

Characterisation and selection of multiple pre-treated lignocellulosic materials for bioethanol production using near infrared spectroscopy

Rosario del P. Castillo^{1*}, Jaime Baeza^{1,2}, Carolina Parra¹, Heriberto Franco^{1,3}, Samuel Peña¹, Eduardo Troncoso¹, Andrea Andrade¹, Edgardo Esparza^{1,2} and Juanita Freer^{1,2}

¹Biotechnology Center, University of Concepcion, Concepcion, Chile

²Faculty of Chemical Sciences, University of Concepcion, Concepcion, Chile

³Faculty of Forestry Sciences, University of Concepcion, Concepcion, Chile

*Corresponding author: rosariocastillo@udec.cl

Introduction

Lignocellulosic biomass (LB) has been recognised as a sustainable alternative to existing starch and sucrose-based bioethanol production. LB recalcitrance, due to cellulose, hemicelluloses and lignin arrangements, is largely responsible for the high cost of LB conversion. Cellulose and hemicelluloses can be converted to ethanol by separate hydrolysis and fermentation (SHF) or by simultaneous saccharification and fermentation (SSF).¹ To disrupt the LB structure and increase the yield of the hydrolysis, a pre-treatment step is needed. The purpose of the pre-treatment is to remove lignin and hemicelluloses, reduce cellulose crystallinity and increase porosity. Physical, physico-chemical, chemical and biological processes are used as pre-treatments.¹ They have specific advantages and disadvantages and the cost and performance in subsequent hydrolysis and fermentation is heavily influenced by this pre-treatment. Characterisation of pre-treated materials is needed to evaluate the substrate quality and its ethanol yield. Traditionally, chemical analyses of the lignocellulosic components have been performed by acid hydrolysis followed by gravimetric determination of lignin and chromatographic determination of sugars. These methods provide highly accurate data but they are laborious, time-consuming and expensive; consequently, sample throughput is limited. Near infrared (NIR) spectroscopy has been reported in diverse applications in production of biofuels. Some of these reports include determination of glucans and ethanol in fermentation of substrates from wheat and rye,³⁻⁴ prediction of degradability of wheat straw⁵ and discrimination of biomass for energy purposes⁶ among others. However, few studies on chemical composition of pre-treated LB,⁷⁻⁸ especially for bioethanol production from woody biomass, have been reported and there are no reports of a possible ethanol yield prediction from these materials after a SSF process. Given that bioethanol production aims to be a multivariable process with respect to the lignocellulosic feedstock, this work attempts to use NIR spectroscopy to characterise simultaneously samples from multiple pretreatments and lignocellulosic species and estimate their ethanol yield after a SSF process. Selection of samples as suitable for obtaining high ethanol yields using a classification method has been also investigated.

Materials and Methods

Lignocellulosic materials and pre-treatments

Samples included *Eucalyptus globulus* chips pre-treated by steam explosion⁹ (EG-SE) and Organosolv process (EG-O),¹⁰ *Pinus radiata* chips pre-treated by sulphite alkaline/anthraquinone delignification (PR-ASA) and wheat straw pre-treated by steam explosion⁹ (WS-SE). EG-SE and WS-SE were obtained using 100g of LB (dry weight) in a 4L steam explosion reactor at temperatures and time conditions of 200-220°C for 5-10 min and 172-230°C for 3-17 min respectively. EG-O samples were obtained in a 1L Parr reactor using 100g of wood and 700mL of pulping liquor containing ethanol:water mixtures from 25:75 to 75:25% (v/v) and 0.83-2.00% H₂SO₄ (w/w of dry wood) at 155-205°C and 26-70 min. PR-ASA were obtained in a Parr 4843 reactor using 100g of wood and 400mL of pulping liquor containing 25g of Na₂SO₃/NaOH (50/50, 60/40 and 70/30 g/g) and anthraquinone 0.1- 0.3% w/w of wood at 120-185°C and 30-120 min. Pre-treated materials were washed with distilled water, filtered under vacuum and the solid residues stored at 4°C for subsequent analysis.

Chemical characterisation of pre-treated materials

Chemical composition of the pre-treated materials was determined by the Puls method.¹¹ Pulps (300 mg) were hydrolysed with 3mL 72% (v/v) H₂SO₄ at 30°C for 60 min followed by dilution to 3% (v/v) H₂SO₄. The hydrolysed material was heated at 121°C in an autoclave for 60 min and the resultant product filtered. Glucose and hemicellulosic sugars (quantified as a mixture of xylose, mannose and galactose) in the filtrate were measured by normal phase HPLC using Merck Hitachi equipment with refractive index detector, an

Reference paper as:

R. del P. Castillo, J. Baeza, C. Parra, H. Franco, S. Peña, E. Troncoso, A. Andrade, E. Esparza and J. Freer (2012). Characterisation and selection of multiple pre-treated lignocellulosic materials for bioethanol production using near infrared spectroscopy, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 351-355.

Aminex HPX-87H column at 45°C, mobile phase H₂SO₄ 5mM and a flow rate of 0.6 mLmin⁻¹. The solid fraction, insoluble lignin, was dried and weighed. Soluble lignin in the filtrate was analysed by UV-VIS spectrophotometry at 205nm. Total lignin was calculated as the sum of soluble and insoluble lignin. All analyses were carried out in triplicate.

SSF process and ethanol quantification

SSF was carried out according to technical report NREL/TP-510-42630¹², using a concentration of 10% of the pre-treated materials (w/v), a thermally-acclimatised (40°C) *Saccharomyces cerevisiae* IR₂-9a (6.0 gL⁻¹) and cellulases (Celluclast; Novozymes, NC, USA) supplemented with β-glucosidase (Novozymes, NC, USA) enzymes. Enzymatic loadings were 20 FPU of cellulases and 20 CBU of β-glucosidase per g of pre-treated material. SSF was performed at 40°C for 72 h with sampling every 12 hours. The content of the released ethanol was analysed by gas chromatography (GC) on a Perkin-Elmer autosystem XL Headspace using a FID detector and a HPSMS column (30 m). The GC programme was: 50°C×3min, 10°C/min, 100°C×1min; 25°C/min, 125°C×1min. Injector and detector temperatures were 200 and 300°C respectively. Ethanol concentration was expressed in grams of ethanol produced per litre of the SSF solution (gL⁻¹) and as the percentage of the obtained ethanol with respect to the theoretical ethanol content in pulps (yield (%)), using the ratio 0.51g theoretical ethanol/g glucose in pulp. Maximal ethanol concentration of samplings was used for construction of NIR models.

NIR spectral measurements

Pulps were milled using a Moulinex A5052HF (sieving between 40 and 60 mesh). Spectra were acquired using a Perkin Elmer Identichex FT-NIR spectrometer in the range of 1000 to 2500nm at 2nm intervals; 32 scans per spectrum in reflectance were collected and converted to Kubelka-Munk units. Mean centering was used as a pre-processing technique and transformation by multiplicative scatter correction (MSC) and/or derivatives were applied for each model. Two spectra were acquired per sample and the average was used the construction of the models.

Chemometric analysis

Principal component analysis (PCA) was applied to the spectra in order to visualise possible clustering of samples by species or pre-treatment. Partial least squares (PLS) regression was used to build predictive models for quantification of glucans, hemicellulosic sugars, lignin and ethanol (gL⁻¹ and yield (%)). Predictive models were validated by cross-validation (leave-one-out method) and by external validation using an exclusion set. Discriminant-PLS (PLS-DA) was used to classify the pulps according to their ethanol concentration after the SSF process. Classification rates were evaluated through the percentage of misclassified samples in modeling and validation sets. Analyses were carried out using Pirouette 4.0 software (Infometrix Inc).

Results and Discussion

PCA mainly showed a separation of *P. radiata* samples from the others and some suggestion of grouping by pre-treatment; however, PLS models can be obtained using all the pre-treated samples. Parameters of the PLS models are shown in Table 1. These results show low errors of calibration (RMSEC) and validation (RMSECV, RMSEP) and high correlation coefficients between the NIR predictions and the measured values obtained by reference methods ($r > 0.9$) in all the cases, indicating accuracy of the models and their capacities to predict the chemical composition and released ethanol from the different pre-treated samples, even for different species. Since RMSECV and RMSEP have the same order of magnitude and values of RPD are relatively high (RPD values > 3 are considered adequate for analytical purposes in most of the NIR applications for agricultural products¹³), the predictive abilities of these models could be considered excellent, except in the case of ethanol yield (%) for which RPD and correlation coefficient values indicate that this model should be improved. Figure 1 shows the measured versus the NIR predicted values for the chemical characterisation and ethanol concentrations. Ethanol yield (%) is directly related to the conversion of potential glucose to ethanol i.e., high yield (%) values means high conversion rates but not necessarily samples with high glucose content. It basically indicates performance of the SSF process which is influenced by the presence of enzyme inhibitors or enzyme adsorption to the lignin part of the substrate; however, the gL⁻¹ yield is proportional both to the glucose content obtained in the pre-treatment and its conversion to ethanol. Thus, the ethanol (gL⁻¹) model could be useful to predict a possible ethanol concentration in pulps after the SSF process under the SSF conditions used. An identification of the important wavelengths for the models developed was carried out using their correlation spectra (Figure 2). Since LB samples are made up

Reference paper as:

R. del P. Castillo, J. Baeza, C. Parra, H. Franco, S. Peña, E. Troncoso, A. Andrade, E. Esparza and J. Freer (2012). Characterisation and selection of multiple pre-treated lignocellulosic materials for bioethanol production using near infrared spectroscopy, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 351-355.

of lignin, cellulose, hemicelluloses and extractives, increases in biomass directed to one of these components will mean less biomass directed toward the others and clear negative correlations among them were observed. Table 3 shows a summary of the principal bands and their R^2 values. Absorbances in the region between 1668 and 1672 nm and at 1154, 2154 and 2388 nm have been associated with lignin

Table 1. PLS parameters of NIR models for chemical characterisation and ethanol yield quantification of pulps.

Model	Glucans (%)			Hemicellulosic sugars (%)			Lignin (%)		
Pre-treatment	MSC			MSC			MSC		
Parameter	Sample set			Sample set			Sample set		
	Cal	CV	Ext Val	Cal	CV	Ext Val	Cal	CV	Ext Val
N	38	38	14	23	23	7	26	26	10
Range	42.5-84.3	42.5-84.3	50.1-84.2	0.4-22.2	0.4-22.2	0.7-21.0	8.2-40.2	8.2-40.2	7.4-35.3
Mean	61.1	61.1	64.5	9.2	9.2	7.0	18.9	18.9	22.2
SD	10.7	10.7	11.5	6.4	6.4	6.6	9.4	9.4	8.6
Comp	5	5	5	6	6	6	5	5	5
RMSEC	2.8	-	-	1.3	-	-	1.9	-	-
RMSECV	-	3.4	-	-	2.1	-	-	2.7	-
RMSEP	-	-	3.6	-	-	0.9	-	-	2.1
R	0.9692	0.9457	0.957	0.9845	0.9441	0.9940	0.9831	0.9571	0.9686
RPD	-	3.1	3.2	-	3.0	7.3	-	3.5	4.1

Model	Ethanol (g/L)			Ethanol yield (%)		
Pretreatment	MSC, 1 st derivative (25 points)			MSC, 1 st derivative (25 points)		
Parameter	Sample set			Sample set		
	Cal	CV	Ext Val	Cal	CV	Ext Val
N	19	19	10	19	19	9
Range	4.7 - 46.8	4.7 - 46.8	7.9-45.7	20.4-96.9	20.4-96.9	20.5-90.6
Mean	24.3	24.3	24.4	63.0	63.0	61.7
SD	12.7	12.7	13.2	20.8	20.8	24.6
PLS comp	9	9	9	12	12	12
RMSEC	1.0	-	-	1.7	-	-
RMSECV	-	3.1	-	-	7.7	-
RMSEP	-	-	3.7	-	-	10.2
R	0.9984	0.9690	0.9568	0.9989	0.9258	0.9002
RPD	-	4.1	3.6	-	2.7	2.4

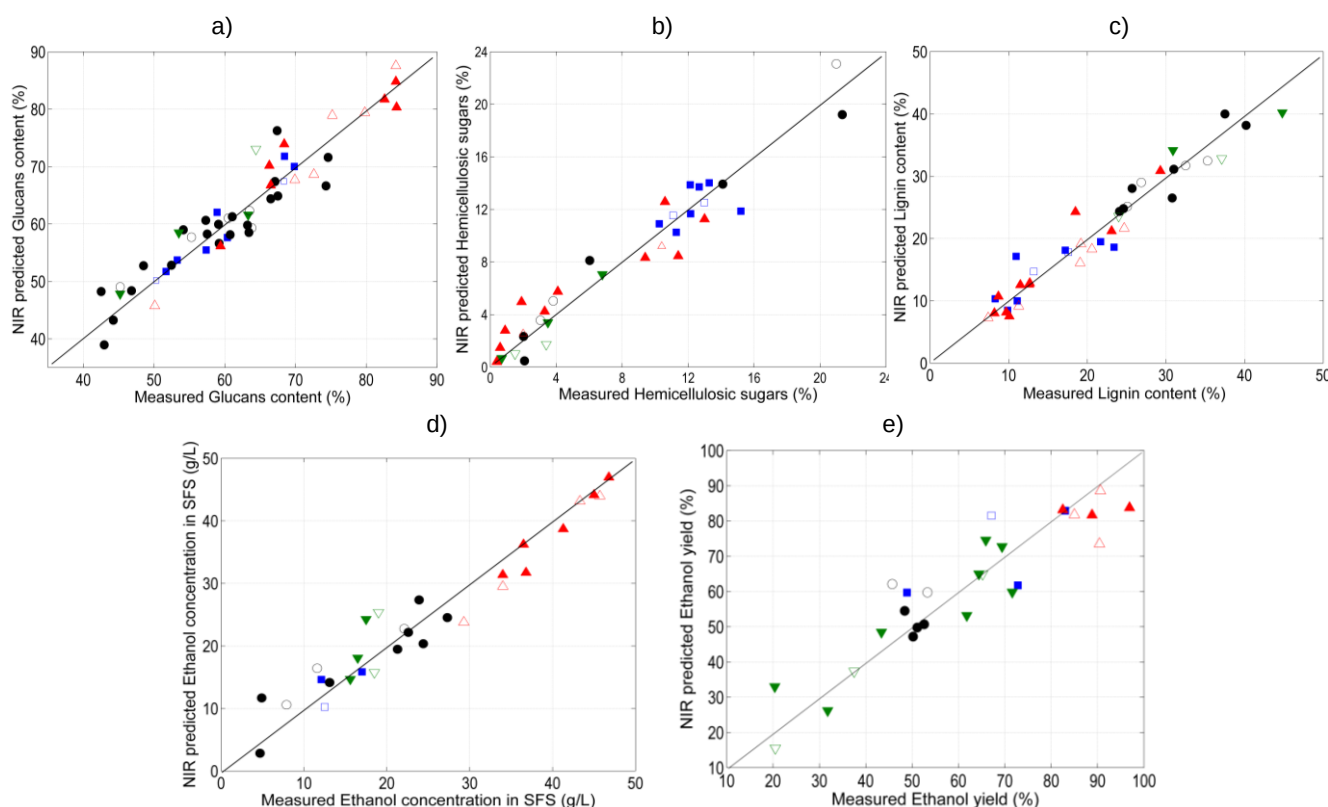


Figure 1. Measured vs NIR predicted values for chemical characterisation and ethanol concentrations. ●WS-SE, ▼EG-SE, ■PR-ASA, ▲EG-O. Empty symbols denote external validation samples.

Reference paper as:

R. del P. Castillo, J. Baeza, C. Parra, H. Franco, S. Peña, E. Troncoso, A. Andrade, E. Esparza and J. Freer (2012). Characterisation and selection of multiple pre-treated lignocellulosic materials for bioethanol production using near infrared spectroscopy, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGovern, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 351-355.

content in both pine and eucalypt wood and 2384 nm with cellulose;¹³ in this study, these wavelengths did not exhibit a high correlation with either and absorbance at 2384nm was strongly correlated with lignin. The highest R^2 for ethanol models show a negative correlation with lignin content at several wavelengths.

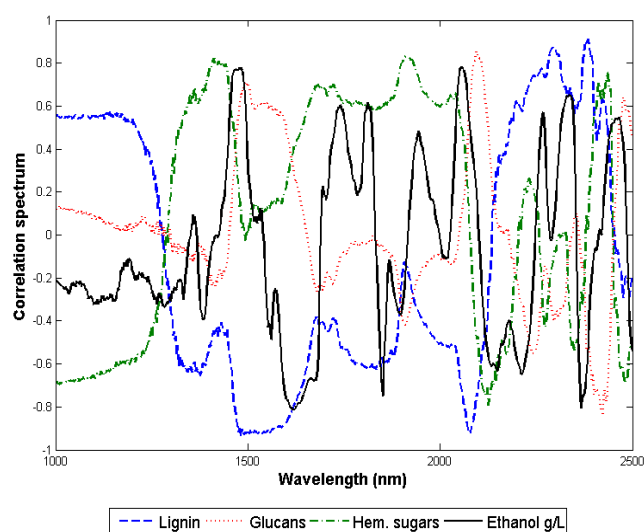


Figure 2. Correlation spectra of PLS models

Table 3. Principal bands of model correlation spectra.

Wavelength (nm)	R^2 of correlation vector			
	Glucans	Lignin	Hem. Sug.	Ethanol (g/L ⁻¹)
1410	0.05(-)	0.20(-)	0.68	0.00
1478	0.34	0.81(-)	0.01	0.60
1548	0.38	0.84(-)	0.01	0.10(-)
1564	0.34	0.82(-)	0.02	0.16(-)
1624	0.06	0.64(-)	0.17	0.64(-)
1918	0.13(-)	0.04(-)	0.69	0.00
2054	0.01	0.51(-)	0.25	0.61
2078	0.36	0.85(-)	0.00	0.23
2094	0.74	0.57(-)	0.24(-)	0.04
2366	0.01(-)	0.41	0.07(-)	0.65(-)
2384	0.27(-)	0.83	0.00	0.33
2422	0.69(-)	0.40	0.46	0.00

Results of PLS-DA classification of the pulps are summarised in Table 2; it was only possible to find acceptable classifications for class 1 samples (yields less than 65%) for the model of ethanol yield (%) when a threshold of 65% of ethanol yield was used, possibly due to the small number of samples with high yield (%). This was not the case for the g/L⁻¹ model for which an excellent classification of samples was obtained in both classes.. For future improvements over these results, a large number of samples is proposed.

Table 2. PLS-DA misclassification matrix for ethanol concentration and ethanol yield prediction.

Data set	Model a (MSC/2 nd derivative (25p))				Model b (MSC/1st derivative (25 p))	
	Real class	Predicted class			Predicted class	
		1(<30g/L ⁻¹)	2(≥30g/L ⁻¹)		1(<65%)	2(≥65%)
Modeling data	1	9(100%)	0		8(80%)	2(20%)
Set	2	0	3(100%)		4(44%)	5(56%)
External data	1	7(100%)	0		4(80%)	1(20%)
Set	2	0	3(100%)		1(14%)	6(86%)

Conclusions

NIR models for chemical characterisation, ethanol yield determination and classification of samples from multiple pre-treatments and lignocellulosic materials were obtained. Use of these models facilitates the fast selection of samples with suitable characteristics to obtain high ethanol yields.

Acknowledgements

Fondecyt Postdoctoral 3100078, Fondecyt 1070492 and INNOVA Bio Bio 05B1416L8 projects, Chile.

References

1. C.E. Wyman, *Handbook on Bioethanol*. CRC, London (1996)
2. B. Liebman, A. Friedl and K. Varmuza, *Anal. Chim. Acta* **642**, 171-178 (2009).
3. B. Liebman, A. Friedl and K. Varmuza, *Biochem. Eng. J.* **52**, 187-193 (2010).
4. S. Bruun, J. Wagner Jensen, J. Magid, J. Lindedam and S. Balling, *Ind. Crop. Prod.* **31**, 321-326 (2010).
5. N. Labbé, S. Lee, H. Cho, M. Jeong and N. André, *Bioresour. Technol.* **99**, 8445-8452 (2008).
6. M.A. Sanderson, F. Agblevor, M. Collins and D.K. Johnson, *Biomass Bioenergy* **11**, 365-70 (1996).
7. S. Kelley, R. Rowell, M. Davis, Ch. Jurich and R. Ibach, *Biomass Bioenergy* **27**, 77- 88 (2004).
8. J. Saddler, L. Ramos and C. Breuil, "Steam Pretreatment of lignocellulosic residues", in *Bioconversion of Forest and Agricultural Plant Residues*, Ed by J. Saddler, C.A.B International, Wallingford, UK, pp. 73-91 (1993).
9. X. Pan, N. Gilkes, J. Kadla, J. K. Pye, S. Saka, D. Gregg, K. Ehara, D. Xie, D. Lam, J. Saddler, *Biotechnol. Bioeng.* **94**, 851-861 (2006).
10. J. Puls, K. Poutanen, H. Körner and L. Viikari, *Appl. Microbiol. Biotechnol.* **22**, 416-23 (1985).

Reference paper as:

R. del P. Castillo, J. Baeza, C. Parra, H. Franco, S. Peña, E. Troncoso, A. Andrade, E. Esparza and J. Freer (2012). Characterisation and selection of multiple pre-treated lignocellulosic materials for bioethanol production using near infrared spectroscopy, in: *Proceedings of the 15th International Conference on Near Infrared Spectroscopy*, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 351-355.

11. P.C. Williams, "Implementation of Near-Infrared Technology", in *Near Infrared Technology in the Agricultural and Food Industries*, 2nd Edition, Ed By P.C. Williams and K. Norris, American Association of Cereal Chemists. St Paul, Minnesota, USA. (2001).
12. National Renewable Energy Laboratory, Technical Report NREL/TP-510-42630. USA (2008).
13. G. Hodge and W.C. Woodbridge. *J. Near Infrared Spectrosc.* **18**, 367-380 (2010).

Reference paper as:

R. del P. Castillo, J. Baeza, C. Parra, H. Franco, S. Peña, E. Troncoso, A. Andrade, E. Esparza and J. Freer (2012). Characterisation and selection of multiple pre-treated lignocellulosic materials for bioethanol production using near infrared spectroscopy, in: Proceedings of the 15th International Conference on Near Infrared Spectroscopy, Edited by M. Manley, C.M. McGoverin, D.B. Thomas and G. Downey, Cape Town, South Africa, pp. 351-355.