

# Multivariate Correlation between Analysis Data on Dissolved Organic Material from Scots Pine (*Pinus sylvestris*) Chips and their Autohydrolysis Pre-Treatment Conditions

Joni Lehto,<sup>a,\*</sup> Raimo Alén,<sup>a</sup> and Petteri Malkavaara<sup>b</sup>

Various chemometric techniques were used to establish the relationship between the autohydrolysis conditions prior to pulping and the chemical compositions of the soluble organic materials removed from Scots pine (*Pinus sylvestris*) wood chips. The aqueous chip pre-treatments (autohydrolysis) were administered at 130 °C and 150 °C for 30, 60, 90, and 120 min, and the hydrolysates obtained were characterized in terms of total carbohydrates (various mono-, oligo-, and polysaccharides together with uronic acid side groups), volatile acids (acetic and formic acids), lignin, and furans (furfural and 5-(hydroxymethyl)furfural). Based on the analytical data gathered, a relatively accurate model for pine chip autohydrolysis was developed.

*Keywords:* Autohydrolysis; Scots pine; Carbohydrates; Volatile acids; Lignin; Furans; Biorefining; Principal component analysis

*Contact information:* a: Laboratory of Applied Chemistry, Department of Chemistry, University of Jyväskylä, P.O. Box 35, FI-40014 University of Jyväskylä, Finland; b: GCP Research Services Ltd, Karkutie 8 B 4, FI-40900 Jyväskylä, Finland; \*Corresponding author: joni.t.lehto@jyu.fi

## INTRODUCTION

Among the many recent biorefinery concepts, one of the most promising is the method for fractionating wood feedstocks that involves the hot-water pre-treatment (autohydrolysis) of wood chips under pressure prior to delignification (Sixta and Schild 2009; Sixta 2006). Autohydrolysis is of special interest because water is the only reagent, making this approach an environmentally friendly and inexpensive process (Garrote *et al.* 1999; Ramos *et al.* 2002; Teo *et al.* 2010). Additionally, autohydrolysis causes no corrosion problems and generally has only a minor negative influence on the strength properties of pulp. The combined overall effect of autohydrolysis time and temperature can be represented by a single numerical value, the so called “P-factor” (pre-hydrolysis-factor), which is comparable to the “H-factor” commonly used in pulping for similar purposes (Sixta 2006; Tunc and van Heiningen 2009). In this pre-treatment process, wood hemicelluloses are partly dissolved and carbohydrate-containing hydrolysates are produced (Sears *et al.* 1971; Yoon *et al.* 2008; Paredes *et al.* 2008; Amidon and Liu 2009; Li *et al.* 2010; Alén 2011). In addition to dissolved carbohydrates, the hydrolysates contain various amounts of other organic components (Tunc and van Heiningen 2011), such as aliphatic carboxylic acids (“volatile acids”, *e.g.*, acetic and formic acids), furans (*e.g.*, 2-furaldehyde or furfural and 5-(hydroxymethyl)furfural (HMF)), and heterogeneous fractions of lignin- and extractive-derived materials.

Several multivariate analysis techniques, such as principal component analysis (PCA), principal component regression (PCR), and projection to latent structures (PLS), have proven to be useful tools for the evaluation of spectral and chemical data obtained from different wood fractionation processes (Hyötyläinen *et al.* 1998; Malkavaara and Alén 1998; Malkavaara *et al.* 2000; Schultz *et al.* 1985; Schultz and Burns 1990). For example, these methods have been applied to the prediction of lignin content and the composition of carbohydrates in wood samples, as well as to the prediction of Klason lignin, xylose, and glucose in pulps.

The main aim of this study was to investigate, by a chemometric approach, the relationships between the autohydrolysis conditions applied and the chemical compositions of the organic materials removed from Scots pine (*Pinus sylvestris*) wood chips by autohydrolysis. For this purpose, the hydrolysates obtained under varying conditions were analyzed in detail with respect to their main chemical constituents.

## EXPERIMENTAL

### Autohydrolyses

Laboratory-scale autohydrolysis experiments on screened (maximum thickness 7 mm, maximum width 13 mm, and minimum width 7 mm) industrial chips from Scots pine (*Pinus sylvestris*) and silver birch (*Betula pendula*) were carried out in stainless steel autoclaves set in oil baths. The chips were heated at two temperatures (130 °C and 150 °C) and for four treatment times (30, 60, 90, and 120 min). The liquid-to-wood ratio was 5 L/kg. In each case, a heating period of 30 min was added to these times. The treatments covered the P-factor (PF) range from 10 to 238.

At the end of each treatment, the autoclave was removed from the oil bath and cooled rapidly in cold tap water. The hydrolysate was then separated from the treated chips *via* filtration, and its pH was immediately measured using an Orion Research 410 A pH-meter.

### Analytical Determinations

The total carbohydrate (TC) and uronic acid contents were determined *via* acid methanolysis (Sundberg *et al.* 1996; Bertaud *et al.* 2002) and various instruments: an Agilent 6890 Series gas chromatography device equipped with an HP-5 analytical column (30 m x 0.32 mm I.D. with a film thickness of 0.25 µm) and a flame-ionization detector (GC-FID, operated at 290 °C). The column temperature program consisted of 2 min at 100 °C, 2.5 °C/min to 190 °C, 12 °C/min to 290 °C, and 5 min at 290 °C.

After the dilution of one portion of the hydrolysate with ultra-high quality (UHQ) water until the absorbance (A) was in the 0.3 to 0.8 range, the dissolved total lignin (TL) was determined using a Beckman DU 640 UV/Vis spectrophotometer at 205 nm. The absorptivity values used for the pine and birch lignins were 120 L/(gcm) and 110 L/(gcm), respectively (Swan 1965). The volatile acids (acetic and formic acids, TA) were determined, as described previously (Käkölä *et al.* 2008), using a Dionex chromatography system. The furanoic compounds (2-furfural and HMF, TF) were determined according to methods developed earlier (Lehto and Alén 2012), using Waters HPLC equipment.

## Data Analysis

The analytical data were subjected to principal component analysis (PCA) and projection to latent structures (PLS) regression calculations using the non-linear iterative partial least squares (NIPALS) algorithm (Hill and Lewicki 2006). Significant ranks of the models were determined by means of cross-validation. Each value of the analytical data was the mean of two replicate determinations (see Section 3.2, samples a and b).

The following data pretreatment procedures were applied: mean-centering and scaling to unit variance. In the mean-centering, a column mean was subtracted from each data point in the matrix, whereas the scaling to unit variance was established by dividing each mean-centered data point by column standard deviation.

The mean of each variable in the mean-centered data was zero, and the method adjusted for differences in the offset between high and low values. It was therefore used to focus on the fluctuating part of the data, and left only the relevant variation for analysis. In case of scaling to unit variance, the mean of each variable was zero, and standard deviation was one, and therefore the data were analyzed on the basis of correlations instead of covariances, as is the case with the mean-centering (van der Berg *et al.* 2006).

The PLS regression models for the P-factor were calculated using the mean-centered analytical data scaled to unit variance, while the PLS models for dissolved solids (DS) were calculated using the mean-centered analytical data as the X matrix. For the PLS models, the X matrix consisted of variables TC, TA, TL, and TF, and the Y vector was the modeled parameter, P-factor, or DS.

All computations were carried out on a personal computer using the Unscrambler® X software package (Unscrambler® 2011).

## RESULTS AND DISCUSSION

### Principal Component Analysis

The analytical data are presented in Table 1. The PCA model based on these data clearly demonstrated the differences amongst the samples, both with respect to the hydrolysis treatment parameters (*e.g.*, the incremental increases in hydrolysis time/temperature) and the wood species (Fig. 1).

For this reason, the samples formed two distinct groups. The loading values of the variables are presented in Fig. 2. As expected, the TC, DS, and P-factor were the most influential variables in the sample grouping.

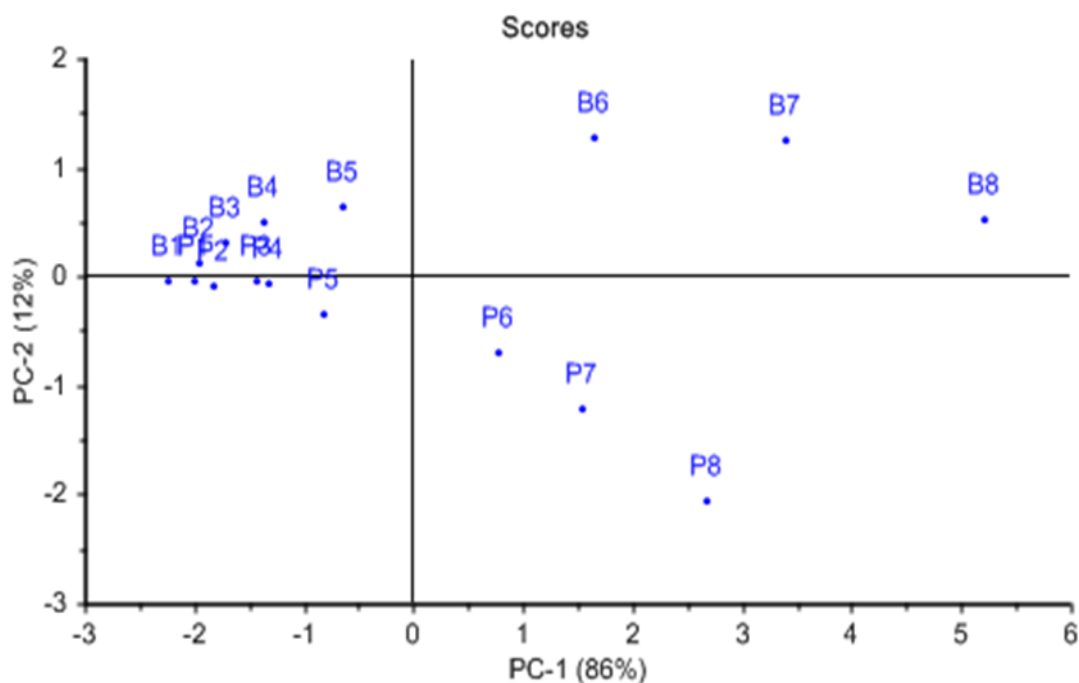
The PCA model, using only the data on the pine samples, was also calculated. In this model (Figs. 3 and 4), three phases of the autohydrolysis process could be distinguished, corresponding to samples P1 through P2, P3 through P5, and P6 through P8. In addition, the heterogeneity within each group increased as the P-factor value increased and was mainly due to TL and TF.

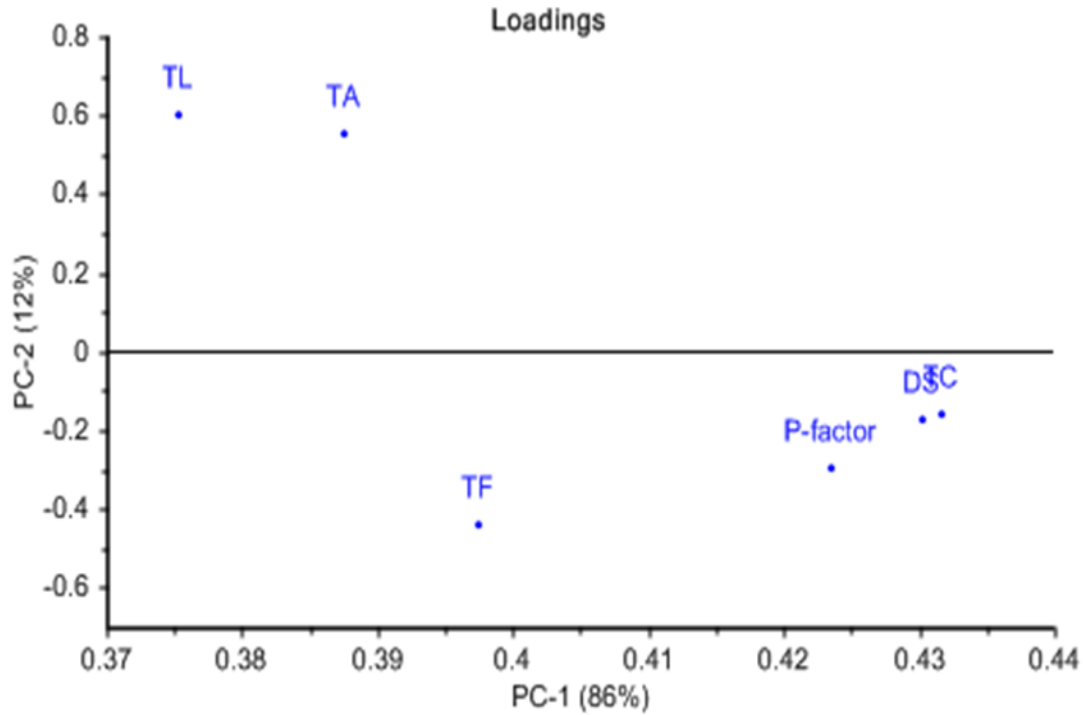
The coefficients of determination ( $R^2$ ) for the first and second principal components of both the PCA models are presented in Table 2.

**Table 1.** Analytical Data Subjected to PCA Calculations

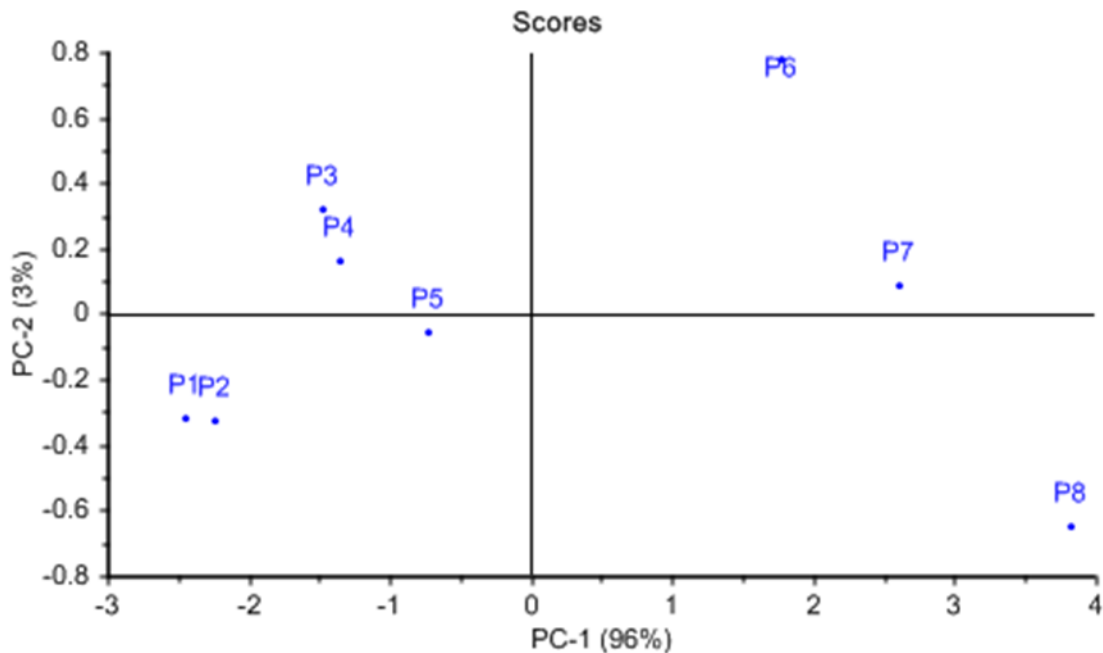
Pine samples. With the exception of P-factor, all the other values are expressed as (g *10 <sup>-3</sup> )/g.						
Symbol*	P-factor	TC	TA	TL	TF	DS
P1	10	15.29	1.51	6.97	0.02	27.73
P2	20	19.97	1.58	7.08	0.03	32.95
P3	30	28.66	1.99	8.53	0.06	45.38
P4	41	32.51	2.49	8.05	0.03	49.53
P5	60	45.77	2.90	8.01	0.17	65.85
P6	119	83.01	4.32	10.86	0.49	106.63
P7	179	96.19	5.08	10.26	0.69	120.00
P8	238	116.54	4.91	10.59	1.14	143.35
Birch samples						
Symbol	P-factor	TC	TA	TL	TF	DS
B1	10	4.95	1.91	5.79	0.01	13.95
B2	20	8.29	2.39	7.50	0.03	18.68
B3	30	9.36	2.73	9.46	0.03	21.35
B4	41	13.89	3.66	11.17	0.05	28.78
B5	60	30.31	4.97	13.59	0.12	45.95
B6	119	73.09	12.25	19.49	0.35	97.50
B7	179	126.85	16.39	21.31	0.43	150.80
B8	238	147.18	17.94	24.46	1.25	172.60

\*TC refers to the total amount of carbohydrates, TA to the total amount of volatile acids, TL to the total amount of lignin, TF to the total amount of furans, and DS to the total amount of dissolved solids.

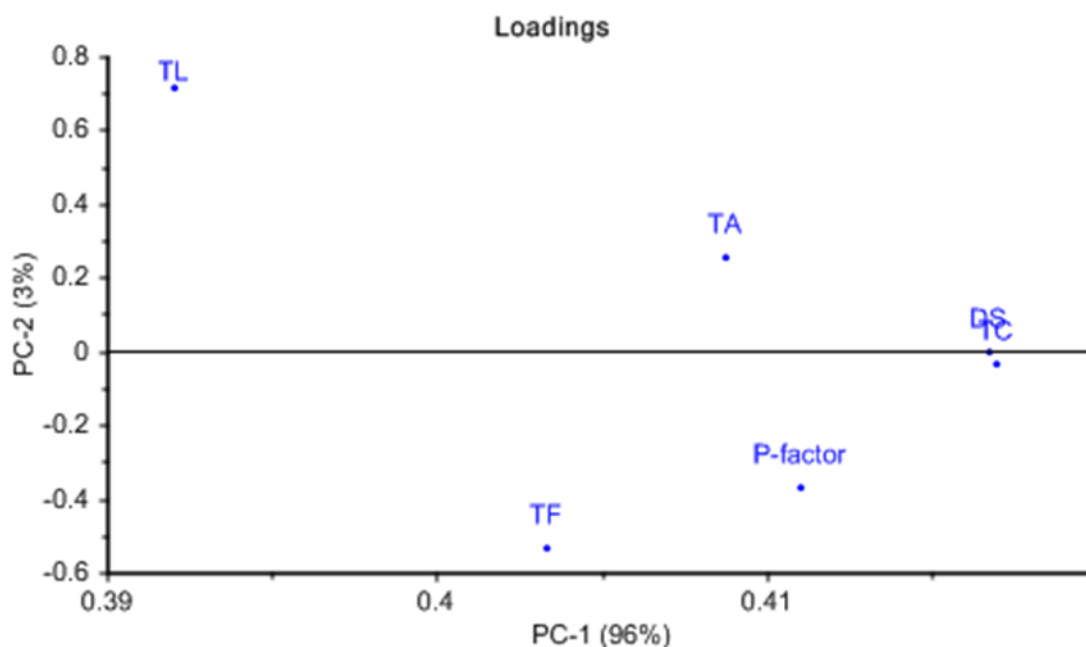
**Fig. 1.** PCA score plot between pine (P1 through P8) and birch (B1 through B8) samples using the first and second principal components (PCs). For abbreviations, see Table 1.



**Fig. 2.** Loading plot for the PCA model with all pine and birch samples using the first and second PCs. For abbreviations, see Table 1.



**Fig. 3.** PCA score plot for pine samples P1 through P8 using the first and second principal components (PCs). For abbreviations, see Table 1.



**Fig. 4.** Loading plot for the PCA model describing pine samples using the first and second PCs. For abbreviations, see Table 1.

**Table 2.** Coefficient of Determination ( $R^2$ , in Cumulative %) for the PCA Models

Rank	PCA with pine and birch samples	PCA with pine samples
1	86	96
2	98	99

### Projection to Latent Structures Regression

For the pine sample data, two PLS models, one for the P-factor and one for the DS, were constructed, and Table 3 gives the cross-validation results of these PLS regression models. The predicted *vs.* determined values for the P-factor and DS are presented in Figs. 5 and 6, respectively.

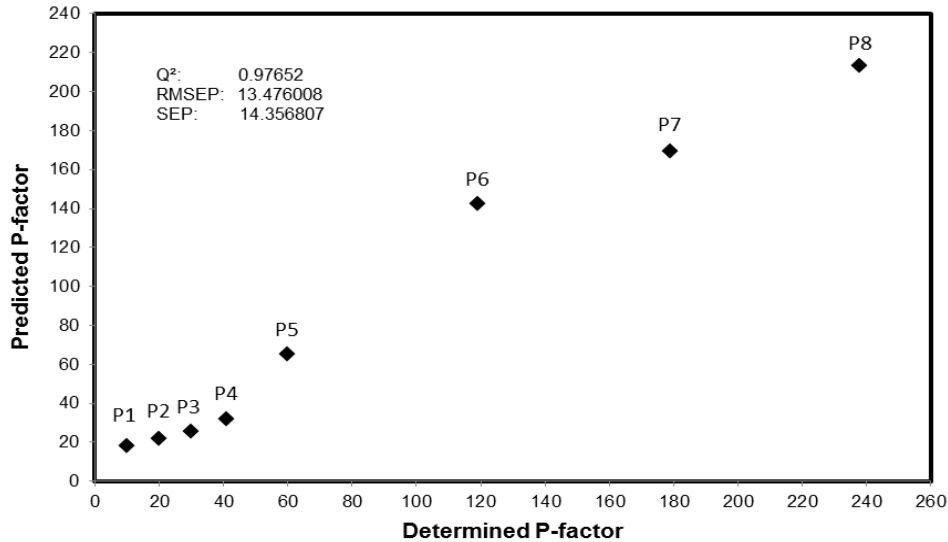
**Table 3.** Cross-Validation Results of the PLS Regression Models

Variable	$Q^2$	RMSEP	SEP	Rank
P-factor	97.6	13.5	14.4	2
DS*	99.9	0.001	0.002	1

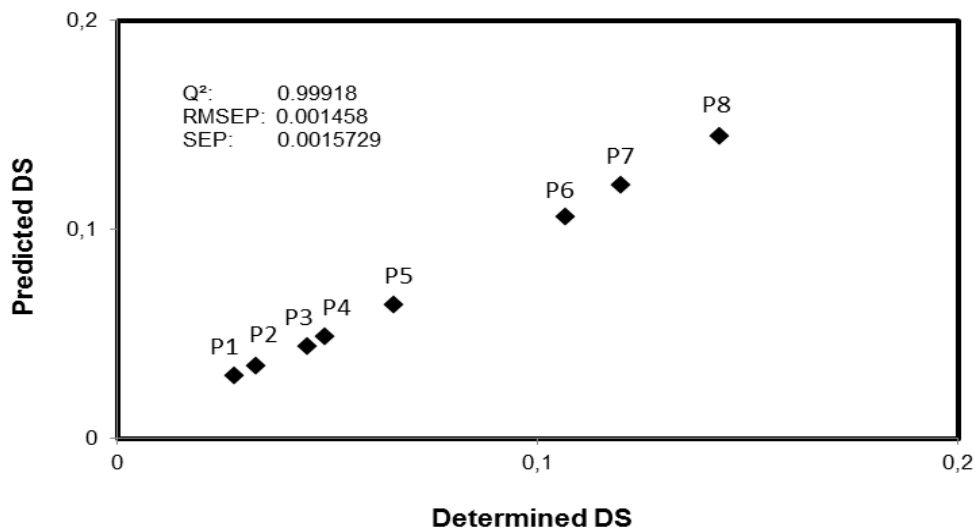
\*DS is the total amount of dissolved solids.

$Q^2$  is the cross-validated coefficient of determination (in cumulative %); RMSEP is the root mean standard error of prediction; and SEP is the standard error of prediction (in original units) (Davies and Fearn 2006).

It should be pointed out that, according to the loading values of the first latent variable of the PLS model for P-factor, the most influential variables were TC and TA, whereas in the case of the second latent variable, the corresponding variables were TF and TL. Furthermore, the loading values in the case of the PLS model for DS indicated that the most influential variable was TC alone.



**Fig. 5.** Predicted vs. determined P-factor values for the pine samples (see Table 1).  $Q^2$  is the cross-validated coefficient of determination (in cumulative %); RMSEP is the root mean square error; and SEP is the standard error of a prediction (in original units).



**Fig. 6.** Predicted vs. determined DS values for the pine samples. For abbreviations, see Fig. 5.

To illustrate the validity of the models, the P-factor and DS values of the individual samples were predicted using the corresponding PLS models. The results of the prediction calculations are presented in Tables 4 and 5.

In addition, other PLS models were constructed in a similar manner, using the data from the individual samples. The predicted vs. determined values for these models for P-factor and DS are presented in Figs. 7 and 8, respectively. The PLS models performed with reasonable accuracy in the validation of the models. The  $Q^2$ , RMSEP, and SEP values for the highly ranked and individual sample-based models are presented in Table 6. These values were relatively similar to those presented in Table 3 and Figs. 5 and 6. The RMSEP and SEP values for the P-factor model were slightly smaller, as expected. The bias of the models was small (0.117 and 0.00005 for the P-factor and DS models, respectively), as seen in the similarity between the RMSEP and SEP values.

**Table 4.** Predicted vs. Determined P-Factor Values based on Data from Individual Samples

Sample	Predicted values	Determined values
P1a	12	10
P1b	19	10
P2a	24	20
P2b	17	20
P3a	23	30
P3b	28	30
P4a	38	41
P4b	29	41
P5a	61	60
P5b	66	60
P6a	124	119
P6b	130	119
P7a	170	179
P7b	176	179
P8a	251	238
P8b	225	238

Calculations were made using the model and incorporating two latent variables. The symbols *a* and *b* refer to parallel samples.

**Table 5.** Predicted vs. Determined DS Values (% o.d.w.) based on Data from Individual Samples

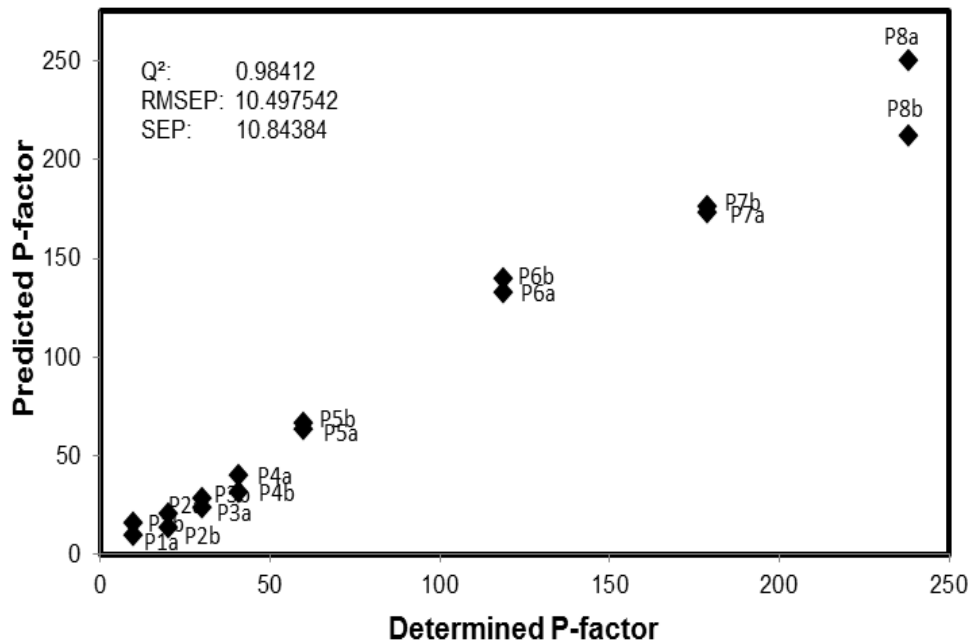
Sample	Predicted values	Determined values
P1a	2.8	2.7
P1b	3.0	2.9
P2a	3.6	3.3
P2b	3.4	3.3
P3a	4.3	4.5
P3b	4.5	4.6
P4a	5.0	5.0
P4b	4.8	4.9
P5a	6.5	6.7
P5b	6.3	6.4
P6a	10.5	10.5
P6b	10.7	10.9
P7a	11.8	12.1
P7b	12.3	11.9
P8a	14.5	14.4
P8b	14.3	14.3

Calculations were made using the model and incorporating one latent variable. The symbols *a* and *b* refer to parallel samples.

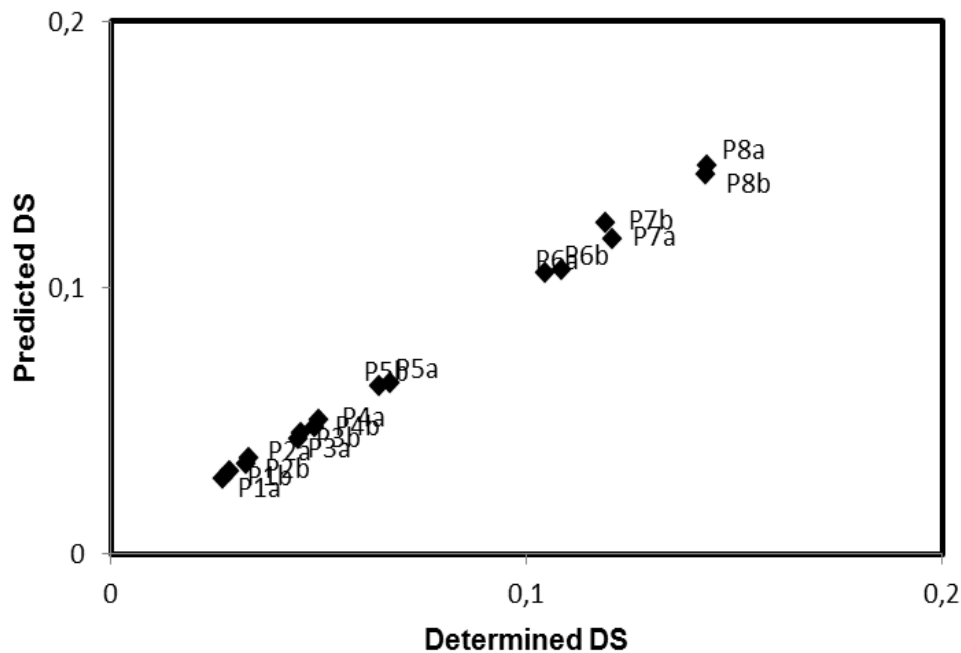
**Table 6.** Cross-Validation Results ( $Q^2$ , RMSEP, and SEP Values along with Rank of the Model) of the PLS Regression Models Calculated based on Data from Individual Samples

Variable	$Q^2$	RMSEP	SEP	Rank
P-factor	98.4	10.5	10.8	2
DS	99.7	0.002	0.002	1





**Fig. 7.** Predicted vs. determined P-factor values based on data from individual samples. The calculations were made using the model and incorporating two latent variables. For abbreviations, see Fig. 5.



**Fig. 8.** Predicted vs. determined DS values based on data from individual samples. The calculation was made using the model and incorporating one latent variable. For abbreviations, see Fig. 5.

## CONCLUSIONS

1. One of the most promising integrated ways to fractionate wood feedstocks is by the autohydrolysis of wood chips in hot water and under pressure prior to delignification. This chemometric approach could provide experimenters with a useful tool for a further understanding of the factors that cause changes in the autohydrolysis system.
2. A relatively accurate multivariate model for pine chip autohydrolysis was developed based on the analytical research data gathered, which will benefit future planning of autohydrolysis processes.
3. The complex hydrolysates produced during the autohydrolysis of wood contain a wide range of carbohydrate- and lignin-derived degradation products, which are characteristic of wood feedstocks. For this reason, the model is also greatly dependent on the feedstock material utilized. Therefore, in addition to the reported model for pine, a similar model for birch will be developed in the forthcoming investigations.

## ACKNOWLEDGMENTS

Financial support within the framework of Future Biorefinery (FuBio) research program (Forestcluster Ltd) is gratefully acknowledged.

## REFERENCES CITED

- Alén, R. (2011). *Biorefining of Forest Resources 20*, R. Alén (ed.), Paperi ja Puu Oy, Helsinki.
- Amidon, T. E., and Liu, S. (2009). "Water-based woody biorefinery," *Biotechn. Adv.* 27(5), 542-550.
- Bertaud, F., Sundberg, A., and Holmbom, B. (2002). "Evaluation of acid methanolysis for analysis of wood hemicelluloses and pectins," *Carbohydr. Polym.* 48(3), 319-280.
- Davies, A. M. C., and Fearn, T. (2006). "Back to basics: Calibration statistics," *Spectr. Eur.* 18(2), 31-32.
- Garrote, G., Domínguez, H., and Parajó, J. C. (1999). "Mild autohydrolysis: An environmentally friendly technology for xylooligosaccharide production from wood," *Chem. Technol. Biotechnol.* 74(11), 1101-1109.
- Hill, T., and Lewicki, P. (eds.). (2006). *Statistics: Methods and Applications: A Comprehensive Ref. for Sci., Industry, and Data Mining*, StatSoft, Inc., Tulsa, OK.
- Hyötyläinen, J., Knuutinen, J., and Malkavaara, P. (1998). "Transport of high molecular mass lignin material in the receiving water system of a mechanical pulp mill," *Chemosph.* 36(3), 577-587.
- Käkölä, J. M., Alén, R. J., Isoaho, J. P., and Matilainen, R. B. (2008). "Determination of low-molecular-mass aliphatic carboxylic acids and inorganic anions from kraft black liquors by ion chromatography," *J. Chromatogr. A.* 1190(1-2), 150-156.

- Lehto, J., and Alén, R. (2012). "Purification of hardwood-derived autohydrolysates," *BioResources* 7(2), 1813-1823.
- Li, H., Saeed, A., Jahan, M. S., Ni, Y., and van Heiningen, A. (2010). "Hemicellulose removal from hardwood chips in the pre-hydrolysis step of the kraft-based dissolving pulp production process," *J. Wood Chem. Technol.* 30(1), 48-60.
- Malkavaara, M., and Alén, R. (1998). "A spectroscopic method for determining lignin content of softwood and hardwood kraft pulps," *Chemom. Intell. Lab. Syst.* 44(1-2), 287-292.
- Malkavaara, M., Harjula, P., Alén, R., and Knuutinen, J. (2000). "Chemometric investigation on structural changes in pine kraft lignin during pulping," *Chemom. Intell. Lab. Syst.* 52(2), 117-122.
- Paredes, J. J., Jara, R., Shaler, S. M., and van Heiningen, A. (2008). "Influence of hot water extraction on the physical and mechanical behavior of OSB," *Forest Prod. J.* 58(12), 56-62.
- Ramos, L., Kristenson, E. M., and Brinkman, U. A. T. (2002). "Current use of pressurized liquid extraction and subcritical water extraction in environmental analysis," *J. Chromatogr. A.* 975(1), 3-29.
- Schultz, T., Templeton, M., and McGinnis, G. (1985). "Rapid determination of lignocellulose by diffuse reflectance Fourier transform infrared spectrometry," *Anal. Chem.* 57(14), 2867-2869.
- Schultz, T., and Burns, D. (1990). "Rapid secondary analysis of lignocellulose: Comparison of near infrared (NIR) and Fourier transform infrared (FTIR)," *TAPPI J.* 73(5), 209-212.
- Sears, K. D., Beélik, A., Casebier, R. L., Engen, R. J., Hamilton, J. K., and Hergert, H. L. (1971). "Southern pine prehydrolyzates: Characterization of polysaccharides and lignin fragments," *J. Polymer Sci.: Part C* 36(1), 425-443.
- Sixta, H., and Schild, G. (2009). "A new generation kraft process," *Lenzing. Berichte* 87, 26-37.
- Sixta, H. (ed.) (2006). *Handbook of Pulp*, Wiley-VCH Verlag GmbH, Weinheim.
- Sundberg, A., Sundberg, K., Lillandt, C., and Holmbom, B. (1996). "Determination of hemicelluloses and pectins in wood and pulp fibres by acid methanolysis and gas chromatography," *Nord. Pulp Pap. Res.* 11(4), 216-219.
- Swan, B. (1965) "Isolation of acid-soluble lignin from the Klason lignin determination," *Svensk Papperstidn.* 68, 791-795.
- Teo, C. C., Tan, S. N., Yong, J. W. H., Hew, C. S., and Ong, E. S. (2010). "Pressurized hot water extraction (PHWE)," *J. Chromatogr. A.* 1217(16) 2484-2494.
- Tunc, M. S., and van Heiningen, A. R. P. (2011). "Characterization and molecular weight distribution of carbohydrates isolated from the autohydrolysis extract of mixed southern hardwoods," *Carbohydr. Polym.* 83(1), 8-13.
- Tunc, M. S., and van Heiningen, A. R. P. (2009) "Autohydrolysis of mixed southern hardwoods: Effect of P-factor," *Nord. Pulp. Pap. Res. J.* 24(1), 46-51.
- Unscrambler® X, version 10.1 (64-bit), CAMO Software AS, Oslo, Norway, 2011.

van der Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K., and van der Werf, M. J. (2006). "Centering, scaling, and transformations: Improving the biological information content of metabolomics data," *BMC Genomics* 7, 142 <http://www.biomedcentral.com/1471-2164/7/142>.

Yoon, S.-H., and van Heiningen, A. (2008). "Kraft pulping and papermaking properties of hot-water pre-extracted loblolly pine in an integrated forest products biorefinery," *TAPPI J.* 7(7), 22-27.

Article submitted: August 7, 2013; Peer review completed: October 9, 2013; Revised version received and accepted: November 4, 2013; Published: November 7, 2013.