Gunnar Almkvist*, Shahin Norbakhsh, Ingela Bjurhager and Kurt Varmuza

Prediction of tensile strength in iron-contaminated archaeological wood by FT-IR spectroscopy – a study of degradation in recent oak and *Vasa* oak

DOI 10.1515/hf-2015-0223
Received October 16, 2015; accepted February 16, 2016; previously published online March 18, 2016

Abstract: Oak from the Swedish warship *Vasa* and recent oak that was aged after impregnation with iron(II) chloride has been analyzed by FT-IR spectroscopy and submitted to tensile strength testing. The aim was to investigate correlations between FT-IR bands in the fingerprint region, chemical degradation, and tensile strength in iron contaminated oak. The concentration of carboxylic functions increased and the acetyl groups in the hemicellulose fraction were decreasing as a function of degradation time. These changes are accompanied by reduced tensile strength and elevated content of oxalic acid (OA) in both *Vasa* wood and the impregnated recent oak samples. To evaluate the possibility to predict tensile strength from spectral data, chemometric modeling by partial least-squares (PLS) regression was applied. The strategy of repeated double cross validation (rdCV) allowed a realistic prediction of tensile strength. Overall, chemical changes and mechanical performances of iron contaminated wood are strongly correlated and thus FT-IR spectroscopy is suited to predict the strength properties of the degraded wood.

Correction note: Correction added after online publication on March 18, 2016: The original third sentence of the Abstract was: The concentration of carboxylic functions and the acetyl groups in the hemicellulose fraction were decreasing as a function of degradation time.

***Corresponding author: Gunnar Almkvist,** Department of Chemistry and Biotechnology, Swedish University of Agricultural Sciences (SLU), Box 7015, SE-750 07, Uppsala, Sweden,
e-mail: gunnar.almkvist@slu.se
Shahin Norbakhsh: Department of Chemistry and Biotechnology, Swedish University of Agricultural Sciences (SLU), Box 7015, SE-750 07, Uppsala, Sweden
Ingela Bjurhager: The Ångström Laboratory, Department of Engineering Sciences, Uppsala University, Box 534, SE-751 21 Uppsala, Sweden
Kurt Varmuza: Institute of Statistics and Mathematical Methods in Economics, Vienna University of Technology, Wiedner Hauptstr. 8-10/E105, 1040 Vienna, Austria

Keywords: degradation, FT-IR, iron compounds, oak, PLS, rdCV, tensile strength, *Vasa*, waterlogged wood

Introduction

Degradation and aging of wood is a continuous process that affects wood polymers depending on the environmental conditions. Waterlogged wood artefacts contaminated by iron compounds are generally susceptible to chemical degradation as iron ions are active catalysts that lead to depolymerization of wood polymers through Fenton type of reactions (Emery and Schroder 1974). This, in turn, may have huge implications on both the chemistry and mechanical performance of wooden artefacts in the conserved state. The role of iron-related degradation reactions has been elaborated through the research projects on the Swedish 17th century warship *Vasa*. In the *Vasa* wood, iron compounds [partly mobile Fe(II) ions] are wide-spread as corrosion products originating from the numerous original bolts, which all had disappeared during the time in Stockholm harbour (1628–1961) (Håfors 2001; Almkvist and Persson 2011). After the salvage and raising of the ship in 1961 and subsequent polyethylene glycol conservation treatment, the load-bearing oak beams are now influenced by increased acidity due to oxalic acid (OA), depolymerization of polysaccharides, and decreased mechanical strength (Almkvist and Persson 2008a; Lindfors et al. 2008; Bjurhager et al. 2012). This is alarming in view of the hull's stability, as the *Vasa* is carrying her own weight, estimated to be almost 1000 metric tonnes.

Previous studies have shown that incorporation of iron(II) ions into recent oak wood and exposure to oxygen induced chemical degradation in terms of strength loss and increased acidity. Reduction of tensile strength (TS) went hand in hand with oxygen consuming reactions where OA and carbon dioxide were released (Norbakhsh et al. 2013, 2014) – evident signs of oxidative degradation. These processes were most intense during the first months of oxygen exposure and declined exponentially with time. Thus, the processes in iron contaminated *Vasa* oak can be simulated by reactions of recent wood.

Traditional methods of both mechanical testing (e.g. TS, in the axial direction) and chemical analyses are relatively sample-demanding methods (gram level) compared to spectroscopic methods, which are less destructive or even non-invasive. The usefulness of FT-IR spectroscopy in wood degradation chemistry has been frequently demonstrated (Moore and Owen 2001). In particular, the region 1800–1600 cm^{-1} includes bands of different functions, which are closely related to the chemical state of the wood. For example, the pH level in lignocellulosic materials can be estimated based on the band ratios in the mentioned region of interest (Pappas et al. 1999).

In the present paper, FT-IR spectra of iron contaminated wood are recorded and evaluated to learn more about the degradation chemistry. Core samples from the *Vasa* were examined together with recent oak that had been aged after impregnation with aqueous iron(II) solutions. The expectation was that it would be possible to determine the preservation status of the *Vasa* wood and that IR data could serve as a decision basis for future treatment strategies. The number of *Vasa* samples that had been analyzed for TS previously was limited due to the destructive action of sampling. Therefore, TS data from Bjurhager et al. (2012) and Norbakhsh et al. (2013) were compiled and spared material from these studies was submitted to FT-IR spectroscopy. Multivariate linear partial least-squares (PLS) models are developed for the prediction of TS from measured IR data. The optimum complexities of the models (numbers of PLS components) and the prediction performances are carefully estimated by the strategy repeated double cross validation (rdCV) (Filzmoser et al. 2009).

Materials and methods

Sample preparation and analyses: In total 17 samples were available for this study, 11 from the *Vasa*, and six from a recent European oak (*Quercus robur* L.) impregnated with iron(II) chloride (Table 1). The *Vasa* samples originated from four sampling positions on the ship (see Bjurhager et al. 2012) divided into subsamples of different depth from the surface (denoted V1–V4). The natural variation of annual rings and morphology may influence the mechanical properties of wood and therefore obscure variation caused by other parameters. The *Vasa* samples (heartwood) were visibly sound with the characteristic ring porous structure of oak clearly discernible. They were assumed not to differ significantly in their original mechanical properties. Samples close to the surface with significant content of polyethylene glycol were omitted due to their heterogeneity, former microbial degradation and significant interference of PEG in FT-IR spectra. Samples of recent oak impregnated with iron(II) chloride (denoted MO) of different age were available from previous studies (Norbakhsh et al. 2013). TS results for the included samples V1, V2, V3 and MO were calculated from sub-samples within each section or treatment to avoid local heterogeneities (e.g. annual ring influences)

Table 1: Sample information, tensile strength (TS), oxalic acid concentration (OA) and polyethylene glycol (PEG) content.

No.	Sample id. no.	Origin	TS (StD) [MPa]	OA [$mg \cdot g^{-1}$]	PEG [%]
	Vasa samples				
1	V1_B	Beam, hold, 65695:15–35 mm	33 (9)	4.1	<0.5
2	V1_C	Beam, hold, 65695:35–50 mm	48 (17)	1.8	0.5
3	V2_B	Planking, orlop deck, 65698:15–35 mm	65 (25)	0.9	3.0
4	V2_C	Planking, orlop deck, 65698:35–50 mm	66 (23)	1.0	6.5
5	V3_B	Knee, lower gun deck, 65701:30–50 mm	35 (7)	3.3	<0.5
6	V3_C	Knee, lower gun deck, 65701:50–80 mm	55 (22)	3.7	<0.5
7	V4a_B	Block of oak, 65542: Da112-030	35[a]	8.4	<0.5
8	V4b_B	Block of oak, 65542: Da148-030	48[a]	5.4	1.5
9	V4b_C	Block of oak, 65542: Da148-050	36[a]	6.7	<0.5
10	V4b_D	Block of oak, 65542: Da149-090	48[a]	3.3	1.3
11	V4b_E	Block of oak, 65542: Da147-110	44[a]	4.0	0.7
	Iron treated reference oak exposed to air				
12	MO_0	0 days in air	108 (18)	0.2	–
13	MO_50	50 days in air	86 (8)	0.3	–
14	MO_200	200 days in air	75[b]	1.3	–
15	MO_300	300 days in air	71 (13)	1.8	–
16	MO_1500	4 years in air	59[b]	3.9	–
17	Oak_1	Reference oak (no iron impregnation)	108 (15)	0.2	–

[a]V4 measured on single samples with estimated standard errors of 15%.

[b]Estimated, see experimental section.

TS results are average values of multiple subsamples in section or treatment, and in the axial direction of the samples. Analyses of OA and PEG were performed on homogenized and merged sub-samples.

(StD given in Table 1). TS testing of V1, V2, and V3 were performed on small sized [0.3 mm (R) by 3 mm (T) by 20 mm (L)] specimens with ca. 30 subsamples in each section. V4 and MO samples were dog-bone shaped samples [4 mm (R) by 2 mm (T) by 170 mm (L)]. The V4 samples are individual sub-samples, for which the TS had been determined (estimated standard errors 15%). Experimental procedures and further details are described elsewhere (Bjurhager et al. 2012; Norbakhsh et al. 2013). For iron impregnated recent oak, TS figures for MO_200 and MO_1500 (Table 1) were estimated from the correlation between exposure time and TS as presented in Norbakhsh et al. (2013), assuming a declining TS over time after exposure to oxygen.

Material for chemical analyses was taken from sections of parallel core samples (V1–V3) or from individual homogenized sub-samples and (V4 and MO). Samples were gently milled to a particle size of approximately 0.05 mm and placed in a desiccator containing P_2O_5 (Sigma Aldritch, St. Louis, MO, USA) as a drying agent. The content of OA and PEG was analyzed by aqueous extraction of wood material. 20 mg of sample material was extracted in 2 ml D_2O (Sigma Aldritch, St. Louis, MO, USA, 99.9%) for 2 days. The concentration of PEG in the extracts was analyzed by quantitative ^{1}H- NMR (Pauli et al. 2005). 20 µl of a 30 mM solution of sodium 3-(trimethylsilyl)propanoate-d_4 (TMSP) (Cambridge Laboratories, Cambridge, UK) were added as a reference to the extracts (600 µl). Instrument: Bruker DRX600 NMR (Bruker-Daltonik, Bremen, Germany); 30°C. Baseline correction, phasing, and peak integration were applied manually. The concentration of PEG was quantified by comparing the relative intensities of PEG signal (δ 3.70) with the TMSP signal (δ 0.00). For details and instrument parameters, see Almkvist and Persson (2008a). The concentration of oxalate in the extracts was determined by HPAE chromatography in the accredited laboratory Innventia AB, Stockholm, Sweden. Based on included double samples and references, the accuracy for determination of OA and PEG was ca. ±10%.

The FT-IR spectra were recorded via the KBr technique (2 mg dry material in 200 mg oven dried KBr) (Aldrich, p.a.) and two pellets were prepared for each sample. As a complement, some of the samples were pre-extracted in water to remove water soluble extractives. Milled material (100 mg) was soaked in 1 ml water for 1 day. The wood was separated by centrifugation and the extraction was repeated twice as described above. The extracts were freeze-dried and the solid material was dried in a desiccator. Microcrystalline cellulose (Aldrich), birch xylan (Fluka), oak holocellulose and milled-wood birch lignin (MWL) (Innventia AB, Stockholm, Sweden) served as references. Instrument: Perkin Elmer Spectrum 100 FT-IR equipped with a lithium tantalite ($LiTaO_3$) detector. The instrument was adjusted to an iris of 8.94 mm, 32 scans (4000–400 cm^{-1}); resolution of 4 cm^{-1} with data provided in steps of 1 cm^{-1}. The background of KBr pellets was subtracted for each analyzed sample. Each sample was averaged from acquisition on two KBr pellets.

Pre-processing of FT-IR spectra: The FT-IR spectra were baseline corrected by applying a spline function where the absorbance was set to zero at 3600, 2400, 1900, 800 and 400 cm^{-1}. Two major types of spectral data set were prepared: 1) deconvoluted spectra where the intensities of underlying bands were estimated, and 2) whole spectra modified by different pre-processing techniques.

For deconvolution a series of signals was assigned based on reference spectra and according to the literature (Figure 1 and Table 2). The acetyl groups (~1740 cm^{-1}), carboxylic groups (~1700 cm^{-1}), and adsorbed water (~1635 cm^{-1}) were each represented by two separate signals and served for optimising the curve fitting in this region. The deconvolution of spectra was performed by the Fityk® software (Wojdyr 2010), where

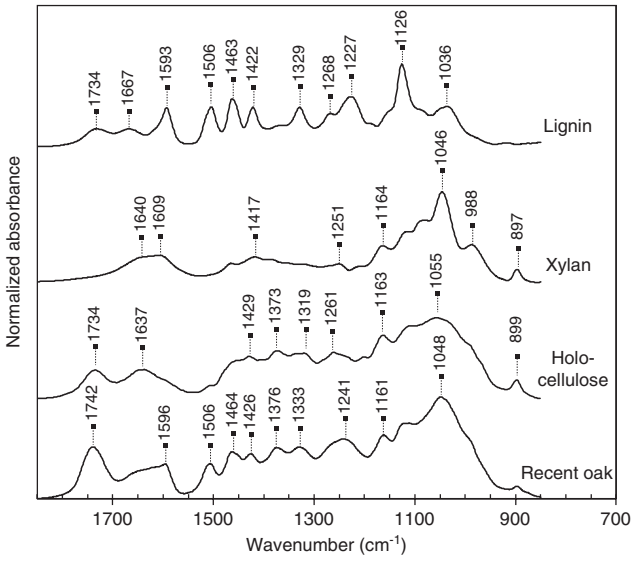

Figure 1: FT-IR spectra of recent oak and reference compounds.

the bands were adjusted by Gaussian functions, and for curve-fitting the Levenberg-Marquadt algorithm was applied. In an iterative process, the widths and positions of the deconvoluted peaks were optimized for the best fit. In this process, the position and width of each band were locked to be able to apply the same deconvolution model. For the sake of the best curve fitting around 1600–1500 cm^{-1}, small additional contributions from imaginary signals at 1565, 1548 and 1531 cm^{-1} were preliminarily applied. Later, the deconvoluted peak areas of these three bands were discarded. The region below 1200 cm^{-1} was not assigned or deconvoluted due to its complexity and many overlapping bands. The peak areas were normalized to the area of the lignin skeleton vibration at 1507 cm^{-1}, which can serve as a kind of internal standard (Faix and Böttcher 1992). An extra variable was added, X_{18}, expressed as a ratio between the sum of signals originating from carboxylic functions and a sum of signals related wood polymers in the region 1376–1270 cm^{-1}, see Table 2 and the discussions below.

$$X_{18} = ([1710]+[1690]+[1591])/([1376]+[1333]+[1270]) \quad (1)$$

Whole spectra in the region of 1800–1200 cm^{-1} were used for chemometric modeling after normalisation and application of spectral philtres including 1st and 2nd derivatives and standard normal variate transformations (SNV), see Table 3. For normalized spectra, the bands at 1507 cm^{-1} were set to 1. Derivatives were calculated by the Savitzky-Golay method with 13 points (6 at each side of the central point) and 2nd order polynomial. SNV transformations were performed by centering each spectrum and then it was scaled by its own StD to minimise spectral scattering (Barnes et al. 1989). All spectral philtres were calculated by the programme SIMCA-P+12.0®.

Chemometric methods: Exploratory data analysis was made by principal component analysis (PCA), and examination of the Pearson correlation coefficient (r) between single X-variables and y (TS, OA). Then linear regression models for the prediction of response y from experimental x-variables (IR absorbances and transformed data) were made by PLS regression. PLS is frequently applied in chemometrics for linear multivariate calibration models; the various PLS methods are well documented (Wold et al. 2001; Filzmoser et al. 2009). PLS is applicable for data with the number of X-variables (m) larger than the number of samples (n), and for highly correlating

Table 2: Assignments of FT-IR bands used in fitting procedure of investigated samples, and for chemometric models.

Region (cm^{-1})	Assignment	Signal positions (cm^{-1})
1760–1710	C=O stretch of acetyl group in xylan[a]	1755, 1742
1710–1600	C=O stretch of carboxylic acid groups[a]	1710, 1690
	C=O stretch in lignin	1670
	H-O-H deformation of adsorbed water[a]	1650, 1624
1600–1500	C=C stretch in lignin	1591, 1507[b]
1500–1400	CH deformation in lignin	1462
	Aromatic skeletal vibr., CH deform. in lignin, CH_2 scissoring in cellulose	1425
	Possible degradation parameter, iron-polysaccharide complex	*1400*
1400–1200	CH deformation in cellulose and xylan	1376
	C-O stretch in lignin (syringyl), CH and OH wagging in cellulose and xylan	1333
	C-O stretch in lignin and xylan, CH and OH wag. in cellulose and xylan	1270
	C-O stretch in O-C-R groups, acetyl group in xylan	1242
	C-O stretch in lignin, CH and OH wag. in cellulose and xylan	1226
	OH plane deformation in cellulose and xylan	1202

[a]Represented by two bands in order to optimise curve fitting.
[b]Signal 1507 cm^{-1} was used for normalisation and not in the chemometric modeling.
Signal positions according to Schwanninger et al. (2004), Faix (1991) and Fengel (1991). For the signal at 1400 cm^{-1} see discussion in the text.

Table 3: Overview of datasets used for chemometric modeling after different pre-processing.

Pre-processing	Abbrev.	X matrix (samples *X* variables)
Deconvolution (integr. individual signals)	IR17+IR18[a]	17×17 (18)[a]
Normalized spectra[b]	IR600	17×600
1st derivative spectra[b]	IR600d1	17×600
2nd derivative spectra[b]	IR600d2	17×600
SNV spectra[b]	IR600snv	17×600

[a]In data set IR18 an additional variable included according to eq. 1,
[b]Full spectral data, 1200–1800 cm^{-1}.

X-variables. The complexity of a PLS regression model is controlled by the number of PLS components, *A*. PLS components are orthogonal latent variables which have maximum covariance with *y*, and are defined as linear combinations of all *X*-variables. An optimum value for *A* avoids underfitting (a too simple model) as well as overfitting (model best adapted to the calibration data used for creating the model), and thus gives optimum performance of the prediction of *y* for new cases (test data). Here, the strategy of rdCV was applied (Filzmoser et al. 2009; Varmuza and Filzmoser 2014), which provides a separate estimation of an optimum *A* and a conservative estimation of the prediction performance. rdCV has been implemented (Varmuza et al. 2013) in the open source software R (R-project 2014).

The prediction of *y* is the standard error of prediction, SEP, defined as the StD of prediction errors, $e_i=y_i-\hat{y}_i$, with y_i for the given (experimental, "true") value, and $\hat{y}_i$ the predicted (modeled) value for a property of a sample *i*. In rdCV all $\hat{y}_i$ are test set predicted results (Varmuza and Filzmoser 2009). SEP is given in the units of *y*; because the prediction errors are usually normally distributed, ±2 SEP defines a 95% confidence interval for predicted *y*'s. The number of samples, *n*=17, in the available data is small, and therefore, the obtained results have to be cautiously interpreted only as rough estimations.

Only a brief description of rdCV is given here; for details see (Filzmoser et al. 2009; Varmuza et al. 2013). In rdCV the *n* available samples are split randomly into s_{TEST} parts (segments). Since *n* is small, the relative large value s_{TEST}=8 was used, resulting in 7 segments each with two samples and one segment with three samples. In the outer CV loop one of the segments was defined as test set (two or three samples), and the others formed a calibration set (14 or 15 samples). An inner CV loop within the calibration set (with s_{CALIB}=5 segments) an optimum *A* was estimated, and the resulting model was applied to the test set samples. After finishing the outer CV loop a test set predicted *y* is obtained for each sample, and a SEP could be calculated as the StD of the *n* prediction errors. Because SEP depends on the (random) split of the samples into segments, the procedure was repeated 50 times with different random splits. The resulting 50 SEP values provide a more complete picture than a single value and the mean served for estimation for a single final performance measure. For comparison of the models, the distribution of the 50 SEP values is represented by a boxplot for each model.

Firstly, models were created by PLS-rdCV based on all *X*-variables in the different data sets defined in Table 3. Secondly, three methods for mathematical variable selection were applied before PLS-rdCV. The methods applied: (1) selection of variables exhibiting maximum absolute Pearson correlation coefficients with *y* (typically 3–100 variables selected); (2) stepwise variable selection. The latter method has been implemented in a new R-function (Varmuza and Filzmoser 2014) as a combination of the forward and the backward strategy with BIC (Bayes information criterion) as selection criterion. (3) Selection by sequential replacement – as implemented in the R-library "leaps" and based on criterion BIC (Varmuza and Filzmoser 2014) – was used for deconvoluted spectra (18 variables). More sophisticated variable selection methods, like genetic algorithms were not applicable because of the small number of samples. Prominent aim of variable selection is an improved prediction performance (carefully tested by rdCV); furthermore, a small number of variables increase the chance for a reasonable interpretation of the model.

Results and discussion

Sample information, TS and content of OA and PEG are given in Table 1. The concentration of iron was previously analyzed on parallel samples: for *Vasa* samples (1.5–10 mg·g^{-1}) and for impregnated recent oak (ca. 6 mg·g^{-1}). The IR spectra of recent oak, *Vasa* oak and iron impregnated recent oak all displayed typical band patterns with underlying bands that are found in the spectra of reference components (i.e. hardwood hemicelluloses, polysaccharides in general, and lignin) (Figure 1 and Table 2). The bands can be assigned to typical vibrations of the wood components (Faix 1991; Fengel 1991; Moore and Owen 2001; Schwanninger et al. 2004). The region of interest in this study was 1800–1200 cm^{-1} containing the main characteristic bands from unconjugated carbonyl (C=O) stretch at 1742 cm^{-1} (acetyl groups attached to xylan), aromatic skeletal vibration at 1507 and 1597 cm^{-1} (lignin), signals related to C-H deformation in the region 1500–1300 cm^{-1} (lignin), and C-O-C vibrations in the region 1300–1200 cm^{-1}. Over all, the spectra of *Vasa* samples and iron impregnated samples displayed relatively small differences compared to recent wood. However, some parts of the spectra indicated more variability and were hence of interest to explain different properties of the samples and variation within the sample group. For the *Vasa* samples, the bands at 1800–1600 cm^{-1} include multiple signals that are affected and more variable as compared to reference oak. The signal at 1400 cm^{-1} is not discernible in recent oak but appears to be different in the *Vasa* samples. The bands at 1300–1200 cm^{-1} tend to be suppressed in the *Vasa* samples and in MO1500. The potential interference from PEG signals was not taken into account in the deconvolution model. However, minor contributions from PEG are observable in the V2 spectra of at 1250 cm^{-1} (Figure 2) in accordance with the elevated content of PEG in these samples (Table 1).

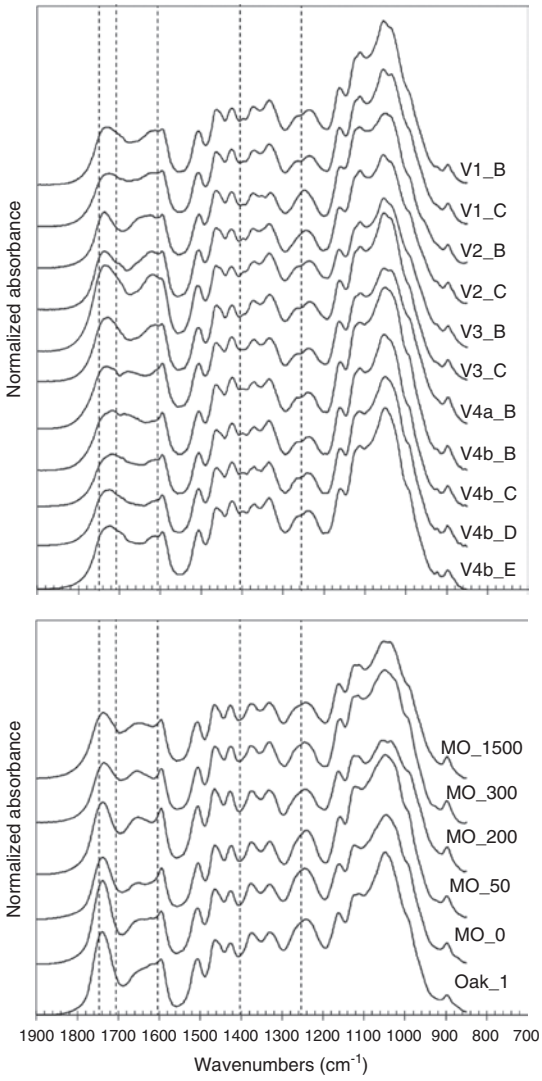

Figure 2: FT-IR spectra of investigated *Vasa* samples (upper) and modeled oak (below).
For sample information and abbreviations, see Table 1.

Principal component analysis (PCA)

Essential data of PCA are presented in Table 3. Figure 3 shows the PCA score and loading plot for the 1st and 2nd principal component (PC1, PC2, resp.) for IR17 (m=17 X-variables from deconvoluted IR bands). The score plot indicates a clear dependence of the location of the samples with the TS value, thus demonstrating a relationship between the 17 variables and TS. In the corresponding loading plot a clear clustering of the variables is visible according to their correlation with TS. This data structure is promising for TS prediction. PCA results for the data sets with 600 variables (in the range 1800–1200 cm^{-1}, Table 3) show a similar clustering of the samples according to their TS values (not shown). From PCA scores and loadings in Figure 3, it is clear that the signals at 1742, 1242, and 1376 cm^{-1} are most prominent in the more preserved samples (higher TS) in opposite to the signals 1710, 1690, and 1400 cm^{-1} which are associated to samples with a state of high degree of degradation (low TS and high OA). These correlations are also summarized in Figure 4 presenting the Pearson correlation coefficients, r, of deconvoluted bands with TS and OA, respectively. The largest absolute correlation coefficients were found for TS, while for OA the correlation coefficients are considerable smaller and have an opposite sign to those of TS. Consequently, TS and OA are negatively correlated (r=-0.78), confirming the fact that

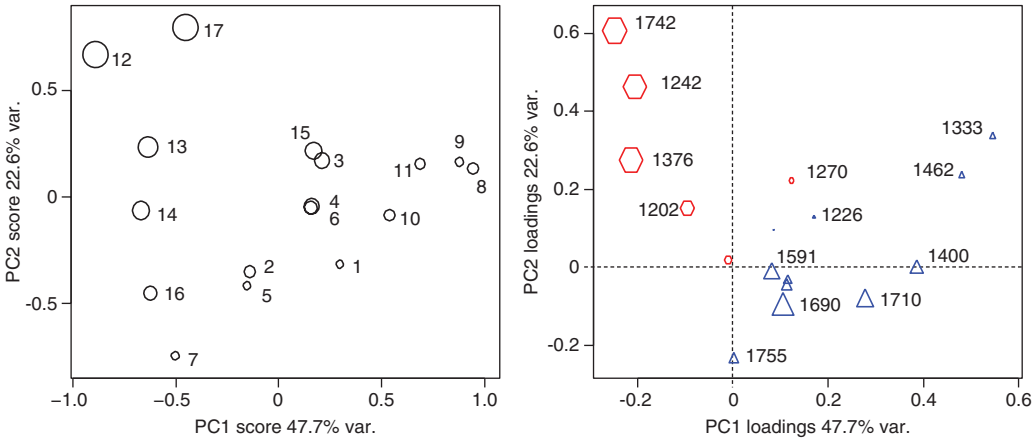

Figure 3: PCA of data set IR17 (n=17 samples, m=17 variables from IR band areas.
Variances preserved in first and second principal component (PC1, PC2) are 47.7% and 22.6%, resp. Left: score plot with the diameters of the circles proportional to TS, and sample numbers (Table 1) Right: loading plot with wave numbers in cm^{-1}; red hexagons for positive correlation coefficients with TS, blue triangles for negative correlation coefficients with TS, and the diameters proportional to the absolute value of the correlation coefficient.

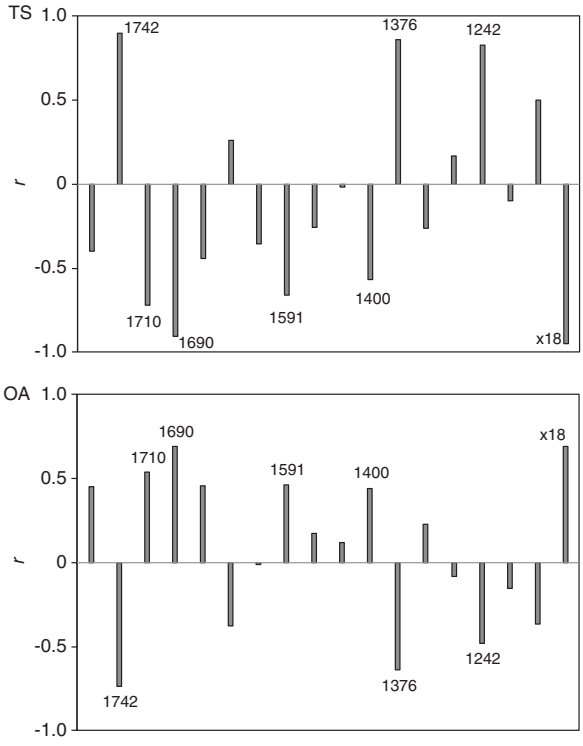

Figure 4: Pearson correlation coefficients, r, between TS (tensile strength) and OA (oxalic acid concentration) with the 17IR band areas listed in Table 2 and variable X_{18} as defined in equation 1. Numbers at the bars are wavenumbers (cm^{-1}) of deconvoluted signals.

decreased TS, as a result of chemical degradation, goes hand in hand with formation of OA both in *Vasa* oak and in modeled oak (Bjurhager et al. 2012; Norbakhsh et al. 2014).

Deacetylation

The band at 1742 cm^{-1}, associated to the acetyl group of xylan, display high correlations to both TS and OA (r=0.90 and r=-0.74, resp.). Acetyl groups associated to xylan are readily released by increased acidity in wood (Fengel and Wegener 1984), which seems to be the most plausible explanation among the samples in this study (Figure 5). Nevertheless, the high correlation of TS to the band at 1742 cm^{-1} indicates that deacetylation has important implications in terms of intermolecular binding of xylan in the ultrastructure and thus to the mechanical performance of wood. Xylan with less acetyl groups could lead to higher

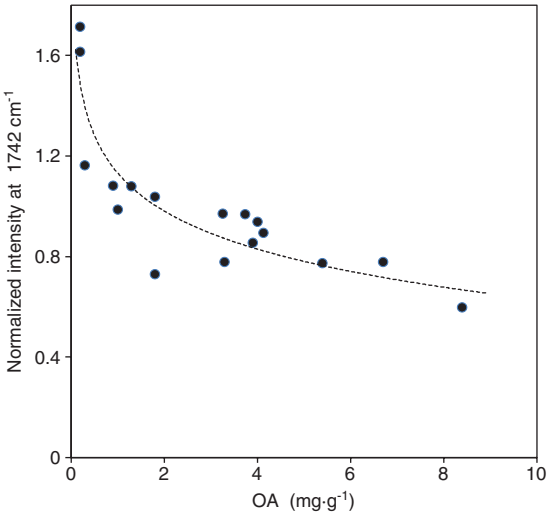

Figure 5: The intensity of band 1742 cm^{-1} as a function of OA, clearly shows the loss of acetyl groups with increasing OA content.

affinity of the remaining hemicelluloses binding to each other and to the cellulose microfibrils as shown by Kabel et al. (2007). A second, but potentially more severe effect of the increased acidity is depolymerization of xylan by acid hydrolysis. It is likely that the effects of high acidity (originating from OA), as indicated by band 1742 cm^{-1}, also affects the polysaccharide fraction including cellulose (Lindfors et al. 2008) towards a lower degree of polymerization. Furthermore, the acetyl group of xylan is a precursor for acetic acid. The content acetic acid was not analyzed in this study, but it is well known that degraded *Vasa* oak contains higher levels than recent oak (Almkvist and Persson 2008a). Hence, the acidic condition is further strengthened by deacetylation.

The relatively high correlation of TS to the band at 1376 cm^{-1}, assigned to C-H deformations in cellulose and hemicellulose, (r=0.86), could indicate changes or loss of the polysaccharide fraction as a result of degradation reactions. A similar effect can be attributed to the band at 1242 cm^{-1} (r=0.83), corresponding to the C-O-R vibrations in the acetyl group. The decrease in intensity of this signal in more degraded samples could indicate changes in the linkage between lignin and xylan (Marchessault 1962). As PEG might contribute to the band at 1250 cm^{-1}, such changes cannot be clearly evaluated for the some of the *Vasa* samples (Table 1). However, for the iron impregnated oak samples (without PEG), the intensity of 1242 cm^{-1} was continuously decreased over time along with reduction of TS and increased concentration of OA; thus this band could be an important indicator for degradation.

Carboxylic functions

The bands at 1700 cm^{-1} display large absolute correlation coefficients for TS (1690 cm^{-1} : r=-0.91) indicating an increased number of carboxylic groups formed during degradation (Moore and Owen 2001). The relatively high correlation to the signal at 1591 cm^{-1} (TS; r=-0.66), which is more pronounced in degraded samples, may be explained by the presence of carboxylate groups contributing to the underlying lignin skeleton signal at this position. Both stretching vibrations from carboxylate C=O (Kacurakova et al. 1999) and iron-carboxylate complex in polysaccharides (Zhbankov 1966) generate signals at 1600 cm^{-1}. Xie et al. (2012) reported an increased band intensity at 1600 cm^{-1}, which occurs parallel to loss of acetyl groups in wood veneers treated with Fenton's reagent. By such treatment, oxidation of polysaccharides by hydroxyl radicals is assumed, which elevates the carboxylic or carboxylate group contents.

The band at 1400 cm^{-1} is weak but observable in many of the *Vasa* samples (Figure 2). This signal has no obvious assignment related to hardwood components or PEG. However, Remazeilles et al. (2000, 2004) observed an absorption band at 1400 cm^{-1} in gall ink degraded paper and proposed an iron complex with cellulose oxidation products as a plausible explanation. They also observed increased signal intensity at 1700 cm^{-1} corresponding to the evolution of carboxylic functions as mentioned above. Indeed, in the current study, the intensity of the signals at 1700, 1591 (carboxylic and carboxylate, resp.) and 1400 cm^{-1} are correlated indicating their common association to iron catalysed degradation in the context of polysaccharides, and the assignment of these signals as typical markers for iron catalysed degradation in wood. This relationship is also successfully expressed by the proposed signal ratio between signals related to carboxylic functions and the sum of signals in the region 1370–1270 cm^{-1} (Eq. 1). This ratio (X_{18}) provided a higher absolute correlation to TS than did the individual deconvoluted bands. Therefore, variable X_{18} was also included into the PLS-modeling (see below).

Depolymerized xylan

It is evident that xylan in *Vasa* oak is partly depolymerized and water soluble (Almkvist and Persson 2008a; Almkvist and Persson 2008b), implying that chemical degradation has taken place, mostly after the salvation of the hull. By extracting *Vasa* wood material with water prior to FT-IR acquisition, information on soluble low-molecular species was obtained. Small amounts of freeze-dried extracted material were analyzed and compared with difference spectra of extracted and non-extracted material and to pure de-acetylated hardwood xylan (Figure 6). The extracted material displayed a typical xylan band pattern with additional bands around 1700 cm^{-1}, 1600 cm^{-1} (shoulder) and 1400 cm^{-1}. These bands in the extracted material, as discussed above, strongly indicate that oxidized functions are associated to the xylan backbone and especially to the low-molecular xylan oligomers, which are extractable by water. The dominating absorption at 1045 cm^{-1} and small sharp band at 897 cm^{-1} are assigned to the linkages (C-O-C) between xylose units in xylan (Kacurakova et al. 1999). The relatively low intensity of these bands in the extracts as compared to the reference material is therefore indicative for shorter xylan oligomers. The presence of carboxylic functions as a part of low-molecular xylan oligomers shows that oxidative reactions have taken place, but acid hydrolysis can also play a significant role in xylan degradation.

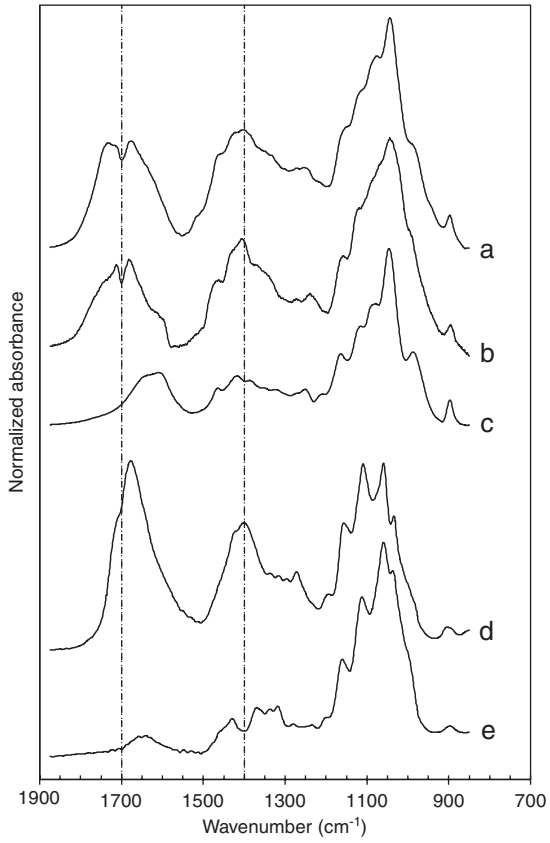

Figure 6: FT-IR spectra of (a) extracted material from *Vasa* oak (V1_C), (b) difference spectrum of extracted and non-extracted material (both from V4a_B), and (c) hardwood xylan reference.
In addition, FT-IR spectra of historic gall-ink paper (Rouchon, personal collection, reference B28-10-3A) from (d) degraded part and (e) non-degraded part, reproduced by courtesy of Remazeilles et al. (2000).

Oxalic acid (OA)

OA was present to different extent in the investigated samples (Table 1). Recent oak was impregnated with 0.1 M OA/oxalate solutions at pH 2 (typical for *Vasa* oak) and then dried before FT-IR analysis. The achieved content of OA was approximately 10 mg·g^{-1} but no significant bands related to OA could be revealed (not shown). By adding excess of OA/oxalate to both recent oak and iron containing *Vasa* oak, significant intensity increment at 1740, 1650, 1245 and 723 cm^{-1} was visible (not shown). For the investigated samples all these signals are less pronounced, even in samples with higher content of OA. Thus OA itself does not provide any significant spectroscopic bands, probably due to its relative low concentration compared to the dominating wood polymers. However, the effects from OA as an inducer of acid hydrolysis or the degradative conditions releasing OA, are indirectly indicated by IR bands, especially those of the acetyl group (Figure 5).

Multivariate models by PLS-rdCV

Data in the context of PLS-rdCV modeling in terms of TS are compiled in Table 3 including different approaches of variable selection. The modeling results are summarized in Table 4. In general, for a successful predictive model, the SEP-values must be considerably smaller than the StD of *y*; otherwise the predictive model is useless. Figure 7 illustrates examples of the outcome with different sets

Table 4: Summary of the result from rdCV-PLS modeling; number of variables (m), standard error of prediction (SEP), Pearson correlation coefficient (r) between experimental and predicted TS, and optimum number of PLS components estimated by rdCV (A).

	m	SEP	*r*	*A*
Deconvoluted bands, 1800–1200 cm^{-1}				
All bands/variables[a]	18	8.1	0.938	2
Variables, selection stepwise	3	6.3	0.962	3
Variables, selection by sequential replacement	6	4.3	0.983	6
Single variable with highest correlation (x_{18}-ratio)	1	8.3	0.966	–
Whole spectra, 1800–1200 cm^{-1}				
Normalized	600	10.5	0.892	1
Normalized with variable selection (highest correlation)	50	8.9	0.923	1
1st derivative	600	9.1	0.920	1
1st derivative with variable selection (highest correlation)	50	8.8	0.924	1
2nd derivative	600	10.4	0.894	1
2nd derivative with variable selection (highest correlation)	50	8.4	0.932	1
SNV transformation	600	8.4	0.931	2
SNV transformation with variable selection (highest corr.)	50	7.7	0.942	1

[a]Including variable ratio defined in eq. 1.
For a useful prediction model, SEP must be considerably smaller than the standard deviation of experimental TS (23.7).

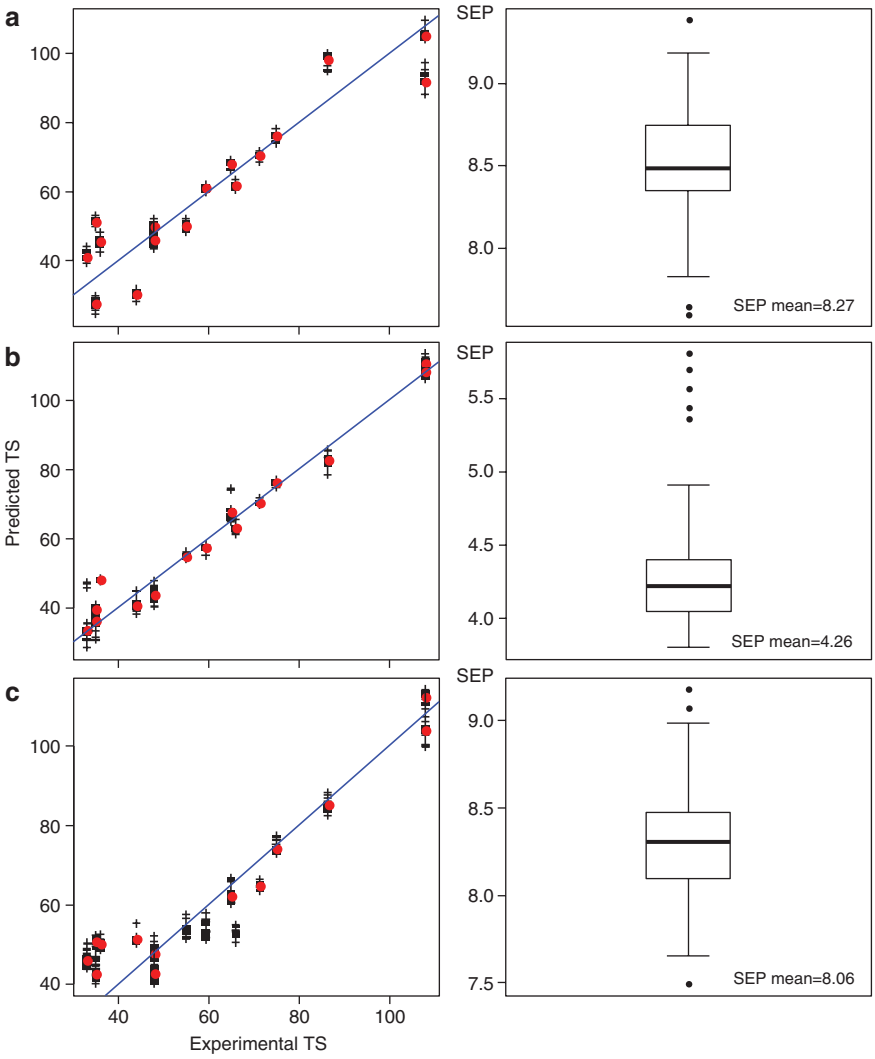

Figure 7: PLS-rdCV regression models for prediction of TS using the data set with 18 variables (17 selected bands and the ratio defined in Eq. 1); (a) one single *X*-variable (X_{18} – ratio) possessing the highest absolute correlation coefficient with TS, (b) the best subset of x-variables among the 18 variables as found by the sequential replacement selection method (containing the bands 1742, 1624, 1400, 1226, 1202, and X_{18} (ratio), and (c) all 18 *X*-variables. Symbol '+' (in black) for 50 repetitions in rdCV with different random splits into the cross validation segments; symbol 'o' (in red) for the mean.
The (blue) line is the 45° line. The boxplots show the distribution of the 50 SEP values obtained in the rdCV repetitions.

of *X*-variables from the data of deconvoluted spectra. At the left hand side the predicted TS is plotted versus the experimental TS. Here, the predicted *y* for the 50 repetitions are plotted in black, and their mean in red; furthermore, the 45° line is also given. At the right hand side, the distributions of the 50 SEP values are plotted as boxplots. It is concluded that the median of SEP for TS is lowest (4.3) for a set of *X*-variables selected by sequential replacement (Figure 7b), which is ca. 18% of the StD of the experimental TS values (23.7), demonstrating that the model gives reasonable (at least semi-quantitative) predictions. The best 95% confidence interval for predicted TS is estimated as ±8.6 (±2 SEP).

The deconvoluted IR-bands provided an advantageous outcome for both exploration and prediction compared to the use of whole spectra. The four data types of the whole spectra (normalized, 1st and 2nd derivatives and SNV transformed) gave similar SEP with no clear sequence of data type with respect to prediction performance (Table 4). The latent information of individual vibration signals is obviously not unmasked in the full spectra due to inherent scattering or bias among the spectra. The fact is that, even though the preparation and acquisition were made with care, the IR analyses were performed at several independent occasions, which may have contributed to such effects. This seems to be,

at least partly, out-ruled by the use of deconvoluted data. PLS-rdCV modeling applied on OA was tested as well. Preliminary results revealed that SEP for OA was higher than the StD of the experimental values and thus a prediction model for OA could not be established (not shown).

Conclusions

The state of degradation in iron contaminated *Vasa* oak was studied by means of FT-IR. Overall, degradation in terms of broken chemical bonds, evolution of oxidized functions, loss of material or structural alteration were reflected by the spectroscopic data. The correlations between IR-bands and changes in tensile strength and content of OA, were most clearly demonstrated by the evolution of oxidized functions and deacetylation in the samples (at 1700 cm^{-1} and 1740 cm^{-1}, resp.). Thus, both oxidative reactions and acid hydrolysis seem to be important for the degradative processes taken place in the *Vasa* oak. In particular, the hemicelluloses fraction was affected by these reactions. The degradation of hemicelluloses is obviously in close correlation to the mechanical performance of the wood upon ageing. Since only a limited number of samples were available and the morphological variation of the included samples was not known, this work shows the principal applicability of multivariate modeling on mechanical properties from IR data. Nevertheless, the prediction of axial TS by PLS-rdCV on FT-IR data seems to be promising. Complementary studies should be performed to investigate possible correlations between chemical degradation, as indicated by IR spectroscopy, and other mechanical properties that are important for the preservation of the *Vasa* (e.g. creep behaviour and mechanical strength in other directions).

Acknowledgments: Financial support from the National Maritime Museums and Swedish National Heritage Board (Grant/Award Number: '3.2.2-3411-2012') is gratefully acknowledged. We thank Dr. Gulaim Seisenbaeva for technical support with the FT-IR instrument. We thank Peter Filzmoser (TU Vienna) for support in statistics. Dr. Celine Remazeilles, Dr. Veronique Rouchon and Dr. Jacky Bernard, are acknowledged for their contributions to Figure 6. We acknowledge Anders Reimann and Eva-Lisa Lindfors (Innventia AB, Sweden), for providing reference material used in this study.

References

Almkvist, G., Persson, I. (2008a) Analysis of acids and degradation products related to iron and sulfur in the Swedish warship *Vasa*. Holzforschung 62:694–702.

Almkvist, G., Persson, I. (2008b) Degradation of polyethylene glycol and hemicellulose in the *Vasa*. Holzforschung 62:64–70.

Almkvist, G., Persson, I. (2011) Distribution of iron and sulfur and their speciation in relation to degradation processes in wood from the Swedish warship *Vasa*. New J. Chem. 35:1491–1502.

Barnes, R.J., Dhanoa, M.S., Lister, S.J. (1989) Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. Appl. Spectrosc. 43:772–777.

Bjurhager, I., Halonen, H., Lindfors, E.L., Iversen, T., Almkvist, G., Gamstedt, E.K., Berglund, L.A. (2012) State of degradation in archeological oak from the 17th century *Vasa* ship: Substantial strength loss correlates with reduction in (holo)cellulose molecular weight. Biomacromolecules 13:2521–2527.

Emery, J.A., Schroder, H.A. (1974) Iron-catalyzed oxidation of wood carbohydrates. Wood Sci. Technol. 8:123–137.

Faix, O. (1991) Classification of lignins from different botanical origins by FT-IR spectroscopy. Holzforschung 45:21–27.

Faix, O., Böttcher, J.H. (1992) The influence of particle-size and concentration in transmission and diffuse reflectance spectroscopy of wood. Holz. Roh. Werkst. 50:221–226.

Fengel, D. (1991) Possibilities and limits of FTIR spectroscopy for the characterization of cellulose. 2. Comparison of various pulps. Papier 45:97–102.

Filzmoser, P., Liebmann, B., Varmuza, K. (2009) Repeated double cross validation. J. Chemometr. 23:160–171.

Håfors, B. Conservation of the Swedish Warship *Vasa* from 1628. The Vasa Museum, Stockholm, 2001.

Kabel, M.A., van den Borne, H., Vincken, J.P., Voragen, A.G.J., Schols, H.A. (2007) Structural differences of xylans affect their interaction with cellulose. Carbohyd. Polym. 69:94–105.

Kacurakova, M., Wellner, N., Ebringerova, A., Hromadkova, Z., Wilson, R.H., Belton, P.S. (1999) Characterisation of xylan-type polysaccharides and associated cell wall components by FT-IR and FT-Raman spectroscopies. Food Hydrocolloid. 13:35–41.

Lindfors, E.-L., Iversen, T., Lindström, M. (2008) Polysaccharide degradation in waterlogged oak wood from the ancient warship *Vasa*. Holzforschung 62:57–63.

Marchessault, R.H. (1962) Applications of infrared spectroscopy to the study of wood polysaccharides. Spectrochim. Acta 18:876–876.

Moore, A.K., Owen, N.L. (2001) Infrared spectroscopic studies of solid wood. Appl. Spectrosc. Rev. 36:65–86.

Norbakhsh, S., Bjurhager, I., Almkvist, G. (2013) Mimicking of the strength loss in the *Vasa*: model experiments with iron-impregnated recent oak. Holzforschung 67:707–714.

Norbakhsh, S., Bjurhager, I., Almkvist, G. (2014) Impact of iron(II) and oxygen on degradation of oak – modeling of the *Vasa* wood. Holzforschung 68:649–655.

Pappas, C., Rodis, P., Tarantilis, P.A., Polissiou, M. (1999) Prediction of the pH in wood by diffuse reflectance infrared Fourier transform spectroscopy. Appl. Spectrosc. 53:805–809.

Pauli, G.F., Jaki, B.U., Lankin, D.C. (2005) Quantitative H-1 NMR: Development and potential of a method for natural products analysis. J.Nat. Prod. 68:133–149.

R-project. R: a language and environment for statistical computing (2014). R Development Core Team, Foundation for Statistical Computing, www.r-project.org, Vienna, Austria.

Remazeilles, C., Quillet, V., Bernard, J. (2000) FTIR techniques applied to iron gall inked damaged paper. In: Proceedings of the 15th WCNDT, Roma, http://www.ndt.net/article/wcndt00/papers/idn323/idn323.htm.

Remazeilles, C., Rouchon-Quillet, V., Bernard, J. (2004) Influence of gum arabic on iron gall ink corrosion part I: a laboratory samples study. Restaur. 25:220–232.

Schwanninger, M., Rodrigues, J.C., Pereira, H., Hinterstoisser, B. (2004). Effects of short-time vibratory ball milling on the shape of FT-IR spectra of wood and cellulose. Vib. Spectrosc. 36:23–40.

Varmuza, K., Filzmoser, P. Introduction to Multivariate Statistical Analysis in Chemometrics. CRC Press, Boca Raton, FL, USA, 2009.

Varmuza, K., Filzmoser, P. (2014) Repeated double cross validation (rdCV) – a strategy for optimizing empirical multivariate models, and for comparing their prediction performances. In: Current Applications of Chemometrics. Ed. Khanmohammadi, M., Nova Science Publishers, Hauppauge, NY, USA. pp. 15–32.

Varmuza, K., Filzmoser, P., Dehmer, M. (2013) Multivariate linear QSPR/QSAR models: rigorous evaluation of variable selection for PLS. Comput. Struc. Biotech. J. 5:1–10.

Wojdyr, M. (2010) Fityk: a general-purpose peak fitting program. J. Appl. Crystallogr. 43:1126–1128.

Wold, S., Sjostrom, M., Eriksson, L. (2001) PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. 58:109–130.

Xie, Y.J., Klarhofer, L., Mai, C. (2012) Degradation of wood veneers by Fenton reagents: effects of 2,3-dihydroxybenzoic acid on mineralization of wood. Polym. Degrad. Stabil. 97:1270–1277.

Zhbankov, R.G. Chapter V. (1966) Oxidation products of cellulose. Salts of oxidation products of cellulose. In: Infrared Spectra of Cellulose and Its Derivatives. Consultants Bureau, New York.