



HAL
open science

A comparison of various infrared (IR) spectroscopy techniques for calibrating poplar wood chemical composition and saccharification potential

Vincent Segura, Kévin Ader, Redouane El Malki, Violaine Murciano,
Jean-Paul Charpentier

► To cite this version:

Vincent Segura, Kévin Ader, Redouane El Malki, Violaine Murciano, Jean-Paul Charpentier. A comparison of various infrared (IR) spectroscopy techniques for calibrating poplar wood chemical composition and saccharification potential. 16. International Conference on Near Infrared Spectroscopy. NIR 2013, Institut National de Recherche en Sciences et Technologies pour l'Environnement et l'Agriculture (IRSTEA). Montpellier, FRA., Jun 2013, La Grande Motte, France. hal-01268468

HAL Id: hal-01268468

<https://hal.science/hal-01268468>

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A comparison of various infrared (IR) spectroscopy techniques for calibrating poplar wood chemical composition and saccharification potential.

V. Segura^a, K. Ader^{a,b}, R. El Malki^a, V. Murciano^a, J P. Charpentier^{a,b}

^a INRA, UR0588 Amélioration Génétique et Physiologie Forestières, Orléans, France

^b INRA, Plateau technique Génomique, Orléans, France

Corresponding author : segura@orleans.inra.fr

Introduction

The development and use of sustainable biofuels from renewable resources, such as lignocellulosic biomass, has become a priority in order to tackle the increasing demand for energy, together with the concerns about the negative effects of greenhouse gas emissions. Realizing this potential will require the simultaneous development of high yielding biomass production systems and bioconversion technologies that efficiently convert biomass into usable forms of energy, such as bioethanol (Ragauskas et al., 2006; Rubin, 2008).

With a conservative average yield of ~ 10 dry Mg ha⁻¹ year⁻¹ of lignocellulosic biomass, short-rotation coppice poplar has a strong potential as a bioenergy feedstock (Ragauskas et al., 2006; Sannigrahi et al., 2010). Nevertheless, current poplar varieties have not been bred for this purpose, and thus a significant improvement in quantity and quality of biomass production is expected to be achieved by integrating biorefinery related selection criteria into poplar breeding programs. Biofuel yield results from both the amount of biomass to be converted into biofuel, and from the efficiency of the conversion. Wood chemical properties have been shown to significantly affect biomass saccharification, *i.e.* its conversion into simple fermentable sugars (Davison et al., 2006; Studer et al., 2011). It is thus crucial to understand better the genetic architecture of wood chemical properties. IR spectroscopy is relevant for this purpose because it allows a high-throughput indirect characterization of the physical and chemical properties of the samples of interest, the number of which is typically very large in breeding programs. It has successfully been applied to predict lignin content (Zhou et al., 2011) and lignin monomer composition (Robinson and Mansfield, 2009) in hybrid poplars. In the context of biofuels, calibrations for the saccharification of various biomasses have recently been reported (Sills and Gosset,

2012a; 2012b), highlighting the potential interest of this technique for screening the saccharification potential of biomass samples in breeding programs.

In this context, we have recently started to use IR spectroscopy for predicting poplar biomass chemical properties and saccharification potential. With the aim of defining a relevant way of collecting IR spectra, the present study reports calibration models for 4 different conditions of spectra acquisition, including mid-infrared (MIR) vs. near-infrared (NIR).

Material and methods

The samples used in the present study consisted of 479 wood powders (50 μ m – 1 mm) that came from the milling of ~ 50 cm 1-year-old stems from a F1 mapping population tested in the nursery of INRA Orleans. A calibration set comprising 100 samples was selected from preliminary NIR spectra (8000-4000 cm⁻¹) collected in rotating cups with a spectrum 400 Perkin Elmer spectrophotometer. These samples were selected to be representative of the NIR spectral variability using the Kennard-Stone algorithm (Kennard and Stone, 1969).

Four different conditions of acquisition of spectra were tested on this calibration set: (i) NIR (8000-4000 cm⁻¹) spectra averaged over 64 scans collected with a rotating system in a quartz cup (min 10 g of wood powder); (ii) NIR (8000-4000 cm⁻¹) spectra averaged over 3 independent collections of scans in a glass vial (min 2 g of wood powder); (iii) NIR (8000-4000 cm⁻¹) spectra from a single collection of scans in a glass vial (min 2 g of wood powder); and (iv) MIR (1800-900 cm⁻¹) spectra from a single collection of scans (several mg of wood powder).

The 100 samples from the calibration set were also phenotyped for 12 quantitative traits related to their chemical

composition and saccharification potential (Table 1). For chemical composition, the amount of extractives and total (soluble and insoluble) lignins was assessed following standard protocols, and the amount of holocellulose (cellulose and hemicellulose) was deduced by difference, considering that the sum of these tree components was equal to 1 (Carrier et al., 2011). For saccharification potential, the amount of sugars and glucose was measured before and after enzymatic hydrolysis by a set of enzymes extracted and purified from the ascomycete fungus *Trichoderma reesei* using an automated assay (Navarro et al., 2010). From these measured phenotypes, new phenotypes were calculated in order to get a better insight into the factors affecting saccharification potential (Table 1).

Prior to the calibration, the spectra were statistically pre-treated in order to improve the signal quality. Statistical pre-treatments consisted of normalization and/or derivation, yielding 8 modes for each acquisition condition: "raw", no pre-treatment; "norm", normalization (centering and scaling); "der 1", first derivative using a 37 points Savitzky-Golay filter; "der 2", second derivative using a 61 points Savitzky-Golay filter; "norm_der1", normalization followed by first derivative; "norm_der2", normalization followed by second derivative; "der1_norm", first derivative followed by normalization; "der2_norm", second derivative followed by normalization. For each mode of each acquisition condition and each phenotype, the calibration procedure was carried out in 2 steps: (i) selection of relevant spectral bands from the IR spectra matrix in order to decrease its complexity using the Competitive Adaptive Reweighted Sampling (CARS) approach (Li et al., 2009); (ii) on the resulting dataset, calibration model construction using partial least squares (PLS) regression in a 4-fold cross-validation scheme repeated 500 times (Monte Carlo Cross-Validation, MCCV), where the optimal number of latent variables was determined using a Wold's R criterion equal to one (Wold, 1980). The best model among all 8 modes was determined for each combination of phenotype and acquisition condition on the basis of both the MCCV statistics (R^2_{cv} : cross-validation R^2 ; $RMSE_{cv}$: root mean square error of cross validation; and RPD_{cv} : Ratio of standard error of performance in cross validation to standard deviation) and the number of latent variables. Afterwards, for each phenotype the best models of each acquisition condition were compared and ranked according to the same criteria (MCCV statistics and number of latent variables). All analyses were carried out in R (R Development Core Team, 2012).

Results and discussion

The best calibration models for each acquisition condition are presented in Tables 2 and 3 for the phenotypes related to wood chemical composition and saccharification potential, respectively. Whatever the acquisition condition, the calibration models for the amount of extractives had RPD_{cv} values close to 2. For lignins and holocellulose content, model quality was poorer, with RPD_{cv} values ranging between 1.55 and 1.95. For saccharification potential related traits, the best calibration

models were obtained whatever the acquisition condition for the soluble sugars (Soluble_Sug, Soluble_Gluc_Soluble_Gluc_Prop), with RPD_{cv} values ranging between 2.42 and 3.91, which was consistent with the results for the extractives. The calibration models for the proportion of solubilized glucose were of similar quality for all acquisition conditions, with RPD_{cv} values higher than 2.5. For all the other phenotypes, the calibration model quality was in general poorer with RPD_{cv} ranging between 1.28 and 2.20. Thanks to the prior spectral band selection with the CARS algorithm, all selected calibration models were characterized by a fairly low number of latent variables, ranging between 1 and 7.

For each phenotype, we then compared and ranked the selected models between the different acquisition conditions. A graphical representation of the rank frequencies per acquisition condition is presented in Figure 1. With a rank of 1 for half of the phenotypes (6) and of 2 for 4 phenotypes, an average spectrum over 3 acquisitions in a glass vial seemed to be the most promising approach for calibrating our phenotypes. It was substantially better than a single collection in a glass vial which reached the rank 1 only once and the rank 2 twice. The rotating cup system performed surprisingly worse than 3 acquisitions in glass vial, despite the fact it integrates spectra over multiple points. Nevertheless, this acquisition condition was, as expected, better than a single collection in a glass vial. The MIR condition consisted also of a single scan but it ranked better than a single scan in a glass vial, with a rank of 1 or 2 for half of the phenotypes. This acquisition condition had a comparable performance to the collection in a rotating cup, and thus it gave models of slightly worse quality than 3 acquisitions in vial, suggesting that in the case of MIR an average of 3 independent collections appears to be the most promising acquisition technique.

These comparisons should nevertheless be taken with care because the 12 phenotypes under study were not completely independent, with significant correlations between some of them (data not shown). It would thus be interesting to extend and confirm the present results with other phenotypes that would not be correlated with the ones used in the present study. However, the present results still allow us to conclude that, despite the fact it involves the advantage of analyzing only a small amount of material, spectral acquisition on a single point in a glass vial does not seem satisfactory for our purpose, and a spectrum averaged over multiple scans either in a rotating cup or preferably in a glass vial should be recommended.

Acknowledgements

The present work was carried within the FUTUROL project funded by OSEO Innovation. We thank David Navarro and Jean-Guy Berrin for their contribution in the acquisition of phenotypes related to the saccharification potential.

References

- Carrier M, Loppinet-Serani A, Denux D, Lasnier J-M, Ham-Pichavant F, Cansell F, Aymonier C (2011). Thermogravimetric analysis as a new method to determine the lignocellulosic composition of biomass. *Biomass and Bioenergy*, **35**, 298-307.
- Davison BH, Drescher SR, Tuskan GA, Davis MF, Nghiem NP (2006). Variation of S/G ratio and lignin content in a Populus family influences the release of xylose by dilute acid hydrolysis. *Applied biochemistry and biotechnology*, 129-132, 427-435.
- Kennard RW, Stone LA (1969). Computer aided design of experiments. *Technometrics*, **11**, 137-148.
- Li H, Liang Y, Xu Q, Cao D (2009). Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. *Analytica Chimica Acta*, **648**, 77-84.
- Navarro D, Couturier M, da Silva G, Berrin J-G, Rouau X, Asther M, Bignon C (2010). Automated assay for screening the enzymatic release of reducing sugars from micronized biomass. *Microbial Cell Factories*, **9**, 58.
- R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ, Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, Tschaplinski T (2006). The path forward for biofuels and biomaterials. *Science*, **311**, 484-489.
- Robinson AR, Mansfield SD (2009). Rapid analysis of poplar lignin monomer composition by a streamlined thioacidolysis procedure and near-infrared reflectance-based prediction modeling. *The Plant Journal*, **58**, 706-714.
- Rubin EM (2008). Genomics of cellulosic biofuels. *Nature*, **454**, 841-845.
- Sannigrahi P, Ragauskas AJ, Tuskan GA (2010). Poplar as a feedstock for biofuels: A review of compositional characteristics. *Biofuels, Bioproducts and Biorefining*, **4**, 209-226.
- Sills DL, Gossett JM (2012a). Using FTIR to predict saccharification from enzymatic hydrolysis of alkali-pretreated biomasses. *Biotechnology and Bioengineering*, **109**, 353-362.
- Sills DL, Gossett JM (2012b). Using FTIR spectroscopy to model alkaline pretreatment and enzymatic saccharification of six lignocellulosic biomasses. *Biotechnology and Bioengineering*, **109**, 894-903.
- Studer MH, DeMartini JD, Davis MF, Sykes RW, Davison B, Keller M, Tuskan GA, Wyman CE (2011). Lignin content in natural Populus variants affects sugar release. *Proceedings of the National Academy of Sciences*, **108**, 6300-6305.
- Wold H (1980). Model construction and evaluation when theoretical knowledge is scarce. Theory and application of partial least squares. In: J Kmenta, JB Ramsey (Eds.), *Evaluation of Econometric Models*, (pp. 47 – 74). New York, Academic Press.
- Zhou G, Taylor G, Polle A (2011). FTIR-ATR-based prediction and modelling of lignin and energy contents reveals independent intra-specific variation of these traits in bioenergy poplars. *Plant Methods*, **7**, 9.

Figures and tables

Table 1. List and definition of the 12 quantitative traits phenotyped on the 100 calibration samples of *P. nigra* wood.

Trait Name	Trait Identifier	Formula
Chemical properties		
Extractives content (%)	Extract	
Lignin content (%)	Lignin	
Holocellulose content (%)	Holocellulose	100 – (Extract + Lignin)
Sacharification Potential		
Soluble Sugars		
Total Sugars (µmol /mg)	Soluble_Sug	
Glucose (µmol /mg)	Soluble_Gluc	
Proportion of glucose (%)	Soluble_Gluc_Prop	Soluble_Gluc / Soluble_Sug
Solubilized Sugars		
Total Sugars (µmol /mg)	Solubilized_Sug	
Glucose (µmol /mg)	Solubilized_Gluc	
Proportion of glucose (%)	Solubilized_Gluc_Prop	Solubilized_Gluc / Solubilized_Sug
Hydrolyzed Sugars		
Total Sugars (µmol /mg)	Hydrolyzed_Sug	Solubilized_Sug – Soluble_Sug
Glucose (µmol /mg)	Hydrolyzed_Gluc	Solubilized_Gluc – Soluble_Gluc
Proportion of glucose (%)	Hydrolyzed_Gluc_Prop	Solubilized_Gluc_Prop – Soluble_Gluc_Prop

Table 2. Calibration models for the phenotypes related to chemical properties of *P. nigra* wood. For phenotype description see table 1. # comp: number of latent variables in the model; R^2_{train} : training R^2 ; R^2_{cv} : cross-validation R^2 ; $RMSE_{cv}$: root mean square error of cross validation; RPD_{cv} : Ratio of standard error of performance in cross validation to standard deviation; # lambda: number of wave numbers selected by CARS. For the cross validation statistics, the average over 500 replications is indicated together with the standard deviation in parenthesis.

Trait	Condition	Treatment	# comp	R^2_{train}	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	# lambda	rank
Extract	NIR-COUP	der1_norm	7	0.85	0.79 (0.01)	1.25 (0.04)	2.20 (0.07)	37	1
	NIR-VIAL1	norm_der1	5	0.80	0.74 (0.02)	1.39 (0.04)	1.97 (0.06)	40	4
	NIR-VIAL3	norm_der2	3	0.79	0.75 (0.01)	1.36 (0.04)	2.02 (0.05)	20	2
	MIR	der2_norm	2	0.77	0.74 (0.01)	1.38 (0.02)	1.98 (0.03)	59	3
Lignin	NIR-COUP	der2_norm	4	0.72	0.65 (0.02)	0.95 (0.03)	1.69 (0.05)	335	1
	NIR-VIAL1	der2_norm	4	0.71	0.62 (0.02)	0.98 (0.03)	1.64 (0.04)	204	4
	NIR-VIAL3	der2	3	0.66	0.62 (0.01)	0.98 (0.02)	1.63 (0.03)	37	3
	MIR	norm	6	0.73	0.63 (0.02)	0.97 (0.03)	1.66 (0.05)	205	2
Holocellulose	NIR-COUP	der1	5	0.75	0.67 (0.02)	1.42 (0.04)	1.76 (0.05)	35	2
	NIR-VIAL1	norm_der1	6	0.81	0.73 (0.02)	1.28 (0.04)	1.95 (0.06)	40	1
	NIR-VIAL3	norm	5	0.70	0.66 (0.01)	1.45 (0.03)	1.71 (0.04)	79	3
	MIR	norm	1	0.60	0.58 (0.01)	1.61 (0.02)	1.55 (0.02)	12	4

Table 3. Calibration models for the phenotypes related to saccharification potential of *P. nigra* wood. For phenotype description see table 1. # comp: number of latent variables in the model; R^2_{train} : training R^2 ; R^2_{cv} : cross-validation R^2 ; $RMSE_{cv}$: root mean square error of cross validation; RPD_{cv} : Ratio of standard error of performance in cross validation to standard deviation; # lambda: number of wave numbers selected by CARS. For the cross validation statistics, the average over 500 replications is indicated together with the standard deviation in parenthesis.

Trait	Condition	Treatment	# comp	R^2_{train}	R^2_{cv}	$RMSE_{cv}$	RPD_{cv}	# lambda	rank
Soluble_Sug	NIR-COUP	der1_norm	7	0.93	0.89 (0.01)	0.02 (0.00)	3.06 (0.11)	759	4
	NIR-VIAL1	norm	4	0.90	0.89 (0.00)	0.03 (0.00)	3.00 (0.06)	39	3
	NIR-VIAL3	norm	3	0.91	0.90 (0.00)	0.02 (0.00)	3.15 (0.05)	10	2
	MIR	norm	5	0.95	0.93 (0.00)	0.02 (0.00)	3.91 (0.11)	50	1
Soluble_Gluc	NIR-COUP	der1_norm	5	0.89	0.87 (0.01)	0.01 (0.00)	2.81 (0.06)	155	3
	NIR-VIAL1	der2	5	0.88	0.85 (0.01)	0.01 (0.00)	2.63 (0.07)	17	4
	NIR-VIAL3	norm_der1	5	0.90	0.88 (0.01)	0.01 (0.00)	2.92 (0.06)	11	2
	MIR	norm	6	0.94	0.92 (0.01)	0.00 (0.00)	3.50 (0.11)	98	1
Soluble_Gluc_Prop	NIR-COUP	der2_norm	6	0.90	0.83 (0.01)	0.06 (0.00)	2.46 (0.09)	552	4
	NIR-VIAL1	der1	4	0.86	0.83 (0.01)	0.06 (0.00)	2.42 (0.06)	37	3
	NIR-VIAL3	der2_norm	6	0.89	0.85 (0.01)	0.05 (0.00)	2.57 (0.08)	182	2
	MIR	norm_der2	5	0.88	0.86 (0.01)	0.05 (0.00)	2.65 (0.08)	61	1
Solubilized_Sug	NIR-COUP	norm_der1	5	0.75	0.69 (0.02)	0.04 (0.00)	1.82 (0.05)	17	3
	NIR-VIAL1	norm	6	0.76	0.69 (0.02)	0.04 (0.00)	1.81 (0.05)	22	4
	NIR-VIAL3	norm	6	0.78	0.73 (0.02)	0.04 (0.00)	1.94 (0.05)	48	1
	MIR	norm	5	0.75	0.70 (0.02)	0.04 (0.00)	1.84 (0.05)	20	2
Solubilized_Sug	NIR-COUP	der2	6	0.76	0.68 (0.02)	0.03 (0.00)	1.79 (0.06)	46	2
	NIR-VIAL1	der1	4	0.71	0.66 (0.02)	0.03 (0.00)	1.72 (0.04)	34	3
	NIR-VIAL3	der2	6	0.82	0.74 (0.02)	0.02 (0.00)	1.97 (0.07)	90	1
	MIR	norm_der1	6	0.72	0.66 (0.02)	0.03 (0.00)	1.73 (0.04)	16	4
Solubilized_Gluc_Prop	NIR-COUP	der2	5	0.89	0.86 (0.01)	0.03 (0.00)	2.65 (0.07)	66	3
	NIR-VIAL1	der2	3	0.87	0.84 (0.01)	0.04 (0.00)	2.55 (0.06)	59	4
	NIR-VIAL3	norm_der1	4	0.89	0.87 (0.01)	0.03 (0.00)	2.74 (0.07)	65	1
	MIR	norm	3	0.88	0.86 (0.01)	0.03 (0.00)	2.65 (0.06)	427	2
Hydrolyzed_Sug	NIR-COUP	der1_norm	7	0.72	0.60 (0.03)	0.04 (0.00)	1.59 (0.05)	54	3
	NIR-VIAL1	norm_der2	6	0.73	0.62 (0.02)	0.04 (0.00)	1.64 (0.05)	50	2
	NIR-VIAL3	der1	7	0.72	0.63 (0.02)	0.04 (0.00)	1.66 (0.05)	26	1
	MIR	norm	3	0.49	0.39 (0.03)	0.05 (0.00)	1.28 (0.03)	459	4
Hydrolyzed_Gluc	NIR-COUP	norm_der2	6	0.81	0.75 (0.02)	0.03 (0.00)	2.00 (0.07)	78	2
	NIR-VIAL1	der1_norm	5	0.78	0.73 (0.01)	0.03 (0.00)	1.92 (0.05)	23	3
	NIR-VIAL3	der2_norm	6	0.85	0.79 (0.01)	0.02 (0.00)	2.20 (0.07)	65	1
	MIR	norm_der1	3	0.74	0.71 (0.01)	0.03 (0.00)	1.86 (0.04)	32	4
Hydrolyzed_Gluc_Prop	NIR-COUP	der2	4	0.69	0.63 (0.02)	0.05 (0.00)	1.65 (0.04)	56	3
	NIR-VIAL1	norm	6	0.71	0.64 (0.02)	0.04 (0.00)	1.68 (0.05)	22	2
	NIR-VIAL3	norm	5	0.74	0.68 (0.02)	0.04 (0.00)	1.77 (0.05)	15	1
	MIR	der1_norm	3	0.64	0.61 (0.01)	0.05 (0.00)	1.62 (0.03)	4	4

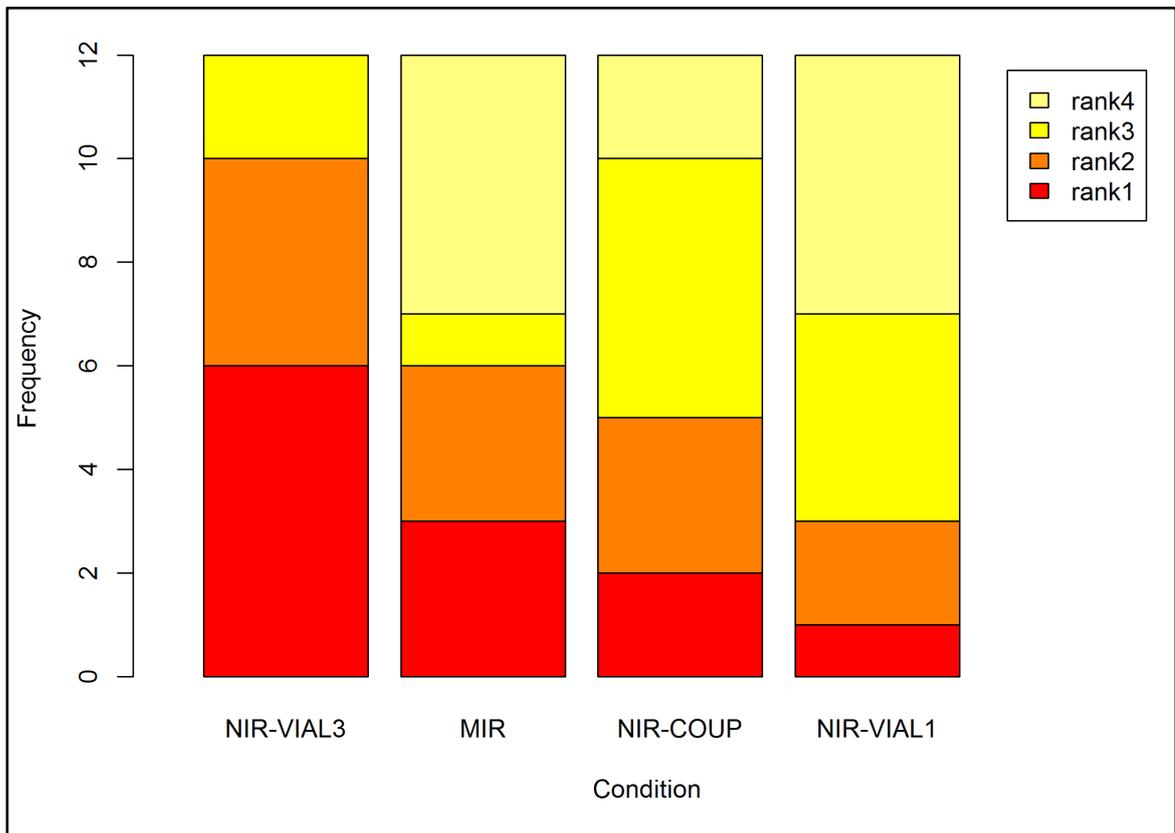


Figure 1. Graphical representation of the frequency of calibration model ranks depending on the spectra acquisition condition over 12 phenotypes related to the chemical properties and saccharification potential of *P. nigra* wood.