Carbohydrate Polymers 326 (2024) 121611

Contents lists available at ScienceDirect

Carbohydrate Polymers

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

Profiling of historical rag papers by their non-cellulosic polysaccharide composition

Hajar Khaliliyan [a], Justine Lin [a], Paul Jusner [a], Sonja Schiehser [a], Markus Bacher [a], Mirjana Kostić [b], Thomas Rosenau [a], Antje Potthast [a], Stefan Böhmdorfer [a,*]

[a] Institute of Chemistry of Renewable Resources, Department of Chemistry, University of Natural Resources and Life Sciences, Vienna (BOKU), Konrad-Lorenz-Strasse 24, 3430 Tulln an der Donau, Austria

[b] University of Belgrade, Faculty of Technology and Metallurgy, Belgrade, Serbia

ARTICLE INFO

Keywords:
Crystallinity
Hemicellulose
Methanolysis
NIR
Pectin
PLS

ABSTRACT

Hemicellulose and pectin are noteworthy components of historical European rag papers, and have not been studied in detail so far. Rag papers were made from used textiles, and fiber-based utilities, such as ropes and bags. These had been prepared until the mid-19th century from plant-based fibers. Their polysaccharide composition could relate to their condition and history. This information can be expected to hold importance for the preservation and conservation of historical objects.

We investigated a collection of rag papers of different age for their composition of non-cellulosic polysaccharides, and compared the findings with modern rag papers and wood pulps. Furthermore, a non-destructive determination of the hemicellulose and pectin content by near-infrared spectroscopy was developed.

Historical rag papers had a lower hemicellulose/pectin content than pulps; the fractions of rhamnose, galactose, and arabinose were higher, while xylose was lower. In modern rag papers, xylose tended to be at the higher end of the range, which suggests a degradation of hemicelluloses/pectin over time or a change in raw materials and manufacturing. Rag papers also showed higher crystallinity than wood pulp papers. These findings provide insights into rag paper characteristics and offer potential classification methods.

1. Introduction

Historical European rag papers (rag) are made from fiber-based utilities, such as textiles, ropes and bags. These were produced from plant-based fibers, a sustainable choice driven by the lack of alternatives until the mid-19th century. Mainly flax and hemp fibers were used (Bloom, 2017), which consist of the four main constituents of plant biomass: cellulose (74–87 %), hemicellulose (2–15 %), pectins (1–4 %), and lignin (2.2–5 %) (Ansell & Mwaikambo, 2009; Espejo et al., 2010). The differences in plant species, plant cell types, and plant origin are reflected in the different carbohydrate composition of rag papers (see Tables S1–S3). These variations in chemical composition of plant fibers also affect the fibers' structural properties (Moriana et al., 2014). Due to different specific raw materials, production processes, and ageing of the historical rags, the percentage of their main compounds and their derivatives can be substantially different between rag paper samples (Becker et al., 2021; Čabalová et al., 2017; Erhardt & Tumosa, 2005).

During the process of rag papermaking, plant fiber derived materials are suspended, fermented, and beaten to obtain a fiber slurry, which is turned into paper sheets. Furthermore, the product undergoes additional physical treatments, such as sizing with gelatin, to achieve desired properties (Cséfalvayová et al., 2010; Erhardt & Tumosa, 2005). Well-made rag papers are in general superior in strength and permanence to wood pulp-based papers, which can be explained by their different overall composition. A relation between hemicellulose content, strength, and the degradation processes of paper was found (Becker et al., 2021; Čabalová et al., 2017).

Although hemicelluloses do not have a strictly defined chemical structure, there are four prominent groups: xyloglucans, xylans, mannans, and glucomannans, each named after its backbone structure. In grass cell walls, there is an additional class, namely mixed linkage glucans (β-(1→3,1→4)-glucans) (Carpita & Gibeaut, 1993; Scheller & Ulvskov, 2010). In plant cells, hemicellulose contributes to the cell walls' structural strength (Rubin, 2008; Scheller & Ulvskov, 2010).

* Corresponding author at: Institute of Chemistry of Renewable Resources, Konrad Lorenz-Straße 24, 3430 Tulln an der Donau, Austria.
E-mail addresses: hajar.khaliliyan@boku.ac.at (H. Khaliliyan), sonja.schiehser@boku.ac.at (S. Schiehser), markus.bacher@boku.ac.at (M. Bacher), kostic@tmf.bg.ac.rs (M. Kostić), thomas.rosenau@boku.ac.at (T. Rosenau), antje.potthast@boku.ac.at (A. Potthast), stefan.boehmdorfer@boku.ac.at (S. Böhmdorfer).

https://doi.org/10.1016/j.carbpol.2023.121611
Received 27 July 2023; Received in revised form 12 November 2023; Accepted 15 November 2023
Available online 17 November 2023

Hemicellulose in the secondary plant cell wall has different structural functions from that in the primary walls; with its distinct helical screw conformations it binds to cellulose and lignin and holds these macromolecules together (Kirui et al., 2022; Simmons et al., 2016). According to the recent cellulose-sliding model, hemicellulose might also allow some sliding of multiple cellulose microfibrils which could be crucial to the mechanical properties of the primary cell walls (Zhang et al., 2021). Pectin is a crucial component of all land plants being located at the middle lamella. It consists of a range of galacturonic acid-rich polysaccharides. It provides an environment for the deposition, slippage, and extension of the cellulose-matrix network and is an adhesive material between cells (Willats et al., 2001). Pectin also has extensive contact with cellulose, like hemicellulose, but its structural functions are mainly achieved by influencing the patterning and spacing of cellulose microfibrils (Kirui et al., 2021). Another potential role of pectin (in particular of homogalacturonan) is to act as a repellent after deacetylation, which pushes cellulose microfibrils apart in the primary cell walls (Haas et al., 2020).

Semi-controlled rotting (retting) is part of the production process of rag paper. It causes the fiber bundles to detach from the surrounding cells of the stem (McDougall, 1993). Earlier studies indicate that xylose contributes to the tensile index and strain at break compared to an untreated sample (flax fiber cells and bleached birch pulp fibers) (McDougall, 1993; Pere et al., 2019). However, hemicellulose and pectin are more sensitive to oxidation than cellulose, which results in early yellowing of paper and loss of strength of the sample (Ahn et al., 2019; Area & Cheradame, 2011; Korntner et al., 2015).

Preserving historical documents and books is a vital concern for conservators and conservation scientists. Various interior and exterior influences can alter an object's structure and cause the degradation of a historical object. For rag papers, their hemicellulose content, water content, degree of polymerization and crystallinity of the cellulose have a crucial role in further degradation (Area & Cheradame, 2011). Therefore, it is essential to study these variables to better understand the behavior and the current condition of historical rag papers (Ahn et al., 2018).

Several chemical analyses are necessary to assess a rag paper's composition, and some of them are destructive, for example the determination of the monosaccharide composition, the molecular weight distribution, or oxidized functionalities (Becker et al., 2021; Budischowsky et al., 2022; Röhrling et al., 2002). When historical objects are to be analyzed, establishing non-destructive methods is necessary. With this in mind, several spectroscopy analyses were already used, such as MIR (mid-infrared) spectroscopy in the Attenuated Total Reflectance (ATR) mode of archaeological bast fibers (Coletti et al., 2021), ATR-MIR spectroscopy for discrimination between natural and modified fibers (Geminiani et al., 2022), MIR and UV/Vis for evaluation of degradation of model paper (Łojewski et al., 2010), and NIR (near-infrared) spectroscopy for the determination of the degree of polymerization of historical paper (Liu et al., 2021). The spectra recorded with MIR and NIR both show the presence of functional groups. For multivariate modelling and sample classification, NIR is a better choice than MIR, since each functional group is indicated by several detected vibrations, and this redundancy is beneficial for multivariate analyses. The interpretation of the recorded spectra is of course not as direct as for MIR. The deeper penetration depth of NIR radiation means that the recorded spectra show the properties of the bulk of the sample and are less dominated by the surface characteristics.

Some compounds which contribute to a paper's stability or instability are present only in small amounts, are not assessed routinely, and can be challenging to detect. There are few comprehensive studies about the non-destructive characterization of European historical rag papers with a sufficiently large sample set, which investigated sample properties such as acidity, degree of polymerization (DP), the molecular weight of cellulose, grammage, tensile strength, as well as contents of ash, aluminum, carbonyl and carboxyl groups, rosin, protein, lignin, and fiber furnish (Henniges et al., 2009; Strlič et al., 2020). The content of non-cellulosic polysaccharide or crystallinity of the rag papers has not been investigated yet. Examples related to the characterization of polysaccharides contained in historical writings are relatively rare (Becker et al., 2021; Čabalová et al., 2017; Erhardt & Tumosa, 2005).

The aim of this work was to assess if non-cellulosic polysaccharides are involved in natural rag processes. This should become visible in a comparison between modern and historical rag papers. For both, comprehensive values for their chemical composition are not available. Therefore, the project studies the composition of the carbohydrate fraction in rags to establish typical values, investigates potential correlations and their causes, and establishes a non-destructive analytical method for quantifying hemicellulose and pectin by their NIR spectra. The monomer composition is determined by methanolysis followed by GC-FID (gas chromatography with flame ionization detection), cellulose crystallinity is determined by solid state NMR (nuclear magnetic resonance spectroscopy), and molecular weight parameters of the samples are determined by GPC-MALLS (gel permeation chromatography with multi-angle laser light scattering detection). The characteristics that distinguish historical and modern rags from pulp samples are identified by data visualization tools, exploratory data analysis and with principal component analysis (PCA) (Bro & Smilde, 2014) of the NIR spectra to facilitate a quick determination of a rag paper's overall state. To allow the non-destructive analysis of sugar content in historical rag samples by NIR spectroscopy, we establish a multivariate calibration method, such as PLS (partial least squares regression), by the correlation between the reference data of the methanolysis results and the recorded spectra (Rinnan et al., 2009).

2. Materials and methods

Solvents were of the highest available purity and used as received.

2.1. Samples

A total of 165 historical rag paper samples of different ages (16th to 19th century) were studied. They come from several European sources and feature a variety of colors, smells, writings, and unique watermarks. According to the results of our own microscopic evaluation, the historic rag papers (Fig. S1) were mainly produced from flax and hemp fibers, which is in agreement with literature (McCrady, 1992). Paper samples were stained with phloroglucinol and Astra blue and were observed in polarized light mode using an Olympus DX 1000 digital microscope. It was established in previous studies that rag papers typically have a low lignin content, usually below 3 % (Bloom, 2017; Henniges et al., 2009). The molecular weight distributions (M_w), carbonyl and carboxyl distributions were determined previously in our group with gel permeation chromatography (GPC) in dimethylacetamide/lithium chloride (DMAc/LiCl) (Henniges et al., 2009).

In addition, we studied 111 pulp and pulp paper samples (summarized as "pulp") that were recently manufactured from different wood sources and according to different pulping processes (hardwood (HW), softwood (SW), sulfite, and kraft process). The pulp samples were selected arbitrarily and encompassed various stages of bleaching.

Seven contemporary (modern) rag paper samples from different sources were also investigated (five unsized papers prepared from known fibers including one cotton, one linen, two with linen and hemp (Gangolf Ulbricht, Berlin, Germany), one hemp (University of Iowa), and two rag paper samples of unknown composition sized with gelatin and alum (one from a private collection and the other one from Moulin de Fleurac)). In contrast to the historical rag samples in this study, modern rags were prepared recently and are not aged.

2.2. Methanolysis

The monosaccharide composition of hemicellulose/pectin and

amorphous cellulose was determined by methanolysis, which was based on the protocol developed by Sundberg et al. (1996). Even if cellulose was not degraded completely by methanolysis, most of the glucose still originated from amorphous cellulose and was therefore not considered for calculating total hemicellulose and pectin content. The sample (4–6 mg, duplicates) was placed into a screw-capped vial, and acidic methanol (2 ml, 2 M HCl, Merck, Darmstadt, Germany) was added. The samples were heated for 4 h at 100 °C. After cooling the samples to room temperature, 200 μL of anhydrous pyridine (Sigma Aldrich/Merck, Darmstadt, Germany) and 100 μL of internal standard (sorbitol, Supelco/Merck, Darmstadt, Germany) were added. The methanol was evaporated under a stream of nitrogen at 45 °C. Then, the samples were silylated with *N*,*O*-bis(trimethylsilyl)trifluoroacetamide (BSTFA, Merck, Darmstadt, Germany) and analyzed by gas chromatography according to Becker et al. (2013, 2021).

2.3. ^{13}C CP/MAS NMR

Solid state ^{13}C cross-polarization/magic angle spinning (CP/MAS) NMR spectra were obtained on a Bruker Avance III HD 400 spectrometer (Bruker, Germany) with a resonance frequency of 400.13 MHz for ^{1}H and 100.61 MHz for ^{13}C. The system is equipped with a 4 mm dual broadband CP/MAS probe. ^{13}C spectra were accumulated at room temperature (20 °C) by applying either the total sideband suppression (TOSS) sequence with a MAS rate of 5 kHz or the direct CP sequence at 12 kHz. In both cases, a CP contact time of 2 ms, a recycle delay of 2 s, and SPINAL-64 ^{1}H decoupling were used. The acquisition time was fixed at 49 ms, and the spectral width was set at 250 ppm. All ^{13}C chemical shifts were referenced externally against the carbonyl signal of glycine at 176.03 ppm. The samples were swollen in deionized water overnight before obtaining the spectra. Peak deconvolution of the C4 resonance (91–79 ppm) was doneaccording to Wickholm et al. (1998) and Zuckerstätter et al. (2013) using the "Peak Analyzer" tool of the OriginPro2021 software (OriginLab Corporation, USA), which allows nonlinear least-squares fitting. For more details, see our previous study (Jusner et al., 2022).

Crystallinity indices (CI) were calculated by the following equation, where the crystalline peaks of the C4 resonance of cellulose [Iα, Iβ, Iαβ, and paracrystalline signal (PC)] are divided by the sum of all fitted curve integrals contributing to the C4 resonance [Iα, Iβ,Iαβ, PC, accessible surface I (ASI), accessible surface II (ASII), inaccessible surface (IAS)] and the hemicellulose C4 signal (HC).

$$CI = I\alpha + I\beta + I\alpha\beta + PC \Big/ \sum I\alpha + I\beta + I\alpha\beta + PC + ASI + ASII + IAS + HC$$

2.4. GPC-MALLS

GPC analysis was conducted using MALLS (multiple-angle laser light scattering) detection. Data evaluation was performed with Chromeleon 7.3 (Thermo), Astra 4.7 (Wyatt), and GRAMS/32 (Thermo Fisher Scientific) software. GPC conditions and details of the analysis were described previously (Potthast, 2015; Potthast et al., 2005). In summary, 10–15 mg of air dried rag or pulp samples are disintegrated in a blender with 0.25 l of deionized water for 20 s. The fibers are collected by filtration, transferred into a vial, and shaken with 3 ml of DMAc for 24 h. This solution is then filtrated again, mixed with 1 ml of DMAc/LiCl (9 %) solution and shaken for 24 h, followed by vortexing to ensure thorough mixing. Afterwards, the dissolved sample was diluted with DMAc to achieve the desired concentration.

2.5. NIR measurements

NIR spectra of the rag samples were recorded with an MPA Multi-Purpose Analyzer by Bruker (Billerica, MA, USA) equipped with a fiber optic probe with a TE-InGaAs detector for measurement in diffuse reflection mode. The recorded wavenumber range was 10,000 cm^{-1} to 4000 cm^{-1} with a spectral resolution of 8 cm^{-1}, and the number of scans for a single sample spectrum was 32. Three spectra were recorded for one sample to account for the heterogeneous surface of the papers. For pulp samples, NIR spectra were accumulated with an integration sphere with a Pb-S detector with 16 cm^{-1} resolution and 64 scans for each sample. The pulp samples were stored in a desiccator before measurement.

2.6. Data treatment

Data distribution graphs, rain cloud plots (Allen et al., 2019), scatter plot matrices, correlation plots, and parallel coordinate plots were prepared with OriginPro 2021 (OriginLab Corporation, USA) to show the similarities and differences between the historical rags and compare them with pulp samples. Multivariate calibration was performed using Matlab (ver. R2020a, Mathworks, Natick, MA, USA) software, PLS Toolbox (ver. 8.8.1., Eigenvector Research, Manson, WA, USA), and Bruker OPUS QUANT2 software.

3. Result and discussion

3.1. Non-cellulosic polysaccharide content

The total content of hemicelluloses and pectin in the historical and modern rag samples determined by methanolysis is shown in Fig. 1. For comparison, the result for pulps is given in Fig. S2, left. As seen from the distribution graph, the content of non-cellulosic polysaccharide in these samples is in the range of 6–31 mg/g with an average of 17 and a standard deviation (SD) of ±6 mg/g. In contrast, the range of the total hemicellulose and pectin content for 111 pulp samples is several times broader. It was 8–200 mg/g, with an average of 70 ± 44 mg/g, which is twice the highest value found for rag samples. Except for one modern rag paper (cotton fiber) with a very low hemicellulose and pectin content, the total neutral as well as acidic sugar units in non-cellulosic polysaccharides content in modern rag papers agree well with historical rag papers. The most abundant acidic carbohydrate unit in rags is galacturonic acid, which is associated mainly with pectin, followed by 4-*O*-

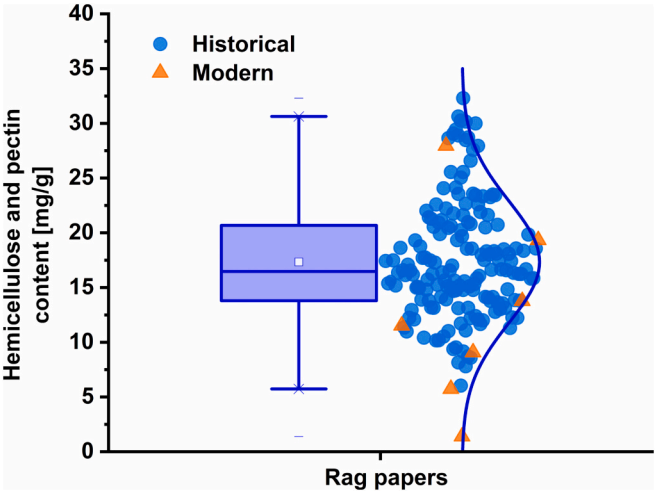

Fig. 1. Distribution plot of the total hemicellulose and pectin content of 165 historical and 7 modern rag paper samples. The graphs show the data as box plot (left) and individual data points (right) with a matching normal distribution function. The box indicates the 2nd and the 3rd quartile, the horizontal line the median, and the square the mean. The length of whisker is determined by the outermost data point within 1.5 × interquartile range. Except for one modern sample with very high sugar content, the modern samples tend to have a low content of degradable sugars; depolymerisation during ageing might contribute substantially to the degradable sugars in historical samples.

methyl glucuronic acid, which is mainly associated with the hemicellulose xylan. The other observed monomers are neutral sugars. The average acidic and neutral carbohydrate composition of rag and pulp samples is shown in Table S4.

3.2. Monosaccharide composition

The distribution of the individual monosaccharides of both rags and pulp samples is shown in Fig. 2 as a rain cloud plot (for comparison with the carbohydrate composition of flax, hemp and cotton fiber see Tables S2–S3). In most rag paper samples, arabinose (Ara), galacturonic acid (GalA), rhamnose (Rha), xylose (Xyl), mannose (Man), galactose (Gal), and glucose (Glc) were detected (Fig. 2, top). Since the historical rag papers came from different sources and are of different age, their monosaccharide composition varied. Glucose from the methanolysis-degradable part of cellulose was always the largest portion with an average of 53 mg/g. For pulp samples, the carbohydrate composition was markedly different (Fig. 2, bottom; note the different scales!). Compared to historical rag papers, most of the pulp samples contain a distinctly higher amount of xylose and mannose, and a lower amount of glucose with an average of 33 ± 14 mg/g; arabinose, galacturonic acid, 4-*O*-methyl glucuronic acid (4OMeGlcUA), and galactose were detected for only few samples (mostly softwood pulps), and rhamnose was not detected in pulp samples at all (Table S4). Even though wood contains a low amount of rhamnose (about 2 mg/g) (Willför, Sundberg, Hemming, & Holmbom, 2005; Willför, Sundberg, Pranovich, & Holmbom, 2005), its complete absence in pulp samples is due to the production processes that remove all rhamnose. The raincloud plot reveals distinct patterns in the glucose content of both rags and pulp, with rags exhibiting a trimodal distribution with maxima at 26, 43 and 62 mg/g and pulp samples displaying a bimodal distribution with maxima at 20 and 51 mg/g. The presence of three maxima in the glucose distribution of rags may be attributed to variations in paper condition and fiber source, while the bimodal distribution in pulp is caused by samples with low glucose content that are dissolving grade pulps from hardwood (HW) and some softwood (SW) sulfite pulps.

Seven modern, unaged rag papers were analyzed as well. These samples are an interesting point of reference, since they show neither signs of ageing nor of usage (Table 1). The observed monosaccharide composition was in the range of the historical samples (Table S4). The paper prepared from cotton was exceptional, since it contains only xylose and glucose. The two modern rag papers with undeclared fiber source were the only modern papers that contained rhamnose and galactose, which are also found in historical papers. However, rhamnose and galactose weren't detected in unsized rag papers with single fiber source. We expected to observe these two monomers in annual plant fiber papers since they contain higher amounts of pectic polysaccharides, which comprise galacturonic acid, arabinose, rhamnose and galactose (Voragen et al., 2003). Traditional methods of rag paper manufacturing with retting followed by extensive alkaline washing can lead to a reduction of the amount of pectin in rag papers. However, the quantity of remaining non-cellulosic polysaccharide is also influenced by the quality of the fiber utilized in paper manufacturing. This is also applicable to modern rag papers, which may or may not contain pectic polysaccharide-derived sugars, depending on the quality of the used raw material: fresh fibers or recycled materials. In the case of modern rags, they may be sourced from contemporary linen fabric that has undergone processing steps to convert raw flax fibers into textile linen; this could have well eliminated non-cellulosic polysaccharides. Conversely, fibers used in modern and historical rag papers that contain sugars derived from pectic polysaccharides may have been more recalcitrant or less treated, for example, not sourced from textiles, and therefore still retain some acidic sugar units in non-cellulosic polysaccharides. In addition, if acidic sugar units in non-cellulosic polysaccharides were able to survive the fiber processing and paper production stages and remain in the final paper sheet, the sizing agents applied after sheet formation and drying might have prevented their decay during the use and storage of the paper sheet. It can be assumed that sizing agents create a protective layer on paper, which aids in paper preservation. This protective barrier shields the fibers in paper from degradation (Barrett et al., 2016). Therefore, if acidic sugar units in non-cellulosic polysaccharides are present in the paper after the papermaking process, they are more likely to be preserved in papers that have been sized, as compared to those that have not been sized or have a deteriorated sizing layer (Dupont, 2002).

The lower amount of xylose and higher amount of methanolysis-degradable glucose for historical rag papers compared to modern rag and pulp samples is likely the result of a degradation process during ageing. Our findings agree with other published data. Erhardt and Tumosa (2005), in a study with 42 samples (six modern pulps and 36 rag papers), observed a similar effect: newer rag papers (post-1800) contained much more xylose and less glucose than older rag papers. It was proposed that historic production processes could have removed many more of the monomeric sugar components. In more recent historic manufacturing, the pulp was soaked for shorter periods, and biological fermentation was minimized or avoided. Becker et al. (2021) also showed a similar trend with 21 samples (17 pulp samples and 4 rag papers).

The observed differences between the three types of paper (pulp, historical, and modern rag samples) are also visible in a scatter plot matrix that correlates all determined parameters in pairs (Fig. S3). Here again, one can observe that some modern rag papers have a content of certain monosaccharides (e.g., glucose) similar to historical papers and some (e. g. xylose) are more similar to pulp samples. Some grouping is already apparent, but it is hard to track the position of every sample in this data space. Therefore, we examined the data set with the density-based clustering (DBC) method to see the possible clusters among the samples.

3.3. Looking for similarities

To see if rag papers can be summarized in smaller classes that might show similar properties, we identified samples with similar monosaccharide composition by unsupervised Density-Based Clustering (DBC). This approach assumes that similar samples form clusters of denser population in a data space (i.e., the samples are close to each other). These classes would also be separated by sparse areas of low sample density. The density-based approach finds these high-density regions, clusters, of any shape and identifies points as noise (Class 0) when they are isolated in low-density areas (Ester et al., 1996). We applied DBC on the monosaccharide data of 14 pulp samples, 165 historical rags, and seven modern rags (Fig. S4). These 14 samples of pulp samples (softwood, sulfite process, dissolving pulp with low hemicellulose content) were selected since they were in the total hemicellulose and pectin content range of historical rag papers. The application of DBC to the data space formed by the monosaccharide content of rhamnose, xylose, and glucose showed that the historical rags can be divided into three classes. The other sugar monomers do not contribute to clustering. Unexpectedly, this method assigned both modern rag papers prepared from single fibers and pulp samples as "noise" (class 0), and two modern sized rag papers of undeclared composition to cluster 2. This observation indicated that the differences between modern and historical rag papers in carbohydrate composition are so pronounced that they are detected as a separate class even by an unsupervised classification algorithm.

The detected classes are related to the monosaccharide content. The relationship is more readily seen in a parallel coordinates graph (Fig. 3) instead of the three separate plots that DBC yields (Fig. S4). This type of graph is not intended to show quantitative values of individual samples, but emphasizes groups within a sample set. Modern rag papers (orange, in the foreground) show a different pattern than the historical rags in shades of blue with different lightness. For wood pulps (black, in the background), a very different monosaccharide composition is apparent. Rhamnose is absent in pulp and single fiber rags, and this monomer

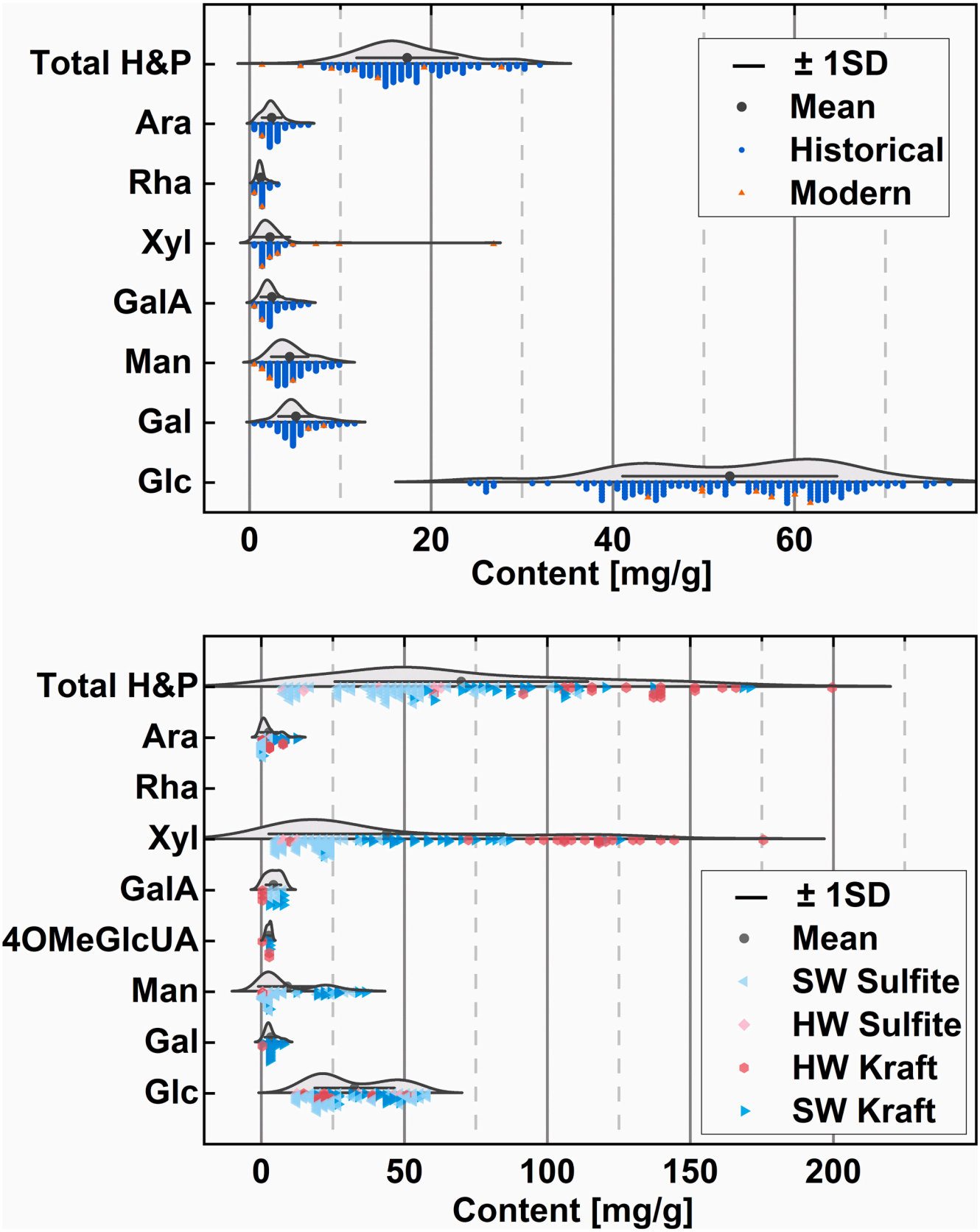

Fig. 2. Rain cloud plot of monosaccharide content in historical and modern rag samples (top) and pulp samples (bottom). Total hemicellulose and pectin content (Total H&P) was calculated without glucose. HW: hardwood; SW: softwood; Ara: arabinose; Rha: rhamnose; Xyl: xylose; GalA: galacturonic acid; 4OMeGlcUA: 4-*O*-methylglucuronic acid; Man: mannose; Gal: galactose; Glc: glucose.

Table 1
Carbohydrate composition of the studied modern rag paper samples (mg/g) according to methanolysis.

Samples	Total H&P[a]	Arabinose	Rhamnose	Xylose	Galacturonic acid	Mannose	Galactose	Glucose
Cotton paper	1.3	nd[b]	nd	1.3	nd	nd	nd	61.9
Linen paper	27.9	nd	nd	26.8	nd	1.0	nd	44.0
Hemp/linen paper a	13.7	1	nd	10.2	0.6	1.8	nd	59.8
Hemp/linen paper b	9.0	nd	nd	7.6	nd	1.5	nd	49.6
Hemp book paper	5.7	nd	nd	5.0	nd	0.6	nd	50.1
Rag paper a	11.5	nd	0.9	1.8	nd	2.6	6.1	55.7
Rag paper b	19.3	nd	1.4	2.8	1.7	4.7	8.5	57.4

[a] Total hemicellulose and pectin content.
[b] Not detected.

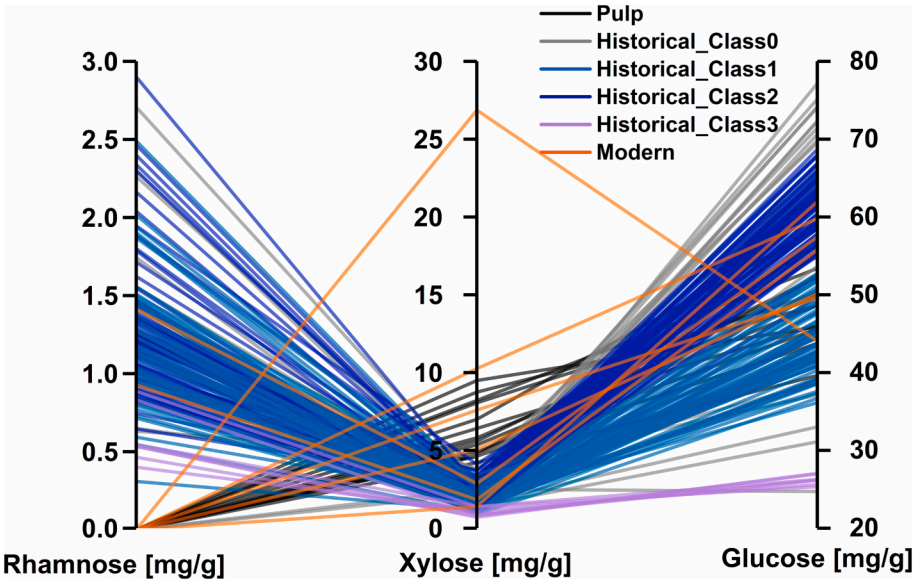

Fig. 3. Parallel coordinate plot of 165 historical rag samples, 7 modern rag samples, and 14 pulp samples; class 0 is considered noise by the density-based clustering method.

content is always low for rag class 3 and high for classes 2 and 1. The trend is the opposite for xylose: the amount of xylose for pulp and single modern fiber papers is always the highest. The modern rag paper with the most xylose is linen rag paper. Furthermore, the xylose content in the historical rags and the two modern rag papers with sizing agents are similar. Considering the glucose content, the classes of historical rags samples are clearly separated. The two modern rag papers with heterogeneous composition were assigned to class 2 of historical rags, which has the highest content of degradable cellulose (glucose). Class 1 has the second-highest, and historical rag class 3 has the lowest glucose content. For pulps, the glucose content is in the range of classes 1 and 2. The samples not assigned to any class (class 0) cover the range from lowest to highest content of rhamnose, xylose, and glucose.

The formation of these clusters could reflect the condition of the samples. Some samples are more prone to hydrolysis, which results in higher glucose content, while newer samples in probably better condition exhibit the opposite trend (Becker et al., 2021). However, it is essential to note that unsuitable storage conditions of the samples, such as inappropriate relative humidity or temperature, provides good condition for growing mold and would thus result in lower glucose or total hemicellulose and pectin content (Erhardt & Mecklenburg, 1994). Our observations indicate that mold-contaminated samples generally have lower hemicellulose and pectin content (class 3). The classification by DBC, as visualized in Fig. 3, might therefore be related to the actual condition of the historical samples. In addition, we also examined the ratio of these monomers to extract other potential reasons for this clustering (Fig. S5). The most discriminating feature is the ratio of rhamnose to other monosaccharides, e.g., glucose and xylose. Here, we can see some trends between the sample classes, such as the different average ratio (Fig. S5 top and middle), that support the observed classification. It is also interesting to observe the contrasting positions of two contemporary-sized rags on the graphs with the remaining modern rags, and that pulp samples show completely different ratios. We could not relate these ratios reliably to the fiber source of the samples since the ratio of hemicellulose to pectin monomers in original fibers and treated fibers seems to depend heavily on the rag paper production processes.

3.4. Condition of the cellulose

3.4.1. Crystallinity

The crystallinity of cellulose can be determined through different techniques. The results of these techniques cannot be unified because the methods detect different chemical and physical properties of cellulose. The determination by solid-state ^{13}C CP/MAS NMR is a model-based method that reflects the relative extent of the crystalline cellulose instead of an absolute value (Ling et al., 2019). It is appropriate for determining cellulosic polymer chains' supramolecular arrangement (Peter, 2021). Here, we used it to determine the crystallinity of cellulose in the samples (see Table 2). From the respective subplot in Fig. 6, we can see that the CI of historical rag papers ranges from 0.60 to 0.64 (average of CI = 0.62 ± 0.01), the CI of modern rag paper is slightly higher (0.64 ± 0.02), and for pulp samples markedly lower with 0.52 to 0.58 (average of CI = 0.54 ± 0.01) (see Table 2). This confirms the typically higher CI of the rag papers produced from annual plants in comparison to pulp samples (Henniges et al., 2011). As illustrated in Fig. S6, the C4 resonance of the crystalline core part consists of the

Table 2
Paper samples used and their crystallinity (CI) index with standard deviation as determined by solid-state ^{13}C NMR spectroscopy and peak deconvolution (Jusner et al., 2022).

Samples name (number of samples)	Flax paper (N = 1)	Hemp paper (N = 1)	Cotton paper (N = 1)	Historical rags paper (N = 10)	Modern rags paper (N = 7)	Pulps (N = 14)
Crystallinity index	0.64	0.64	0.69	0.62 ± 0.01	0.64 ± 0.02	0.54 ± 0.01

crystalline part of cellulose Iα, Iβ and a combination of the allomorphs Iα and Iβ (Iαβ), and the para-crystalline phase (PC). The C4 resonance of the amorphous part of cellulose represents two contributions from accessible crystallite surfaces (ASI and ASII), and a broad signal assigned to inaccessible surfaces (IAS). If hemicelluloses (mostly xylan) are present in the sample, an additional C4 resonance signal (HC) is imposed over the overall C4 peak shape. With regard to the accessibility to water molecules or other solvents, non-ordered amorphous cellulose has similar properties to accessible crystallite surfaces (Jusner et al., 2022). The graphs in Fig. S6 show that the crystalline part of historical rag samples is greater than that of pulp samples. Based on these results, the CI of both types of samples (10 historical and seven modern rags, 14 pulps) were calculated (see Table 2 for the comparison of the average of CI of the samples and Fig. 6 for the results in the context of other parameters). The relative intensity for each peak position and full width at the half maximum (FWHM) are presented in SI Table S5–7.

3.4.2. *Molecular weight distribution*

A major part of the historical rag samples (110 out of 165), and all modern rag samples were subjected to GPC to determine their molecular weight (Fig. 4). A corresponding graph for pulp sample is presented in Fig. S7. The observed range in historical rag paper samples for M_w was 75–319 (kg/mol) with an average of 211 ± 87 (kg/mol). Modern rags agree well, with a tendency towards higher values. Only the M_w of the modern sample made from cotton is out of the range with an M_w of 833 (kg/mol); this sample was already noted as exceptional due to its extremely low hemicellulose and pectin content. This is an indication that the historical samples in this study indeed do not contain cotton fibers; also, no cotton fiber was observed during microscopy. The range of the M_w for this collection of pulp samples (108 out of 111) is 155–687 (kg/mol) with an average of 383 ± 154 (kg/mol).

Examples of the molar mass distribution of each investigated paper, historical and modern rag as well as pulp, are shown in Fig. 5. In historical rag paper samples, significant chain cleavage must have occurred, resulting in a considerable decrease in weight-average molecular mass compared to modern rag (Fig. 5). We can also see the effect of a deteriorating condition by the hemicellulose peak, e.g., the lower peak at the left side of the graph, which is narrower for historical samples in contrast to modern rags. This is another indication of the occurrence of polymer degradation in rag papers during natural ageing. It would be worthwhile to investigate the underlying mechanisms in future work.

A scatter plot matrix (Fig. 6) allows examination of the relationship between characteristics of the paper samples, including glucose content, xylose content, crystallinity index (CI), and molecular weight (M_w). The CI showed a clear correlation with xylose and surprisingly did not correlate with degradable glucose at all. The former is expected, since the NMR method assigns the hemicellulose (HC) peak, which includes xylan, to the non-crystalline part of cellulose; thus, a larger HC peak decreases the crystallinity value in NMR. The unexpected lack of correlation between degradable glucose and crystallinity might either be a true representation of the samples' composition or simply be caused by the two different perspectives that methanolysis and NMR provide on a sample. In methanolysis, glucose and xylose are both detected with equal precision. In NMR, the hemicellulose signal (which includes xylan) is distinct, and thus reliably deconvoluted. The signal for amorphous cellulose in NMR (IAS), however, is very broad and is superimposed by several other peaks. It is therefore less precisely defined in peak deconvolution. In contrast, if we assume that the measurements are correct and comparable across methods, the lack of correlation means that the degradable glucose released during methanolysis and the amorphous cellulose detected by NMR are two different fractions of cellulose.

Apart from that, no correlations were observed across parameters. Samples of different origin separated according to the monosaccharide content as determined by methanolysis. Single fiber modern rag papers behaved as outliers for most of the parameters, which can be expected from their more defined hemicellulose/pectin composition. These

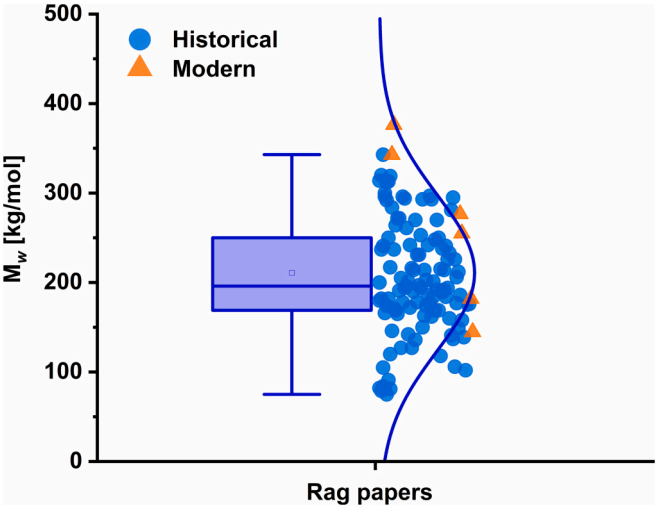

Fig. 4. Distribution plot of M_w of 110 out of 165 historical and 7 modern rag papers. The graphs shows the data as box plot (left) and individual data points (right) with a matching normal distribution function. The box indicates the 2nd and the 3rd quartile, the horizontal line the median, and the square the mean. The length of whisker is determined by the outermost data point within 1.5 × interquartile range. Except for one modern sample with a molecular weight in the lowest quartile, the modern rag papers tend towards high molecular weight.

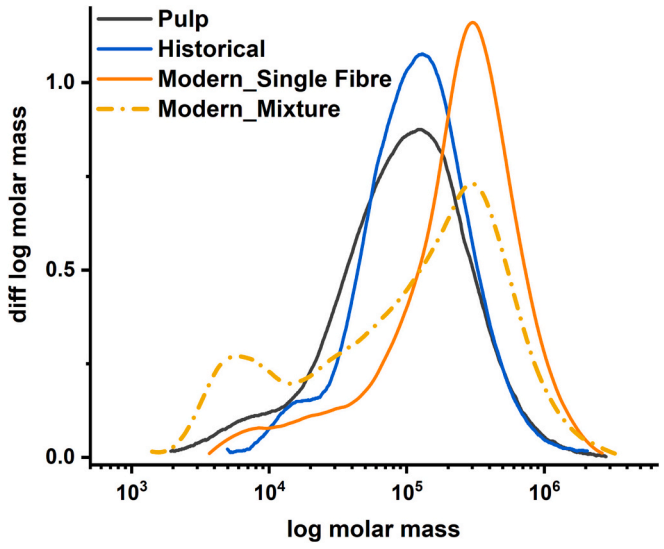

Fig. 5. Exemplary molar mass distributions of a pulp paper, a historical rag paper, and two modern rag papers with mixed (flax, hemp, and cotton according to microscopy) and single (hemp) fiber sources.

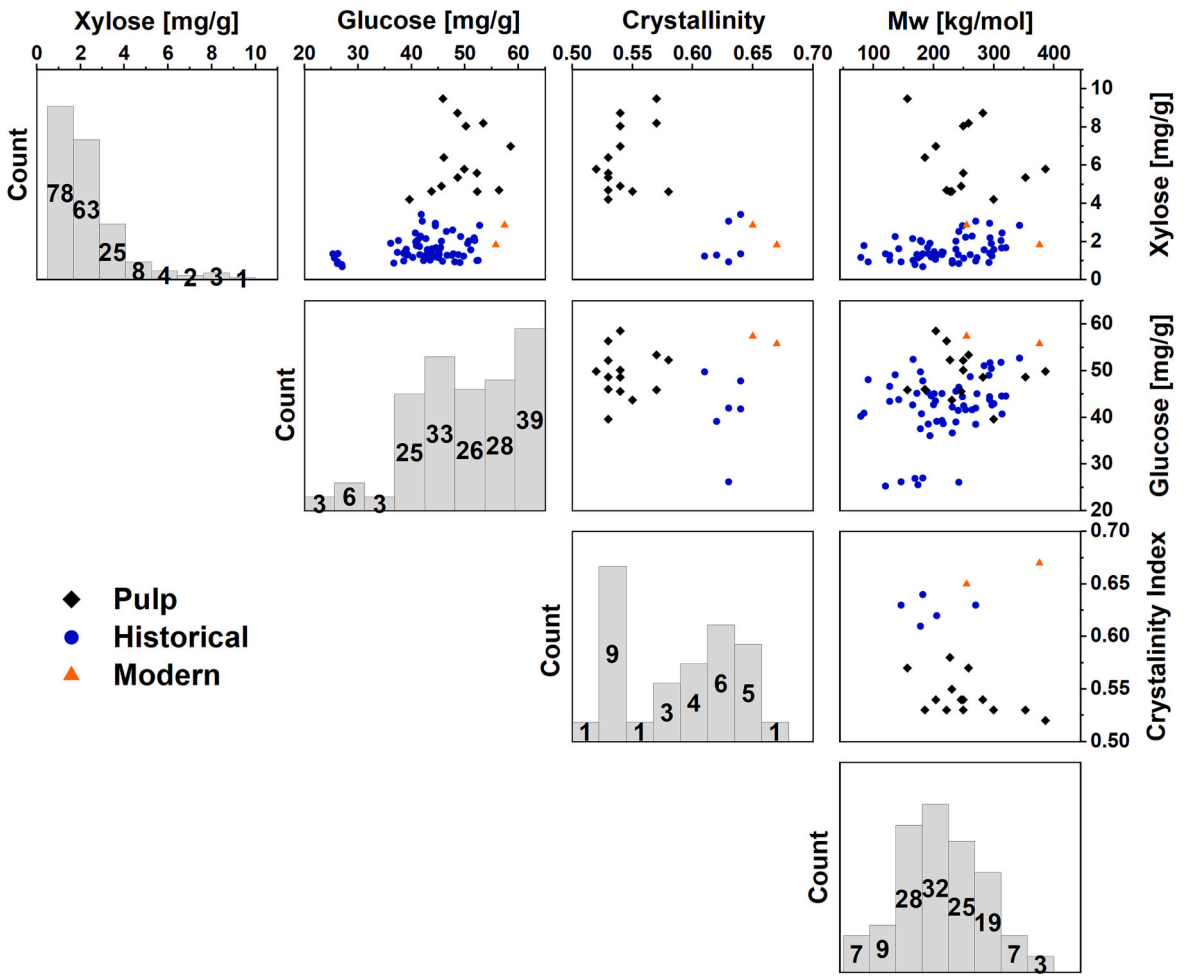

Fig. 6. Scatter plot matrix of historical rags (10 samples for crystallinity index and 110 samples for other parameters), pulp samples (14), and modern samples (2), to inspect chemical parameters for correlations (not all the parameter values were available for each sample, therefore the number of samples in each correlation varies). The numbers in the histograms give the number of samples in each bin.

samples have an extremely high amount of xylose (linen) or a high amount of glucose, a low amount of xylose, and a very high M_w (cotton). Crystallinity was consistently higher for rag papers than for pulps.

3.5. NIR spectroscopy and multivariate data evaluation

NIR spectroscopy is an excellent analytical tool that requires minimal sample preparation and is non-destructive. In the widely used mid-infrared spectroscopy (MIR), the presence of functional groups and structural elements is measured and not the presence of a single, specific compound. In contrast, near-infrared (NIR) spectra represent the entirety of a molecule, as it detects the sum of different vibrations and overtone vibrations. To comprehend and interpret a spectrum band assignment is necessary, and a correlation with reference data is required for quantitative work (Sandak et al., 2016; Schwanninger et al., 2011). Ideally, a property such as the content of hemicellulose and pectin can then be determined by a quick, non-destructive measurement instead of an hour-long sample preparation that requires precious sample material and several reagents. Since raw NIR spectra can be distorted by interferences, such as noise, it is essential to alleviate these interferences by preprocessing, which must be optimized during model development. In this work, the first derivative Savitzky-Golay filter (with an 11 point window width using a second order polynomial), Standard Normal Variate (SNV), and mean centering were selected as the most effective preprocessing methods (Rinnan et al., 2009). The NIR spectra of all paper samples pre-processed with first derivative Savitzky-Golay are shown in Fig. S8, and Table S8 provides major bands that are essential variables for detecting different sample classes (shown in loading plots and variable importance for projection (VIP) scores plot).

To determine a quantitative parameter, such as hemicellulose and pectin content, the values of the reference analyses have to be correlated with the NIR spectra. There are two types of correlation for calibration: univariate calibration, in which the relationship $y = f(x)$ is obtained with only one spectral variable, e.g., the absorbance at a single wavelength, and multivariate calibration, where the relation is obtained with data from several wavelengths or the whole spectrum, e.g., absorbance at every wavelength. The multivariate approach is superior since it can take many variables into account to understand the relationships between them. It aims at recognizing the variables that contribute the most to the overall data variability, which is essential for understanding complex data or processes and reduce the contribution of noise (Khaliliyan et al., 2020). Also, correlated variables are isolated and can be considered in deeper investigations for model development and analysis. Rather than just displaying statistics, multivariate data can be presented in various forms, making patterns easier to spot, interpret and control for outliers (Bro, 2003). In this project, Partial Least Squares (PLS) was used as a multivariate calibration method to create a model that predicts hemicellulose and pectin content from NIR spectra. The PLS method consists of the correlations between two matrices: X and Y. X represents the independent variables, e.g., spectral data, and Y the dependent variables, e.g., concentrations. With this, one can predict a chemical parameter (a variable in the Y matrix) from an NIR spectrum

(values in the X matrix).

In a preliminary step, Principal Component Analysis (PCA) was used to find notable differences between the NIR spectra of different classes of the studied samples. In this method, only NIR spectra (first derivative) were analyzed to check for the presence of clusters. Fig. 7B–D shows that the samples are scattered due to very heterogenous sample sources. While the previously determined samples classes are not obvious in this representation, PCA shows some discrimination between historical and modern rag samples. The plot of Q residuals and Hotelling's T^2 statistic (Fig. 7A) provides a consolidated view of both residual and score outliers (as well as inliers); Hotelling's T^2-test is a multivariate version of Student's *t*-test with a preferred confidence limit of 95 %, and Q residuals show how satisfactorily each observation corresponds to the confidence level of the PCA model. Using this plot, outliers can be identified (Fig. 7): the samples above the threshold lines of 1:1 (dashed lines) are generally considered to be both score and residual outliers. For the investigated sample set, only one sample was a proper outlier both in terms of residuals and scores, namely the modern single fiber rag paper made from cotton.

In this PCA model, the scores capture 85.52 % of the total x variance with p = 0.95, and the variance that remains in the residual is 14.48 % with p = 0.95. The first principal component (PC1) captures most of the

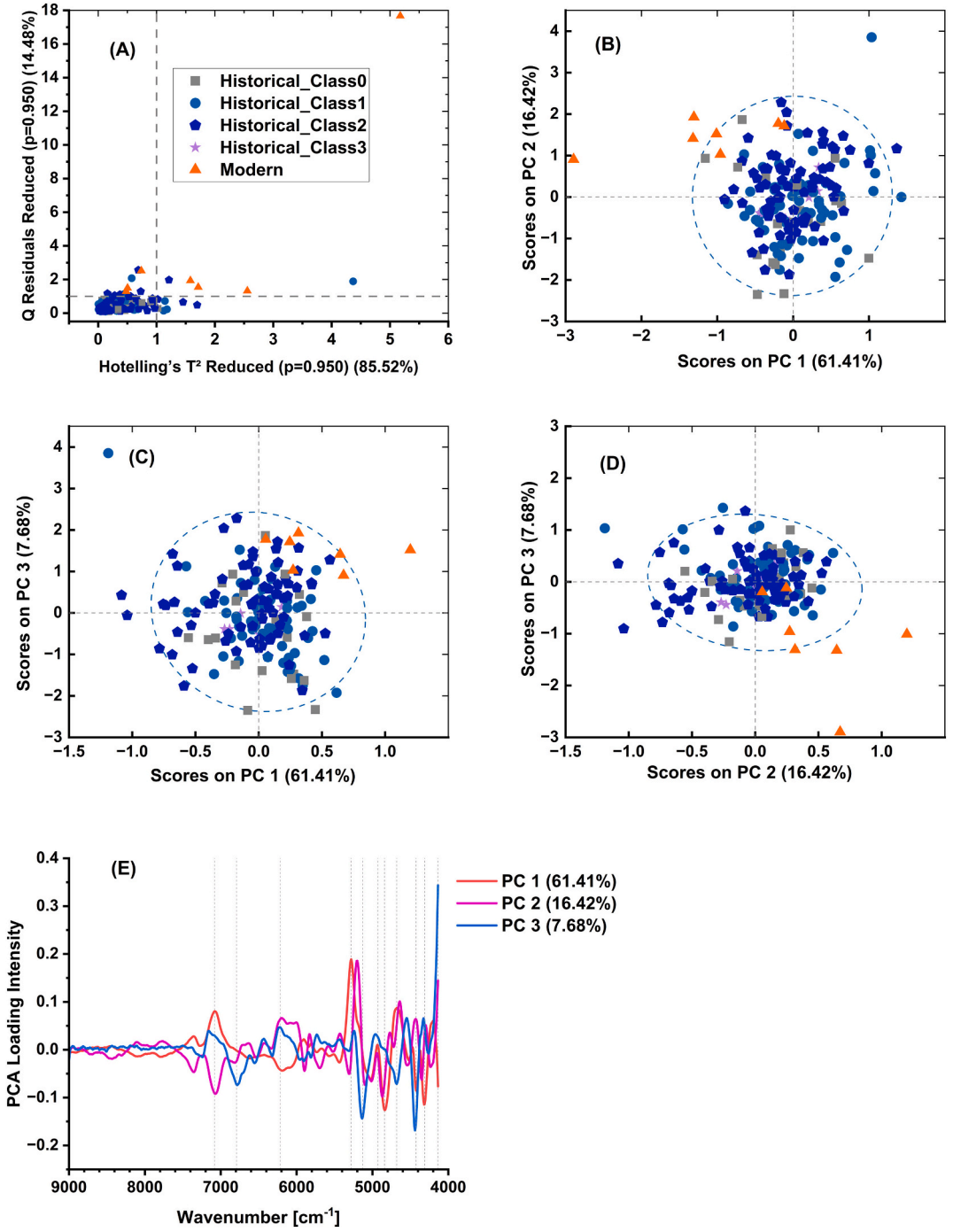

Fig. 7. (A) plot of Q residuals vs. Hotelling's T^2 statistic. This plot provides the trend of the residual as well as outliers and inliners. The samples above the threshold lines of 1:1 can be considered to be both score and residual outliers; (B), (C) and (D): PCA scores plots of the rag samples (165 historical and seven modern samples) for NIR spectra with Savitzky-Golay 1st derivative, SNV, and mean centering preprocessing. (E): Loadings of the first 3 principal components.

variance in the data (61.41 %), followed by the orthogonal second principal component (PC2, 16.42 %) and the third principal component (PC3), which retained 7.68 % of data variation (Fig. 7B–D). The explained variations appear to be related to the water content and C—H and O—H stretching vibrations, as we can observe from the loading plot (Fig. 7E). For this reason, it was important to store samples in a desiccator before recording their spectra. Both physical and chemical changes could be the cause for these differences. In the loading plots, which show the most influential wavelength, the most important variables are highlighted as: 7080, 6792, 6216, 5280, 5128, 4928, 4840, 4680, 4424, 4312, and 4136 cm^{-1}, which are related to O—H stretching of water and cellulose, C—H stretching of cellulose, hemicellulose, and other extractives (Schwanninger et al., 2011; Toscano et al., 2017) (Table S8). The classes in the parallel coordinate plot (Fig. 3) and density-based clustering (Fig. S4) are distributed across all regions in PCA plots and cannot be discerned in any combination of these principal components.

When pulp samples are included, the PCA model score and loading plots (Fig. S9) show that modern rag papers still associate with historic rag papers and not with pulps, and that they are not considered outliers any more based on their Q residuals vs. Hotelling's T^2 statistic plot (Fig. S9 top left). Most differences between rags and pulp samples are related to the wavenumbers: 7368, 7088, 6784, 6088, 5872, 5280, 5120, 4920, 4784, 4696, 4504, 4432, 4352, 4248, 4200, and 4136 cm^{-1} (Fig. S9, bottom right) (see Table S8). In this case, PCA can easily separate the pulps from rags.

Multivariate calibration PLS models were built to predict the content of hemicellulose and pectin (Fig. 8). The prediction results of rhamnose, xylose and glucose content are shown in Fig. S10. Different preprocessing methods were tested to find the most beneficial preprocessing options. To determine the quality of a model, the sample set was split into a calibration (Cal) and cross validation (CV) set to build the models. The aim was to minimize the root mean squares error of cross-validation (RMSECV, accounting for prediction performance) and to maximize correlation of the cross validated (R^2CV) and the reference values (R^2Cal). Since some parts of the spectra have no relevant information for building the models, they were excluded. In the end, the spectral range from 9000 cm^{-1} to 4100 cm^{-1} was used.

The final model for the determination of the total hemicellulose and pectin content (Fig. 8, left) had a RMSECV of 3.3 mg/g, which is in the range of the uncertainty of the reference method. Higher values are underestimated, and lower values overestimated, which is illustrated by the difference between the green line (prediction) and the red line (perfect agreement). The model was based on 6 latent variables (LV) and had an RMSEC (root mean square error of estimation of calibration) of 2.8 mg/g, and an R^2Cal of 0.71 and R^2CV of 0.60, which are reasonable values for such a heterogeneous sample set. Three more models were developed: for the determination of rhamnose content, xylose content and glucose content (see Table 3 and Fig. S10). The RMSE of all three models is in the range of reference methods (methanolysis). Overall, the calibration for total hemicellulose and pectin performs well enough for quantification, while the calibrations for rhamnose, xylose and glucose are more suitable for screening.

The wavelengths that affect the model can be investigated by the variable importance for projection (VIP) method (Andersen & Bro, 2010). For the hemicellulose and pectin models (Fig. 8, right), the VIPs are 8913, 8798, 8458, 7251, 7101, 6807, 6742, 6171, 5526, and the relevant range is from 5315 to 4160 cm^{-1}. These bands have been assigned to hemicellulose, cellulose, other extractives, and water content (Schwanninger et al., 2011). The number of essential variables (VIP scores plot) for the rhamnose and xylose PLS models is lower than for the other two models (Fig. S10 top and middle right). The important wavenumbers are: 6684 (assigned to hemicelluloses band) and from 5315 to 4136 cm^{-1} (assigned to hemicellulose, cellulose, and water content band) (Schwanninger et al., 2011). In contrast, almost all variables are essential for the model to predict the glucose content (Fig. S10, bottom right).

With the capability to determine the hemicellulose and pectin content and composition non-destructively, future samples can be

Table 3
Statistics for the established PLS models. Bias was less than 1 % for all obtained models both for Cal and CV.

Models	Total hemicellulose & pectin	Rhamnose	Xylose	Glucose
Latent variables	6	5	4	7
R^2Cal	0.71	0.64	0.63	0.63
R^2CV	0.61	0.59	0.57	0.47
RMSEC [mg/g]	2.8	0.26	0.47	7.4
RMSECV [mg/g]	3.3	0.29	0.50	8.9

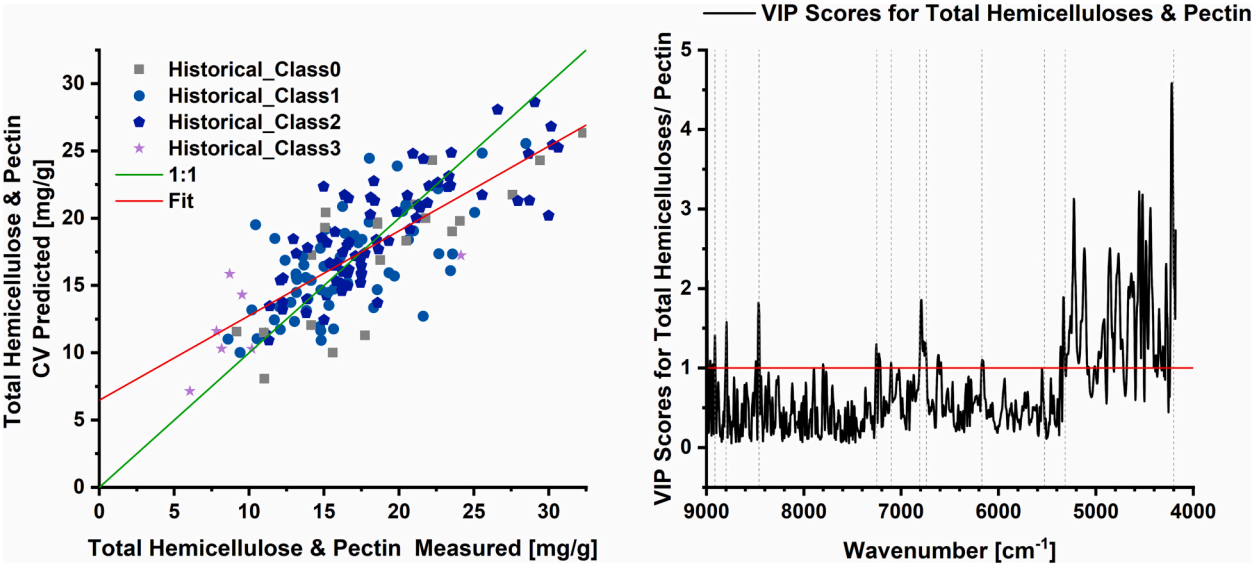

Fig. 8. Left: Performance of the PLS model (NIR, Savitzky-Golay 1st derivative, SNV, and mean centering) for the total hemicellulose and pectin content. Ideally, the PLS model would give the same results as the reference method, and the values would be positioned on the green line. The red line is a linear regression of the cross-validation prediction results. Right: Relevant wavenumbers (higher than 1) according to VIP scores.

characterized with little effort. With time, the underlying meaning of the observed sample classes should become apparent, which can be expected to offer insights into the condition and history of individual rag samples.

4. Conclusion

The polysaccharide composition of historical and modern rag papers was studied by several analytical methods to obtain an overview over typical values of rag paper with regard to hemicellulose and pectin content, the portion of individual monosaccharides, crystallinity, and molecular weight distribution. The polysaccharide composition of historical and modern rag papers can indeed clearly differ. Modern rag papers contained rhamnose, galactose and galacturonic acid only when they were sized, while these carbohydrates were absent in unsized modern samples. Furthermore, modern rag papers showed high contents of xylose compared to historical samples, which hints at a degradation of xylose over time. The hemicellulose and pectin content in rag papers is typically much lower than in industrial pulps (6–31 mg/g for rag paper, compared to 8–200 mg/g for pulp). Rhamnose was typically found in rag papers and never present in pulps; also, the concentrations of galactose and arabinose were typically higher in rag papers.

In an unsupervised density-based clustering based on their xylose, rhamnose, and glucose content, three distinct classes were indicated within the rag papers. A straightforward explanation for this grouping could be based on the permanence of the samples—the more stable or unstressed the samples, the higher the content of xylose and the lower the content of glucose according to methanolysis. Also, one class consisted of samples that were contaminated with mold; this class showed an unusually low content of carbohydrates according to methanolysis. It can be speculated that the polysaccharides were consumed by the mold. Not all modern rag samples were placed into one of the classes, and sized modern rag papers agree better with historical rag papers.

NMR-determined crystallinity was higher in rag papers than in pulp, and a negative correlation with xylose was found. Crystallinity and methanolysis-degradable cellulose did not show any correlation, which could also be a side-effect of the spectral modelling required for the determination of crystallinity. The average molecular weight (M_w) of rags was 211 ± 87 (kg/mol). Modern rags usually had higher average molecular weight.

In the NIR spectra of the samples, the major difference between the spectra of pulp and rags were connected to the intensities of water bands. A multivariate calibration was established from this sample set to determine the content of hemicellulose and pectin monosaccharides non-destructively by their NIR spectra. This allows determining the content of non-cellulosic polysaccharides, and thus the classification of rag samples, within a few minutes without sacrificing sample material.

CRediT authorship contribution statement

A.P., T.R. and S.B. conceptualized and designed the study. H.K., and J.L. performed the experimental work. P.J., M.B., S.S., and M.K. provided expertise. H.K. and S.B. processed the data. A.P. coordinated and supervised the work. H.K. and S.B. wrote and visualized the manuscript, which was reviewed and edited by all the authors. All the authors approved the final version of the manuscript.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Data will be made available on request.

Acknowledgments

The authors acknowledge the financial support from the Austrian Biorefinery Centre Tulln (ABCT) and the Country of Lower Austria. Hajar Khaliliyan expresses her gratitude to Austropapier and the companies Austrocel, Delfort, Lenzing AG, Mondi, and SAPPI for funding. The help of Ute Henniges, Stuttgart State Academy of Art and Design, with the rag samples and helpful discussion is gratefully acknowledged. H. Khaliliyan and P. Jusner thank the BOKU doctoral school ABCM for financing. DI Dr. Elfriede Hogger is acknowledged for the introduction to Olympus Microscope. Open access funding is provided by the University of Natural Resources and Life Sciences, Vienna (BOKU).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.carbpol.2023.121611.

References

Ahn, K., Schedl, A., Zweckmair, T., Rosenau, T., & Potthast, A. (2018). Fire-induced structural changes and long-term stability of burned historical rag papers. *Scientific Reports, 8*(1), Article 1. https://doi.org/10.1038/s41598-018-30424-7

Ahn, K., Zaccaron, S., Zwirchmayr, N. S., Hettegger, H., Hofinger, A., Bacher, M., … Rosenau, T. (2019). Yellowing and brightness reversion of celluloses: CO or COOH, who is the culprit? *Cellulose, 26*(1), 429–444. https://doi.org/10.1007/s10570-018-2200-x

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization (4:63). *Wellcome Open Research.* https://doi.org/10.12688/wellcomeopenres.15191.1

Andersen, C. M., & Bro, R. (2010). Variable selection in regression—A tutorial. *Journal of Chemometrics, 24*(11–12), 728–737. https://doi.org/10.1002/cem.1360

Ansell, M. P., & Mwaikambo, L. Y. (2009). The structure of cotton and other plant fibres. In *Handbook of textile fibre structure* (pp. 62–94). Elsevier. https://doi.org/10.1533/9781845697310.1.62.

Area, M. C., & Cheradame, H. (2011). Paper aging and degradation: Recent findings and research methods. *BioResources, 6*(4), Article 4.

Barrett, T., Ormsby, M., & Lang, J. B. (2016). Non-destructive analysis of 14th–19th century European handmade papers. *Restaurator. International Journal for the Preservation of Library and Archival Material, 37*(2), 93–135. https://doi.org/10.1515/res-2015-0017

Becker, M., Ahn, K., Bacher, M., Xu, C., Sundberg, A., Willför, S., Rosenau, T., & Potthast, A. (2021). Comparative hydrolysis analysis of cellulose samples and aspects of its application in conservation science. *Cellulose, 28*(13), 8719–8734. https://doi.org/10.1007/s10570-021-04048-6

Becker, M., Zweckmair, T., Forneck, A., Rosenau, T., Potthast, A., & Liebner, F. (2013). Evaluation of different derivatisation approaches for gas chromatographic–mass spectrometric analysis of carbohydrates in complex matrices of biological and synthetic origin. *Journal of Chromatography A, 1281*, 115–126. https://doi.org/10.1016/j.chroma.2013.01.053

Bloom, J. M. (2017). Papermaking: The historical diffusion of an ancient technique. In H. Jöns, P. Meusburger, & M. Heffernan (Eds.), *Mobilities of knowledge* (pp. 51–66). Springer International Publishing. https://doi.org/10.1007/978-3-319-44654-7_3.

Bro, R. (2003). Multivariate calibration: What is in chemometrics for the analytical chemist? *Analytica Chimica Acta, 500*(1), 185–194. https://doi.org/10.1016/S0003-2670(03)00681-0

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods, 6*(9), 2812–2831. https://doi.org/10.1039/C3AY41907J

Budischowsky, D., Sulaeva, I., Hettegger, H., Ludwig, R., Rosenau, T., & Potthast, A. (2022). Fluorescence labeling of C1-oxidized cellulose. Part 1: Method development. *Carbohydrate Polymers, 295*, Article 119860. https://doi.org/10.1016/j.carbpol.2022.119860

Čabalová, I., Kačík, F., Gojný, J., Češek, B., Milichovský, M., Mikala, O., Tribulová, T., & Ďurkovič, J. (2017). Changes in the chemical and physical properties of paper documents due to natural ageing. *BioResources, 12*(2), Article 2.

Carpita, N. C., & Gibeaut, D. M. (1993). Structural models of primary cell walls in flowering plants: Consistency of molecular structure with the physical properties of the walls during growth. *The Plant Journal, 3*(1), 1–30. https://doi.org/10.1111/j.1365-313X.1993.tb00007.x

Coletti, F., Romani, M., Ceres, G., Zammit, U., & Guidi, M. C. (2021). Evaluation of microscopy techniques and ATR-FTIR spectroscopy on textile fibers from the Vesuvian area: A pilot study on degradation processes that prevent the characterization of bast fibers. *Journal of Archaeological Science: Reports, 36*, Article 102794. https://doi.org/10.1016/j.jasrep.2021.102794

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*(5), 1784–1790. https://doi.org/10.1016/j.talanta.2010.07.062

Dupont, A.-L. (2002). Study of the degradation of gelatin in paper upon aging using aqueous size-exclusion chromatography. *Journal of Chromatography A, 950*(1), 113–124. https://doi.org/10.1016/S0021-9673(02)00010-9

Erhardt, D., & Mecklenburg, M. (1994). Relative humidity re-examined. *Studies in Conservation, 39*(sup2), 32–38. https://doi.org/10.1179/sic.1994.39.Supplement-2.32

Erhardt, D., & Tumosa, C. S. (2005). Chemical degradation of cellulose in paper over 500 years. *26*(3), 151–158. https://doi.org/10.1515/rest.2005.26.3.151

Espejo, T., Duran, A., Lopez-Montes, A., & Blanc, R. (2010). Microscopic and spectroscopic techniques for the study of paper supports and textile used in the binding of hispano-arabic manuscripts from Al-Andalus: A transition model in the 15th century. *Journal of Cultural Heritage, 11*(1), 50–58. https://doi.org/10.1016/j.culher.2009.01.007

Ester, M., Kriegel, H.-P., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *American Association for Artificial Intelligence*, 226–231.

Geminiani, L., Campione, F. P., Corti, C., Luraschi, M., Motella, S., Recchia, S., & Rampazzi, L. (2022). Differentiating between natural and modified cellulosic fibres using ATR-FTIR spectroscopy. *Heritage, 5*(4), Article 4. https://doi.org/10.3390/heritage5040213

Haas, K. T., Wightman, R., Meyerowitz, E. M., & Peaucelle, A. (2020). Pectin homogalacturonan nanofilament expansion drives morphogenesis in plant epidermal cells. *Science, 367*(6481), 1003–1007. https://doi.org/10.1126/science.aaz5103

Henniges, U., Kostic, M., Borgards, A., Rosenau, T., & Potthast, A. (2011). Dissolution behavior of different celluloses. *Biomacromolecules, 12*(4), 871–879. https://doi.org/10.1021/bm101555q

Henniges, U., Schwanninger, M., & Potthast, A. (2009). Non-destructive determination of cellulose functional groups and molecular weight in pulp hand sheets and historic papers by NIR-PLS-R. *Carbohydrate Polymers, 76*(3), 374–380. https://doi.org/10.1016/j.carbpol.2008.10.028

Jusner, P., Bacher, M., Simon, J., Bausch, F., Khaliliyan, H., Schiehser, S., Sumerskii, I., Schwaiger, E., Potthast, A., & Rosenau, T. (2022). Analyzing the effects of thermal stress on insulator papers by solid-state 13C NMR spectroscopy. *Cellulose, 29*(2), 1081–1095. https://doi.org/10.1007/s10570-021-04338-z

Khaliliyan, H., Schuster, C., Sumerskii, I., Guggenberger, M., Oberlerchner, J. T., Rosenau, T., … Böhmdorfer, S. (2020). Direct quantification of lignin in liquors by high performance thin layer chromatography-densitometry and multivariate calibration. *ACS Sustainable Chemistry & Engineering*. https://doi.org/10.1021/acssuschemeng.0c03950

Kirui, A., Du, J., Zhao, W., Barnes, W., Kang, X., Anderson, C. T., … Wang, T. (2021). A pectin methyltransferase modulates polysaccharide dynamics and interactions in Arabidopsis primary cell walls: Evidence from solid-state NMR. *Carbohydrate Polymers, 270*, Article 118370. https://doi.org/10.1016/j.carbpol.2021.118370

Kirui, A., Zhao, W., Deligey, F., Yang, H., Kang, X., Mentink-Vigier, F., & Wang, T. (2022). Carbohydrate-aromatic interface and molecular architecture of lignocellulose. *Nature Communications, 13*(1), Article 1. https://doi.org/10.1038/s41467-022-28165-3

Korntner, P., Hosoya, T., Dietz, T., Eibinger, K., Reiter, H., Spitzbart, M., … Rosenau, T. (2015). Chromophores in lignin-free cellulosic materials belong to three compound classes. Chromophores in cellulosics, XII. *Cellulose, 22*(2), 1053–1062. https://doi.org/10.1007/s10570-015-0566-6

Ling, Z., Wang, T., Makarem, M., Santiago Cintrón, M., Cheng, H. N., Kang, X., … French, A. D. (2019). Effects of ball milling on the structure of cotton cellulose. *Cellulose, 26*(1), 305–328. https://doi.org/10.1007/s10570-018-02230-x

Liu, Y., Fearn, T., & Strlič, M. (2021). Quantitative NIR spectroscopy for determination of degree of polymerisation of historical paper. *Chemometrics and Intelligent Laboratory Systems, 214*, Article 104337. https://doi.org/10.1016/j.chemolab.2021.104337

Łojewski, T., Miśkowiec, P., Missori, M., Lubańska, A., Proniewicz, L. M., & Łojewska, J. (2010). FTIR and UV/vis as methods for evaluation of oxidative degradation of model paper: DFT approach for carbonyl vibrations. *Carbohydrate Polymers, 82*(2), 370–375. https://doi.org/10.1016/j.carbpol.2010.04.087

McCrady, E. (1992, November). The great cotton-rag myth. https://cool.culturalheritage.org/byorg/abbey/ap/ap05/ap05-5/ap05-503.html (Text.Article).

McDougall, G. J. (1993). Isolation and partial characterisation of the non-cellulosic polysaccharides of flax fibre. *Carbohydrate Research, 241*, 227–236. https://doi.org/10.1016/0008-6215(93)80109-R

Moriana, R., Vilaplana, F., Karlsson, S., & Ribes, A. (2014). Correlation of chemical, structural and thermal properties of natural fibres for their sustainable exploitation. *Carbohydrate Polymers, 112*, 422–431. https://doi.org/10.1016/j.carbpol.2014.06.009

Pere, J., Pääkkönen, E., Ji, Y., & Retulainen, E. A. (2019). Influence of the hemicellulose content on the fiber properties, strength, and formability of handsheets. *BioResources, 14*(1), Article 1.

Peter, Z. (2021). Order in cellulosics: Historical review of crystal structure research on cellulose. *Carbohydrate Polymers, 254*, Article 117417. https://doi.org/10.1016/j.carbpol.2020.117417

Potthast, A. (2015). Comparison testing of methods for gel permeation chromatography of cellulose: Coming closer to a standard protocol. *Cellulose, 22*. https://doi.org/10.1007/s10570-015-0586-2

Potthast, A., Rosenau, T., Kosma, P., Saariaho, A.-M., & Vuorinen, T. (2005). On the nature of carbonyl groups in cellulosic pulps. *Cellulose, 12*(1), 43–50. https://doi.org/10.1023/B:CELL.0000049347.01147.3d

Rinnan, Å., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry, 28*(10), 1201–1222. https://doi.org/10.1016/j.trac.2009.07.007

Röhrling, J., Potthast, A., Rosenau, T., Lange, T., Ebner, G., Sixta, H., & Kosma, P. (2002). A novel method for the determination of carbonyl groups in cellulosics by fluorescence labelling. 1. Method development. *Biomacromolecules, 3*(5), 959–968. https://doi.org/10.1021/bm020029q (Scopus).

Rubin, E. M. (2008). Genomics of cellulosic biofuels. *Nature, 454*(7206), 841–845. https://doi.org/10.1038/nature07190

Sandak, A., Sandak, J., & Riggio, M. (2016). Assessment of wood structural members degradation by means of infrared spectroscopy: An overview: Assessment of Wood Degradation by Means of Spectroscopy. *Structural Control and Health Monitoring, 23*(3), 396–408. https://doi.org/10.1002/stc.1777

Scheller, H. V., & Ulvskov, P. (2010). Hemicelluloses. *Annual Review of Plant Biology, 61*(1), 263–289. https://doi.org/10.1146/annurev-arplant-042809-112315

Schwanninger, M., Rodrigues, J. C., & Fackler, K. (2011). A review of band assignments in near infrared spectra of wood and wood components. *Journal of Near Infrared Spectroscopy, 19*(5), 287–308. https://doi.org/10.1255/jnirs.955

Simmons, T. J., Mortimer, J. C., Bernardinelli, O. D., Pöppler, A.-C., Brown, S. P., deAzevedo, E. R., … Dupree, P. (2016). Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. *Nature Communications, 7*(1), Article 1. https://doi.org/10.1038/ncomms13902

Strlič, M., Liu, Y., Lichtblau, D. A., De Bruin, G., Knight, B., Winther, T., … Brereton, R. G. (2020). Development and mining of a database of historic European paper properties. *Cellulose, 27*(14), 8287–8299. https://doi.org/10.1007/s10570-020-03344-x

Sundberg, A., Sundberg, K., Lillandt, C., & Holmbom, B. (1996). Determination of hemicelluloses and pectins in wood and pulp fibres by acid methanolysis and gas chromatography. *Nordic Pulp & Paper Research Journal, 11*(4), 216–219. https://doi.org/10.3183/npprj-1996-11-04-p216-219

Toscano, G., Rinnan, Å., Pizzi, A., & Mancini, M. (2017). The use of near-infrared (NIR) spectroscopy and principal component analysis (PCA) to discriminate bark and wood of the most common species of the pellet sector. *Energy & Fuels, 31*(3), 2814–2821. https://doi.org/10.1021/acs.energyfuels.6b02421

Voragen, A. C. J., Rolin, C., Marr, B. U., Challen, I., Riad, A., Lebbar, R., & Knutsen, S. H. (2003). Polysaccharides. In *Ullmann's Encyclopedia of Industrial Chemistry*. John Wiley & Sons, Ltd.. https://doi.org/10.1002/14356007.a21_a25.pub2

Wickholm, K., Larsson, P. T., & Iversen, T. (1998). Assignment of non-crystalline forms in cellulose I by CP/MAS 13C NMR spectroscopy. *Carbohydrate Research, 312*(3), 123–129. https://doi.org/10.1016/S0008-6215(98)00236-5

Willats, W. G. T., McCartney, L., Mackie, W., & Knox, J. P. (2001). Pectin: Cell biology and prospects for functional analysis. *Plant Molecular Biology, 47*(1), 9–27. https://doi.org/10.1023/A:1010662911148

Willför, S., Sundberg, A., Hemming, J., & Holmbom, B. (2005). Polysaccharides in some industrially important softwood species. *Wood Science and Technology, 39*(4), 245–257. https://doi.org/10.1007/s00226-004-0280-2

Willför, S., Sundberg, A., Pranovich, A., & Holmbom, B. (2005). Polysaccharides in some industrially important hardwood species. *Wood Science and Technology, 39*(8), 601–617. https://doi.org/10.1007/s00226-005-0039-4

Zhang, Y., Yu, J., Wang, X., Durachko, D. M., Zhang, S., & Cosgrove, D. J. (2021). Molecular insights into the complex mechanics of plant epidermal cell walls. *Science, 372*(6543), 706–711. https://doi.org/10.1126/science.abf2824

Zuckerstätter, G., Terinte, N., Sixta, H., & Schuster, K. C. (2013). Novel insight into cellulose supramolecular structure through 13C CP-MAS NMR spectroscopy and paramagnetic relaxation enhancement. *Carbohydrate Polymers, 93*(1), 122–128. https://doi.org/10.1016/j.carbpol.2012.05.019