

Original Work

Timothy Barrett, Mark Ormsby* and Joseph B. Lang

Non-Destructive Analysis of 14th–19th Century European Handmade Papers

DOI 10.1515/res-2015-0017

Received September 22, 2015; revised March 28, 2016; accepted April 5, 2016

Abstract: Non-destructive analysis of 1,578 paper specimens made between the fourteenth and the nineteenth centuries was undertaken in an effort to better understand changes in paper composition over time and how these variations might affect paper stability during long term natural ageing. Gelatine content and colour were determined using UV/Vis/NIR spectrometry. Residual metals were measured using XRF. These components included potassium and sulfur as elements indicative of alum concentration; iron as a typical paper contaminant; and calcium, which is often associated with compounds such as calcium carbonate that can serve as alkaline reserves. The research demonstrated that papers that are lighter in colour (closer to white) are associated with higher levels of gelatine and calcium, and lower levels of iron. The survey also showed significant decreases in gelatine and calcium concentration over time with the largest differences coinciding with the rise of printing around 1500. The drop in pH over the centuries observed by other researchers is explained by the decreases in these two components rather than by a rise in alum concentration, which remained fairly stable. The craftsmanship of the specimens was evaluated using materials and workmanship ratings which showed better quality paper associated with higher gelatine and calcium concentrations and colour closer to white. Poorer quality papers were associated with higher iron levels and greater thickness.

Keywords: handmade paper, gelatine, alum, chemometrics, XRF

***Corresponding author: Mark Ormsby**, Preservation Programs, National Archives and Records Administration, 8601 Adelphi Road, Rm. 1800, College Park, MD 20740, USA,
E-mail: mark.ormsby@nara.gov

Timothy Barrett, Center for the Book, 216 NH, University of Iowa, Iowa City, IA 52242, USA,
E-mail: timothy-barrett@uiowa.edu

Joseph B. Lang, Department of Statistics and Actuarial Science, 207 SH, University of Iowa, Iowa City, IA 52242, USA, E-mail: joseph-lang@uiowa.edu

1 Project overview

William Barrow's 1974 study of 1,470 book papers made between 1507 and 1949 was pioneering in its analysis of historical specimens to reveal new information about paper permanence and durability (W.J. Barrow Research Laboratory 1974). The Barrow Lab's plots showing a distinct decline in pH over the centuries, and increasing incidence of alum addition and decreasing incidence of calcium carbonate addition over the same time period, were intriguing because of the new light they shed on the causes of paper degradation. What is truly revolutionary about the Barrow work was not the new data per se, but the convincing demonstration of a new research methodology. Barrow showed that, when papermakers' archives are non-existent, the historical paper specimens themselves can reveal important information about how they were made and the impact of various processes and ingredients on paper stability. The challenge for the researcher is to find ways to decode the messages hidden within the paper. The Barrow study is, therefore, the inspiration for the present research.

We had several goals in the present study that we hoped would amplify the Barrow work. We planned to include papers from the 1400s known to be exceptionally stable. We hoped to test for gelatine concentration, a material generally considered to be a common ingredient in papers made between the fifteenth and the eighteenth century. And because the Barrow data documented the presence of calcium carbonate and alum using qualitative spot tests, we hoped to test for the same compounds using quantitative measurements.

In the time since Barrow's work, new non-destructive, quantitative techniques for analysing paper properties have been developed. Portable x-ray fluorescence (XRF) instruments have been applied in numerous conservation applications (Shugar and Mass 2012). Near infrared spectroscopy (NIR) combined with chemometrics has been used to study gelatine and other paper properties (Lichtblau et al. 2008; Henniges et al. 2009; Cséfalvayová et al. 2010).

In addition, several studies have explored the role of gelatine in paper preservation. Barrett found indications of gelatine in higher concentration in historical papers in good condition (Barrett 1989; Barrett and Mosier 1995). Other researchers found that gelatine-sized papers aged more slowly than non-sized papers based on changes in degree of polymerization, particularly for papers containing alum, but that for one type of modern gelatine the sized papers fared worse in terms of pH and yellowing (Dupont 2003; Missouri et al. 2006). Using historical specimens, Stephens et al. concluded that high gelatine content specimens were in good condition (based on degree of polymerization and yellowness index) while low gelatine specimens ranged from poor to good condition

(Stephens et al. 2008). Kolbe reported that gelatine sizing slowed iron gall ink induced ageing, although Potthast et al. found no evidence of a protective effect (Kolbe 2004; Potthast et al. 2008). Gelatine appeared to reduce copper catalysed corrosion of a model paper, although the mechanism was unclear (Ahn et al. 2015). These previous studies were based on relatively small numbers of historical specimens as well as accelerated ageing of modern samples.

This project augments Barrow’s research by analysing 1,578 historical papers made between the fourteenth and the nineteenth centuries. We were particularly interested in studying changes in papermaking practices over time and evaluating how these variations in materials and techniques might influence paper stability. Because we employed exclusively non-destructive methods, we were able to study more earlier papers than those Barrow examined, including a large number of fifteenth-century papers and five specimens from the fourteenth century. We also gathered quantitative measurements of gelatine sizing and the concentrations of a variety of metals that may influence paper stability. The data allow us to look at both trends over time as well as the impact of individual variables on paper stability. The number of specimens tested by Barrow and the corresponding numbers analysed during our research are shown in Figure 1, sorted by century.

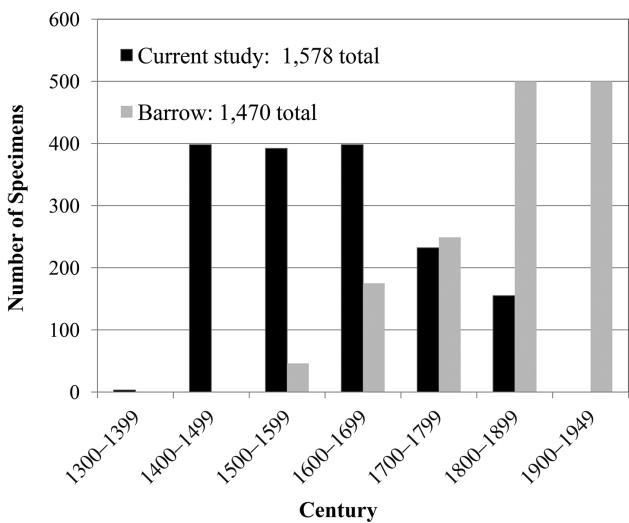


Figure 1: Barrow was particularly interested in the deterioration of modern machine-made papers, and his collection emphasized specimens from the 19th and early 20th centuries. This project complements his work by analyzing older, handmade specimens. While Barrow studied only book papers, this work includes approximately 360 printed leaves and 100 manuscript books and leaves.

Analysis of this wealth of quantitative data must consider the fact that these specimens were handmade over several centuries by artisans with different skills, resources, economic pressures, and motivations. Throughout the history of European hand papermaking, mill owners constantly struggled with the tension between the quality and quantity of the paper they produced. Good quality paper had to be made more slowly and carefully, but it could be sold for a higher price. Poorer quality paper could be made more quickly with lower raw materials costs and less skilled (and we assume cheaper) labour, but it would not command as high a price. As the demand for paper increased between the fifteenth and the eighteenth centuries, papermakers were naturally driven to improve daily production rates. Depending on the size, weight and quality of the paper type being made, a skilled three-person team could produce between fifteen hundred and three thousand sheets a day (Houghton 1699). These seem like staggering numbers to modern hand papermakers, but at the time they were routine. Reynard gives an enlightening description of how eighteenth century French papermakers dealt with increased production, and the resulting lessening of quality, by inventively adding to the number of paper grades they sorted the finished sheets into and provided to their customers (Reynard 2000).

Judging the quality of materials and workmanship in paper was (and remains) a qualitative exercise undertaken by specialists familiar with the papermaking process. While our study employed various instruments to collect quantitative data, numerical grades for materials and workmanship (M&W) had to be assigned by individuals familiar with the skills necessary to make paper of varying qualities. Principal Investigator Timothy Barrett and research assistant Jessica White were both trained in hand papermaking and were responsible for assigning these M&W grades to each specimen tested.

Poor quality was evidenced by stray foreign fibres, straw, bits of debris, lumps, clumps, and signs of quick or unskilled sheet forming or couching. At the other extreme were papers that appeared uniform in high-quality rag fibre content; had a minimum of stray fibres or debris, and in transmitted light showed exceptional formation quality, evidence of careful couching, and freedom from knots or clumps (Figure 2). We took care not to assign M&W grades based on colour. That is, a browned specimen otherwise showing signs of excellent materials selection and worker skills would receive a high score (4 or 5), while a very light-coloured sheet with characteristics like those shown at the left in Figure 2 would still receive a low score of 1 or 2.

Those who routinely handle historical papers tend to associate browned paper with brittleness or lack of durability and light-coloured sheets with stability or strength. Analysis of data on thirty-eight historical specimens in a range of condition from very poor to excellent supports this perception (Stephens et al. 2008). The

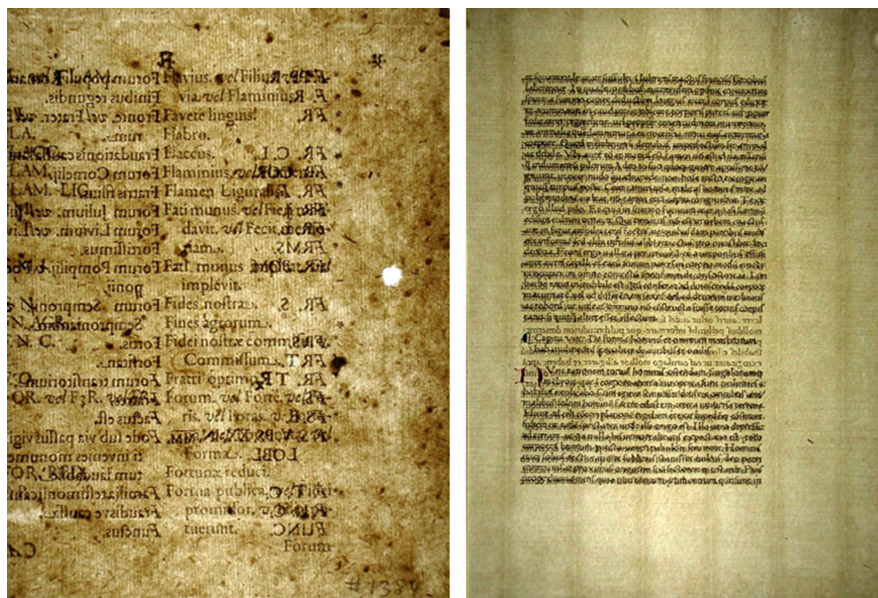


Figure 2: Photographed using transmitted light, these two specimens illustrate the range of materials and workmanship ratings. The poor quality paper at left was assigned grade 1 due to evidence of stray foreign fibers, straw, bits of debris, lumps, clumps, and signs of quick or unskilled sheet forming or couching. The grade 5 paper at right is uniform in high-quality rag fibre content, has a minimum of stray fibres or debris, and illustrates exceptional formation quality and careful couching.

data indicate that darker, redder, or yellower colours can be associated with lower pH, degree of polymerization, and zero-span (fibre strength) values. While these results show a relationship between colour and chemical and physical measurements that are, in turn, associated with permanence and durability, the same trends may not be apparent in experiments with modern papers. For example, using accelerated ageing of gelatine and alum sized papers, Dupont showed that colour was not a reliable indicator of pH or molecular weight. She also observed that pH was not a reliable indicator of molecular weight (Dupont 2003, 163).

While one would generally expect superior papermaking craftsmanship to produce more durable papers, there are many examples of old, poorly made papers that are nevertheless in stable condition today. By combining the M&W ratings with data from the instrumental analysis we hoped to obtain a more complete picture of the choices papermakers made, how these practices changed over time, and the combined effect of ingredients and craftsmanship on the permanence and durability of paper.

2 Specimen selection, instrumentation and methods, and data collection

2.1 Specimen selection

Our goal was to select specimens according to the criteria outlined below in order to illustrate changes in papermaking materials and techniques over time.

1. Specimens are evenly distributed by
 - a. country (of original paper mill),
 - b. date (of writing or printing),
 - c. quality (of materials and workmanship),
 - d. book size (i. e., not all folios).
2. Specimens are in original bindings or other evidence should confirm that there is no history of “aqueous intervention” such as washing or resizing.
3. Specimens have been stored, on average, in similar conditions.

Approximately 4,000 specimens were considered during the initial stages of this study from collections at the Newberry Library and The University of Iowa Libraries. Roughly 2,440 specimens were rejected and 1,578 specimens selected for analysis. These included printed books and leaves as well as manuscript books and leaves. As we acknowledge below, this collection of specimens was not evenly distributed according to the criteria outlined above. Considerably more staff and time would have been necessary to assemble such a specimen set. Even if the effort had been expended, we still would have faced the reality of significantly different paper and book production rates in various countries over time. This can be seen in the Figure 3 plots that show, by date, the percentage of gelatine by weight for the samples from each country. Each open circle on the plots represents a different specimen analysed. (More on gelatine concentration follows under Results below.) In Figure 3 it is evident that we tested few papers from fifteenth-century England but a relatively high number from fifteenth-century Italy and Germany, primarily because at that time a great deal of book production took place in Italy and Germany and very little in England. For mid-seventeenth century and later dates we had access to fewer Italian and German books and more British books.

While attaining the ideal distribution of specimens was not possible, we attempted to address potential biases reflected in the specimen pool by using multiple data analysis methods in addition to simple chronological plots. We included the M&W rating in part to attempt to evaluate a potential bias toward

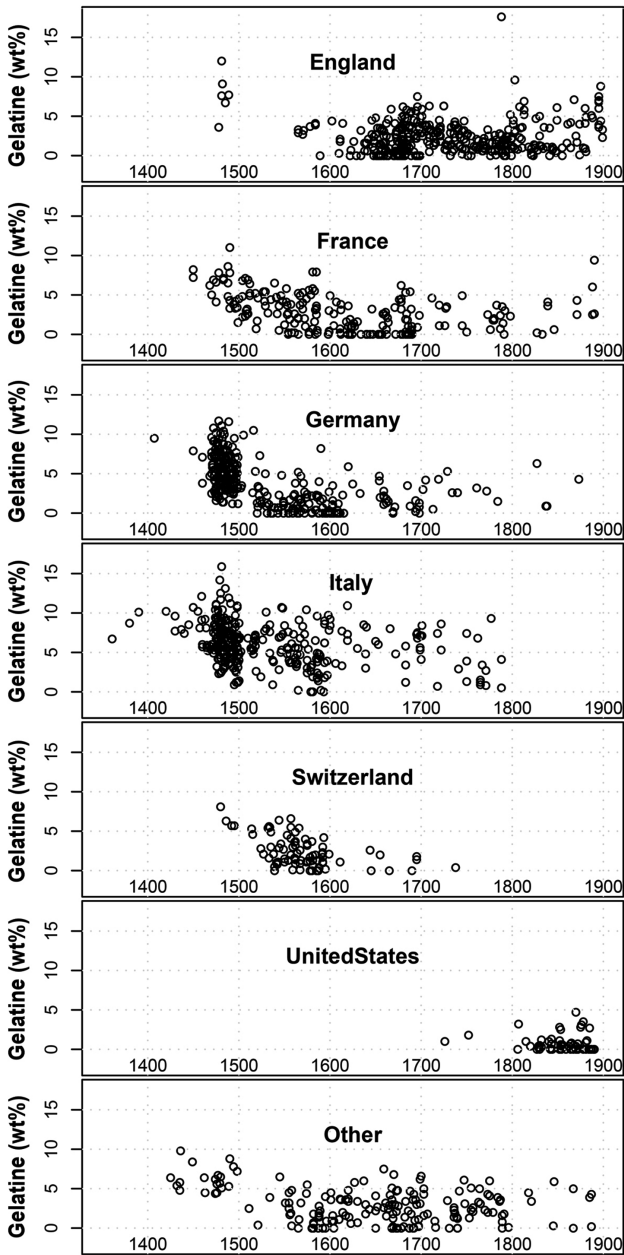


Figure 3: Gelatine concentration of specimens according to year plotted by country of origin. Note that the date range extends to 1900.

higher quality papers. For example, the group of oldest specimens may be weighted toward expensive, highly-valued papers that were more likely to be preserved in a library, archive, or museum rather than inferior papers that were less well cared for and deteriorated long ago. The chosen specimens represented a fairly normal distribution of paper quality, from worst to best (Figure 4), with the highest count falling in the middle at M&W grade 3. The specimens are grouped into 50-year periods in Figure 5, again showing a fairly normal distribution over time.

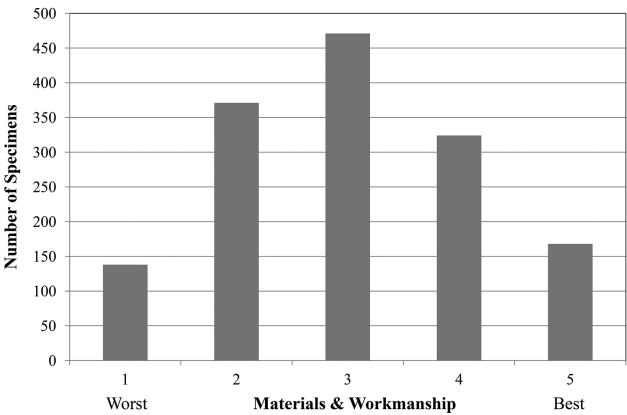


Figure 4: This histogram displays the number of specimens for each materials and workmanship grade according to the criteria described in Figure 2.

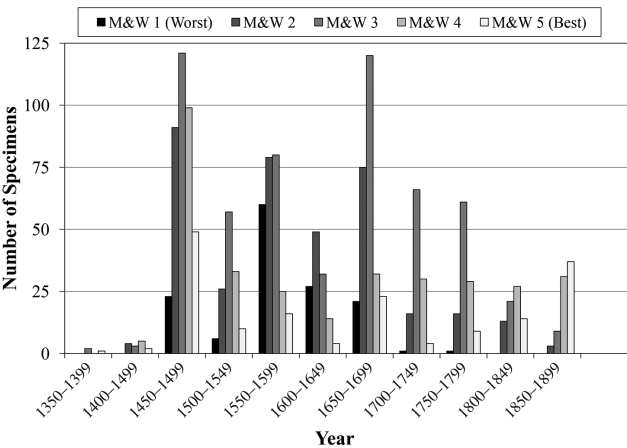


Figure 5: The histogram is similar to Figure 4 but displays the distribution of materials and workmanship ratings within 50-year periods.

The M&W rating reflects the initial quality of the specimen, but we have no knowledge of the environment the paper was exposed to over centuries of storage. A high quality paper may be severely discoloured and weak if stored in hot, humid conditions. Thus, in selecting specimens we cannot meet the third criteria listed above, similar storage conditions. However, in 229 of the books tested we found two or more papers within the same book that were clearly in different condition, usually with one noticeably darker than the other (Figure 6). Often, but not always, the mould surface indicated a different maker or mill and/or significant differences in formation quality, knots, clumps, etc. Sometimes both sheets appeared to be from the same moulds, but their condition was clearly different. For this special subset we have two pieces of paper that display

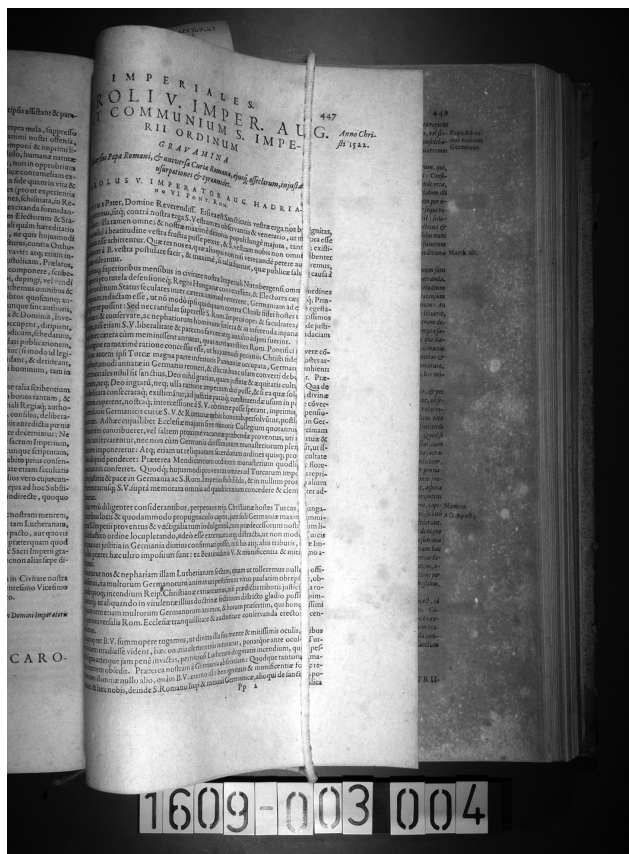


Figure 6: Typical example of two specimens in different condition found within the same book.

very different apparent condition and were stored in the same environment. This fact allows us to look for differences in paper properties between the two sheets that might explain the contrasting condition. We note that Baker, in her analysis of nineteenth-century hand and machine papermaking in America, explains this phenomenon by citing discussions in papermaking manuals of “Tuesday paper” and “Saturday paper.” The Tuesday paper was sized with a fresh batch of gelatine that was then used every day thereafter. The Saturday paper was sized in the same solution, which by the end of the week was dirty and/or perhaps loaded up with alum to make the size last through the week without spoiling. It is possible that something similar to this routine was in use throughout the history of the craft (Baker 2010).

Returning to the selection criteria listed above, the overall composition of our specimen pool was determined, therefore, by several factors: (1) the nature of the collections from which it was drawn, (2) the history of book production in Europe, and (3) our attention to selecting a mix of good- and poor-quality papers. We acknowledge that equal distribution of specimens across centuries and by country was not possible. It is important to keep in mind the location bias we see in Figure 3 because it can play a role in the results, especially any results plotted over time. Even if the specimen pool was perfectly balanced, a collection of 1,578 papers can only present an incomplete picture of European papermaking of the period. Nevertheless, we feel our plots across time can be of use in understanding changes in papermaking materials and technique and their implications for paper stability if those apparent trends are consistent across different data analysis methods.

2.2 Instrumentation and methods

2.2.1 XRF spectrometer

Measurements of calcium (Ca), potassium (K), sulfur (S), and iron (Fe) concentration were gathered using a Bruker Tracer III-V portable XRF spectrometer. These elements have an anticipated positive (Ca) or negative (K, S and Fe) association to paper stability. The instrument was calibrated using inductively coupled plasma optical emission (ICP-OES) measurements from a set of 40 historical specimens used in earlier research (Stephens et al. 2008). These calibration standards exhibited a range of materials and workmanship from poor to good and purposely included examples with clumps, uneven formation (thick and thin areas), etc. in an attempt to anticipate most of the paper qualities we would encounter during analyses in libraries.

To compensate for variations in paper thickness/density or more gross imperfections at the exact location of an XRF analysis, we incorporated a thin film of fiberglass resin impregnated with chromium (Cr) and bromine (Br) compounds. Using a special apparatus, the paper specimen being tested was held gently against the nose of the instrument and backed by the Cr/Br thin film. The Cr 5.4 KeV emission line is significantly attenuated by paper, but the Br 12 KeV emission line has high transmission. By evaluating the relative strength of these two lines we determined correction factors for the thickness/density of the unknowns. The XRF measurements (mg/cm^2) were converted to ppm using the ICP-OES calibration curves.

Measurements of aluminium were also made to evaluate alum content, but quantitative results were not possible because of the instrument's limited sensitivity to lighter elements. As discussed below, K and S concentrations were used to study alum content. Details on the XRF instrument, the accessory for positioning the unknown and the Cr/Br thin film, and our calibration methodology have been published (Barrett et al. 2012).

2.2.2 UV/Vis/NIR spectrometer

An Analytical Spectral Devices (ASD) QualitySpec Pro ultraviolet-visible-near-infrared (UV/Vis/NIR) spectrometer was used to gather data on gelatine concentration and to evaluate the colour of each specimen. A chemometric model for gelatine was developed using the NIR range. The model was calibrated using gas chromatography/mass spectroscopy (GC-MS) analysis of amino acid concentration in the 40 historical specimens used with the XRF calibration (Stephens et al. 2008). CIELAB colour values were derived from the visible range data. Calibration curves were determined from measurements of the 40 specimens using an X-Rite 968 spectrophotometer. See Appendices I and II for details.

2.2.3 Handheld micrometer

A Mitutoyo No. 2046F handheld micrometer with dial increments of 0.01 mm was used for all specimen thickness determinations. We recorded the thickness in millimeters of the specimen and, in books, the thickness of ten leaves (the leaf analyzed and the following nine leaves). These latter data were gathered because of the natural variation in paper thickness in a single book.

2.2.4 Data collection

Measurements were conducted in special collections or conservation facilities under ambient conditions at The University of Iowa and the Newberry Library. After basic information on each specimen was logged in using a Filemaker Pro template, a representative leaf was selected for testing. XRF, UV/Vis/NIR, and thickness data were then collected. Using a Nouvir “Transilluminator” fibre optic light sheet, each specimen was viewed by transmitted light and assigned a grade for materials and workmanship. Each specimen was then photographed under the same reflected light source. Specimen log in procedures and data are detailed at <http://paper.lib.uiowa.edu>.

We collected data at five locations on each specimen (Figure 7). One 120-second XRF scan was done at the centre of the fore edge of the leaf as close to the edge as possible (dotted arrow). This was followed by four 30-second XRF scans in the

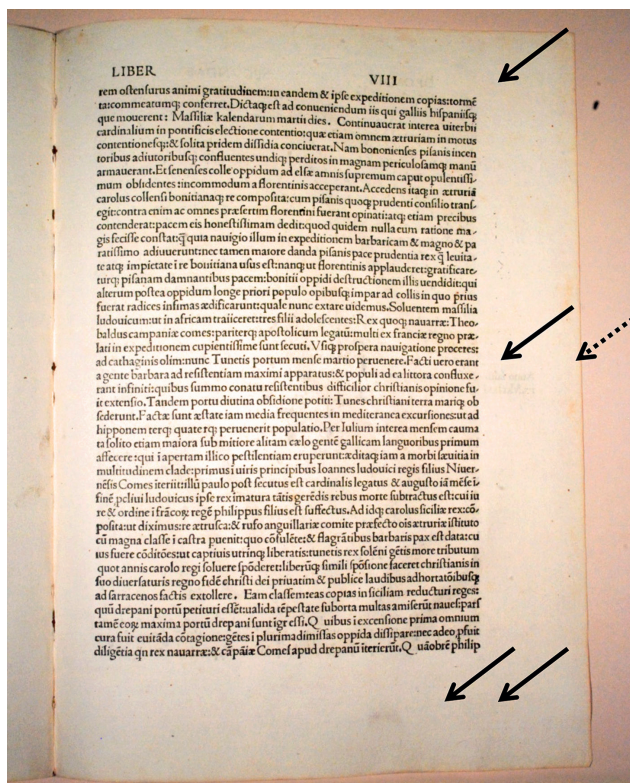


Figure 7: Typical locations (indicated by arrows) of XRF and UV-VIS-NIR analyses. The dotted arrow indicates the edge or exterior sampling position.

margins; one midway between the fore edge centre and the printed area, one midway between the upper right-hand corner and the printed area, one likewise at the lower right-hand corner, and finally one in the lower margin, midway between the edge and the printed area as far from the lower right corner as the XRF accessory would permit (about 5 cm). The four “interior” margin analyses were later combined into a single “interior” value for a total of 120 sec of sampling time, which could then be compared to the single 120-second analysis at the very edge of the leaf. This amalgamation was done to investigate whether airborne components such as sulfur accumulate at the edge of the leaf relative to the interior of the sheet. All plots shown below use only interior data. Interior versus edge comparisons are discussed at <http://paper.lib.uiowa.edu>. For the UV/Vis/NIR measurements, spectra were collected at two sites approximately 1 cm apart at each of the five positions described above for a total of 10 analyses on each specimen. At each location the instrument was set to gather fifty spectra which were averaged to give a single spectrum with reduced noise. XRF and UV/Vis/NIR analysis of a specimen typically required 15 min.

Details about each specimen, including photographs, date and place of origin, leaf type, analysis results, and other information are available at the website <http://paper.lib.uiowa.edu>. The site also contains supplementary technical details about instrumental and statistical analysis methods.

3 Results and discussion

3.1 Chronological plots

Figure 8 shows the weight percent gelatine concentration for the 1,578 specimens over time. The black dots represent the observed data reported by the instrument and NIR model, and the light grey circles indicate the range of error in the analysis method due to instrumental and modelling error. The inner line shows the estimated mean as a function of year, and the two outer curves indicate the reliability of these mean estimates. In other words, with 95% confidence the actual means are between the two outer curves. Thus, to get a more reliable sense of the trend for the weight percent gelatine concentration over the centuries, in Figure 8 one should consider not only the inner line but also the swath described by the space between the two outer curves. Details of the statistical analysis, including calculation of the R and p coefficients, are included in Appendix III.

The table top display across the top of the graph depicts the mean levels over the centuries. The thickness of the table top gives the 95% confidence interval for the mean. Non-overlapping table tops imply that the difference

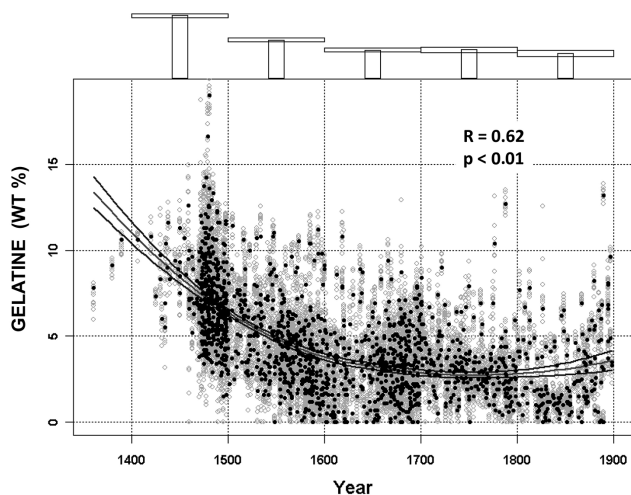


Figure 8: Weight percent gelatine as measured by the NIR model from the 14th to 19th century. The black dots are the average of interior gelatine measurements from a specimen. Light gray circles are the range of error in the analysis method due to instrumental and modeling error. The inner line shows the estimated mean as a function of year, and the two outer curved lines indicate the reliability of the mean estimate. The tabletop plot across the top shows the mean level across the centuries (at reduced scale to preserve space). No mean is shown for the 14th century because of insufficient data. The thickness of the tabletop indicates the 95% confidence interval for the mean, and non-overlapping tabletops imply a statistically significant difference. The decline in mean gelatine concentration between the 15th, 16th, and 17th centuries was statistically significant.

between the means is statistically significant. The table top units coincide with the vertical axis units but their scale has been compressed for display purposes.

It is important to note that a statistically significant difference does not necessarily imply a significant practical difference. In other words, within certain mathematical parameters, we can have confidence an observed difference is not a chance occurrence; however, whether that difference has real-world implications is open to discussion. To that end, we first discuss chronological plots for evidence of changes in paper properties over time. We then address possible impacts of these changes.

Figures 8–12 show chronological plots of gelatine, calcium concentration, paper thickness (for single sheets and 10 sheets), and L^* . In each graph the variable decreases across the centuries with a noticeable change apparent around 1500. This difference is visible in the tabletop plots, which indicate whether the change in the means over the centuries is statistically significant. On the L^* scale from white to black in Figure 12 the means start out light, become darker, and then lighter again as discussed further below.

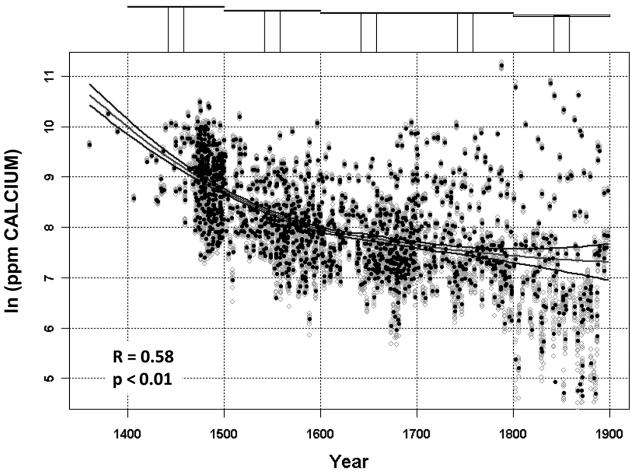


Figure 9: The calcium concentration measured by XRF is plotted according to year. To more conveniently display the wide range of values the natural logarithm of the calcium concentration (measured in ppm) is shown. The fitted curves and tabletop plots are as described in Figure 8. Based on the tabletop plots there was a statistically significant decrease in the mean calcium concentration over the first three centuries.

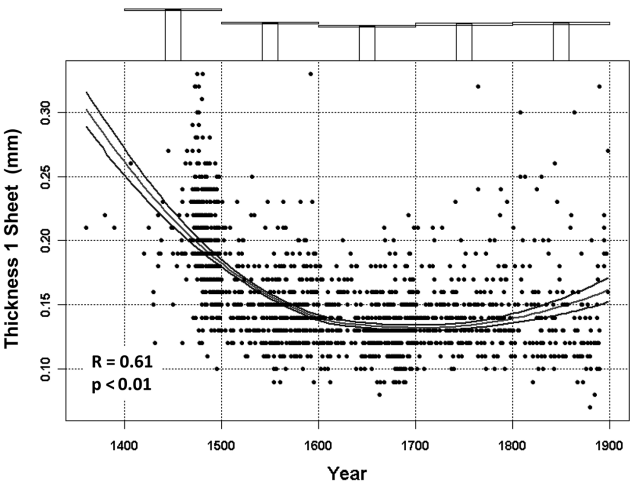


Figure 10: The thickness of specimens is shown according to year with fitted curves and tabletop plots as described in Figure 8. The discrete levels in the vertical axis result from the minimum increments on the micrometer. The decrease in thickness was statistically significant over the first three periods in the tabletop plot.

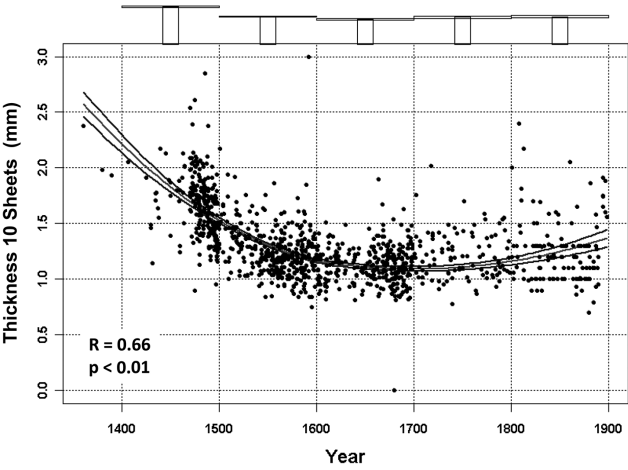


Figure 11: The discrete micrometer increments are less apparent in this plot of the thickness of the specimen together with the following 9 sheets in the book. The fitted curves and tabletop plots are as described in Figure 8. The tabletop confidence intervals indicate a statistically significant decrease in thickness over the first three periods.

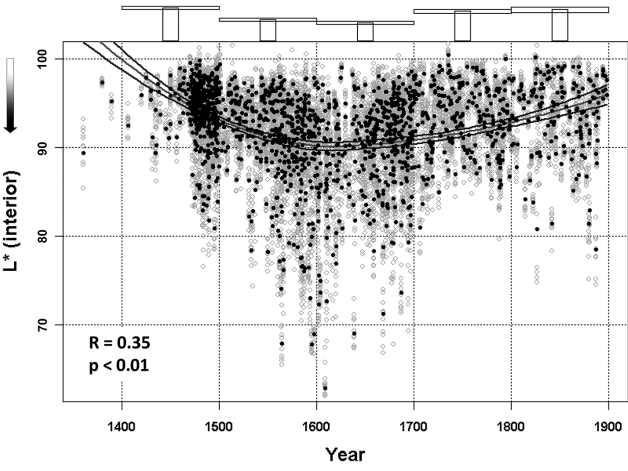


Figure 12: L^* (lightness of specimens relative to white reference) of specimens according to year with fitted curves and tabletop plots as described in Figure 8. Over the first three centuries there was a statistically significant decrease in mean L^* , but in the 18th century the mean increased.

Figure 13 shows the average gelatine concentration by M&W rating in four 100-year periods. For each quality rating the amount of gelatine used dropped substantially over time, and the greatest change was between the fifteenth and sixteenth centuries. The error bars give approximate 95 % confidence intervals

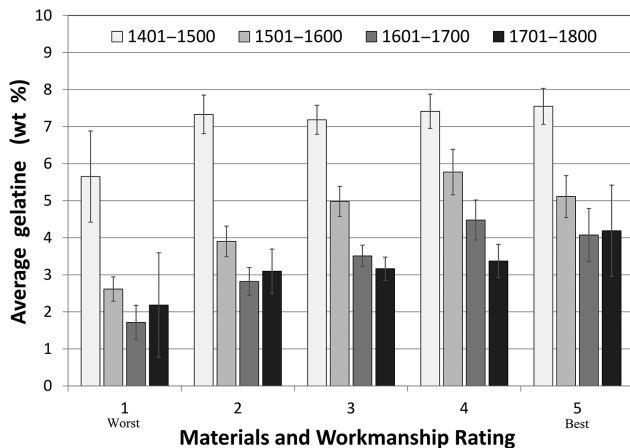


Figure 13: The average gelatine concentration over four centuries is plotted for each materials and workmanship rating. Because of the many changes in papermaking processes in the 19th century only data from before 1800 are shown. At each grade there was a large drop between the first two centuries shown followed by a smaller but still statistically significant decrease between the second and third periods.

for the means. When the overlap between two intervals is less than about 25% the difference can be deemed statistically significant. Thus, the marked change around 1500 is consistent, regardless of the quality of the materials and workmanship. This consistency gives us confidence that the pronounced change in the chronological plot for gelatine content (Figure 8) is real and not simply due to a bias toward higher quality papers in the oldest specimens.

Most of the specimens from the fifteenth and sixteenth centuries were made in Italy, France, and Germany (Figure 3). In Figure 14 the average gelatine concentrations for specimens from these three countries are plotted in 50-year periods. For the French and German specimens there were statistically significant decreases in gelatine concentration through the first four periods. The Italian papers did not have a statistically significant decrease until the third period. There were too few Italian specimens to draw statistically valid comparisons in the last two periods. Overall, however, these results support the general conclusion that papermakers used less gelatine over the course of these centuries and that the trend apparent in Figure 8 is not due to a location bias in the specimen selection.

In Figures 8–12 we see similar trends in four variables measured with three separate instruments, and all four variables have a statistically significant difference between the fifteenth and sixteenth centuries. If we assume that these trends do indeed represent changes in the craft that took place across Europe, what historical events would explain them? We would offer that early

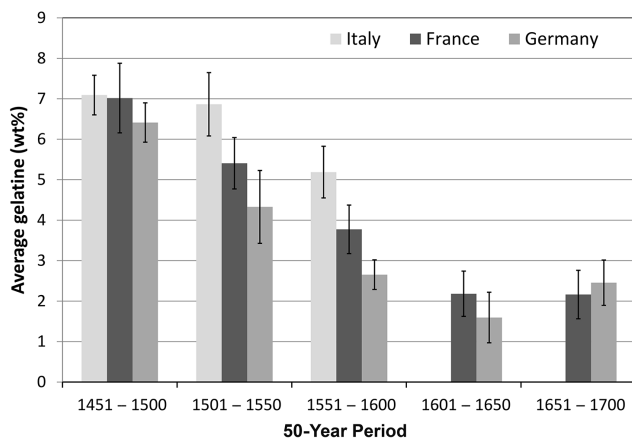


Figure 14: Average gelatine concentrations over 50-year periods are plotted for specimens from Italy, France, and Germany. There were statistically significant decreases in gelatine concentration through the first four periods for the French and German specimens. The Italian papers maintained higher gelatine levels, and there was not a statistically significant decrease until the third period.

printers, beginning with Gutenberg in 1455, were by their type designs and with hand-rubricated letters, endeavouring to print and sell imitation hand-copied manuscript books. Printers likely made this effort because manuscript books - the better of which were on parchment - were the only models at the time for how a book should appear, feel, and function in the hands. Up until the turn of the sixteenth century, and even beyond, bookbinders also continued to design books that evoked monastic wooden-board structures. In this atmosphere, we believe it is plausible that papermakers of the period were attempting to make not paper per se, but essentially a form of imitation parchment.

Paper of that era, made from old, well-worn linen and hempen rags, was rather weak, soft, and absorbent after drying. To improve the strength of the paper somewhat it could be made thicker. But the application of gelatine size, followed by burnishing with a polished stone, transformed it into a very believable substitute for parchment: tough, abrasion resistant, smooth, and able to accept ink without bleeding. The colour of the finished paper would almost certainly have been lightened by calcium compounds which entered the paper from a number of sources, whether intentionally added or not. Calcium was a key ingredient in making parchment, and parchment clippings were a source of high quality gelatine for sizing paper. According to fifteenth- and seventeenth-century accounts, lime was used during beating, probably to help facilitate maceration of the rags by swelling the fibre (Dabrowski and Simmons 2003, Fahy 2003). Additional evidence that papermakers were attempting to emulate parchment is found in sheet paper

dimensions, and in particular, the ratio between the short and the long edge of the sheet - a ratio that matches that of parchment from the period. See <http://paper.lib.uiowa.edu> for more details on paper and parchment dimensions.

The rapid spread of printing and increasing demand for books could indeed have changed all this around 1500. The quickest way for papermakers to lower the price per sheet was simply to make their paper thinner. Cutting back on other ingredients, such as calcium compounds and gelatine, would have also helped lower expenses and therefore the price of the finished paper. Figure 13 shows that the average gelatine concentration for the *poorest* quality paper from before 1500 was comparable or greater than the average gelatine content for the *highest* quality paper in all other periods. These results suggest papermakers in the fifteenth century tended to incorporate higher concentrations of gelatine in their papers, regardless of grade, than did papermakers in subsequent centuries.

With regard to the decline in gelatine addition, we need to remember that early printers were effectively printing on *writing paper* - the only paper available at the time for books and a material designed to properly receive water-based inks and paints. The printers, who were using oil-based inks, had trouble with it. The ink sat on top of the paper and squeezed out under the typefaces, leaving a less-than-sharp imprint. Dampening the paper prior to printing became standard operating procedure (Entlesberger et al. 2011) and helped a great deal, but it was an added step. And the more size in the paper, the longer the step took. Szirmai suggests that by the late fifteenth century papermakers were in fact supplying completely *unsized* paper to the printers, and it was the binders who applied the size later (Szirmai 1999). Papermakers would have welcomed the request to supply paper without the need for the laborious and troublesome sizing step. Bookbinders may have taken on this added step because the paper was otherwise too weak to withstand binding and end use by readers. Being able to write in the margins of a book without ink bleeding may have also been an expected end use for many readers (Blair 2010, Sherman 2008).

While this view constitutes a plausible historical scenario that can explain the trends we see in the chronological plots, in reality we are probably observing a combination of actual significant changes in papermaking around 1500 as well as the influence of the biases in specimen selection, collection, and care habits over the centuries, etc. that we have discussed above.

3.2 Alum and calcium carbonate

Alum of the period was potassium aluminium sulfate. Papermaker's alum, aluminium sulfate, did not become common until the nineteenth century (Brückle 1993). The two compounds were almost certainly used, perhaps for

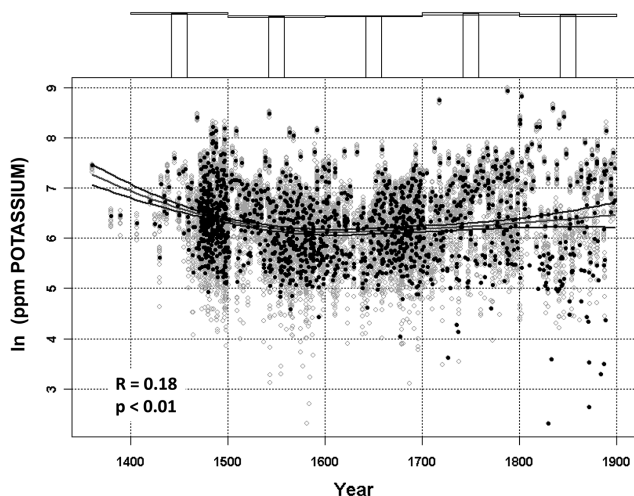


Figure 15: The potassium concentration according to year is shown using natural log of the potassium to more conveniently display the range of values. The fitted curves and tabletop plots are as in Figure 8.

different applications, after 1800. We were unable to obtain a useful XRF calibration for aluminium, primarily because of reduced instrument sensitivity to the lighter elements. Therefore, we studied alum concentration based on measurements of potassium and sulfur as shown in Figures 15 and 16. Note that the R values of 0.18 and 0.17 are much lower than in Figures 8–12 and that the table top plots are nearly flat. In general, these results indicate that there was little change in the amount of potassium and sulfur in the papers over these periods. These almost flat trends contradict the Barrow lab plot (Figure 17) that appears to show dramatically increasing alum use over the centuries (W.J. Barrow Research Laboratory 1974).

We note, however, that Barrow’s plot does not display the *amounts* of alum and calcium carbonate in paper but rather the *percentage of papers containing those compounds*. The Barrow workers used a spot indicator and recorded a positive or negative reaction for the presence of alum and calcium carbonate on each book tested. The plots in our work display actual concentrations.

When considering the related Barrow plot for pH (Figure 18), many have mentally combined it with the previous plot (Figure 17) and interpreted them to indicate that more alum was being added over the centuries because the pH does go down, and it goes down significantly. Based on our data we believe these declines in pH were not a result of increasing alum concentrations but rather a result of decreasing concentrations of gelatine and Ca compounds (Figures 8 and 9).

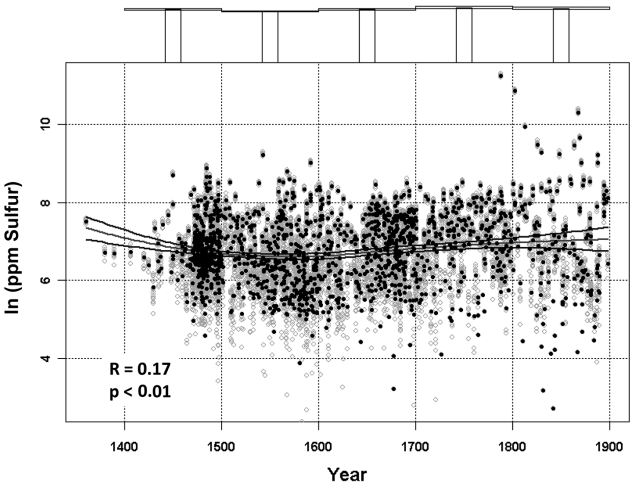


Figure 16: The sulfur concentration according to year is shown using the natural log of the sulfur measurement. Note that the vertical axis covers a narrower range than in Figures 9 and 15. The fitted curves and tabletop plots are as in Figure 8.

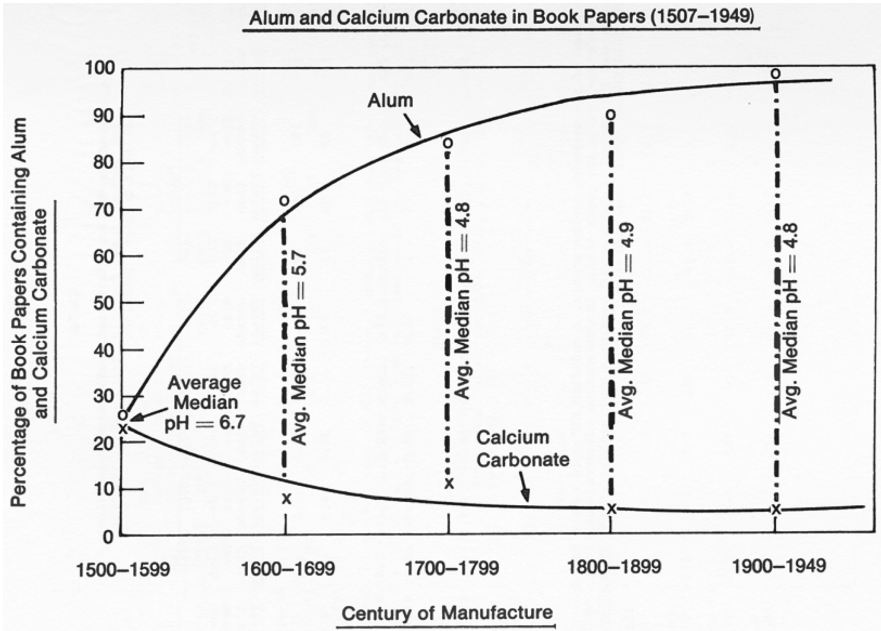


Figure 17: Data from Barrow’s work showing change according to year in percentage of book papers containing alum and calcium carbonate. Note that the date range is different from Figures 8–13. Chart reproduced from W. J. Barrow Research Laboratory 1974 publication cited in references.

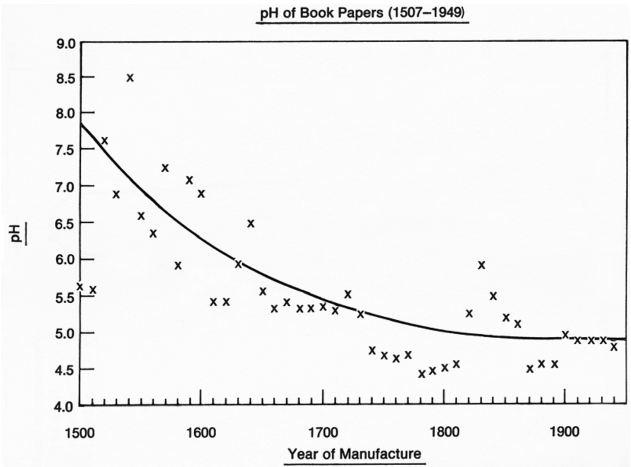


Figure 18: Barrow’s data showing the change in pH according to year. Again, the date range is different from Figures 8–13. Reproduced from W. J. Barrow Research Laboratory 1974 publication.

Our view is based on previous work demonstrating that gelatine can act as a pH buffer (Baty and Barrett 2007). Also, calcium carbonate is a common alkaline reserve added to modern papers designed for long-term applications to counter acidic compounds that may be in the paper or may enter it in the future. Figure 19 shows that higher gelatine concentrations were generally associated with higher Ca concentrations, thus contributing a likely combined pH buffering effect.

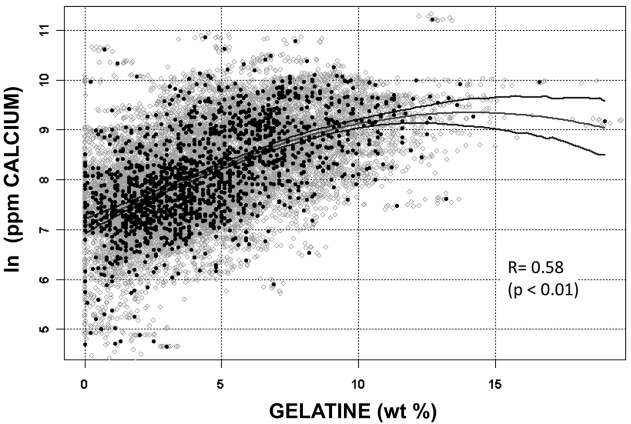


Figure 19: Specimen calcium concentration is plotted versus gelatine. Note that the natural log of the ppm calcium concentration is shown to compress the wide range of values. The fitted curve was calculated as described in Figure 8.

3.3 Non-chronological plots

As mentioned, the oldest and rarest specimens in the collection may be weighted toward more white and less red or yellow papers because they were handled and stored more carefully than later works on paper. Therefore, we pursued a number of non-chronological analyses of the data. Figures 20 and 21 use “ornament” or “violin” plots to illustrate potassium and sulfur trends with M&W. The top and bottom of each ornament represent the highest and lowest values. The grey swelled areas in the middle show the distribution of data between the 25th and the 75th percentiles, and the bars with dots in the middle show the medians. The ornament plots show that there was not a statistically significant difference in the means of potassium and sulfur over the range of M&W ratings. That is, the amount of alum added in making paper did not appear to be associated with intended quality of the finished sheet.

On the other hand, iron, calcium, and gelatine content appear to be more closely associated with differences in paper quality based on Figures 22–25. The lower M&W scored (poorer quality) papers tended to be darker in colour (Figure 22). This result was in spite of our efforts to assign M&W grades without concern for specimen colour. Sheets made with apparent attention to materials

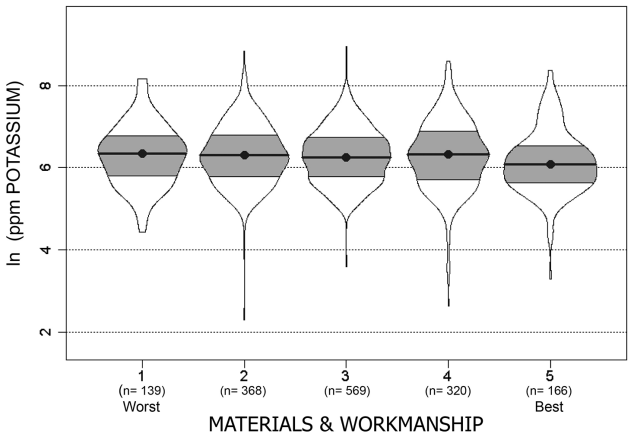


Figure 20: These ornament plots show the potassium concentration (using the natural log to compress the range) versus the materials and workmanship rating. These ornaments can be thought of as vertically oriented, smoothed histograms, with a mirror image to create the ornament shape. The top and bottom represent the high and low values, respectively. The gray swelled areas show the distribution between the 25th and 75th percentiles, and the median is indicated by the dots on the bars. There was not a statistically significant difference among the means over the 5 ratings.

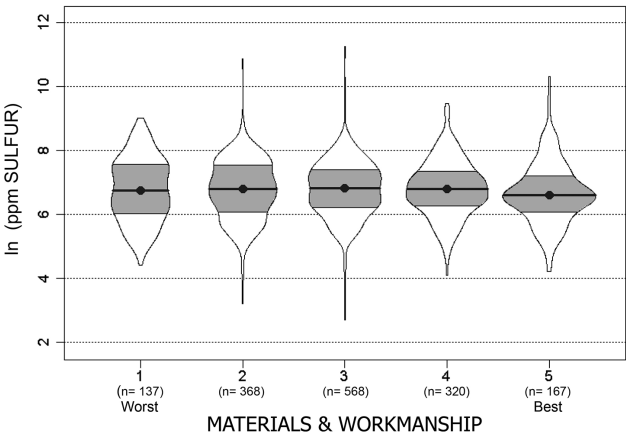


Figure 21: The natural logarithm of the sulfur concentration is plotted versus materials and workmanship using the ornament shapes as described in Figure 20. There was not a statistically significant difference among the 5 means.

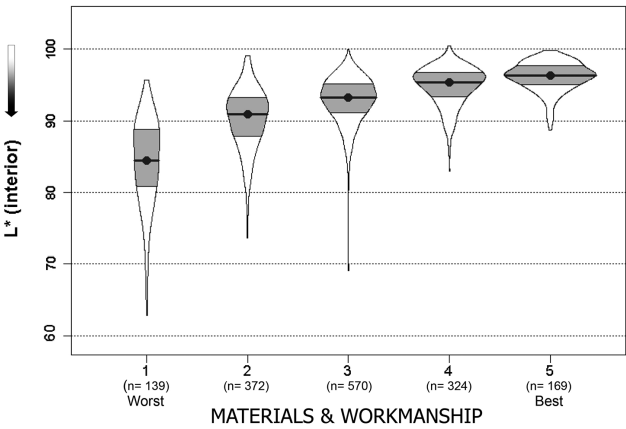


Figure 22: The L* values for the specimens are plotted versus materials and workmanship using the ornament shapes as described in Figure 20. The long, narrow shape of rating 1 indicates a broad range of L* values compared to the bulging centre of rating 5. The means at all ratings were statistically different.

selection and preparation, and with skilled workers at sheet forming and couching stations, were lighter in colour.

This colour trend in Figure 22 may be due to the higher concentrations of iron in the poorer quality sheets illustrated in Figure 23. There was a statistically significant difference in the means for all five M&W levels. The data suggest that

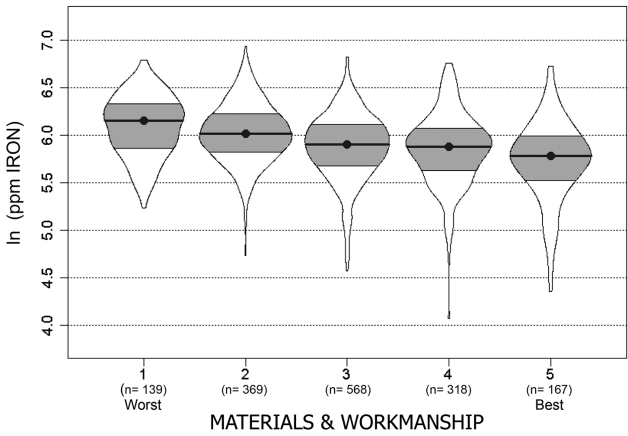


Figure 23: Using the ornament plots as in Figure 20, this graph displays the natural logarithm of the iron concentration versus materials and workmanship rating. There was a statistically significant difference in the means at all 5 ratings.

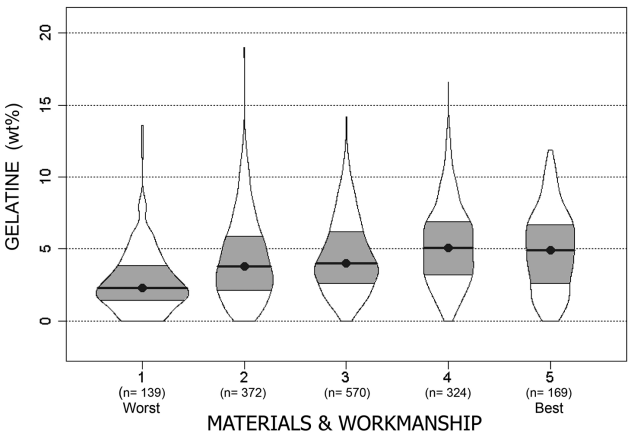


Figure 24: The lower mean gelatine concentration for materials and workmanship rating 1 was statistically significant compared to the other four means. The ornament shapes are as described in Figure 20.

in mills making cheaper, lower-quality paper, there was less attention paid to water quality. Rusty, iron-fitted equipment may have been common. In mills where high quality white paper was made, on the other hand, high quality, debris- and iron-free water was essential. Likewise, exposure to any source of

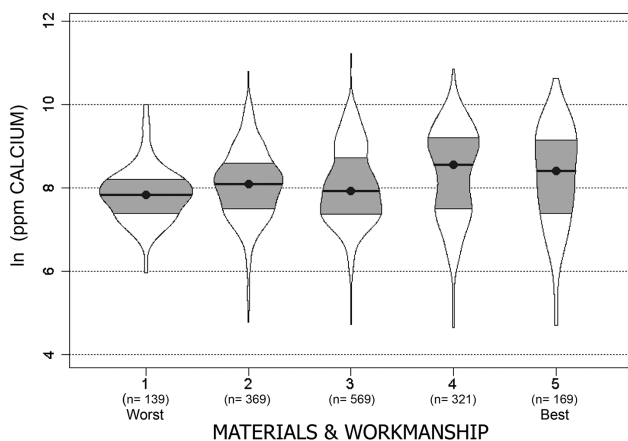


Figure 25: The natural logarithm of the calcium concentration is plotted versus materials and workmanship using the ornament shapes as described in Figure 20. The lower mean for rating 1 was statistically significant compared to the means for ratings 4 and 5.

rusting iron in equipment would have been avoided whenever possible. We are not aware of any pre-nineteenth century tests for iron in water, but red stains on rocks in water sources, vats, stampers, buckets or other equipment would have been a reliable indicator of high iron concentration in the water supply.

The data in Figure 24 suggest that the time and expense associated with higher gelatine concentration was appropriate when making the highest quality sheets compared with the lowest. This trend was not as strong as with iron because only the M&W 1 mean was statistically different from the other M&W ratings. Likewise, higher calcium concentration (Figure 25) may be associated with the higher M&W grades compared to the lowest because the difference between the M&W 1 mean and the M&W 4 and 5 means is statistically significant.

Curiously, in Figure 26 we see that poorer quality papers tended to be thicker than specimens with higher M&W scores. This result is surprising because thicker paper generally requires more pulp and, therefore, costs the papermaker more money per sheet. One explanation is that the extended beating time needed for better formation quality tends to produce a higher density, more compact sheet, while shorter beating times are cheaper and produce a bulkier sheet. Such extended beating times, if used to produce shorter fibre and better formation quality could well be associated with better quality papers because of the added expense associated with prolonged beating. A related issue is that extended beating times result in pulp that drains more slowly.

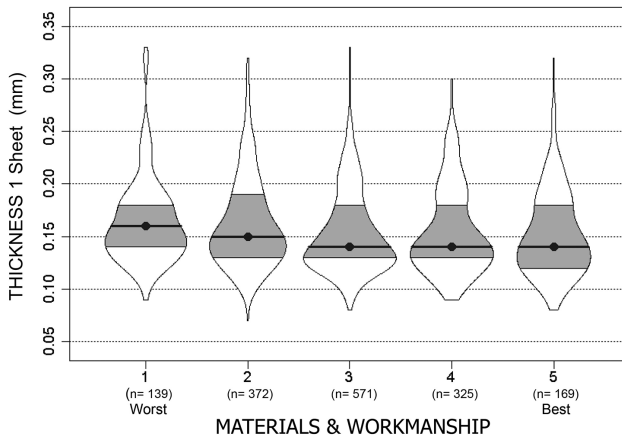


Figure 26: The specimen thickness is plotted versus materials and workmanship rating using the ornament shapes as described in Figure 20. The decrease in mean thickness as the rating increased was statistically significant at all five levels.

Thus, there could have been an inclination to make high quality paper thinner so it would drain more quickly at the vat and help keep up the total number of sheets possible per hour or day. In summary, there are reasonable explanations for overall thinner paper in the higher M&W graded sheets.

Returning to Figures 8, 9, and 12, it appears that as the average amount of gelatine and calcium decreased, the papers tended to be darker, but then the papers become lighter again toward the eighteenth century. The introduction of chlorine bleach around 1800 is an important development that likely began to artificially lighten the overall colour of papers made, at least until the introduction of the first papers made from wood pulp around 1840–1850. Not chemically purified, these early wood pulp sheets tended to discolour quickly. To explore this behaviour more carefully, the plots in Figure 27 use only the data from books printed before 1800, and they examine colour changes over time as a function of M&W, gelatine, calcium, and iron. The three columns show average values for L^* , a^* (red to green scale), and b^* (yellow to blue scale), respectively.

Beginning at the top left in Figure 27, the graph shows the average L^* value over four centuries for printed books. The shade of the marker indicates the M&W rating (light for the highest grade 5; dark for the poorest grade 1), and the size of the marker indicates the average gelatine concentration. (M&W 1 for the eighteenth century is not shown because there was only one specimen in that category.) Moving vertically upward in this top-left plot, in each century the

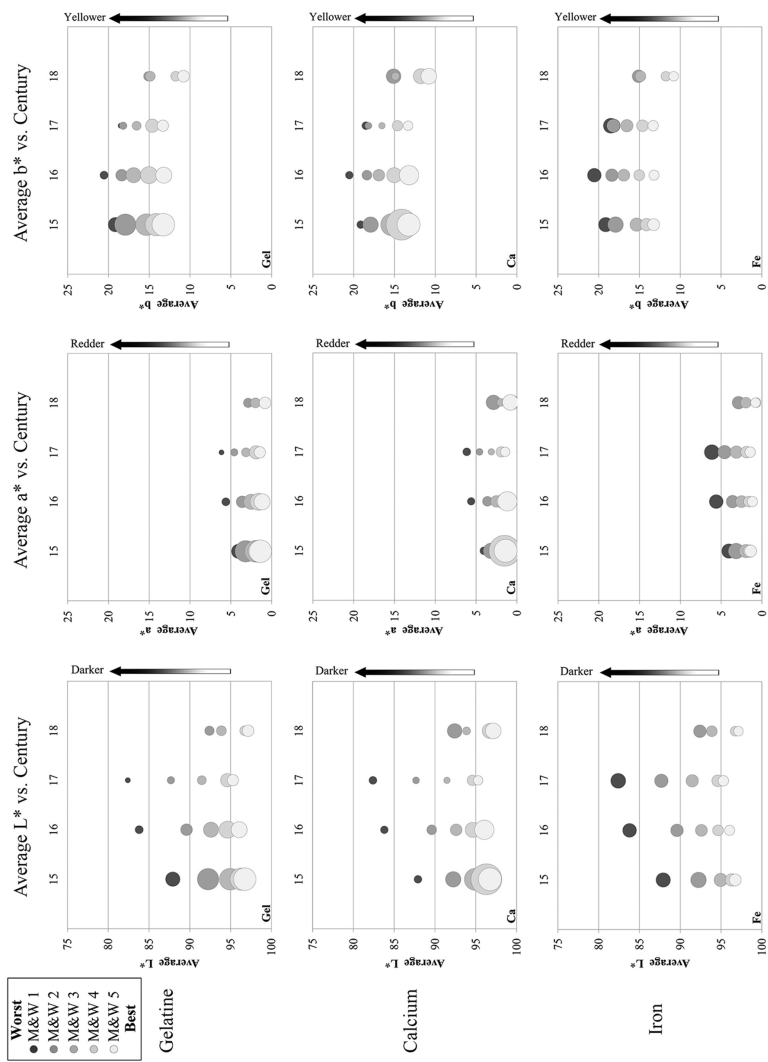


Figure 27: The three columns show the change in L^* , a^* , and b^* values for printed books over four centuries. In the first column the same L^* data is plotted in all three graphs, but the markers indicate different properties. In the top figure the marker size is proportional to the mean gelatine concentration, and in the middle and bottom rows the marker size is proportional to the mean calcium and iron concentrations, respectively. The shade of the marker indicates the materials and workmanship rating with darkest signifying grade 1 (worst) and lightest representing grade 5 (best). The a^* and b^* data are plotted similarly in the second and third columns.

average gelatine concentration generally decreases with M&W, i. e., the markers are generally smaller as the M&W drops from 5 to 1. Moving horizontally, the average gelatine concentration generally decreases over the centuries (the markers are smaller), and as it drops the specimens generally become darker (moving upward on the white-to-black). The same data are plotted again in the middle graph in the left column, but the marker size is proportional to the calcium concentration. The overall trends are similar to gelatine. In the eighteenth century the gelatine and calcium levels increase somewhat (the markers are larger), and the specimens are correspondingly closer to white.

The results for iron are plotted in the bottom row in Figure 27. Compared to gelatine and calcium, there was less variation in the Fe concentrations, i. e., the markers were similar in size and trends were less visually apparent.

The middle column in Figure 27 shows the trends for a^* with the redness of the specimen increasing vertically. There was less spread in the a^* values than with L^* , and the differences are most apparent with M&W 1. When comparing data from the same century, the specimens with more Fe tended to be redder, as did specimens with less gelatine and calcium.

The range of b^* values in the right column of Figure 27 was somewhat larger. Specimens with more Fe tended to be more yellow. Specimens with less gelatine and calcium also tended to be more yellow.

The relationships inferred from Figure 27 and other graphs do not take into account the effect of different storage conditions on the colour of the specimens. As noted above, however, some specimens in very different condition were found within the same book (Figure 6). For these pairs the difference in their L^* values cannot be due to differences in storage since they have been exposed to the same environmental conditions. Even though the specifics of the storage environments are unknown, each of these books effectively represents an individual, long-term natural ageing experiment. As such, this subset is perhaps the most important of the entire project. (There may be exceptions to this assumption, as when two books are disbound and rebound together again as a single book. But evidence of two or more books bound as one was rare in our specimen set, and to qualify for this subset the two specimens tagged had to be from the same “book.”)

In the top graph in Figure 28, each point represents the differences in L^* and gelatine concentration for a pair of specimens from the same book. Where the points are on the positive vertical axis the specimen with more gelatine was closer to white than the specimen with less gelatine. The plot includes only the points where the difference was greater than the confidence interval for the gelatine model (see Appendix I) and the uncertainty in the L^* value (based on the 95% confidence interval calculated from the standard deviation of the 8

interior readings). For most of the pairs the specimen with the higher gelatine concentration was closer to white than the specimen with less gelatine.

Similarly, the middle plot in Figure 28 shows the differences in L^* and calcium concentration for pairs from the same book. The plot includes more points than the top graph because there were more pairs with statistically significant differences in Ca. Again, in over 80% of the pairs the specimen

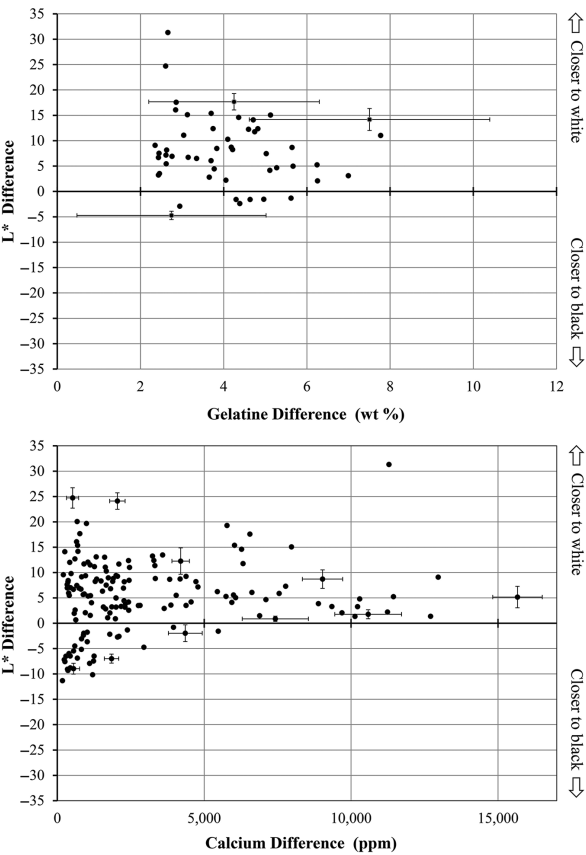


Figure 28: These three plots compare paper color in terms of L^* with gelatine, calcium, and iron content for pairs of specimens taken from individual printed pre-1800 books. Each point represents the difference in L^* and gelatine/calcium/iron content between the two sheets, which have been exposed to the same environmental conditions. Where the L^* difference is positive the sheet with higher gelatine/calcium/iron content was closer to white than the specimen with less gelatine/calcium/iron. Representative error bars indicate the 95% confidence intervals for the difference. The calcium plot has more points because the calcium differences were statistically significant for a greater number of pairs.

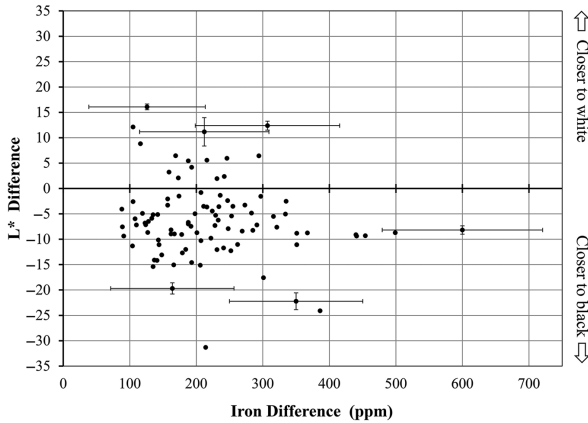


Figure 28: (continued)

with higher calcium concentration was closer to white than the specimen with less calcium. The opposite case is apparent for Fe in the bottom plot where the specimen with more Fe was darker in 85 % of the pairs. This result is consistent with a recent study of selective discolouration in two seventeenth century codices (Bainbridge 2015).

4 Conclusions

This study used non-destructive, quantitative measurement of 1,578 historic specimens to investigate changes in various components over time and the relationship of those data to Materials and Workmanship (M&W) grades. This information may be helpful in understanding the present condition of paper collections and evaluating preservation options. Papers made before 1500 contained higher concentrations of gelatine and calcium than papers made in subsequent centuries. The pre-1500 papers were also thicker and lighter in colour based on L^* (white to black) values. We found an association between higher gelatine and Ca content and colour that was closer to white (L^*) and less red (a^*). The data show an apparent association between more red and more yellow colour and increasing Fe content. No such association can be attributed to K and S in the specimens tested. When data on K and S were used as an indicator of alum concentration, we saw no increase over the 1400 to 1900 period. We attribute the increasingly acidic pH during this time reported by other researchers to a decrease in gelatine and Ca concentration. Analysis of the M&W data suggests that more Ca compounds and more gelatine were used by

papermakers when they produced more carefully made, better quality sheets. Poorer quality papers were made using lower levels of Ca and gelatine, and water and/or equipment that left higher levels of Fe in the finished sheets.

Acknowledgements: We would like to acknowledge the other original co-authors on the research documented at paper.lib.uiowa.edu:

- Robert Shannon, applications physicist, Bruker Elemental
- Irene Brückle, professor of conservation, Stuttgart State Academy of Art and Design
- Michael Schilling, senior scientist, Getty Conservation Institute
- Joy Mazurek, assistant scientist, Getty Conservation Institute
- Jennifer Wade, program director, Deep Earth Section of the Division of Earth Sciences of Geosciences at the National Science Foundation; formerly research chemist, Preservation Research and Testing Division, Library of Congress
- Jessica White, proprietor, Heroes & Criminals Press; formerly research assistant, University of Iowa Center for the Book

In addition to the co-authors, many other individuals made important contributions to various aspects of the project. They include:

- Gary Frost, conservator, University of Iowa Libraries
- John Baty, Johns Hopkins University Conservation Research Associate & Andrew W. Mellon Postdoctoral Fellow
- Yvonne Hilbert, Stuttgart State Academy of Art and Design
- Josefine Werthmann, Stuttgart State Academy of Art and Design
- Lee Marchalonis, University of Iowa Center for the Book
- Heather Wetzel, University of Iowa Center for the Book
- University of Iowa Main Libraries Website Team:
 Paul Soderdahl, Associate University Librarian for Information Technology
 Greg Prickman, head, Special Collections & University Archives
 Nicole Saylor, head, Digital Research & Publishing
 Wendy Robertson, digital scholarship librarian
 Linda Roth, library webmaster
 Kevin McMullen, student assistant

Finally, the following institutions and offices also helped make this research possible by direct financial or logistical support, or by contribution of their staff member's time:

- Institute of Museum and Library Services
- University of Iowa Center for the Book

- University of Iowa Office of the VP for Research
- University of Iowa Museum of Art
- Kress Foundation
- Preservation Programs, National Archives and Records Administration
- The Newberry Library
- Bruker Elemental
- Getty Conservation Institute
- Library of Congress, Preservation Research and Testing Division
- The John D. and Catherine T. MacArthur Foundation Fellows Program

Appendix I: Chemometric modelling of gelatine concentration

An Analytical Spectral Devices LabSpec Pro UV/Vis/NIR spectrometer collected reflectance spectra using a standoff diffuse reflectance probe accessory at approximately 70 degrees relative to the specimen surface. The sampling covered a circular area about 5 mm in diameter. Spectra were gathered over a range of 350–2,500 nm at 1 nm intervals, and each spectrum was an average of 50 readings. A Spectralon® (pressed polytetrafluoroethylene) disk was used as the white reflectance standard. Specimens were placed on top of a Gore-tex® sheet to provide a uniform and consistent background. This thin, flexible sheet of expanded polytetrafluoroethylene bonded to a non-woven polyester felt can be inserted between pages of a book.

The chemometric model was calibrated using 40 historical specimens produced between the fifteenth and eighteenth centuries. Of the 40 specimens, half were categorized as light and half as dark based on visual appearance. To produce a realistic NIR model, the calibration specimens must provide a range of gelatine, paper thickness, and paper density combinations reflecting the variety likely to be found among the unknowns (Duckworth 1998, 162). At the same time collinearity must be avoided, and this was evaluated by calculating linear correlation coefficients for the 40 specimens: density vs. thickness, $R^2 = 0.12$; gelatine vs. density, $R^2 = 0.09$; and gelatine vs. thickness, $R^2 = 0.30$.

Gelatine concentrations were measured by removing samples of paper at 10 locations on each sheet. The gelatine was assumed to be uniformly distributed throughout the paper rather than concentrated at the surface (Hummert et al. 2013, Rouchon et al. 2010). The removed samples were analysed as ethyl chloroformate derivatives by GC-MS (Stephens et al. 2008) using a method which quantifies gelatine based on seven stable amino acids (AA). NIR spectra were

gathered near the sampling locations on both sides of each sheet, so a total of 20 spectra were gathered from a specimen during a given data collection session. To study the repeatability of the measurements and simulate use in the field, a second set of 20 spectra was gathered on another day. Between sessions the system was powered down and disassembled. The same procedure was followed for a third session, so a total of 60 spectra were gathered from each specimen.

The model was developed with GRAMS PLSplus/IQ chemometric software. The multivariate analysis attempts to correlate the 40 measured gelatine values with differences in the broad, overlapping peaks of cellulose and gelatine in the spectra. The model used partial least squares with leave one-out cross-validation on the mean-centred first derivative of the spectra, which were calculated using the gap method with spacing 15. Based on the prediction residual error sum of squares plot and F-ratio values, a model was selected utilizing four factors. Including data from the UV and visible ranges introduced more noise without improving the model.

These calibration spectra were collected with the instrument in an environmentally-controlled room with temperature 22 °C and relative humidity (RH) 50–58 % whereas the ambient RH may vary considerably in the field. Tests on a few specimens as they equilibrated from 20–25 % RH to 50–55 % RH showed the greatest response in the range 1,900–2,000 nm for the raw spectra and 1,775–1,975 nm for the first derivatives. The latter region was excluded from the model to reduce sensitivity to RH changes.

Two ranges were used in the model: 1,058–1,775 nm and 1,975–2,358 nm. The penetration depth of NIR radiation decreases as wavelength increases. Based on tests with and without the backing sheet and estimates of the information depth (Clarke et al. 2002), at the higher wavelength range the thickness of the specimens is great enough that they can be treated as infinitely thick for diffuse reflection measurements. Below about 1,500 nm this approximation begins to break down for a large fraction of the historical specimens, and the NIR radiation likely reaches the backing layer. Since the uniformly reflected signal from the polytetrafluoroethylene is not correlated with the gelatine AA measurements the multivariate model should be able to account for it, particularly since there is relatively little cellulose or gelatine absorption in this range. To evaluate this assumption a model was calculated using only wavelengths greater than 1,650 nm, but there was little difference compared to the model extending down to 1,058 nm. We chose to use the full lower region because the linear correlation coefficient spectrum for gelatine showed both positive and negative correlations over the range, indicating that the region provided useful information for the model (Duckworth 1998, 143).

The horizontal axis in Figure 29 shows the percent by weight of gelatine as measured by the destructive AA analysis, and the vertical axis is the percent gelatine as predicted by the cross-validated NIR model. The error bars show one

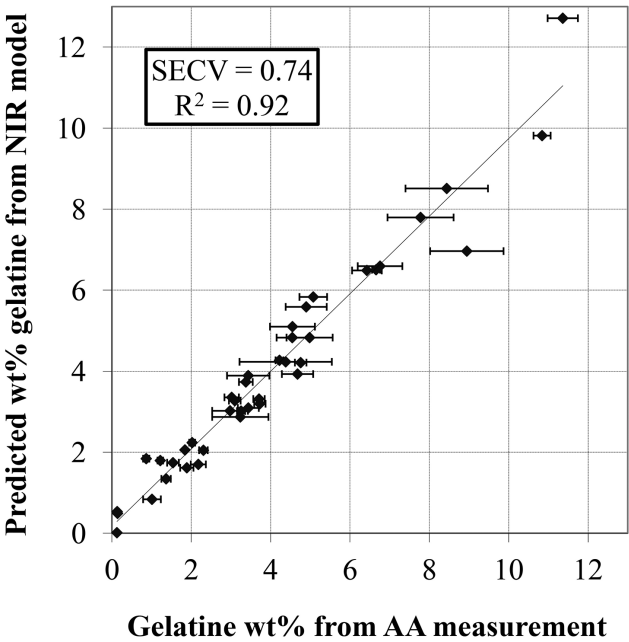


Figure 29: NIR model calibration plot. Error bars indicate standard deviation of triplicate AA measurements.

standard deviation calculated from the triplicate AA measurements. The standard error of cross validation (SECV) was 0.74.

While the R^2 and SECV values give some indication of how well the NIR model fits the calibration data, these quantities do not take into account the uncertainties in the AA measurements, which are illustrated by the horizontal error bars in Figure 29. For this study, the size of the reference value uncertainty was compared to the standard deviation of the NIR predictions. As discussed above, 10 spectra were collected from both sides of a reference specimen during a given data collection session. The average of each set of 10 was calculated for all 3 sessions, yielding 6 spectra per specimen. These averaged spectra yielded 6 gelatine predictions. Figure 30 compares the size of the standard deviation of these 6 NIR predictions with the standard deviation of the AA measurements. For the majority of specimens with gelatine concentrations below 6%, the standard deviation of the NIR prediction is significantly larger than the standard deviation of the AA measurement. Above 6% the standard deviations are comparable, with a few exceptions. Overall, the standard deviation of the replicate NIR readings is a reasonable approximation of the uncertainty in the

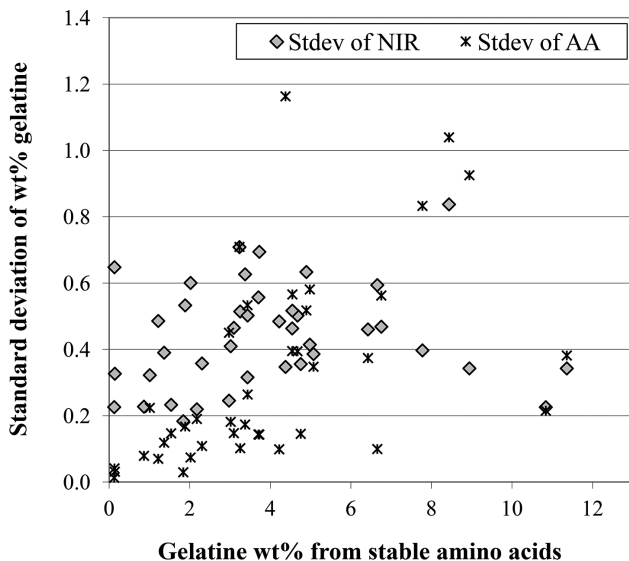


Figure 30: Comparison of standard deviation of destructive AA measurements with standard deviation of replicate NIR predictions.

prediction. That is, the uncertainties in the AA values are assumed to be negligible relative to the standard deviation of the NIR readings.

This approximation is less accurate at higher concentrations (8% and above), and the accuracy of the model could be improved by adding more calibration specimens in this range. In addition, at higher concentrations there may be limitations in the water extraction procedure used to quantitatively extract the gelatine from the paper for subsequent AA analysis. These difficulties may account for the wider error bars on some points at higher levels in Figure 29. The accuracy might also be improved if the model were calibrated using spectra taken at lower humidities such as 25% provided that the specimens can safely be exposed to these conditions (Cséfalvayová et al. 2010). This latter approach was not feasible for this survey project, but it might be possible for analysis of individual items in a conservation lab.

The data from Figure 29 are re-plotted in Figure 31, but all replicate predictions are shown instead of their average. The vertical axis shows the difference between the concentration predicted by the NIR model and the AA measurement. The outer lines are the 95% prediction intervals determined using the repeatability of the measurements. Based on these results, if the NIR measurement on a specimen predicted a gelatine concentration in the range 0 to 6% then there is a 95% probability that the difference between the NIR model value and a destructive AA

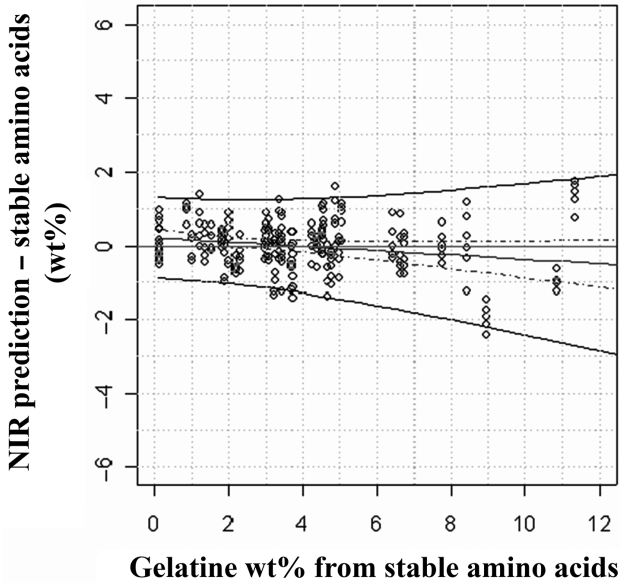


Figure 31: NIR model showing 95% prediction bands.

measurement would be between -1.6 and $+1.3$ percentage points. Between 6 and 8% gelatine there is a 95% probability the difference between the two measurements will be between -2.0 and $+1.5$ percentage points, and between 8 and 12% the difference is between -3.0 and $+2.0$ percentage points.

The gelatine model was developed using a LabSpec Pro instrument. The model was then applied to a set of historical specimens using spectra collected with an Analytical Spectral Devices QualitySpec Pro UV/Vis/NIR spectrophotometer. This instrument is identical to the LabSpec Pro except for the latter's optional battery-powered operation for use in the field. The QualitySpec Pro was initially factory calibrated to match the LabSpec Pro as closely as possible. Readings of a set of historical paper reference specimens were made on both instruments initially and throughout the project. The QualitySpec Pro gelatine predictions were 0.5% lower than the LabSpec values, and the data were corrected by this factor.

Appendix II: Colour measurements

While the UV/Vis/NIR instrument provides data for the visible region, it does not conform to the geometry, illumination, and other specifications of various colour measurement standards. Using the 40 specimens in the calibration set, data in

the visible range of the UV/Vis/NIR spectra were converted to CIELAB values using the D65 illuminant and 2° observer. At the same locations spectra were collected with an X-Rite Model 968 spectrophotometer designed for colour measurement. The 10 readings on each specimen were averaged to give a single value for the whole sheet. Using the X-Rite measurements we calculated linear calibration curves ($R^2 > 0.98$) for L^* , a^* , and b^* to apply to the UV/Vis/NIR data from unknowns.

Appendix III: Statistical analysis

For Figures 8–12, 15, 16, and 19 the centre curve represents a locally smoothed estimate of the mean, as a function of year. The outer curves give the pointwise 95 % confidence intervals for the means over the years. We used LOESS (locally weighted scatter-plot smoothing) in R to compute the smoothed estimate of the mean (R Foundation 2015). The confidence bands were derived using Monte Carlo simulation. The confidence bands account for both the sampling error (the error that results because a sample rather than a census of archived papers is taken) and the measurement error as described below. The black points in the scatter plot represent the observed values. These observed values were measured with some error, so if we had carried out the same measurement procedures on the same samples, we would have come up with slightly different values. Precision in our work was calculated differently for the XRF and the UV/Vis/NIR instrumentations. Using the respective precision parameters, corresponding to each observed datum (solid black symbol), we generated 10 auxiliary data points (light grey circles) that represent data we would expect to see if the same measurement process was repeated 10 times. The light grey circles represent potential values under replicate measurements. The spread of these potential values reflects the lack of precision in the measurements; i. e., the more spread out these potential values are, the less precision. The statistic R is a generalization of the Pearson correlation coefficient. Whereas the correlation measures the strength of the linear relationship between X and Y , the statistic R measures the strength of the functional (linear or nonlinear) relationship between X and Y . If the observed (X, Y) values fall close to a smooth, non-constant function of X , then R will take on a value close to 1. If X and Y are linearly related, then R will be numerically identical to the absolute value of the correlation coefficient. Numerically, R is the correlation between the smoothed estimates of the Y means and the observed Y values. The P-value was computed using a nonparametric bootstrap approach. P-values less than 0.05 indicate a very low probability that the relationship apparent in the plots is due to chance.

Suppose we had an additional specimen from a particular year and that, without collecting a measurement, we wanted to use the data from Figure 8 to predict its gelatine concentration within a range so that there was a 95 % probability that the actual value was within our prediction interval. Since Figure 8 shows a large spread of the concentrations (due to instrumental uncertainty and the variation among the specimens) it is apparent that this prediction interval will be wide. That is, given a single specimen from a particular year, without actually collecting a measurement we can only make an imprecise prediction of its gelatine concentration. We are more interested, however, in the mean value over time periods, which we can estimate with much better precision. The 95 % confidence interval for the mean (as opposed to the prediction interval for a single specimen) is about twice the standard deviation divided by the square root of the number of specimens. Within the century-long periods in Figure 8 the wide variation among specimens will be reflected by the size of the standard deviation. This number is then divided by at least 10 (the square root of 100) because we measured hundreds of specimens within each period. The size of the sample set accounts for the narrow confidence intervals for the mean indicated in Figure 8 and similar plots. Suppose we returned to the libraries, selected a different sample of 1,500 + specimens from this population using the same criteria, measured their gelatine concentrations, determined confidence intervals like those in Figure 8, and repeated this process many times. About 95 % of the confidence intervals would include the true average gelatine concentration for this population.

References

- Ahn, K., Hofmann, C., Horsky, M., Potthast, A.: *How copper corrosion can be retarded – new ways investigating a chronic problem for cellulose in paper*. Carbohydrate Polymers 134 (2015): 136–143.
- Bainbridge, A. W.: *Non-destructive analysis of selective discolouration in two seventeenth century codices*. Journal of Institute of Conservation 38 (2015): 3–13.
- Baker, C. A.: *From the Hand to the Machine: Nineteenth-Century American Paper and Mediums: Technologies, Materials, and Conservation*. Ann Arbor, MI: The Legacy Press, 2010.
- Barrett, T. D.: *Early European papers/contemporary conservation papers: A report on research undertaken from fall 1984 through fall 1987*. The Paper Conservator 13 (1989): 1–108.
- Barrett, T. D., Mosier, C.: *The role of gelatin in paper permanence*. Journal of American Institute for Conservation 34 (1995): 173–186.
- Barrett, T., Shannon, R., Wade, J., Lang, J.: *XRF analysis of paper in open books*. In: Handheld XRF for Art and Archaeology (Studies in Archaeological Sciences), A. N. Shugar, J. L. Mass (eds.), Leuven: Leuven University Press, 2012: 191–214.
- Baty, J., Barrett, T. D.: *Gelatin size as a pH and moisture content buffer in paper*. Journal of American Institute for Conservation 46 (2007): 105–121.

- Blair, A.: *The rise of note-taking in early modern Europe*. Intellectual History Review 20 (2010): 303–316.
- Brückle, I.: *The role of alum in historical papermaking*. The Abbey Newsletter 17 (1993): 53–57.
- Clarke, F. C., Hammond, S. V., Jee, R. D., Moffat, A. C.: *Determination of the information depth and sample size for the analysis of pharmaceutical materials using reflectance near-infrared microscopy*. Applied Spectroscopy 56 (2002): 1475–1483.
- Cséfalvayová, L., Pelikan, M., Kralj Cigić, I., Kolar, J., Strlič, M.: *Use of genetic algorithms with multivariate regression for determination of gelatine in historic papers based on FT-IR and NIR spectral data*. Talanta 82 (2010): 1784–1790.
- Dabrowski, J., Simmons, J. S. G.: *Permanence of early European hand-made papers*. Fibers and Textiles in Eastern Europe 11 (2003): 8–13.
- Duckworth, J. H.: *Spectroscopic quantitative analysis*. In: Applied Spectroscopy: A Compact Reference Guide for Practitioners, J. Workman, Jr, A. W. Sprinsteen (eds.), Chestnut Hill, MA: Academic Press, 1998: 93–165.
- Dupont, A. -L.: *Gelatine sizing of paper and its impact on the degradation of cellulose during aging - a study using size-exclusion chromatography*. Ph.D. Thesis, Amsterdam: University of Amsterdam, 2003.
- Entlesberger, I., Schwanninger, M., Eyb-Green, S., Mayer, M., Baatz, W.: *Atypical discolourations and local differences in paper-surface structures: Investigation of the collection of incunabula of the Karl-Franzens-university, Graz*. Journal of Paper Conservation 12 (2011): 16–24.
- Fahy, C.: *Paper making in seventeenth-century genoa: The account of giovanni domenico peri (1651)*. Studies in Bibliography 56 (2003–2004): 243–259.
- Henniges, U., Schwanninger, M., Potthast, A.: *Non-destructive determination of cellulose functional groups and molecular weight in pulp hand sheets and historic papers by NIR-PLS-R*. Carbohydrate Polymers 76 (2009): 374–380.
- Houghton, J.: *Account of Paper Making*. In: A Collection for the Improvement of Husbandry and Trade, vol. 13 nos. 356–362. London: 1699.
- Hummert, E., Henniges, U., Potthast, A.: *Fluorescence labeling of gelatin and methylcellulose: Monitoring their penetration behavior into paper*. Cellulose 20 (2013): 919–931.
- Kolbe, G.: *Gelatine in historical paper production and as inhibiting agent for iron-gall ink corrosion on paper*. Restaurator 25 (2004): 26–39.
- Lichtblau, D., Strlič, M., Trafela, T., Kolar, J., Anders, M.: *Determination of mechanical properties of historical paper based on NIR spectroscopy and chemometrics – a new instrument*. Applied Physics A 92 (2008): 191–195.
- Missori, M., Righini, M., Dupont, A. -L.: *Gelatine sizing and discoloration: A comparative study of optical spectra obtained from ancient and artificially aged modern papers*. Optics Communications 263 (2006): 289–294.
- Potthast, A., Henniges, U., Banik, G.: *Iron gall ink-induced corrosion of cellulose: Aging, degradation and stabilization. Part 1: model paper studies*. Cellulose 15 (2008): 849–859.
- R Foundation: *The R Project for Statistical Computing*, <https://www.r-project.org> (accessed 3.1.16).
- Reynard, P. C.: *Manufacturing quality in the pre-industrial age: Finding value in diversity*. Economic History Review 53 (2000): 493–516.
- Rouchon, V., Pellizzi, E., Janssens, K.: *FTIR techniques applied to the detection of gelatine in paper artifacts: From macroscopic to microscopic approach*. Applied Physics A 100 (2010): 663–669.

- Sherman, W. H.: *Used Books-Marking Readers in Renaissance England*. Philadelphia: University of Pennsylvania Press, 2008.
- Shugar, A. N., Mass, J. L. (eds.): *Handheld XRF for Art and Archaeology (Studies in Archaeological Sciences)*. Leuven: Leuven University Press, 2012.
- Stephens, C. H., Barrett, T., Whitmore, P. M., Wade, J. A., Mazurek, J., Schilling, J.: *Composition and condition of naturally aged papers*. J. American Institute for Conservation 47 (2008): 201–215.
- Szirmai, A.: *The Archaeology of Medieval Bookbinding*. Aldershot: Ashgate, 1999.
- W. J. Barrow Research: *Laboratory: Physical and Chemical Properties of Book Papers, 1507–1949*. Permanence/Durability of the Book 7. Richmond: W. J. Barrow Research Laboratory, 1974.

Zusammenfassung

Zerstörungsfreie Analyse handgeschöpfter Europäischer Papiere aus dem 14.- 19. Jahrhundert

Um Veränderungen in der Papierzusammensetzung durch die Jahrhunderte nachvollziehen zu können und den Einfluss dieser Variationen auf die Langzeitstabilität von natürlich gealtertem Papier besser zu verstehen, wurden 1578 Papiere, die zwischen dem 14. und dem 19. Jhdt. hergestellt worden waren, zerstörungsfrei analysiert. Gelatinegehalt und Farbe wurden mittels UV/Vis/NIR Spektrometrie und Spuren metallischer Verunreinigungen wurden mittels RFA ermittelt. Die Untersuchungen konzentrierten sich dabei vor allem auf Kalium und Schwefel als Indikatoren von Alaunzusätzen, auf Eisen als typische Verunreinigung von Papier und auf Calcium, das oft in Verbindungen wie Calciumcarbonat im Papier zu finden ist und als alkalische Reserve im Papier fungiert. Die Studie ergab, dass Papiere mit hellerem Farbton (einem höheren Weißgrad) mit einem höheren Gelatine- und Calcium-Gehalt und einem geringeren Eisenanteil in Verbindung stehen. Die Messungen ergaben außerdem eine signifikante Abnahme von Gelatine- und Calcium-Konzentration im Laufe der Jahre; dabei fällt der stärkste Rückgang in die Zeit des sprunghaften Anstiegs der Druckproduktion um 1500. Die Abnahme des pH-Werts über die Jahrhunderte, die in anderen Studien beschrieben wurden, ist dabei eher auf die Abnahme dieser beiden Komponenten als auf die Alaunkonzentration zurückzuführen, welche durch die Jahrhunderte relativ gleichbleibend war. Weiters wurde die handwerkliche Qualität der Papiere bewertet und es konnte festgestellt werden, dass ein Zusammenhang zwischen Papieren besserer Qualität und einem höheren Gelatine- und Calciumgehalt sowie einem höheren Weißgrad besteht. Papiere schlechterer Qualität haben oft einen höheren Eisengehalt und eine größere Papierdicke.

Resumé

Analyse non destructive de papiers européens faits à la main du 14^{ème} –19^{ème} siècle

L'analyse non-destructive de 1.578 échantillons de papier fabriqués entre le 14^{ème} et 19^{ème} siècle a été entreprise dans le but de mieux comprendre les changements dans la composition du papier au fil du temps et comment ces variations pourraient affecter la stabilité du papier au cours du vieillissement naturel sur le long terme. Le contenu en gélatine et la couleur ont été déterminés par spectrométrie UV/Vis/NIR. Les métaux résiduels ont été mesurés par XRF. Ces composants incluaient le potassium et le soufre, comme éléments indicatif de la concentration d'alun; le fer, comme contaminant typique du papier; et le calcium, qui est souvent associé avec des composés tels que le carbonate de calcium qui peut servir de réserve alcaline. Cette recherche a démontré que des papiers qui sont de couleur plus claire (plus proche du blanc) sont associés à des niveaux plus élevés de calcium et de gélatine et à des niveaux inférieurs de fer. Cet état des lieux a également montré des diminutions significatives de concentration en gélatine et en calcium dans le temps, les plus grandes différences coïncidant avec l'essor de l'imprimerie autour de 1500. La baisse du pH au cours des siècles observés par d'autres chercheurs s'explique plus par une diminution de ces deux variables que par une augmentation de la concentration en alun qui est demeuré relativement stable au cours des siècles. Le savoir-faire de ces échantillons a été évalué en utilisant des classifications de matériaux et de procédés de fabrication qui ont montré un papier de meilleure qualité associée à une concentration de gélatine et de calcium supérieure et a une couleur plus proche du blanc. Les papiers de moindre qualité étaient associés à des niveaux de fer plus élevés et une épaisseur plus importante.

Authors

Timothy Barrett is an associate professor in the University of Iowa Center for the Book and the School of Library and Information Science. His many years of research on early European handmade papers have been funded by the NEA, the Kress Foundation, the Institute for Museum and Library Services and as of 2010, a MacArthur Fellowship. Barrett is author of two books, one website, seven videotapes, and thirty-six articles or book chapters on the history, technique, science and aesthetics of hand papermaking.

Mark Ormsby is a scientist in the Preservation Research and Testing Laboratory at the National Archives and Records Administration. He has worked on the re-encasement and monitoring of the Declaration of Independence, U.S. Constitution, and Bill of Rights. He also focuses on environmental monitoring and energy-efficient operation of archival storage areas.

Joseph B. Lang is a professor and chair of the Department of Statistics and Actuarial Science at the University of Iowa. A fellow of the American Statistical Association, his research interests include categorical data analysis and causal inference.