ORIGINAL RESEARCH

Development and mining of a database of historic European paper properties

Matija Strlič · **Yun Liu** · **Dirk Andreas Lichtblau** · **Gerrit De Bruin** ·
Barry Knight · **Thea Winther** · **Irena Kralj Cigić** · **Richard G. Brereton**

Abstract A database of historic paper properties was developed using 729 samples of European origin (1350–1990), analysed for acidity, degree or polymerisation (DP), molecular weight of cellulose, grammage, tensile strength, as well as contents of ash, aluminium, carbonyl groups, rosin, protein, lignin and fibre furnish. Using Spearman's rank correlation coefficient and principal component analysis, the data were examined with respect to methods of manufacture, as well as chemical stability of paper. Novel patterns emerged related to loss of DP and accumulation of carbonyl groups and acidity with time and the role of lignin and rosin, as well as rate of degradation ($k = 10^{-5}$ year^{-1}) at room conditions. In-depth understanding of long-term degradation of lignin and rosin is needed to better understand the relationships between composition and degradation of historic paper. This study highlights the importance of mining significant volumes of analytical data, and its variability, obtained from real historic objects.

M. Strlič · Y. Liu
Institute for Sustainable Heritage, University College London, 14 Upper Woburn Place, London WC1H 0NN, UK
e-mail: yun.liu.14@alumni.ucl.ac.uk

M. Strlič (✉) · I. Kralj Cigić
Heritage Science Lab Ljubljana, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Večna pot 113, 1000 Ljubljana, Slovenia
e-mail: m.strlic@ucl.ac.uk

I. Kralj Cigić
e-mail: irena.kralj-cigic@fkkt.uni-lj.si

D. A. Lichtblau
Lichtblau e.K, Loschwitzer Straße 15A, 01309 Dresden, Germany
e-mail: mail@lichtblau-germany.com

G. De Bruin
Nationaalarchief, Prins Willem-Alexanderhof 20, 2509 LM Den Haag, The Netherlands
e-mail: gerrit.de.bruin@nationaalarchief.nl

B. Knight
The British Library, 96 Euston Rd, London NW1 2DB, UK
e-mail: drbarryknight@hotmail.co.uk

T. Winther
Riksarkivet, Box 12541, 102 29 Stockholm, Sweden
e-mail: thea.winther@riksarkivet.se

R. G. Brereton
School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK
e-mail: r.g.brereton@bris.ac.uk

Introduction

As one of the first globalised industries, papermaking was introduced to Europe in the 12th Century (Burns and Lindgren 1996; Hunter 2011). After the invention of the printing press in the fifteenth Century in Europe and until the advent of digital media, it served as an almost exclusive data carrier. As a consequence, libraries and archives are now tasked with the management of kilometres of shelving. While established conservation practices are enshrined in international standards (BS:4971 2017), and manual and industrial papermaking has itself been the subject of intensive historical research (Clapperton and Henderson 1947; Stromer 1993; Lucas 2005), our knowledge of long-term material degradation is still mainly based on the inference from accelerated degradation experiments using a small selection of model and historical material samples (Zou et al. 1996a; Barański et al. 2005; Strlič and Kolar 2005). To overcome this limitation, our research focuses on the survey of a large collection of historic papers of European origin, presenting the most comprehensive database of chemical and mechanical material properties with a view not only to investigate patterns of raw material use throughout history, but also to explore any interdependencies between composition and degradation that can have a bearing on conservation research and practice.

The database contains 729 historic paper samples of European origin (1350–1990). The samples cover three categories: (i) rag papers (**gelatine**); (ii) groundwood-containing papers (**groundwood**); and (iii) bleached-pulp containing papers (**bleached**). In addition to the date of sample production (**dat**), 14 properties that define paper compositions and state of degradation are of interest: viscometric degree or polymerisation (**DP**) and molecular weight (**MW**, g mol^{-1}) of cellulose, tensile strength (**TS**, N), grammage (**grm**, g m^{-2}), acidity (expressed as **pH**), contents of ash (**ash**, %), aluminium (**Al**, mg g^{-1}), carbonyl groups (**car**, mmol g^{-1}), rosin (**ros**, mg g^{-1}), protein (**pro**, %) and lignin (**lig**, mg g^{-1}), as well as the content of three sources of fibrous material that were mainly used for historic European paper (2)—groundwood fibres produced directly from wood using thermomechanical processes (**woo**); cotton, linen and hemp rags, or native cotton fibres (**cot**); and partly

delignified wood-derived cellulose fibres using various bleaching methods (**cel**).

Materials and methods

Sample selection and categorisation (gelatine, bleached, lignin)

1373 naturally aged paper samples were collected for the purpose of the SurveNIR project, spanning over the fourteenth to the twentieth century. Geographically, the samples were obtained by donation of purchases through antique shops in diverse European countries (Fig. 1a). 67% are known to be of European origin, i.e. the volumes have been published in a European country, it is therefore possible to assume that paper was made in Europe as well. For 4% of samples, the publisher was either US or from elsewhere, while for 30% of samples it isn't possible to unequivocally determine the origin, however, as they were written/printed in European languages it is possible to claim that the majority of papers in the current study came from European sources that exhibit a range of variability in properties and represent the most important paper types in Western libraries, archives and museums, including rag papers, bleached pulp containing papers, groundwood containing papers, post-1990 produced papers and coated papers. Despite significant effort, the samples produced before 1800 are significantly under-represented.

For the purpose of the research presented in the paper, only pre-1990 and non-coated paper samples were selected to reduce the diversity of papermaking practices and because papers produced post-1990 are predominantly alkaline and thus stable. Coated samples were not taken into account as coating technologies need more in-depth historical research.

Initially, the samples were categorised as rag papers (i.e. gelatine-sized, denoted as '**gelatine**') on the basis of the presence of papermaking sieve marks and random orientation of fibres, indicative of manual sheet formation. Classification into '**bleached**' or '**groundwood**' categories was primarily based on fibre furnish analysis, i.e. < 20% of groundwood fibres was considered to be **bleached** paper category and > 20% as **groundwood** paper category. The secondary criterion was the content of lignin, if this was
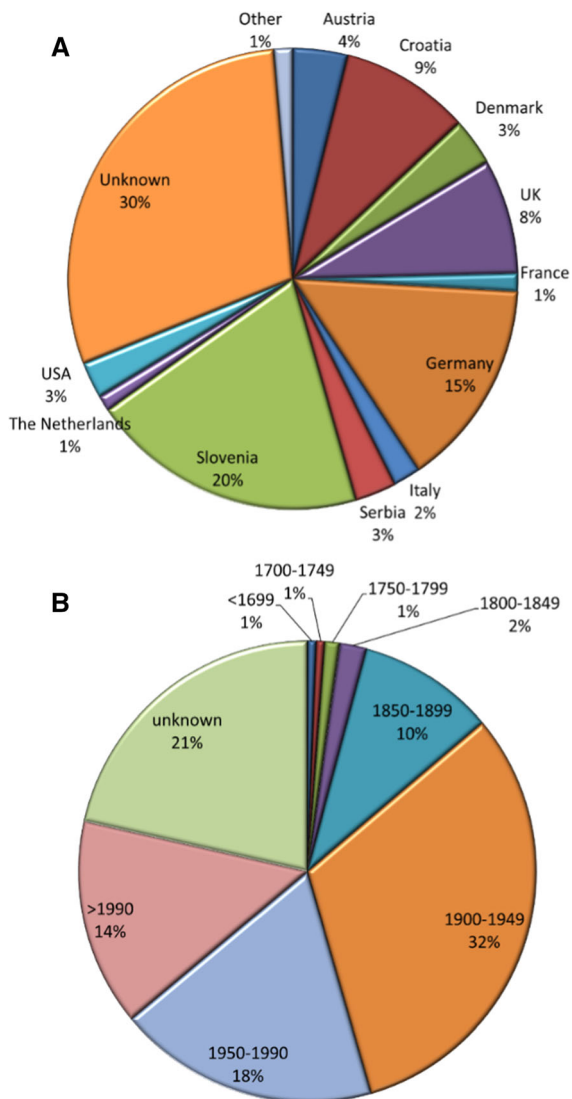
Fig. 1 **a** Origin of SurveNIR samples (country of publication); **b** the distribution of samples according to age

> 50 mg/g, paper was considered to be **groundwood** paper category and if less, it was considered to be **bleached**. Categorisation is thus partly based on conservation experience and partly on chemical analyses.

Samples for analysis were taken from areas without print, ink or visible degradation due to water or mould, in order to achieve representativeness and reduce the interference of minor factors contributing to degradation. To conduct all the analyses, several pages from each book were necessary, taken from the centre of a

book, with 1-cm margins discarded, to avoid any edge effects.

Date of production (dat)

For the purpose of the research presented in the paper, only pre-1990 and non-coated paper samples were selected to reduce the diversity of papermaking practices. The age of samples (Fig. 1b) was deduced from the age a volume has been printed, although paper could have been manufactured prior to printing. For manuscripts, determination of age (**dat**) was often not possible and the database contains 92 out of 137 rag papers that do not have an associated date. There are 597 out of 729 samples in the database with **dat** available.

Where the approximate period of production could be obtained from contextual information (a non-dated paper within a collection of dated papers, manuscripts or printed books), then the median date was taken as **dat**, e.g. 1400–1450 was translated to 1425.

Fibre furnish analysis (cel, cot, woo)

Quantitative fibre furnish analysis was carried out to determine the percentage composition of cellulose, cotton and wood-derived fibres based on BS 7463-1 (1991): groundwood fibres produced directly from wood using thermomechanical processes (**woo**); cotton, linen and hemp rags, or native cotton fibres (**cot**); and partly delignified wood-derived cellulose fibres using various bleaching methods (**cel**).

Approximately 250 mg was sampled from different parts of each paper sheet and prepared according to the standard procedure. The number of crossings of various types of fibres with the counting line was counted under an optical microscope and the measurements were transformed into proportion by weight by application of weight factors. **Woo**, **cot** and **cel** were determined for all the 729 samples in the database.

Degree of polymerisation (DP)

Degree of polymerisation (**DP**) was determined viscometrically based on BS ISO 5351 (2004). Intrinsic viscosity η of each sample was used to calculate **DP** using the Mark–Houwink–Sakurada equation: $DP^{0.85} = 1.1(\eta)$ (Evans and Wallis 1987). Each sample was

measured twice in adjacent areas and the averages of the duplicates were used for further analysis. It is worth noting that ground-wood containing papers were mostly excluded from the viscometric analysis due to the insolubility of lignin in cupriethylenediamine (1 mol L^{-1}), the solvent used for viscometry. **DP** of 446 samples was determined and the standard deviation of **DP** is 36.6.

It is worth noting that although there is little difference between different standards for the determination of intrinsic viscosity, several different versions of Mark–Houwink–Sakurada coefficients are used in the literature which results in deviations in viscometric **DP** (Łojewski et al. 2010). Conversions were applied when comparing the viscometric DP measured and calculated in this research and the viscometric DP reported in literature.

Size exclusion chromatography (MW)

Weight-average molar mass (**MW**, g mol^{-1}) was determined for samples using SEC of cellulose tricarbanilate (CTC) based on the method proposed by (Stol et al. 2002). Samples chosen for SEC were mostly those where **DP** could not be determined, and a random selection of other samples. Following the modified procedure (Clapperton and Henderson 1947), approximately 0.2 mg sample was used for the determination of **MW** relative to polystyrene standards using the universal calibration approach with a Hewlett Packard series 1100 chromatographic system. The determined **MW** values were converted to absolute molar masses using the established correlation between relative **MW** and absolute **MW** determined by SEC coupled to a multi-angle light scattering photometer [SEC-MALS, the procedure using a GaAs laser at 656 nm is reported in detail in Balažic et al. (2008)]. Duplicate determinations were carried out for 616 samples and the averages were used to represent the **MW** of each paper sheet, with a standard deviation of 14159 g mol^{-1}.

It is known that small amounts of **lig** do not interfere with **MW** determination using the CTC procedure (Potthast et al. 2015). The samples were filtered prior to injection to remove undissolved lignin, while low-MW dissolved lignin was disregarded in the integration step. However, it needs to be taken into account that large amounts of lignin could systematically affect the results.

Reducing carbonyl groups (car)

The reducing carbonyl groups (**car**, mmol g^{-1}) in the samples were determined by the colorimetric method proposed by Szabolcs (1961). 10 mg of air dry sample was sampled and weighed in a test tube with 0.5 mL of KOH solution (0.2 mol L^{-1}) and 0.5 mL triphenyltetrazolium chloride solution (0.2%) added. The test tube was heated in a water bath at boiling point for 10 min and cooled down under running water before the sample was vacuum-filtered. The sample was then washed with 10 mL of methanol p.a. and the absorbance of the resulting solution at 546 nm was measured using a Varian Cary® 50 UV–Vis Spectrophotometer (Agilent Technologies, US). The amount of reducing carbonyl groups was determined in mmol g^{-1} for 717 samples based on the calibration curve developed using a solution of glucose (0.1%) with a standard deviation of 0.005 mol g^{-1}.

Acidity (pH)

Acidity of the samples (**pH**) was determined using the modified cold extraction method with optimised sample consumption (TAPPI T509, 2002) (Strlič et al. 2004). 20–50 μg sample was suspended in 5 μL deionised water overnight and the extraction was measured using a micro-combined glass electrode (MI 4152, Microelectrodes, Bedford, NH). **pH** of 725 samples was determined and the standard deviation of pH determinations is 0.08 pH units.

Aluminium content (Al)

Aluminium content (**Al**, mg g^{-1}) was determined using atomic absorption spectroscopy after extraction of 30–50 mg of sample in 15% HNO$_3$ for 1 h in an ultrasonic bath (Sonis 4, Iskra, Kranj, Slovenia) at 65 °C. **Al** was determined in 697 samples and the standard deviation of aluminium content determinations is 0.21 mg g^{-1}.

Rosin content (ros)

To determine the rosin content in papers samples (**ros**, mg g^{-1}), approximately 70 mg sample was sampled from each paper sheet for rosin content analysis. The extraction of rosin acids was carried out twice using 2 mL acidified acetonitrile (90% (v/v) acetonitrile and

10% (v/v) trifluoroacetic acid, 0.1%) in an ultrasound bath (Sonis 4, Iskra, Kranj, Slovenia) for 20 min. Both extracts were collected and centrifuged at 3000 rpm for 5 min (Eppendorf 5408 R, Hamburg, Germany) before analysed using liquid chromatography coupled to mass spectrometry (liquid chromatograph Series 200 [Perkin-Elmer, Shelton, CT, USA), 3200 QTRAP LC–MS/MS system equipped with electrospray ionization source (ESI) (Applied Biosystems/MDS Sciex, Foster City, CA, USA)]. A reversed-phase column Gemini C18 with 3 μm particle size and dimensions 150 × 4.6 mm from Phenomenex (Torrance, California, United States) and isocratic elution (1 mL/min) with mobile phase composed of 70% (v/v) acetonitrile and 30% (v/v) acetic acid (1%) were used. The results were calibrated using external standard solutions of abetic acid (AA) and dehydroabietic acid (DHAA) in acidified acetonitrile (90% (v/v) acetonitrile and 10% (v/v) of 1% acetic acid) for quantification. Rosin content is expressed as the sum of AA and DHAA, the standard deviation of determinations is 0.25 mg g^{-1}.

Samples selected for analysis were those produced in the nineteenth and twentieth Century, excluding samples that were rag samples or alkaline samples (with the exception of a random selection of the latter).

In order to reduce the number of samples without all the key variables determined, and thus enable multivariate data analysis, the amount of **ros** in the **gelatine** samples was estimated to be 0%, as re-sizing that could lead to significant amounts of both **pro** and **ros** in the same paper was a rare practice. In total, **ros** was determined in 715 samples in the database.

Protein content (pro)

Protein content (**pro**, %) was determined as weight percentage of dry paper basis using the reagents and procedures described by Cséfalvayová et al. (2010). The procedure consisted of gelatine extraction from approximately 5 mg sample in 1 ml HCl (0.1 mol L^{-1}) at 100 °C for 1 h and further gelatine hydrolysis with an aliquot of HCl (6 mol/L) at 100 °C for 18 h to obtain free amino acids. Excess HCl was removed by drying, following which the residue was redissolved in the HPLC eluent and buffered to pH 9.9. Derivatization was performed in the automated injection module of the HPLC by reacting with 9-fluorenylmethylchloroformate (FMOC). The derivates formed are stable at room temperature, separated at 40$^{°}$C and detected at 262 nm. The values of a specific marker for gelatine, the amino acid hydroxyproline (which remains stable in the course of ageing of paper) indicate the amount of gelatine applied on paper and the amount of gelatine is calculated using the factor 0.126 that represents the weight fraction of hydroxyproline in a typical hide glue. The standard deviation of determinations is 0.23. All samples classified as rag paper (**gelatine**) were analysed using this process, and 10% of all other samples.

In order to reduce the number of samples without all the key variables determined, and thus enable multivariate data analysis, the amount of **pro** in the **groundwood** and **bleached** paper samples from 1850 to 1990 was estimated to be 0%, as re-sizing that could lead to significant amounts of both **ros** and **pro** in the same paper was a rare practice. In total, pro was determined in 699 samples in the database.

Lignin content (lig)

Lignin content (**lig**, mg g^{-1}) of samples was determined based on the UV spectrometry method proposed by Iiyama and Wallis (1988). 1–1.5 mg sample was dissolved in 0.5 mL of the solution made from 2.5 mL acetyl bromide, 10 mL glacial acetic acid, and 0.5 mL perchloric acid (70%), at 70 °C. After sample dissolution, 1 mL NaOH solution (2 mol L^{-1}) and 2.5 mL glacial acetic acid were added, and UV absorption at 280 nm was subsequently measured using a Varian Cary® 50 UV–Vis Spectrophotometer (Agilent Technologies, US). The lignin content was determined based on the calibration curve established using alkali lignin (Sigma-Aldrich, Steinheim, Germany). The standard deviation of determinations is 13.56 mg g^{-1}.

In order to reduce the number of samples without all the key variables determined, and thus enable multivariate data analysis, the amount of **lig** in **gelatine** paper samples where **lig** was not analytical determined, was estimated as the average of **lig** in **gelatine** samples from 1850 to 1990, as re-sizing that could lead to significant amounts of both **ros** and **pro** in the same paper was a rare practice. In total, 556 samples in the database has **lig** available.

Ash content (ash)

Ash content (**ash**, %) was determined as the residue on ignition at 900 °C based on ISO 2144 (2000), with sample size optimised to 0.100 g. The mass of the residue was determined in duplicate for each paper sample and the average was used to calculate the residual as a percentage on oven-dry basis of the samples. **Ash** was determined for 702 samples and the standard deviation of determinations is 1.04%.

Optical brighteners (OB)

Reflectance spectrophotometry was used for the determination of the presence of optical brighteners. Reflectance spectra were collected using a Spectro-densitometer X-Rite 500 and the presence of the blue fluorescence peak at 430–450 nm was taken as the indication of the presence of optical brighteners. Although **OB** is not used in data analysis as presented in this paper, the data are offered as part of the database for 672 samples, where 0 = optical brightener not detected and 1 = optical brightened detected.

Grammage (gra)

The grammage of samples (gra, g m^{-2}) was determined gravimetrically by weighing 10 cm^2 of a sample. **Gra** was determined for 501 samples and the standard deviation of duplicated determinations is 0.95 g m^{-2}.

Tensile strength (TS)

Tensile strength was determined following the modified standard method ISO 1924-2 (1994). 12 strips of 120 × 15 mm of each sample were prepared. The measurements were carried out using a Zwick Proline z0.5 TS instrument (Ulm, Germany), and a 500-N load cell nominal force was applied (type II), and 6 bar jaw pressure.

Each jaw pair had a straight and a concave half to avoid sample breaking, while the test length was 100 mm. Due to the use of already degraded historical papers with low values and higher standard deviation, 12 strips were used instead of 10 as recommended by the standard. The maximum and minimum values were removed from the obtained data set and the remaining 10 values were used to calculate the mean.

TS was determined for 330 samples and the standard deviation for 10 determinations was 2.57 N.

Data analysis

Pairwise sets of variables for all samples as well as **gelatine**, **groundwood** and **bleached** paper categories separately were analysed using Spearman's rank correlation coefficients (Spearman 1904). As some measurements were missing, different numbers of samples were used for each pairwise comparison. Spearman's correlation was used rather than the more usual Pearson correlation because the expected ideal relationship between connected variables was likely to be monotonic rather than linear. Principal component analysis [PCA (Wold et al. 1987; Jolliffe 2002; Mardia et al. 1979; Manly 2004; Jackson 1991)] using the NIPALS algorithm (Wold 1966) was carried out on the resulting database using diverse combinations of variables. All samples with one or more missing variables were deleted and all variables were standardised prior to PCA. PCA was performed on the standardised data matrix using singular value decomposition. The full data set, as well as detailed results of correlation analyses, are provided in supplementary information.

Results and discussion

DP and **MW** of cellulose are chemical properties that give rise to mechanical characteristics of paper such as **TS** (Zou et al. 1996a). The linear correlation between **TS** and **DP** across the categories (Fig. S.10 in supplementary information, $\rho = 0.6130$) confirms that **DP** can be used as indicators of degradation at both macro and molecular levels. Higher **DP** on average have better mechanical properties, regardless of the type of sizing or fibre, although data scatter indicates that other variables may play a significant role. The agreement between **MW** and **DP** (Fig. S.7 in supplementary information, $\rho = 0.7805$) indicates that results of further analysis are likely linearly transferrable between them. The slope of the linear correlation is equal to 0.77 times the molecular weight of a derivatised glucose monomer (519 g mol^{-1}), which agrees with the results obtained from model paper where viscometric **DP** is typically 0.66–1.12 times **DP** calculated from **MW** (Łojewski et al. 2010;

Kes and Christensen 2013). Since **DP** has been the most frequently studied chemical property in paper degradation (Łojewski et al. 2010; Ekenstam 1936; Zou et al. 1996b), we use **DP** as the main variable for further analysis. **MW** is only used where a large number of **groundwood** samples are of particular interest since **DP** could only be determined for a few exceptional samples of **groundwood**.

Exploratory analysis of all samples by PCA reveals significant associations between paper properties and degradation (in current state, i.e. not at time of production), as well as papermaking practices. The loadings plot of PC2 versus PC1 for all samples (Fig. 2a) shows the negative loadings for **DP** and **pro** and postivie loadings of **lig**, **ash**, **ros** and **car** on PC1, suggesting that PC1 focuses on sizing and the quality of raw materials that separates the three paper categories. High **pro** and low **ash** are typical for **gelatine** paper category since a high content of **pro** reflects the practice of surface sizing with gelatine (Adams et al. 2009). High **ros** is associated with high **lig**, consistent with the fact that rosin sizing was invented at about the same time as the processes that allowed wood-derived fibres to be used in papermaking (Hunter 2011). The positive loading for **Al** and the negative one for **pH** indicate that PC2 separates samples on the basis of acidity, and that low **pH** is associated with high **Al** content which would support the hypothesis that alum (source of **Al**), used to precipitate rosin acids on fibres during bulk sizing, has a defining influence on paper acidity of the freshly made papers (Launer 1939).

However, to confirm this association and reduce the influence of sizing, it is useful to look at a single paper category, e.g. **bleached** (Fig. 2b). The loadings plot for PC2 versus PC1 reveals a similar association: low **pH** is associated with high **Al**, and less so with high **ros** and **car**. The role of lignin remains unclear from the loading plots. Figure 2c is again a reflection of the sizing practice, as high **pH** is associated with high **ash**, which is indicative of papermaking practices developed in the 2/2 of the 20th Century (Hunter 2011), where high-value fibre materials were replaced with inorganic fillers and additives, typically calcium and magnesium carbonates resulting in paper pH > 8 (Hunter 2011; Strlič and Kolar 2005). It appears therefore that PC2 separates predominantly older samples from the more recent ones.

The relationships revealed by PCA are quantified by correlation and regression analysis for further insights into the degradation processes of historic papers. Of particular interest are the clear linear correlations between DP^{-1} and **dat** (Fig. 3a) suggesting degradation kinetics for samples from 1850 to 1990, i.e. mainly **bleached** paper. Since the linearity of the correlation complies with the Ekenstam equation (Ekenstam 1936; Zou et al. 1996b), where the change in DP^{-1} is proportional to time, it can be assumed that the papers were on average stored in similar conditions, and that their **DP** at the point of manufacture was close to **DP** 2500, as indicated for papers made ca. 1990. The few exceptions with **DP** up to 4000 may indicate the use of less processed fibres, e.g. **cot**. The slope of the correlation (k = 10^{-5} year^{-1}), i.e. the apparent rate constant, agrees with Zou et al. (1996b) for hydrolytically degraded papers in natural conditions, suggesting hydrolysis be the single predominant mechanism of natural degradation of historic paper (1850–1990).

This hypothesis is further verified by the evident linearity between **car** and DP^{-1} in the larger sample set in Fig. 2b. To confirm that this linearity complies with the intrinsic linear correlation between the concentration of cellulose chains and the number of reducing end groups that indicates hydrolytic splitting being the dominant mechanism of cellulose chain scission (Whitmore and Bogaard 1994), multiple linear regression (MLR) was carried out for **car** with the amount of chain scission of cellulose molecules (**scission**) and the two compositional properties that show the highest Spearman's correlation coefficient with **car**: **lig** ($\rho = 0.5325$, Fig. S.2 in supplementary information) and **ros** ($\rho = 0.5934$, Fig. S.13 in supplementary information). Molecular weights of 8000 g mol^{-1} (Asikkala et al. 2012; Lange et al. 2013) and 296.32 g mol^{-1} are taken for **lig** and **ros** respectively. **Scission** is estimated using number average DP (Kes and Christensen 2013) assuming initial **DP** of 2500.

Table 1 presents the MLR summary data. It is evident that **scission** is most strongly correlated with **car** and is likely to be the major cause of the formation and accumulation of **car**. The 1:1 proportionality between **scission** and **car** suggests the dominant role of hydrolytic splitting in the formation of carbonyl
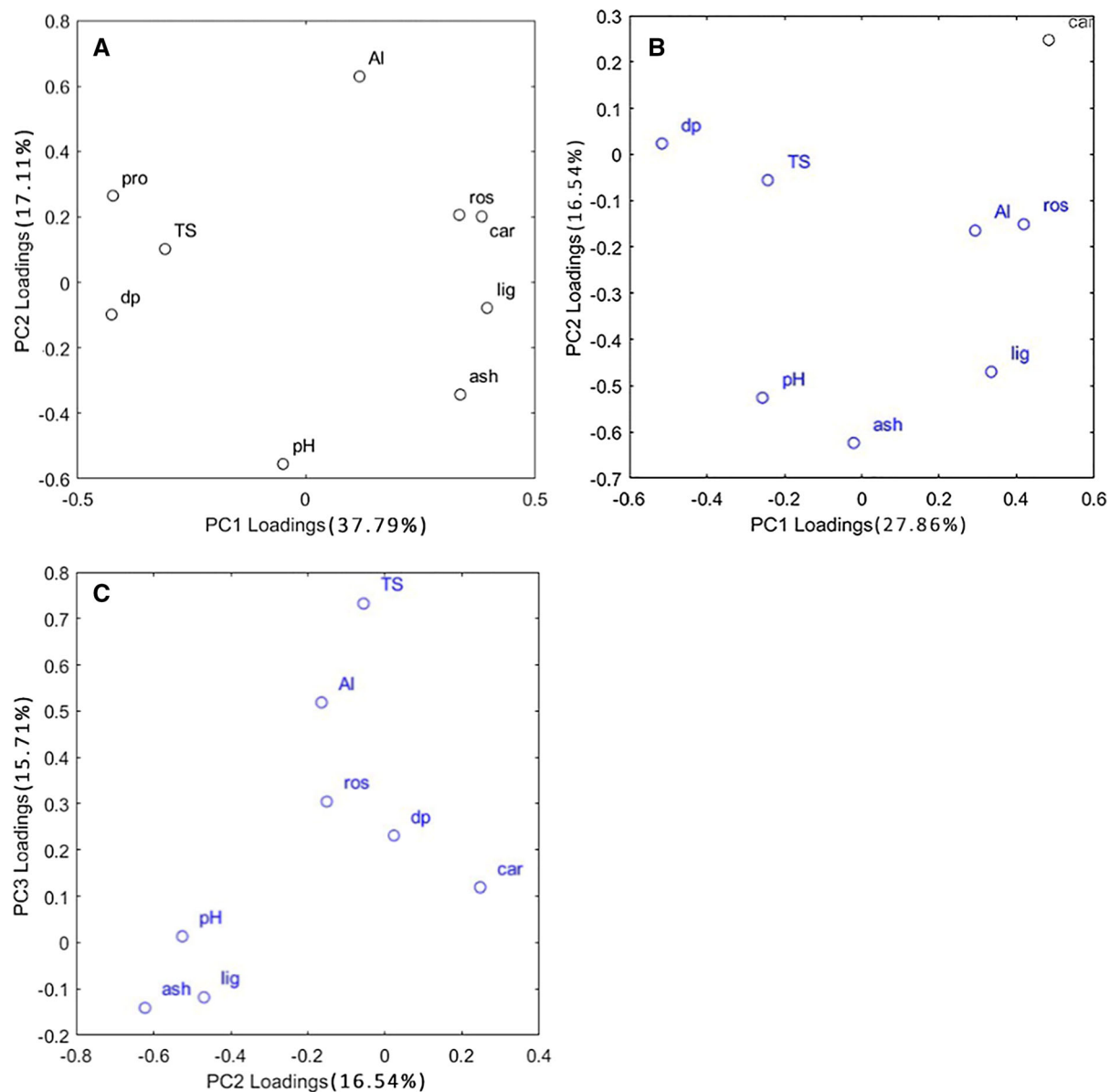
Fig. 2 Loading plots for: **a** PC2 versus PC1 for all samples with determined **pH**, **car**, **lig**, **ash**, **Al**, **DP**, **TS**, **ros**, **pro**; **b** PC2 versus PC1 for **bleached** samples with determined **pH**, **car**, **lig**, **ash**, **Al**, **DP**, **TS**, **ros**; **c** the same as B, except PC3 versus PC2

groups (Whitmore and Bogaard 1994) with their starting content ∼ 0 (Fig. 3B).

The analysis also provides insight into the reactivity of rosin and lignin in historic papers which challenges the possible assumption that the determinable **ros** and **lig** represent their original contents. **Ros** is found to have strong effect on **car**, suggesting that the contents of degraded **ros** and remaining **ros** (as measured, cf.

supplementary information) are correlated. This is consistent with the well-known instability of rosin, which via autoxidation leads to rearrangement of double bonds and formation of endo- and hydroperoxides, epoxides, and hydroxyl and keto groups (Prinz et al. 2002). In the presence of air and light, the structural transformation of abietane acids starts within several hours and leads to good yields of oxidation products in a few days at room temperature
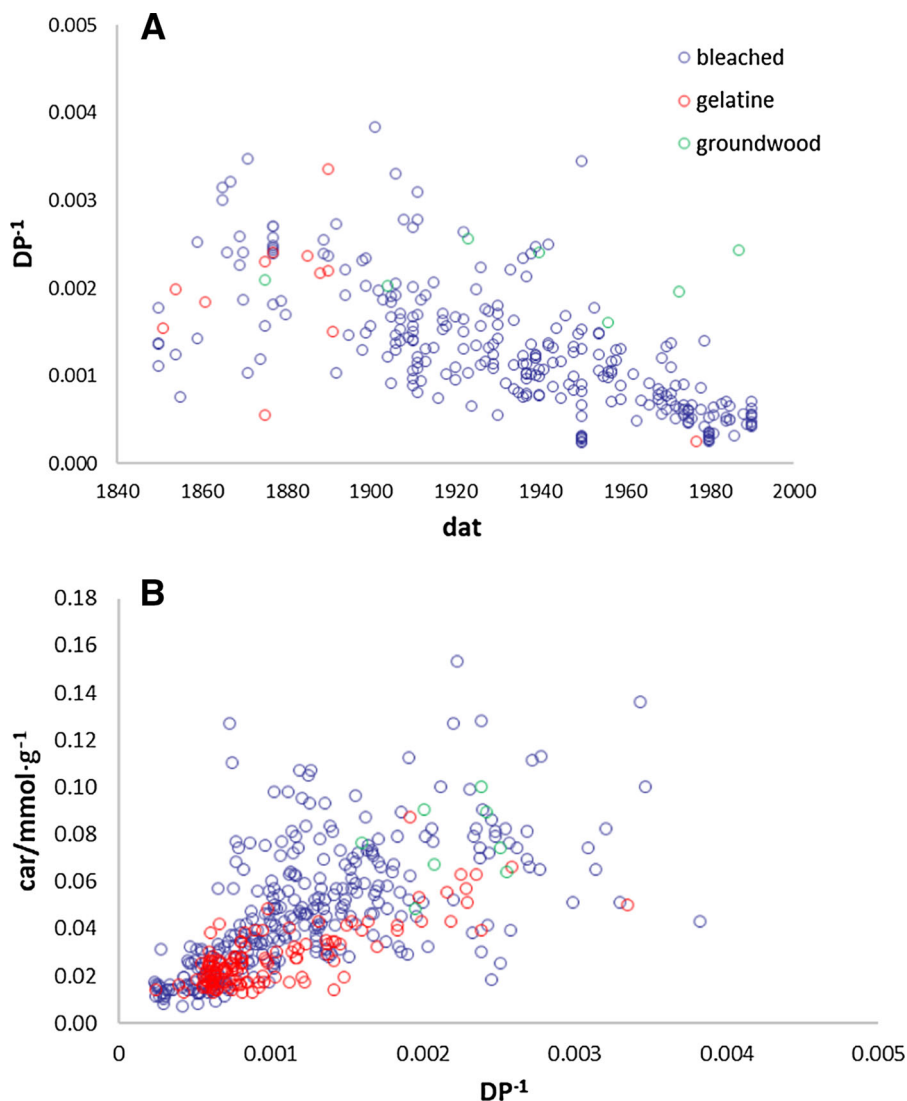
Fig. 3 Scatterplots for (**a**) DP^{-1} and **dat** (N = 290) and (**b**) **car** and DP^{-1} (N = 441), suggesting that acid-catalysed hydrolysis is the main degradation mechanism of historic samples from 1850–1990, leading to the accumulation of **car**

Table 1 Summary of the statistics for the MLR of **lig**, **ros** and **scission** on **car**. N = 295. R-squared = 0.45

Parameters	Estimated coefficients	SE	p value
Intercept	0.0212	0.0020	1.3579×10^{-21}
lig	0.5524	0.3875	0.1551
ros	1.0166	0.2664	0.0002
scission	1.1273	0.1006	1.8884×10^{-24}

(Ren et al. 2015; Enoki 1976; Schuller and Lawrence 1961).

Given the uncertainty in the estimation of the molecular weight of lignin due to the chemical variability and the analytical challenge of structural characterisation (Lange et al. 2013), it is difficult to draw solid conclusions on its contribution to **car** based on MLR. The selected molecular weight of **lig** is an approximation only as the real molecular weight of lignin in degraded paper is unknown. Nevertheless, its actual value does not affect the strength of correlation in Table 1, which provides evidence of the absence of

a linear relationship between **car** and **lig**. Generally, catalysis plays a major role in the oxidation of lignin (Lange et al. 2013). Due to the limited presence of catalysts in historic papers, the amount of degraded lignin is likely to be very low, especially in papers with low lignin content, which are mainly used in the above MLR analysis.

Arguably, acidity is the paper property that is of defining importance for the rate of degradation of paper, in addition to **DP** and environmental parameters such as temperature, humidity etc. (Zou et al. 1996a; Strlič et al. 2015). Therefore, we explore **pH** specifically in more detail for insights into its interactions with degradation of the three category papers.

Table 2 summarises the Spearman's rank correlation coefficients between **pH** and eight individual measured variables. In contrast with what is revealed by PCA (Fig. 2), no uniquely high correlation is found between **pH** and **Al** ($\rho = -0.2466$), suggesting degradation over time has led to much more complex relationships. Indeed, for all samples, **pH** is best related to **car** ($\rho = -0.5942$), **DP** ($\rho = 0.6209$) and **MW** ($\rho = 0.4547$) which show a clear dependence between low **pH** and high extent of hydrolytic degradation. Therefore, **pH** is not only the cause but also the result of paper degradation. Although the change in **pH** may not have been substantial, it is enough to establish a close association with degradation which may have largely affected the original correlation between **pH** and **Al**.

Hydrolysis, leading to both low molecular weight saccharides and their oxidation and formation of low molecular weight degradation products, including organic acids (Nevell and Zeronian 1985), may have the defining effect on **pH**. **lig** and **ros** could also contribute to acidity. Carboxylic acids (Lange et al. 2013) and enoic acids (Prinz et al. 2002) have been found to be the degradation products of rosin and lignin respectively, which can potentially contribute to the observed high correlation coefficients between **pH** and **ros** ($\rho = -0.4610$) as well as **pH** and **lig** ($\rho = -0.4717$). This provides further evidence that rosin and lignin may have degraded over time. Furthermore, the effect of degradation processes on **pH** is likely to be **pH**-dependent, which is mainly reflected in that a strong correlation between **pH** and **car** is obtained for both **bleached** ($pH_{median} = 5.8$, $\rho = -0.5534$) and **groundwood** ($pH_{median} = 5.1$, $\rho = -0.4842$) categories whereas for **gelatine**, the relationship is less clear ($pH_{median} = 6.8$, $\rho = -0.2587$, Table 1).

It is worth noting the few gelatine papers in the studied collection that have relatively low **pH**. These acidic gelatine papers were mainly produced in the nineteenth Century (Fig. S.27 in supplementary information), possibly associated with the increased use of alum in gelatine sizing to prevent gelatine spoilage (Fig. S.28 in supplementary information). Rosin sized papers were occasionally resized with gelatine to produce paper with improved mechanical properties (Garlick 1986) and a few examples of papers containing measurable amounts of **ros** and **pro** exist in the database (supplementary information). It adds a layer of complexity to the systematic change of sizing practices over time which lead to the low dependency of **lig**, **ros** and **Al** on **dat** ($\rho = 0.1598$, -0.2125, -0.0847, respectively).

A closer look at the nineteenth Century further reveals the variability and uncertainty of practices in this transitional and experimental period of paper-making history. Particularly, the transition in frequency of use of **woo**, **cot** and **cel** through time is highlighted by fibre furnish analysis. As shown in Fig. 4, although the use of **woo** fibres is prevalent in **groundwood** category samples (Fig. 4a), there are samples in this category that also contain **cot** fibres (Fig. 4b) which are unusual fibre combinations. The majority of the **bleached** papers contain more than 50% of **cel** fibres, however, a number of samples in this category have a surprisingly low quantity of **cel**

Table 2 Spearman's rank correlation coefficients between **pH** and selected individual measured variables (cf. supplementary information)

Variable	Paper category			
	All	**Bleached**	**Gelatine**	**Groundwood**
Al	− 0.2466	− 0.2750	− 0.2563	− 0.0718
ash	− 0.2505	− 0.1904	0.0811	0.1343
lig	− 0.4717	− 0.3082	− 0.1568	− 0.3471
pro	0.2188	n/a	− 0.1563	n/a
ros	− 0.4610	− 0.5213	n/a	− 0.0632
car	− 0.5942	− 0.5534	− 0.2587	− 0.4842
DP	0.6209	0.6049	0.5230	n/a
MW	0.4547	0.4406	0.2300	0.2981

Fig. 4 Distribution of raw fibre types: (**a**) **woo**, (**b**) **cot** and (**c**) **cel** through time for **gelatine**, **groundwood** and **bleached** paper categories
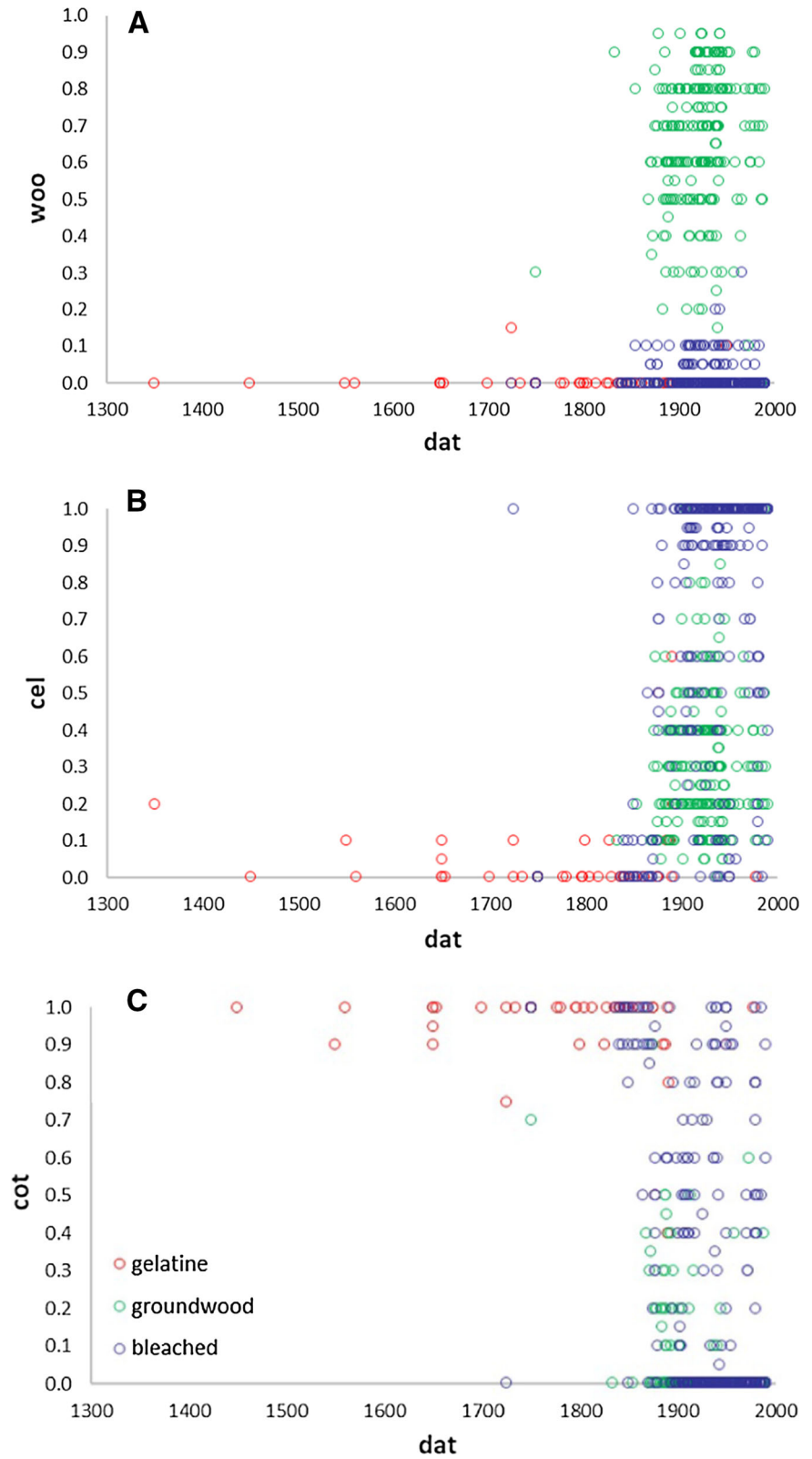
fibres (Fig. 4c), which seem to have been replaced with **cot** (Fig. 4b), mainly between 1850 and 1900. These uncertainties not only provide useful information for historians of technology but also may have a bearing on conservation as the presence of higher quality **cot** fibres in **groundwood** category samples can lead to better than expected mechanical properties for paper produced in the nineteenth Century.

Conclusions

In this research, we carried out comprehensive characterisation and analysis of a large database of historic European paper to reveal the underlying relationships between variables that illuminate the history of papermaking and paper degradation as interlinked processes. The uncertain use of fibre sources and sizing materials reflects the experimentation and exploration for the best practices in papermaking in the 19th Century. Acidity of paper in its current state of degradation is shown not to be only a consequence of alum-rosin sizing but a complex variable closely related to degradation of cellulose and possibly rosin and lignin. The average rate of acid catalysed hydrolysis of cellulose in papers from 1850 to 1990 is shown to be 10^{-5} year^{-1}, which appears to be closely associated with the accumulation of carbonyl groups. Questions emerge related to degradation of rosin and lignin, specifically to their contribution to carbonyl group content and acidity which has so far been largely overlooked in the literature. This research demonstrates the extraordinary value of chemical characterisation of a significant number of historic samples in the studies of their composition and degradation.

Compliance with ethical standards

References

Adams MS, Baker CA, Zachary S (2009) Sizing in nineteenth-century book papers. The Book Paper Group Annual 28:1–6

Asikkala J, Tamminen T, Argyropoulos DS (2012) Accurate and reproducible determination of lignin molar mass by acetobromination. J Agric Food Chem 60:8968–8973

Balažic A, Kolar J, Strlič M, Žagar E (2008) Size exclusion chromatography of cellulose. In: Strlič M, Kolar J (eds) Durability of paper and writing II. National and University Library, Ljubljana, pp 33–34. http://www.science4heritage.org/DPW2/DPW2%20book.pdf . Accessed 6 July 2020

Barański A, Lagan JM, Lojewski T (2005) Acid-catalysed degradation. In: Strlič M, Kolar J (eds) Ageing and stabilisation of paper. National and University Library, Ljubljana, pp 93–109

BS 4971 (2017) Conservation and care of archive and library collections

BS 7463-1 (1991) Fibre furnish analysis of paper, board and pulps. General method

BS ISO 5351 (2004) Pulps. Determination of limiting viscosity number in cupri-ethylenediamine (CED) solution

Burns RI, Lindgren U (1996) Europäische Technik im Mittelalter. 800 bis 1400. Tradition und Innovation, 4th edn. Gebr. Mann Verlag, Berlin, pp 413–422

Clapperton R, Henderson W (1947) Modern paper-making. Blackwell, Oxford

Cséfalvayová L, Pelikan M, Kralj Cigić I, Kolar J, Strlič M (2010) Use of genetic algorithms with multivariate regression for determination of gelatine in historic papers based on FT-IR and NIR spectral data. Talanta 82:1784–1790

Ekenstam A (1936) Über das Verhalten der Cellulose in Mineralsäure-lösungen, II. Mitteil: Kinetisches Studium des abbaus der Cellulose in Säure-lösungen. Ber Dtsch Chem Ges 69:553–559

Enoki A (1976) Isomerization autoxidation of resin acids. Wood Res Bull Wood Res Inst Kyoto Univ 59/60:49–57

Evans R, Wallis A (1987) Comparison of cellulose molecular weight determined by high performance size exclusion chromatography and viscometry. In: 4th International Symposium in Wood Chemistry, Paris – Palais des Congres

Garlick K (1986) A brief review of the history of sizing and resizing practices. AIC Book Paper Group Annual 4:94–107

Hunter D (2011) Papermaking: the history and technique of an ancient craft. Dover Publications, Mineola

Iiyama K, Wallis A (1988) An improved acetyl bromide procedure for determining lignin in woods and wood pulps. Wood Sci Technol 22:271–280

ISO 2144 (2000) Paper, board and pulps—determination of residue (ash) on ignition at 900 degrees C

ISO 1924-2 (1994) Paper and board—determination of tensile properties—part 2: constant rate of elongation method

Jackson JE (1991) A user's guide to principal components. Wiley, New York

Jolliffe IT (2002) Principal components analysis. 2nd edn. Springer, Berlin

Kes M, Christensen BE (2013) A re-investigation of the Mark–Houwink–Sakurada parameters for cellulose in Cuen: a study based on size-exclusion chromatography combined with multi-angle light scattering and viscometry. J Chromatogr A 1281:32–37

Lange H, Decina S, Crestini C (2013) Oxidative upgrade of lignin—recent routes reviewed. Eur Polym J 49:1151–1173

Launer HF (1939) Determination of the pH value of papers. J Res Natl Bur Stand 22:553–564

Łojewski T, Zieba K, Lojewska J (2010) Size exclusion chromatography and viscometry in paper degradation studies. New Mark–Houwink coefficients for cellulose in cupriethylenediamine. J Chromatogr A 1217:6462–6468

Lucas AR (2005) Industrial milling in the ancient and medieval worlds. A survey of the evidence for an industrial revolution in medieval Europe. Technol Cult 46:1–30

Manly BFJ (2004) Multivariate statistical methods: a primer. 3rd edn. Chapman & Hall/CRC, New York

Mardia KV, Kent JT, Bibby J (1979) Multivariate analysis. Academic Press, London

Nevell TP, Zeronian SH (eds) (1985) Cellulose chemistry and its applications. Wiley, New York

Potthast A, Radosta S, Saake B, Lebioda S, Heinze T, Henniges U, Isogai A, Koschella A, Kosma P, Rosenau T, Schiehser S, Sixta H, Strlič M, Strobin G, Vorwerg W, Wetzel H (2015) Comparison testing of methods for gel permeation chromatography of cellulose: coming closer to a standard protocol. Cellulose 22:1591–1613

Prinz S, Müllner U, Heilmann J, Winkelmann K, Sticher O, Haslinger E, Hüfner A (2002) Oxidation products of abietic acid and its methyl ester. J Nat Prod 65:1530–1534

Ren F, Zheng Y, Liu X, Yue X, Ma L, Li W, Lai F, Liu J, Guan W (2015) An investigation of the oxidation mechanism of abietic acid using two-dimensional infrared correlation spectroscopy. J Mol Struct 1084:236–243

Schuller WH, Lawrence RV (1961) Air oxidation of resin acids. III. The photosensitized oxidation of neoabietic acid and the configurations of the pine gum resin acids. J Am Chem Soc 83:2563–2570

Spearman C (1904) The proof and measurement of association between two things. Am J Psychol 15:72–101

Stol R, Pedersoli JL, Poppe H, Th Kok W (2002) Application of size exclusion electrochromatography to the microanalytical determination of the molecular mass distribution of celluloses from objects of cultural and historical value. Anal Chem 74:2314–2320

Strlič M, Kolar J (eds) (2005) Ageing and stabilisation of paper. National and University Library, Ljubljana

Strlič M, Kolar J, Kočar D, Drnovšek T, Šelih V, Susič R, Pihlar B (2004) What is the pH of alkaline paper? e-Preservation Sci 1:35–47

Strlič M, Grossi CM, Dillon C, Bell N, Fouseki K, Brimblecombe P, Menart E, Ntanos K, Lindsay W, Thickett D, France F, de Bruin G (2015) Damage function for historic paper. Part III: isochrones and demography of collections. Heritage Sci 3(1):40

Stromer W (1993) Große Innovationen der Papierfabrikation in Spätmittelalter und Frühneuzeit. Technikgeschichte 60:1–6

Szabolcs O (1961) Eine Kolorimetrische Methode zur. Bestimmung der reduzierenden Carbonyl Gruppen in der Cellulose. Papier 15:41–45

Whitmore PM, Bogaard J (1994) Determination of the cellulose scission route in the hydrolytic and oxidative degradation of paper. Restaurator 15:26–45

Wold H (1966) Estimation of principal components and related models by iterative least squares. In: Krishnaiah PR (ed) Multivariate analysis. Academic Press, New York, pp 391–420

Wold S, Esbensen K, Geladi P (1987) Principal component analysis. Chemom Intell Lab Syst 2:37–52

Zou X, Uesaka T, Gurnagul N (1996a) Prediction of paper permanence by accelerated aging I. Kinetic analysis of the aging process. Cellulose 3:243–267

Zou X, Uesaka T, Gurnagul N (1996b) Prediction of paper permanence by accelerated aging. II. Comparison of the predictions with natural aging results. Cellulose 3:269–279