An in situ Raman spectroscopic method for quantification of polyethylene glycol (PEG) in waterlogged archaeological wood

Abstract: The weakened microstructure of archaeological wood (AW) objects from waterlogged environments necessitates consolidation to avoid anisotropic shrinkage upon drying. Polymer impregnation through submergence or spraying treatments is commonly applied, and for larger and thicker objects, the impregnation period can stretch over decades. Thus, for efficient treatment, continuous monitoring of the impregnation status is required. Today, such monitoring is often destructive and expensive, requiring segments for extraction and chromatographic quantification. This study proposes an in situ Raman spectroscopic method for quantification of polyethylene glycol (PEG) in waterlogged AW. A calibration model was built on standards of PEG, cellulose powder, and milled wood lignin using orthogonal partial least squares (OPLS). The OPLS model had a strong linear relationship, and the PEG content in wood of varying degrees of degradation could be determined. However, the accuracy of the model was low with a root mean square error of prediction of 11 wt%. The low accuracy was traced to the heterogeneity in the calibration and validation set samples with regard to the small probing volume of the confocal instrumental setup.

Keywords: Multivariate calibration, Norway spruce (Picea abies (L.) H. Karst), OPLS, PEG, Raman spectroscopy, waterlogged archaeological wood

Introduction

Archaeological artifacts of wood constitute a large and important part of our historical records. Artifacts ranging from great warships to delicately carved tools offer insights into the societies that crafted them and help us decipher the past. When these are excavated and incorporated into our collective heritage, a process of conservation often has to take place to secure the physical and aesthetic integrity of the object.

When exposed in waterlogged, near-anaerobic environments such as lake, sea, sediment, or peatland, wood is solely degraded by specialized anaerobic bacteria. The speed of decay is extremely slow; therefore, wood can be found seemingly well-preserved after centuries or even millennia of exposure in nature (Björdal 2012). The bacterial decay does, however, cause a substantial mechanical weakening of the material. As the weakened material would collapse upon drying, waterlogged archaeological wood (AW) is commonly impregnated with a consolidator such as the water-soluble polymer polyethylene glycol (PEG; Mortensen et al. 2007).

In larger conservation projects, such as the conservation of the warship Vasa (Håfors 2010) and the Bremen Cog (Hoffman 1989), monitoring of the PEG impregnation status is carried out regularly. This is necessary as the impregnation process often stretches over decades. The impregnation process is slow, as it is driven by passive diffusion and the wood matrix does not accommodate fast diffusion. Without knowledge of the accumulation of PEG in the wood, it is impossible to determine when the process is completed, i.e. when the weakened areas in the wood are consolidated to an adequate level. In the two cases mentioned above, cores were regularly taken from the ships’ wood and the PEG was extracted and quantified using spectrophotometry or chromatography, respectively (Hoffmann and Jones 1989).

If an analytical method performed directly on the material were to be developed instead of techniques dependent on extraction, the destructiveness of PEG quantification, and the consequent damage to cultural heritage objects, could be minimized. A Raman spectroscopic
quantification method with measurements done directly on the material would allow for an inexpensive, simple, and nondestructive analysis (Gierlinger and Schwaninger 2007), which in turn enables a more detailed study of the impregnation procedure.

The attempts of in situ, i.e. nonextractive, vibrational spectroscopic quantification of unlabeled polymers in wood are few. Estimation schemes using Raman spectroscopy were developed for PEG (Jeremic et al. 2006) and melamine-formaldehyde (Gierlinger et al. 2005), where the ratio between wood compounds and the polymer bands was found to be proportional to the polymer content. The proposed method is simple to apply but is developed for microanalysis and produces nonabsolute values that are difficult and risky to base conservation decisions on. A similar study that also used Raman band ratios, but on PEG-impregnated AW, created a calibration based on known mixtures of fresh oak and PEG (Christensen et al. 2006). The technique is of interest due to its ease of application, but few experimental details or validations were provided. A quantification scheme for melamine-formaldehyde in wood was also developed using ultraviolet microscopy and Lamberts-Beers law for quantification (Gindl et al. 2003), but as PEG has a limited ultraviolet resonance (Li et al. 2008), the method is unsuitable.

A further drawback of these studies is the assumption that the content of the wood component used as an internal standard is constant. This assumption does not hold true for most AW, as one of the major constituents of wood, cellulose, may, as a consequence of microbial decay, vary as much as to make up one third of the dry weight to approach almost complete depletion (Rowell et al. 2013; Pedersen et al. 2015).

In this paper, we demonstrate the first steps toward a nondestructive method for PEG quantification in AW, where Raman spectroscopic measurements were conducted directly on the impregnated wood without extensive sample preparation. To handle the heterogeneity of AW chemistry, the method proposed herein was based on a mixture design of cellulose, lignin, and aqueous PEG (4000 g mol⁻¹). A calibration model was built on Raman spectroscopic measurements of the mixture design factors. Multivariate statistical algorithms, principal component (PC) analysis (PCA; Wold et al. 1987), and orthogonal partial least squares (OPLS; Trygg and Wold 2002) were used for the qualitative and quantitative interpretation of data. The calibration model was validated on waterlogged AW samples of Norway spruce (Picea abies (L.) H. Karst) and on sound recent sapwood also of Norway spruce. Both AW and recent wood (RW) samples were impregnated with PEG. The aim was to construct a calibration model capable of predicting PEG content in impregnated AW.

Materials and methods

Preparation of milled wood lignin (MWL)

Lignin was represented by MWL prepared in-house, as developed by Björkman. To extract lignin, a fine extractive-free wood meal was prepared. The extraction of extractives was carried out according to the standard procedure (Holmbom 1999). Recent sawn sapwood of Norway spruce was milled in a Retsch rotor mill to mesh 40. The wood meal (8 g) was placed in a Whatman 603 cellulose thimble (VWR, Göteborg, Sweden) and extracted with 200 ml GC-grade acetone (Fisher Scientific, Göteborg, Sweden) in a Soxhlet set-up for a minimum of 24 circulations. Heating was performed with a water bath at 65°C to 70°C. The Soxhlet set-up was only partially covered with tin foil to avoid hot extraction, as lignin is thermally instable (Watkins et al. 2015).

The average amount of extractives, based on three extractions, amounted to 0.8 ± 0.2 wt%, which is in the same range as what has previously been reported for Norway spruce (Holmbom 1999; Nisula 2018). The first three batches were extracted for another 24 circulations, but only an additional 0.1 wt% decrease was noted; thus, the remaining 10 batches were extracted according to the standard procedure.

MWL was prepared according to Björkman’s procedure, apart from drying of the wood meal with phosphorus pentoxide and milling of the wood meal in toluene. As these procedures could not be accommodated (the drying was performed with silica gel, instead), a slightly coarser and more oxidized product was expected. The crude product was dissolved in acetic acid with 10% of water and then precipitated by dropwise addition to water. The precipitate was collected through centrifugation and washed with water until the pH was neutral. The recovered lignin with a yield of 30% was consistent with descriptions in the literature (Björkman 1956). Raman spectroscopic analysis of the acquired lignin showed a spectrum devoid of peaks not associated with lignin. As reported previously (Agarwal and Ralph 1997), the major peaks of MWL were similar in shape to lignin in wood.

Mixture design

To represent holocellulose (cellulose and hemicellulose), cellulose powder from cotton linters (Sigma-Aldrich, Stockholm, Sweden) was used. Hemicellulose and cellulose have very similar chemical bonds and thus similar Raman scattering properties (Agarwal and Ralph 1997; Gierlinger et al. 2012). A preliminary Raman spectroscopic analysis of cellulose powder suggested a pure product. The calibration standards were mixed according to a simplex lattice mixture design (Eriksson et al. 1998; Figure I; Supplementary Table S1) using varying proportions of cellulose powder, MWL, and PEG-4000 (3500–4500 Da; synthesis grade; Merck, Solna, Sweden). The design was nonconstrained, meaning that all concentrations ranged from 0 to 100 wt%.

A seven-level simplex lattice design with 22 points was calculated in JMP® software version 13 (SAS Institute, Inc., Cary, NC, USA). To evaluate the precision of the measurements, the center and end points of the design were triplicated.
The mixtures were prepared by grinding of MWL and cellulose powder in a porcelain mortar, adding the homogenized mixture to a vial, and then adding an aqueous PEG solution. The amount of water was so that two thirds of the total weight of the suspension was composed of water; this ratio was similar to waterlogged softwood and moderately degraded waterlogged AW (Skaar 1988). After thorough stirring, the content of the vials was transferred to microscope slides and analyzed using the Raman system.

Validation set

The calibration model was evaluated on a validation set of PEG impregnated wood samples. The validation set was the sample material of AW and recent sound wood (RW), both of Norway spruce. The RW came from the same sapwood block that was used to isolate MWL. From this dry sample, a subsample of $1 \times 3 \times 5$ cm was submerged in an aqueous PEG-4000 solution of 20 wt% for 3 months.

The AW sample consisted of a circular pole, excavated by archaeologists in Motala Ströms (a river in Sweden), in 2010. The pole was found together with other poles driven down into the riverbed, and they are believed to have been a part of a fishing station. The pole was from a young tree of approximately 15 cm in diameter. No dendrochronological examination has been performed on this specific pole, but most poles found at the site were dated to the end of the 11th century (von Arbin 2017). The pole was after its recovery submerged in an aqueous PEG-4000 solution of 20 wt% for 3 months.

The AW sample consisted of a circular pole, excavated by archaeologists in Motala Ströms (a river in Sweden), in 2010. The pole was found together with other poles driven down into the riverbed, and they are believed to have been a part of a fishing station. The pole was from a young tree of approximately 15 cm in diameter. No dendrochronological examination has been performed on this specific pole, but most poles found at the site were dated to the end of the 11th century (von Arbin 2017). The pole was after its recovery submerged in an aqueous PEG-4000 solution of 20 wt% for 3 months.

The calibration model was evaluated on a validation set of PEG impregnated wood samples. The validation set was the sample material of AW and recent sound wood (RW), both of Norway spruce. The RW came from the same sapwood block that was used to isolate MWL. From this dry sample, a subsample of $1 \times 3 \times 5$ cm was submerged in an aqueous PEG-4000 solution of 20 wt% for 3 months.

The AW sample consisted of a circular pole, excavated by archaeologists in Motala Ströms (a river in Sweden), in 2010. The pole was found together with other poles driven down into the riverbed, and they are believed to have been a part of a fishing station. The pole was from a young tree of approximately 15 cm in diameter. No dendrochronological examination has been performed on this specific pole, but most poles found at the site were dated to the end of the 11th century (von Arbin 2017). The pole was after its recovery submerged in an aqueous PEG-4000 solution of 20 wt% for 3 months.

Subsamples were taken from AW and RW, and cross-sections were cut by hand using a double-edged razor blade. These were approximately $5 \times 5$ mm with a thickness of approximately 100 μm. As the longitudinal axis is the main transport route in wood, cross-sections were chosen for this study. The samples were placed on Raman-grade CaF₂ microscopy slide (Crystarn Ltd., Poole, United Kingdom) fitted with eight sticky wells (ibidi Gmbh, Planegg, Germany). To avoid drying out of the test material, the wells that were not in use were filled with water. The calibration standards were analyzed in the same set-up.

Raman instrumentation and measurements

The Raman spectrometer Dilor Labram IV (HORIBA Jobin Yvon S.A.S., Palaiseau, France) with a charge-coupled device detector coupled with an inverted confocal microscope ( Olympus IX70, Tokyo, Japan) was used for analysis. The excitation wavelength used was 632.8 nm (He/Ne laser) with laser intensity at approximately 2 mW. Each acquisition was a triplicate of 20 s that were averaged into one spectrum. A 950 groves/mm grating was used to record a spectrum from 200 to 3200 cm⁻¹. An $10 \times$ objective was used to focus, and the theoretical diameter of the focal spot was calculated as 5 μm. Photoluminescence (PL) was addressed by photobleaching all samples for 6 min before spectral acquisition. The preexposure did not result in any degradation of Raman signals in either RW or AW. The untreated spectra of PEG, cellulose powder, MWL, and AW can be found in Supplementary Material.

The AW and RW samples were analyzed at six and five sites, respectively (Figure 1). For both test materials, the sites were selected at different depths to create a spread in PEG concentrations. In the AW samples, the sites were also chosen at different distances to the exterior of the original pole (as the surfaces exposed to the aquatic environment will be more degraded) to get a spread in lignin and hemicellulose content. Three consecutive sections were taken from each subsampling site to overbridge the representative problem of the heterogeneous material. As the laser spot was small compared to the wood micromorphology (~5 μm), 9 randomized acquisitions of each section were performed. A total of 27 acquisitions were thus performed for each subsampling site.

As the standards of the mixture design were assumed to be more homogenous, only 5 acquisitions of each standard were performed. The 5 acquisitions were treated as separate observations and used individually to construct the calibration model.

PEG extraction

To estimate the PEG content in AW and RW, the samples were extracted with methanol and the PEG content was gravimetrically determined. After Raman analysis, the triplicate sections were dried in a desiccator for 1 week and their dry weight was measured. The sections were then placed in a vial and submerged in 2 ml methanol (GC-MS grade; Merck, Solna, Sweden). The methanol was changed daily until no weight change was noted (after 24 h in a desiccator). No weight change was noted after 72 h extraction.

The weight loss upon extraction is also a result of the removal of extractives and, in the case of the archaeological sample, degradation products. To estimate the error from the extraction of other compounds than PEG, the sample material that had not been PEG impregnated was treated with the same extraction scheme. RW and AW were extracted 0.8 and 1.5 wt%, respectively, with a relative standard deviation (SD) of approximately 50% ($n=6$). This is in accordance with both the extractive contents of spruce (Nisula 2018) and the error found between PEG determination in AW by high-performance liquid chromatography and weight loss upon methanol extraction (Hoffman 1989). As the SD was large, the extractive content was not added to the calculation of PEG content; instead, the error of the gravimetric determination was assumed to be equal to the average extracted plus 2 SD. This equals to an error of 1.6 and 3 wt% for RW and AW, respectively.
Spectral preprocessing

To minimize the influence of real and apparent nonlinear signals stemming from the instrument and from the samples (Greperud 1992), the data were preprocessed. Baseline correction was performed using a rolling circle filter (RCF; Brandt et al. 2006) with a circle radius of 98 cm⁻¹. The subtracted background was mainly due to the PL of lignin.

As Raman spectroscopy is sensitive to polymer conformation, ordering and orientation (Schmidt et al. 2018) spectral discrepancies can be expected between the calibration set and the validation set — a suspension of particles and a waterlogged network of fibers — that are not founded in concentration. Thus, scaling was deemed necessary. The centering of the data, where each variable is centered to zero mean, proved to improve the diagnostics of the calibration modeling notably but not that of the validation set. Therefore, the data were also scaled to a variance of 1 (division of variables by SD). The centering and scaling to SD is called unit variance (UV) scaling and typically employed on data of different scales, giving all variables the same chance to influence the model.

Multivariate calibration and model validation

From the measurements of the calibration set, the OPLS models were constructed in SIMCA 15.0.1 (Sartorius Stedim Biotech, Umeå, Sweden). The cumulative goodness-of-fit of both the spectroscopic data (R²X) and the concentration data (R²Y) and the cumulative goodness of prediction (Q²) were used to validate the models. To test the accuracy of the model, the root mean square error (RMSE) of estimation, cross-validation, and prediction were calculated. The RMSE of prediction was calculated for the validation set (Eriksson 2008).

The predictions made on the validation set were also evaluated with linear regression analysis of the observed vs. predicted values. Apart from the coefficient of determination, which is the same as R²Y, the slope and intercept of linear regression were significance tested with a t-test to detect any systematic errors in the method.

Results and discussion

The aim of this study was to predict PEG content based on analytical response. This is typically done by regression analysis, where the relationship between correlated variables (analytical response) and uncorrelated variables (concentration) is estimated. If it is known how the variables covary, it is possible to predict one based on the other. In multivariate analysis performed herein, the correlated variables were held in matrix X (spectral responses), whereas uncorrelated variables were in matrix Y (PEG concentration).

The design used in this study had all the components — cellulose, MWL, and PEG — in proportions, i.e. the sum of components always added up to 100%. The underlying assumption was that the response depends on the relative proportions of the components in the mixture, not on their absolute amounts. The wt% values that the model predicted were the ratio of PEG to lignin and holocellulose; thus, inorganic material and organics not based on lignin and saccharide chemistry were not part of the model.

PCA of the calibration set

The calibration set consisted of spectra measured on mixtures of MWL, cellulose, and PEG according to a mixture design.

As a first means of evaluating the calibration set, the spread in the spectra were analyzed using PCA. As PCA does not force the separation of observations, it can be used as a qualitative validation tool of PLS-based models (Worley and Powers 2016). Based on the five measurements of each of the 30 mixtures (150 observations), a model was constructed with eight components (R² = 0.995 and Q² = 0.993). Of these, bands of PEG (1141, 1280, and 1478 cm⁻¹; Koenig and Angood 1970), lignin (1273, 1600, and 1658 cm⁻¹; Agarwal and Ralph 1997), and cellulose (1094, 1122, and 1376 cm⁻¹; Agarwal and Ralph 1997) were separated into loading vectors of p1, p2, p3, and p4, respectively (Figure 2). The loading vector of p2 was similar to p1 with the bands of PEG but with a smaller magnitude and broader. As loading vectors contain the weight of each variable (wavenumber) to the PC, they indicate which variables are important to describe the variation along that PC. In this case, it means that the variation in intensity of the bands of PEG was important in describing the change along PC1 and that the variation in intensity of the bands of cellulose and lignin were important in describing the change along PC3 and PC4, respectively. The remaining loading vectors, p5–p8, were variations of the first four but with more noise. Their importance for the model (see eigenvalues in Supplementary Material) was not large, but as they helped improve both the goodness-of-fit and prediction, they were kept.

The model showed a large spread between replicate measurements but a small spread between replicated mixtures (Figure 3). This suggests that the replicate mixtures were similar but that the individual measurements had more variation. Consequently, the five replicates were likely too few to give an accurate representation of the contents of the standards.

The largest spread in the t2 vs. t1 space was caused by the standards of 100 wt% PEG (number 5, 14, and 29). Inspection of the individual spectra showed, apart from variation in the intensity of the PEG bands, a strong variation in the CaF₂ band at 320 cm⁻¹ from the microscopy slide. This variation suggests that the placement of the focal plane and thus the focal volume was the source of
this variation. As this variation was unrelated to PEG concentration, the OPLS algorithm should have been sorted into an orthogonal component.

As evident in Figure 4, the observations were separated according to PEG content. This separation, together with the PCA loading vectors in Figure 2, suggests that OPLS regression should find a high degree of correlation, but the large spread in replicative measurements hints at a low accuracy of the regression.

**OPLS of the calibration set**

OPLS is an expansion of PLS that facilitates easier model interpretation by dividing the systematic variation into variation correlated to Y and variation uncorrelated to Y (Trygg and Wold 2002). OPLS is well applied to a spectroscopic based design, as the general spectral variations can be separated from variation correlating with the Y. OPLS does not assume that the factors are uncorrelated (Stenlund et al. 2009), which makes it directly applicable to the data of the mixture design (Eriksson et al. 1998).

OPLS seeks the maximum covariation between two sets of variables (e.g. Raman spectra and concentrations) and attempts to construct a linear relationship (Miller and Miller 2005). To predict the concentration of PEG (Y matrix variables) based on Raman spectra (X matrix variables), the two matrices of the calibration set were subjected to an OPLS regression. As the cellulose and lignin content of the validation set was not quantified in either AW or RW, only the PEG content was included in the Y data.

The most successful OPLS model was created by applying the RCF baseline subtraction and UV scaling. The spectral range 350–3200 cm$^{-1}$ was used for modeling, and 200–350 cm$^{-1}$ was excluded due to the 320 cm$^{-1}$ band of the CaF$_2$ microscopy slide. Spectral artifacts 479–510, 610–670, and 700–790 cm$^{-1}$, which appear to be the result of the variation of where the RCF brings the spectrum down to zero intensity, were present in the ultraviolet-treated spectra. As these artifacts varied systemically with
PL, they were deemed as mostly introducing noise and were removed from the spectra.

The goodness-of-fit (R²ₓ, R²ᵧ) and prediction (Q²) of the model was 0.962, 0.902, and 0.758, respectively, with 1 + 4 components (predictive + orthogonal) calculated. The high value and the high equality of the numbers suggest a model that can describe the variation in the data well.

The loading vectors of the predictive (p) and the first orthogonal (po1) components are seen in Figure 5 in the fingerprint region (800–1700 cm⁻¹). The figure looks like a distorted Raman spectrum, but in fact it presents the weight of each variable (wavenumber) to the component. A high value means that that wavenumber was important in describing the positive variation in p and a low value means a negative correlation. The distorted appearance was due to the UV scaling, which gives higher weight to small variations.

The loading vector of the predictive component p was mainly characterized by the bands of PEG at 844, 1061, 1141, 1230, 1280, and 1478 to 1485 cm⁻¹ (Koenig and Angood 1970) and a broad band at 2670 to 3000 cm⁻¹. The broad band was difficult to interpret: PEG, cellulose, and water all have bands of hydrogen stretching in this area, making the covariation complex.

The loading vector of the first orthogonal component po1 holds the bands of cellulose at 1097, 1122, and 1376 cm⁻¹ (Figure 5; Agarwal and Ralph 1997). The bands of lignin were more difficult to spot, but there were slight indications of bands at 1297 and 1600 cm⁻¹. However, the score vector to1 separates observations according to lignin content (results not shown), which entails that the bands of lignin have been separated into the po1 component. This separation also suggests that lignin and cellulose could be quantified in the same calibration scheme.

The third and fourth orthogonal components were the variations of the first and second but with more noise. These were not further considered in the analysis.

The RMSE of estimation and RMSE of cross-validation of the model were determined as 12 and 16 wt%, respectively. The accuracy thus was not very high, but the R²ᵧ value of 0.902 suggests a strong linear model (Figure 7). Student’s t-test does not indicate any systematic error.

The separation of PEG and the wood constituents into two separate components suggests that the model

![Figure 5: OPLS loading vectors: predictive (p, black) and orthogonal (po1, blue).](image-url)

The bands of PEG (844, 1061, 1141, 1230, 1280, and 1478–1485 cm⁻¹) are held in predictive loading p, whereas the bands of cellulose were found in the orthogonal loading po1 (1097, 1122, and 1376 cm⁻¹).

![Figure 6: OPLS scatterplot of the t1 and to2 score vector.](image-url)

Coloring is according to PEG content (wt/wt).

![Figure 7: Predicted PEG content of the calibration model plotted against the observed content of the calibration standards.](image-url)

The solid line represents perfect correlation, whereas the dashed line is the actual correlation. In this graph, R² is the same as Rᵧ.
would be able to differentiate between the two analytes in the validation set as well. That the first orthogonal component was able to predict the content of cellulose and lignin also appears promising for further method development. As mentioned earlier, the large variation between replicate measurements and consequently their poor representation of the calibration standards was likely one of the reasons for the poor accuracy of the calibration model.

**PCA of the validation set**

The validation set contained spectral responses collected from PEG-impregnated AW and RW samples. PCA was performed on the individual spectra of the samples from AW and RW. The model was constructed with five components with Q² and R²X values of 0.986 and 0.989, respectively. The PCA model had separated observations along a PEG content gradient along t₄ (Figure 8). As separation according to PEG content was present in PCA models in both the calibration set and the validation set, it is likely that an OPLS regression could be (1) constructed based on the calibration set and (2) used to predict the PEG content of the validation set.

The range in PEG content of the two groups of samples of the validation set, RW and AW, was determined by gravimetry to 21 to 46 and 33 to 75 wt%, respectively (Table 1). Individually, these ranges were narrow, but when grouped together, they span over half of the possible scope (21–75 wt%) and also the part of the scope that is of most interest to conservators. To fully test the model, the two groups were tested both individually and grouped together.

**OPLS of the validation set**

The validation set consisted of AW and RW samples, where the PEG content has been determined after Raman analysis. The PEG content predicted by the OPLS model has been plotted against the content determined gravimetrically (Figure 9) and the statistics of the linear regression analysis are presented in Table 2.

The RMSE of prediction was relatively high, but it was not larger than that of the calibration set. The RMSE of prediction was higher for AW, which was likely due to the greater degree of heterogeneity and PL in decayed wood.

| Sample | PEG (wt%) |
|--------|-----------|
| AW1.1  | 32.5      |
| AW1.2  | 74.8      |
| AW1.3  | 43.9      |
| AW2.1  | 38.4      |
| AW2.2  | 65.1      |
| AW2.3  | 66.9      |
| RW1    | 46.1      |
| RW2    | 30.3      |
| RW3    | 29.9      |
| RW4    | 20.8      |
| RW5    | 33.3      |

*AW, Samples from Motala Ström; RW, samples of RW. The calculated error is 1.6 and 3.0 wt% for RW and AW, respectively.*

---

**Table 1: PEG content as determined by weight loss upon PEG extraction with methanol.**

**Figure 8:** PCA score scatterplot of t₃ vs. t₄ of the validation set. Coloring is according to PEG content in wt/wt.

**Figure 9:** Predicted vs. observed plot of the validation set. Triangles represent RW subsamples, and squares represent AW subsamples. The solid line represents perfect correlation, and the dashed line represents least-squares linear regression of AW and RW grouped together.
Table 2: Statistics of the validation set of the linear regression analysis of the predicted vs. observed values.

| Validation set (number of observations) | RMSE of prediction | $R^2Y$ (P value) | Intercept (P value) | Slope (P value) |
|----------------------------------------|--------------------|------------------|--------------------|-----------------|
| AW (6)                                 | 13 wt%             | 0.795            | 35 (0.001)         | 0.52 (0.017)    |
| RW (5)                                 | 8 wt%              | 0.749            | 20 (0.054)         | 0.58 (0.058)    |
| AW + RW (11)                           | 11 wt%             | 0.816            | 18 (0.013)         | 0.77 (0.000)    |

AW, Samples from Motala Ström; RW, samples of RW. Data are presented both for the individual sample groups, AW and RW, and for the two grouped together. P value is the probability value of the t-test at 95% confidence interval.

There was a clear linearity in the data ($R^2Y = 0.8162$); however, all linear regressions showed a tendency to overestimate the PEG content: the intercept was significantly larger than 0, whereas the slope was significantly smaller than 1. This means that, at low PEG content, the model overestimates, but this systematic error was reduced at higher values.

A discrepancy between the standard mixtures and wood, which could be the cause of the overestimation of PEG, may have arisen due to the directionality of the microfibrils in wood. Microfibrils in the cell walls of wood are highly ordered (Rowell et al. 2013), whereas the cellulose fibers in the calibration set likely had something close to a random order. As not all Raman modes of vibration are accessible from all directions, microfibril direction influences the Raman signal (Gierlinger et al. 2005). Thus, if less cellulose was detected in the wood, due to microfibril directionality, the PEG content would be overestimated – a consequence of the mixture design. As the overestimation declines as cellulose becomes less important in describing the system, it is consistent with the above argumentation.

Conclusions

The quantification of PEG content was performed on both AW and RW PEG impregnated wood samples using a calibration of MWL, cellulose powder, and PEG together with Raman spectroscopy and multivariate algorithms. Validation suggested that our method is reliable, and the results showed that our method is a viable first step toward developing an in situ method for PEG quantification in wood. The lack of accuracy appears to have been mainly caused by the heterogeneity of the calibration set mixtures and validation set samples with regard to the size of the laser spot of the Raman instrument.

As the calibration was built on standards of lignin and cellulose, and the OPLS model separated them into separate components, the model could, upon further development, also be used to predict their content. Validation of such predictions was more difficult to produce due to the heterogenous nature of lignin and holocellulose. However, if a validation technique for the quantification of lignin and holocellulose with the proposed method could be developed, a very detailed picture of the state of microbial degradation and that of the impregnation procedure could be gained from one single analysis.

The adaptability of Raman spectroscopy allows for the method to be applied in many different manners. Affordable and easy-to-use handheld systems could be applied, or advanced systems, such as spatially offset Raman spectroscopy, would allow for continuous measurements through thin walls, thus taking another step toward non-destructive analysis in waterlogged AW.

Other substances aside from PEG are employed in the impregnation of RW and AW. If the chemical structure of the consolidation agent is dissimilar enough from cellulose and lignin, it could likely be quantified with a method similar to the one presented here.

Acknowledgments: We thank Prof. Anette Larsson (Chalmers University of Technology) and Prof. Monica Ek (KTH Royal Institute of Technology) for letting us use the rotor and the ball mills. We also thank Jill Ågren (University of Gothenburg) for her assistance with extractions and The Department of Conservation at University of Gothenburg for their generosity in both materials and time.

Author contributions: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

Research funding: This work was supported by Grant-in-Aid for Scientific Research (C) no. 17K06638.

Employment or leadership: None declared.

Honorarium: None declared.

References

Agarwal, U.P., Ralph, S.A. (1997) FT-Raman spectroscopy of wood: identifying contributions of lignin and carbohydrate polymers in the spectrum of black spruce (Picea mariana). Appl. Spectrosc. 51:1648–1655.

Björkman, A. (1956) Studies on finely divided wood. Part 1. Extraction of lignin with neutral solvents. Sven. Papperstidn. 59:477–485.
Brandt, N.N., Brovko, O.O., Chikishev, A.Y., Paraschuk, O.D. (2006) Optimization of the rolling-circle filter for Raman background subtraction. Appl. Spectrosc. 60:288–293.

Christensen, M., Frosch, M., Jensen, P., Schnell, U., Shashoua, Y., Nielsen, O.F. (2006) Waterlogged archaeological wood – chemical changes by conservation and degradation. J. Raman Spectrosc. 37:1171–1178.

Eriksson, L. Design of experiments: principles and applications. Umetrics Academy, Umeå Learnways AB, Stockholm, 2008.

Eriksson, L., Johansson, E., Wikström, C. (1998) Mixture design – design generation, PLS analysis, and model usage. Chemometr. Intell. Lab. 43:1–24.

Gierlinger, N., Schwanninger, M. (2007) The potential of Raman microscopy and Raman imaging in plant research. J. Spectrosc. 21:69–89.

Gierlinger, N., Hansmann, C., Röder, T., Sixta, H., Gindl, W., Wimmer, R. (2005) Comparison of UV and confocal Raman microscopy to measure the melamine-formaldehyde resin content within cell walls of impregnated spruce wood. Holzforschung 59:210–213.

Gierlinger, N., Keplinger, T., Harrington, M. (2012) Imaging of plant cell walls by confocal Raman microscopy. Nat. Protoc. 7:1694–708.

Gindl, W., Zargar-Yaghubi, F., Wimmer, R. (2003) Impregnation of softwood cell walls with melamine-formaldehyde resin. Biotechnol. Technol. 17:325–330.

Greperline, P.J. (1992) Developments in nonlinear multivariate calibration. Chemometr. Intell. Lab. 15:115–126.

Håfors, B. Conservation of the wood of the Swedish warship Vasa of AD 1628. Evaluation of polyethylene glycol conservation programmes. In: Department of Conservation 2010, University of Gothenburg, Gothenburg, Sweden.

Hoffman, P. (1989) HPLC for the analysis of polyethylene glycols (PEG) in wood. In: Conservation of Wet Wood and Metal: Proceedings of the IC0M Conservation Working Groups on Wet Organic Archaeological Materials and Metals. Eds. Grattan, D.W., MacLeod, I.D. Western Australian Museum, Perth. pp. 41–60.

Hoffmann, P., Jones, M.A. (1989) Structure and degradation process for waterlogged archaeological wood. In: Archaeological Wood. Eds. Rowell, R.M., Barbour, R.J. American Chemical Society, Washington. pp. 35–65.

Holmbom, B. (1999) Extractives. In: Analytical Methods in Wood Chemistry, Pulping, and Papermaking. Eds. Sjöström, E., Alén, R. Springer-Verlag, Berlin. pp. 125–148.

Jeremic, D., Cooper, P., Heyd, D. (2006) PEG bulking of wood cell walls as affected by moisture content and nature of solvent. Wood Sci. Technol. 41:597.

Koenig, J.L., Angood, A.C. (1970) Raman spectra of poly(ethylene glycols) in solution. J. Polym. Sci. Polym. Phys. 8:1787–1796.

Li, N., Ziegemeier, D., Bass, L., Wang, W. (2008) Quantitation of free polyethylene glycol in PEGylated protein conjugate by size exclusion HPLC with refractive index (RI) detection. J. Pharm. Biomed. Anal. 48:1332–1338.

Miller, J.N., Miller, J.C. Statistics and chemometrics for analytical chemistry. Pearson Education Ltd., Harlow, England, 2005.

Mortensen, M.N., Egsgaard, H., Hvilstø, S., Shashoua, Y., Glastrup, J. (2007) Characterisation of the polyethylene glycol impregnation of the Swedish warship Vasa and one of the Danish Skuldelev Viking ships. J. Archaeol. Sci. 34:1211–1218.

Nisula, L. Wood extractives in conifers: a study of stemwood and knots of industrial species. Åbo Akademi University, Åbo, Finland, 2018.

Pedersen, N., Gierlinger, N., Thygesen, L. (2015) Bacterial and abiotic decay in waterlogged archaeological Picea abies (L.) Karst studied by confocal Raman imaging and ATR-FTIR spectroscopy. Holzforschung 69:103–112.

Rowell, R., Petersson, R., Tshabalala, M.A. (2013) Cell wall chemistry. In: Handbook of Wood Chemistry and Wood Composites. Ed. Rowell, R. Boca Raton. pp. 33–72.

Schmidt, U., Müller, J., Koenen, J. (2018) Confocal Raman imaging of polymeric materials. In: Confocal Raman Microscopy. Eds. Toporski, J.,Dieing, T., Holricher, O. Springer International Publishing, Cham. pp. 485–507.

Skaar, C. Wood-Water Relations. Springer-Verlag, Berlin; New York, 1988.

Stenlund, H., Johansson, E., Gottfries, J., Trygg, J. (2009) Unlocking interpretation in near infrared multivariate calibrations by orthogonal partial least squares. Anal. Chem. 81:203–209.

Trygg, J., Wold, S. (2002) Orthogonal projections to latent structures (O-PLS). J. Chemom. 16:119–128.

von Arbin, S. Fiskeanläggningar i Motala ström: Arkeologi Motala. Rapport 2017:120. Arkeologisk undersökning, Östergötlands län, Östergötland, Motala kommun, Motala socken, Kanaljorden 3:45, Motala 188, 2017.

Watkins, D., Nuruddin, M., Hosur, M., Tcherbi-Narteh, A., Jeelani, S. (2015) Extraction and characterization of lignin from different biomass resources. J. Mater. Res. Technol. 4:26–32.

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

Worley, B., Powers, R. (2016) PCA as a practical indicator of OPLS-DA model reliability. Curr. Metabolomics 4:97–103.

**Supplementary Material:** The online version of this article offers supplementary material (https://doi.org/10.1515/hf-2019-0238).