Appl Biochem Biotechnol (2011) 164:194–203
DOI 10.1007/s12010-010-9127-5

Near-Infrared Chemometric Approach to Exhaustive Analysis of Rice Straw Pretreated for Bioethanol Conversion

Yoshiki Horikawa · Tomoya Imai · Rie Takada · Takashi Watanabe · Keiji Takabe · Yoshinori Kobayashi · Junji Sugiyama

Received: 13 August 2010 / Accepted: 16 November 2010 /
Published online: 21 December 2010

Abstract We report a simple analytical procedure combining near-infrared (NIR) spectroscopy with multivariate analysis to detect the saccharification efficiency of pretreated rice straw. Three types of sample preparation methods were tested to develop a powerful calibration model, with the disk sample used as the standard protocol. From the spectra dataset of NaOH-treated biomass, we obtained a good calibration for the saccharification ratio and some major structural components by partial least-squares regression. Adding dataset from hot water and dilute sulfuric acid pretreatments to NaOH sample dataset, an acceptable calibration model to predict the saccharification ratio as well as the glucose, xylose, and lignin contents was generated. NIR has a great potential for rapid screening of saccharification efficiency of pretreated biomass, which would allows us to control the quality of processing toward better bioethanol production.

Keywords NIR spectroscopy · Multivariate analysis · Calibration model · Saccharification ratio

Introduction

The world faces the shortage of energy and the problem of global warming due to the excessive use of non-renewable fuels, which have driven the investigation for alternative energy sources from renewable materials. Lignocellulosic biomass, such as softwood,

Y. Horikawa (✉) · T. Imai · R. Takada · T. Watanabe · J. Sugiyama
Research Institute for Sustainable Humanosphere, Kyoto University, Uji, Kyoto 611-0011, Japan
e-mail: yhorikawa@rish.kyoto-u.ac.jp

K. Takabe
Division of Forest and Biomaterials Science, Graduate School of Agriculture, Kyoto University, Kyoto 606-8502, Kyoto, Japan

Y. Kobayashi
Tsukuba Research Laboratory, Japan Bioindustry Association, Tsukuba 305-8566, Ibaraki, Japan

hardwood, grasses and crop residue, is an attractive feedstock for fuel ethanol production because it contains large amounts of potentially fermentable sugars in the form of cellulose and hemicellulose [1]. The conversion of lignocellulosic biomass to ethanol involves three processes: pretreatment, to enhance the susceptibility of cellulose and hemicellulose in biomass; enzymatic hydrolysis, to recover fermentable sugars from the pretreated biomass; and fermentation, to convert ethanol from the obtained sugars. Unlike starch and sucrose, which can be easily decomposed into carbohydrate components, lignocellulosic materials should be pretreated before enzymatic hydrolysis and fermentation for utilization in a bioconversion process. A number of alternative chemical pretreatments, involving alkali [2–4], acid [5, 6], and hydrothermal [7, 8] processing, have been investigated over the years. The performance of several pretreatments has been evaluated via the saccharification yield by using commercial enzymes. The conventional method to analyze the saccharification efficiency or chemical components of numerous samples under a variety of conditions is the wet chemical method, which is labor-intensive, expensive, and time consuming. Therefore, a rapid and reproducible method is a desirable alternative to the traditional analytical measurement.

Near-infrared (NIR) spectroscopy is a nondestructive, fast, and accurate technique for measuring chemical components on the basis of overtone and combination bands of specific functional groups. The NIR bands are 10–100 times less intense than the corresponding mid-infrared fundamental bands, which enables direct analysis of samples with high absorbance and strong light-scattering properties without dilution or extensive sample preparation. Recently, NIR spectroscopy combined with multivariate statistics has provided chemometric tools such as principal components analysis and partial least-squares (PLS) regression methods that can build the model relationships between large numbers of dependent variables containing complex variations as NIR spectra and independent variations. PLS regression has been particularly successful in creating calibration model, which has been widely applied for estimating chemical components in plant biomass such as corn stover [9, 10], rice straw [11], and loblolly pine [12]. Wolfrum and Sluiter [13] successfully built a calibration model to measure chemical components after dilute acid pretreatment, indicating this technique can dramatically increase the number of samples that can be evaluated.

The objective of this study was to develop a rapid method not only to quantify individual chemical components but also to determine the saccharification efficiency of pretreated rice straw. We initially prepared NaOH-treated samples under various reactive conditions and then attempted to create the calibration model. Hot water- and dilute sulfuric-treated rice straws were further prepared and these dataset was incorporated with the NaOH calibration set. The performance for these calibrations will be discussed.

Materials and Methods

Sample Preparation

Rice straw chips, named “Momiroman” which is a new paddy rice cultivar for feed use [14], were uniformly milled to 150-μm particle size by two steps of roughly milling (Orient mill VM-16; Seishin Enterprise Corp., Tokyo, Japan) followed by finely milling (Bantam mill AP-BL; Hosokawa micron Corp., Osaka, Japan). The samples were treated in various ways: (1) dilute alkaline treatment (0.1–1.0% *w*/*v*) at 90–200°, for 4–15 min. (2) Dilute sulfuric acid treatment (0.5–1.0% *w*/*v*) at 100–200°, for 5 min. (3) Hydrothermal treatment

at 170–230°, for 5–30 min. The data of sample preparations and the corresponding chemical analysis data will be published elsewhere. The solid samples obtained were washed in distilled water several times for chemical component analysis or enzymatic hydrolysis.

Wet Chemical Analysis

The pretreated biomass was subjected to two stages of 72% sulfuric acid hydrolysis at 30 °C for 1 h and then at 121 °C for 1 h, which fractionates the biomass into liquor and a solid component. The hydrolysate liquor includes monosaccharides and acid-soluble lignin, and the solid residue contains acid-insoluble lignin, called Klason lignin, and ash. The hydrolysate liquor was analyzed for monosaccharides by high-performance liquid chromatography (HPLC) using an HPLC system equipped with an Asahipak NH2P-50 4E column, an RF-AXL detector, an autosampler, and a pneumatic controller (Prominence UFLC; Shimadzu Corp., Kyoto, Japan). The acid-soluble lignin content was not determined in this study. The hydrolysate residue was burned in a furnace at 575 °C for 5 h, and the residue obtained was gravimetrically measured as the ash content. The acid-insoluble lignin content was determined by subtracting the ash content from the hydrolysate content.

Enzymatic Hydrolysis

The enzyme used for hydrolysis was Accellerase 1500 (Genencor, Danisco US, Inc. Rochester, NY). The experiments for enzymatic hydrolysis were performed with 100 mg of pretreated biomass in 2 ml of 100 mM acetate buffer (pH 5.0) containing the enzyme (40 FPU). The filter paper activity (FPU) was measured according to the standard procedure recommended by NREL [15]. The mixtures were incubated at 50 °C with 150 strokes/min for 24 h, and the sugar released was estimated as reducing sugar by using the DNS method [16]. Saccharification ratio (%) of pretreated rice straw was estimated using the following equation:

$$\text{Saccharification ratio}(\%) = [\{\text{Reducing sugar}(\text{mg/ml}) \times 0.9\}/\text{initial substrate}(\text{mg/ml})] \times 100$$

NIR Data Acquisition

Three sample types were prepared: wet sample, prepared by washing pretreatments; powder sample, prepared by freeze-drying a wet sample; disk sample, prepared by collecting approximately 0.04 g of each pretreated sample and then molding with a handpress. All NIR spectra were recorded by using a PerkinElmer Spectrum 100 N system ranging from 10,000 to 4,000 cm^{-1} with a spectral resolution of 16 cm^{-1} and acquisition of 32 scans. Each sample was directly placed onto the NIR integrating sphere diffuse reflectance accessory (Perkin Elmer) with a triglycine sulfate detector and scanned in duplicate for monitoring the spectral reproducible. The absorbance spectrum was obtained by rationing the single beam spectrum against the background spectrum using a Teflon-based material (Spectralon; LabSphere, North Sutton, NH). Prior to chemometric analysis, the spectral data were transformed by using a Savitzky–Golay second derivative [17].

Chemometric Analysis

Multivariate analysis of the NIR spectra for the saccharification ratio and quantitative chemical composition prediction was conducted by using commercial software (Unscrambler v.9.8; CAMO Software, Inc., Woodbridge, NJ). For building the multivariate

calibration models, we adopted PLS regression, which is particular successfully in developing multivariate calibration models for NIR spectroscopy. The optimum number of factors used for prediction was determined by cross-validation (leave-one-out). The resulting models were evaluated in terms of coefficient of determination (R^2), the root mean square error of prediction (RMSEP) and the ratio of performance to deviation (RPD) [18]. Though RPD has generally been estimated as the ratio of the standard deviation of the reference data to the standard error of calibration and/or prediction (SEP) [19], RMSEP was used instead of SEP for calculating RPD in this study. Figure 1 summarizes the procedure for constructing the calibration model.

Results and Discussion

Accurate and repeatable NIR spectral acquisition is the most important factor to determine successful application of NIR techniques for fast characterization. To optimize the sample preparation, we examined three types of samples: a wet sample (Fig. 2a *i*); a powder sample (Fig. 2a *ii*); and a disk sample (Fig. 2a *iii*). For wet sample, the strong water absorption bands around 7,000 cm^{-1} and 5,200 cm^{-1} relating to O–H overtone and combination were seen [20], which completely overlapped the signals from biomass structural components (Fig. 2b *i*). In contrast, powder sample measurement by freeze-drying wet biomass allowed the extraction of several spectral features as shown in Fig. 2b *ii*. Among these spectra repeatedly obtained from same sample, the pattern difference is clear, indicating low reproducibility measurement. In the diffuse reflectance mode, it is requisite to obtain repeatable data to make a uniform density distribution of the samples. Accordingly, we prepared the disk sample by using handpress attachment, which had high spectral reproducibility as seen in Fig. 2b *iii*. The disk sample was therefore used throughout the study as the standard protocol.

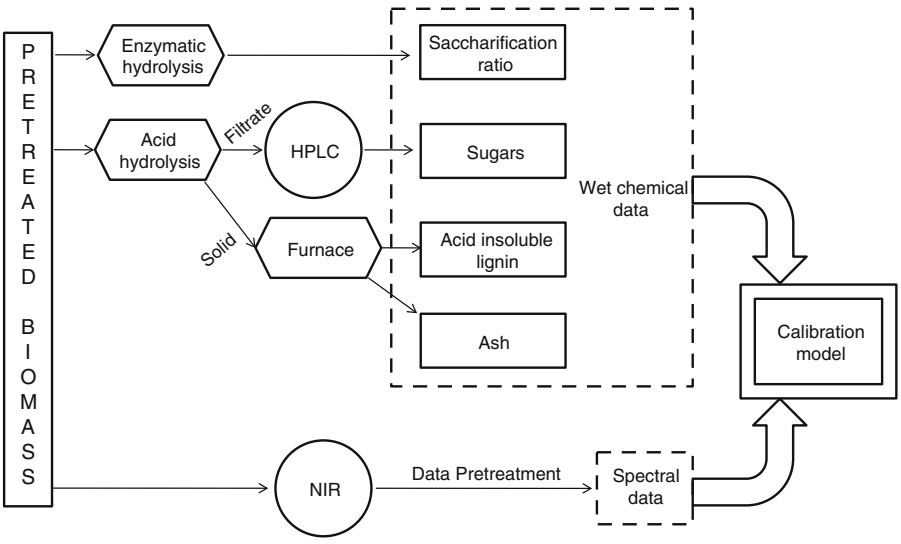

Fig. 1 Schematic illustration of the procedure to construct the calibration model between the wet chemical data and the spectral data

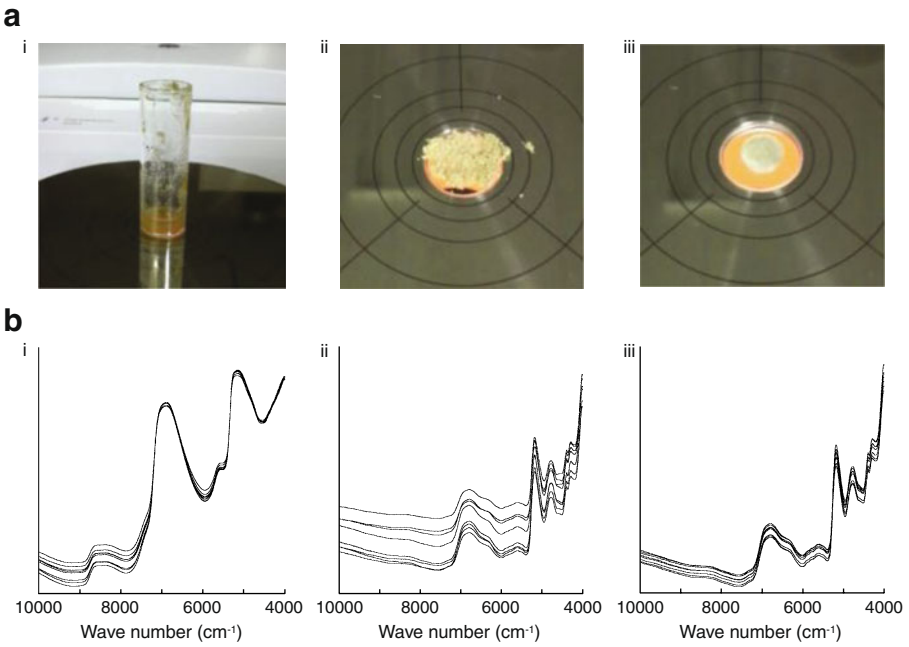

Fig. 2 **a** Three types of samples were examined for spectral reproducibility: (*i*) wet, (*ii*) powder, and (*iii*) disk samples. **b** NIR spectra corresponding to (**a**) ranging from 10,000 to 4,000 cm^{-1}. There are eight spectra repeatedly measured from rice straw treated in 0.1% NaOH at 120 °C for 30 min, involving 40% of saccharification ratio

Although disk samples were used, the difference in absorbance among the samples remained a major problem. Furthermore, the original spectra contained signal and noise, and therefore, pretreatment of the spectra was needed prior to multivariate analysis. So far, varietal mathematical treatments such as baseline offset, normalization, and mathematical pretreatments or a combination of these have been reported to build a reliable calibration model. However, for simplicity, we transformed the spectra by using a Savitzky–Golay second derivative, with which the spectral reproducibility was maintained and the information hidden in the original spectra would be extracted (Fig. 3).

All of the pretreated rice straw exhibit typical features such as the OH absorption bands assigned to the first overtone of the fundamental OH stretching vibration mode ranged from beneath the water stretching bands to 6,000 cm^{-1}. The broad absorption around 6,750 cm^{-1} was ascribed to that from amorphous cellulose, hemicellulose, and lignin. The shoulder bands around 6,720 cm^{-1} are related to hydrogen bonds forming a cellulose crystalline structure [21]. C–H groups from polysaccharide and lignin have been reported in the range of 6,000 to 5,386 cm^{-1}. The bands at 5,980 cm^{-1}, assigned to aromatic skeletal vibration of lignin, were not significant due to fewer amounts of lignin in rice straw. In addition, bands in the region from 5,000 to 4,000 cm^{-1} were ascribed to C–H groups of carbohydrates. In order to construct the PLS calibration model, full spectral range (10,000–4,000 cm^{-1}) was applied throughout this study,

There were 50 samples in the NaOH calibration data set, but as Table 1 indicates, ash and lignin had fewer samples. One or five samples were removed from the constituent

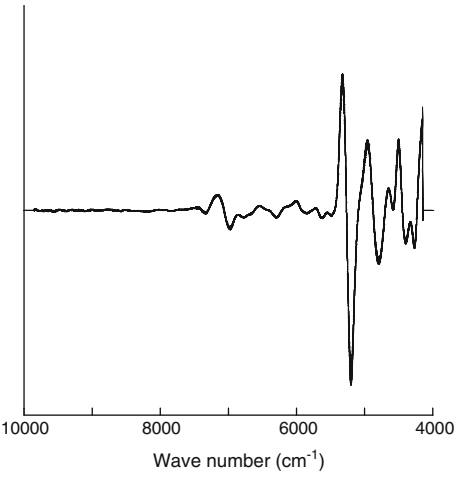

Fig. 3 Second-derivative spectra obtained from Fig. 2a *iii* including eight spectra

models, respectively, because of suspect wet chemistry values. Table 1 shows widely variations on the saccharification ratio and glucose content whereas insufficient variability for the hemicellulose components. The reason for these different of variability may be attributed to the contents of the hemicellulose components hardly changed under the moderate condition of alkaline processing.

Figure 4 shows the relationships between the NIR-predicted values and the wet chemical values measured for the saccharification ratio. Table 2 presents the statistical summary including the number of factors used in the calibration set, coefficient of determination (R^2) and the RMSEP for the cross-validated prediction models. The full cross-validation gave the optimum number of factors, which was considered to be the one minimizing the sum of residuals. As the factor numbers increased, the sum of residuals sometimes decreased without an inflection point. However, excess numbers may induce overfitting, which reflects a noisy signal in the loading spectra. Therefore, the number of factors was determined by referring to the loading spectra. Figure 4 illustrates the good correlation of the saccharification ratio, with an RMSEP of 5.45 and R^2 value of 0.86. As enzymatic

Table 1 Summary of the wet chemical data from the calibration set for the NaOH pretreatments

	Samples	Max.	Min.	Mean	SD
Saccharification ratio	50	88.61	22.73	69.28	14.38
Glucose	50	79.42	40.61	59.65	8.10
Xylose	50	23.66	14.31	20.87	1.83
Arabinose	50	4.85	1.90	3.88	0.73
Mannose	50	0.35	0.09	0.22	0.07
Galactose	50	1.64	0.26	1.11	0.35
Lignin	44	9.66	0.53	5.22	2.30
Ash	49	15.89	2.45	6.26	3.40

SD standard deviation

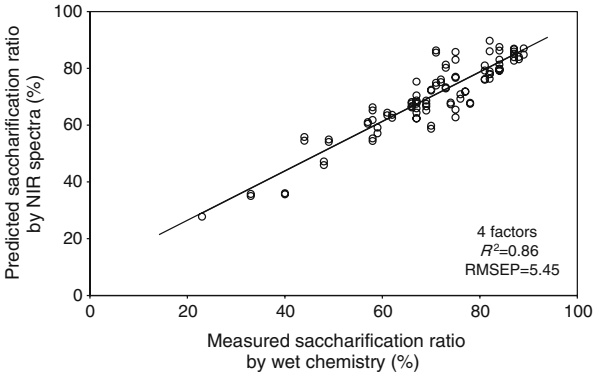

Fig. 4 Relationships between the NIR-predicted values and the wet chemistry values for saccharification ratio after the NaOH pretreatments

hydrolysis following fermentation for bioethanol conversion depends on pretreatments, the saccharification value is the most important index. The result obtained could be used for rapid analysis of the saccharification efficiency of pretreated samples as line-monitoring. For glucose, mostly originating from cellulose, the lower root mean square error and higher R^2 value indicated good predicting capability (Table 2). On the other hand, hemicellulose components such as xylose, arabinose, and mannose had an R^2 value of 0.71, 0.79, and 0.71, respectively, suggesting a weaker correlation than the saccharification and glucose data. These models may be due to insufficient compositional variance from alkaline treatment. Apart from hemicelluloses, lignin displayed acceptable calibration coefficients with an R^2 value of 0.84 and RMSEP of 0.91.

Almost 28 other samples prepared by dilute sulfuric acid and hot water treatments, with which hemicellulose degradation increasing the potential of enzymatic hydrolysis, were added to the NaOH sample set, which subsequently provided the greatest variation of chemical quantity. As observed in Table 3, the hemicellulose components ranged widely compared with the NaOH sample set, especially in the xylose content. The correlation between the wet chemical values and the NIR-predicted values for saccharification ratio is illustrated in Fig. 5 and structural components after various treatments are listed in Table 4. An RPD, the ratio of the standard deviation of the reference data to RMSEP, was also used to assess calibration performance. Determination of the RPD allows comparison of

Table 2 Statistics of the NIR calibration model for the NaOH pretreatments

	Factors	R^2	RMSEP	RPD
Saccharification ratio	4	0.86	5.45	2.64
Glucose	4	0.93	2.14	3.78
Xylose	4	0.71	0.98	1.86
Arabinose	4	0.79	0.34	2.15
Mannose	4	0.71	0.04	1.85
Galactose	4	0.81	0.15	2.29
Lignin	4	0.84	0.91	2.52
Ash	4	0.78	1.60	2.13

Table 3 Summary of the wet chemical data from the calibration set for the various pretreatments

	Samples	Max.	Min.	Mean	SD
Saccharification ratio	78	88.61	15.91	56.86	20.58
Glucose	78	79.42	39.99	56.88	8.14
Xylose	78	23.66	0.20	16.07	7.64
Arabinose	78	4.85	0.03	2.76	1.69
Mannose	78	0.37	0.04	0.20	0.08
Galactose	78	1.64	0.00	0.83	0.53
Lignin	73	28.32	0.34	10.12	7.54
Ash	75	29.00	2.45	10.01	6.75

SD standard deviation

calibration developed for different sample set, the higher the RPD the more accurate the data fitted by the calibration. The saccharification ratio (Fig. 5a) with an RPD of 3.44 were modified compared to that of 2.64 in NaOH sample set, demonstrating the prediction of that was possible even in the alkali- and acid-pretreated samples. The saccharification efficiency was limited by enzymatic accessibility to sugars exposed by chemical pretreatment. The successfully calibration model may be due to the NIR spectra involving the chemical and physical property for sugar digestibility. The calibration of glucose seemed to maintain the acceptable performance as R^2 value of 0.88 and RPD of 2.94. In the case of xylose (Fig. 5c), the calibration accuracy significantly improved an RPD of 4.30 in various sample set from that of 1.86 in NaOH sample set. The accuracy of the other hemicellulose components also became better than that of the NaOH sample set except for mannose and galactose. Rice straw has scarce amounts of these, which explains the unimproved calibration. An SD of the lignin in the various pretreatment set were much higher than that of the NaOH sample set, because acid treatment preferentially decomposes sugars compared with lignin. A powerful correlation (R^2=0.96, RMSEP=1.59) between the lignin content and the NIR spectra was observed even in mixture sample set (Fig. 5d).

The ash content seemed to be not well-predicted in the NaOH sample set, with an R^2 of 0.78 and RMSEP of 1.60. We initially hypothesized that biomass ash obtained from furnace treatment contains no NIR-absorbing functional groups, and therefore, it would be difficult

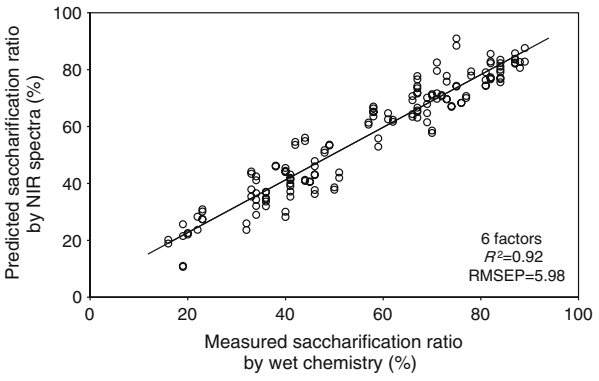

Fig. 5 Relationships between the NIR-predicted values and the wet chemistry values for saccharification ratio after the various pretreatments

Table 4 Statistics of the NIR calibration model for the various pretreatments

	Factors	R^2	RMSEP	RPD
Saccharification ratio	6	0.92	5.98	3.44
Glucose	6	0.88	2.77	2.94
Xylose	3	0.95	1.78	4.30
Arabinose	4	0.88	0.60	2.85
Mannose	4	0.69	0.05	1.78
Galactose	4	0.82	0.23	2.33
Lignin	3	0.96	1.59	4.74
Ash	4	0.93	1.75	3.87

to create a calibration model. However, a practical calibration model could be developed by using various treated samples, as presented in Table 4. Considering NIR inactivity in combination with our results, we infer that the structural ash could be indirectly influenced through correlation with other NIR vibrations, leading to acceptable calibration.

Conclusion

The combination of NIR spectra from the NaOH sample set with multivariate analysis demonstrated that acceptable calibrations can be created for the saccharification ratio and some major structural components by using disk samples. The saccharification ratio has further maintained the higher calibration performance even in the alkali- and acid-pretreated sample set, as well as glucose and lignin. When wide variation was provided by hot water and dilute sulfuric acid treatments, xylose showed a powerful calibration model in addition to ash. This technique allows simple analysis of the saccharification efficiency, and could be applied for quality control of pretreated biomass or to monitor defects and contamination. Furthermore, the method allows simultaneous quantification of sugars and lignin in biomass, enabling easier measurements in structural components analysis.

Acknowledgments This work was supported by the New Energy and Industrial Technology Development Organization (NEDO). The authors express their appreciation to Dr. H. Shinzawa from Research Institute of Instrumentation Frontier, Advanced Industrial Science and Technology (AIST) for the technical advice on chemometric analysis.

References

1. Adney, B., & Baker, J. (1996). National Renewable Energy Laboratory, Golden, CO, USA. from http://www.nrel.gov/biomass/analytical_procedures.html#lap-009.
2. Hames, B. R., Thomas, S. R., Sluiter, A. D., Roth, C. J., & Templeton, D. W. (2003). *Applied Biochemistry and Biotechnology, 105*, 5–16.
3. Hirabayashi, H., Nemoto, H., Ando, I., Kato, H., Oota, H., Satou, H., et al. (2010). *Bull Natl Inst Crop Sci, 11*, 31–47.
4. Hörmeyer, H. F., Schwald, W., Bonn, G., & Bobleter, O. (1988). *Holzforschung, 42*, 95–98.
5. Jin, S. Y., & Chen, H. Z. (2007). *Industrial Crops and Products, 26*, 207–211.
6. Jones, P. D., Schimleck, L. R., Peter, G. F., Daniels, R. F., & Clark, A. (2006). *Wood Science and Technology, 40*, 709–720.

7. Knappert, D., Grethlein, H., & Converse, A. (1980). *Biotechnology and Bioengineering, 22*, 1449–1463.
8. Luchsinger, W. W., & Cornesky, R. A. (1962). *Analytical Biochemistry, 4*, 346–347.
9. Macdonald, D. G., Bakhshi, N. N., Mathews, J. F., Roychowdhury, A., Bajpai, P., & Moo-Young, M. (1983). *Biotechnology and Bioengineering, 25*, 2067–2076.
10. Negro, M. J., Manzanares, P., Ballesteros, I., Oliva, J. M., Cabañas, A., & Ballesteros, M. (2003). *Applied Biochemistry and Biotechnology, 105*, 87–100.
11. Nguyen, Q. A., Tucker, M. P., Keller, F. A., & Eddy, F. P. (2000). *Applied Biochemistry and Biotechnology, 84–86*, 561–576.
12. Rodriguez-Saona, L. E., Fry, F. S., McLaughlin, M. A., & Calvey, E. M. (2001). *Carbohydrate Research, 336*, 63–74.
13. Sánchez, Ó. J., & Cardona, C. A. (2008). *Bioresource Technology, 99*, 5250–5295.
14. Savitzky, A., & Golay, M. J. E. (1964). *Analytical Chemistry, 36*, 1627–1639.
15. Schimleck, L. R., Evans, R., Jones, P. D., Daniels, R. F., Peter, G. F., & Clark, A., III. (2005). *IAWA Journal, 26*, 175–187.
16. Soto, M. L., Domínguez, H., Núñez, M. J., & Lema, J. M. (1994). *Bioresource Technology, 49*, 53–59.
17. Tsuchikawa, S., Yonenobu, H., & Siesler, H. W. (2005). *The Analyst, 130*, 379–384.
18. Williams, P. C., & Sobering, D. C. (1993). *Journal of Near Infrared Spectroscopy, 1*, 25–33.
19. Wolfrum, E. J., & Sluiter, A. D. (2009). *Cellulose, 16*, 567–576.
20. Ye, X. P., Liu, L., Hayes, D., Womac, A., Hong, K. L., & Sokhansanj, S. (2008). *Bioresource Technology, 99*, 7323–7332.
21. Zhao, Y. L., Wang, Y., Zhu, J. Y., Ragauskas, A., & Deng, Y. L. (2008). *Biotechnology and Bioengineering, 99*, 1320–1328.