroscopy and partial least squares regression analysis. Part 2: development and evaluation of the final model, *J. Near Infrared Spectrosc.* 19 (5) (2011b) 331–341, <https://doi.org/10.1255/jnirs.945>.
- [10] P. Fardim, M.M.C. Ferreira, N. Durán, Multivariate calibration for quantitative analysis of eucalypt kraft pulp by nir spectrometry, *J. Wood Chem. Technol.* 22 (1) (2002) 67–81, <https://doi.org/10.1081/WCT-120004435>.
- [11] N. Brown, et al., Non-destructive collection survey of the historical Classense Library. Part I: paper characterisation, *Heritage Science* 8 (1) (2020) 1–11, <https://doi.org/10.1186/s40494-020-00432-w>. Springer International Publishing.
- [12] N. Brown, et al., Characterisation of 19th and 20th century Chinese paper, *Heritage Science* 5 (1) (2017) 1–14, <https://doi.org/10.1186/s40494-017-0158-x>.
- [13] H. Mahgoub, et al., Material properties of Islamic paper, *Heritage Science* 4 (1) (2016) 1–14, <https://doi.org/10.1186/s40494-016-0103-4>. Springer International Publishing.
- [14] M. Oriola, et al., 'Looking beneath Dalí's paint: non-destructive canvas analysis', *Analytical Methods* 6 (1) (2014) 86–96, <https://doi.org/10.1039/c3ay41094c>.
- [15] M. Strlič, et al., Non-destructive characterisation of iron gall ink drawings: not such a galling problem, *Talanta* 81 (1–2) (2010) 412–417, <https://doi.org/10.1016/j.talanta.2009.12.017>.
- [16] U. Henniges, M. Schwanninger, A. Potthast, Non-destructive determination of cellulose functional groups and molecular weight in pulp hand sheets and historic papers by NIR-PLS-R, *Carbohydr. Polym.* 76 (3) (2009) 374–380, <https://doi.org/10.1016/j.carbpol.2008.10.028>. Elsevier Ltd.
- [17] J. Trygg, S. Wold, PLS regression on wavelet compressed NIR spectra, *Chemometr. Intell. Lab. Syst.* 42 (1–2) (1998) 209–220, [https://doi.org/10.1016/S0169-7439\(98\)00013-6](https://doi.org/10.1016/S0169-7439(98)00013-6).
- [18] S. Wold, et al., Orthogonal signal correction of near-infrared spectra, *Chemometr. Intell. Lab. Syst.* 44 (1–2) (1998) 175–185, [https://doi.org/10.1016/S0169-7439\(98\)00109-9](https://doi.org/10.1016/S0169-7439(98)00109-9).
- [19] BS Iso, *Pulps — Determination of Limiting Viscosity Number in Cupri-Ethylenediamine (CED) Solution*, 2004. BS ISO 5351:2004.
- [20] R. Evans, A.F. Wallis, Comparison of cellulose molecular weight determined by high performance size exclusion chromatography and viscometry, in: 4th International Symposium on Wood and Pulping Chemistry April, vols. 27–30, Paris - Palais des Congres, Paris, 1987, 1987.
- [21] Y. Liu, I. Kralj Cigić, M. Strlič, Kinetics of accelerated degradation of historic iron gall ink-containing paper, *Polym. Degrad. Stabil.* 142 (2017) 255–262, <https://doi.org/10.1016/j.polydegradstab.2017.07.010>.
- [22] M. Manso, et al., Investigation of the composition of historical and modern Italian papers by energy dispersive X-ray fluorescence (EDXRF), X-ray diffraction (XRD), and scanning electron microscopy energy dispersive spectrometry (SEM-EDS), *Appl. Spectrosc.* 65 (1) (2011) 52–59, <https://doi.org/10.1366/10-06105>.
- [23] A. Savitzky, M.J.E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (8) (1964) 1627–1639, <https://doi.org/10.1021/ac60214a047>.
- [24] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, *Appl. Spectrosc.* 43 (5) (1989) 772–777, <https://doi.org/10.1366/0003702894202201>.
- [25] T. Inagaki, et al., Difference of the crystal structure of cellulose in wood after hydrothermal and aging degradation: a NIR spectroscopy and XRD study, *Biomacromolecules* 11 (9) (2010) 2300–2305, <https://doi.org/10.1021/bm100403y>.
- [26] J.D. Robert, M.C. Caserio, *Basic Principles of Organic Chemistry*, second ed., W. A. Benjamin, Inc, Menlo Park, CA, 1977.
- [27] T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning*, Springer-Verlag, 2017. Available at: <https://web.stanford.edu/~hastie/ElemStatLearn/>.
- [28] J. Workman Jr., L. Weyer, *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy*, CRC Press, Boca Raton, 2012.
- [29] *Isa, Recommended Environments for Standards Laboratories*, 2006.
- [30] J.T. Paltakari, M.A. Karlsson, Determination of specific heat for dry fibre material. 82nd Annual Meeting, CPPA, Technical Section, 1996, pp. B117–B120.
- [31] K.L. Kato, R.E. Cameron, Structure-property relationships in thermally aged cellulose fibers and paper, *J. Appl. Polym. Sci.* 74 (1999) 1465–1477, [https://doi.org/10.1002/\(SICI\)1097-4628\(19991107\)74:6<1465::AID-APP20>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-4628(19991107)74:6<1465::AID-APP20>3.0.CO;2-3).