

MIRS & Pectin Content Analyses for Cassava Cooking Properties

Montpellier, France, 04/10/2023

Karima MEGHAR, CIRAD, Montpellier, France

Romain DOMINGO, CIRAD, Montpellier, France

Léa OLLIER, CIRAD, Montpellier, France

Santiago ARUFE VILAS, CIRAD, Montpellier, France



This report has been written in the framework of the RTB Breeding project, Quality-component (under CIRAD coordination), as a continuation of work initiated under the RTBfoods project.

To be cited as:

Karima MEGHAR, Romain DOMINGO, Léa OLLIER, Santiago ARUFE VILAS (2023). *MIRS & Pectin Content Analyses for Cassava Cooking Properties*. Montpellier, France: RTB Breeding, Scientific Report, 14 p. <https://doi.org/10.18167/agritrop/00849>

Ethics: The activities, which led to the production of this document, were assessed and approved by the CIRAD Ethics Committee (H2020 ethics self-assessment procedure). When relevant, samples were prepared according to good hygiene and manufacturing practices. When external participants were involved in an activity, they were priorly informed about the objective of the activity and explained that their participation was entirely voluntary, that they could stop the interview at any point and that their responses would be anonymous and securely stored by the research team for research purposes.

Acknowledgments: This work was supported by the RTB Breeding project, through a sub-grant from the International Potato Center (CIP) to the French Agricultural Research Centre for International Development (CIRAD), Montpellier, France, incorporated in the grant agreement INV-041105 between CIP and the Bill & Melinda Gates Foundation (BMGF).

Image cover page © CIRAD

This document has been reviewed by

--	--

Final validation by

Santiago ARUFE VILAS (CIRAD)

23/02/2025

TABLE OF CONTENTS

List of Figures.....	5
List of Tables.....	5
1 Introduction.....	7
2 Materials and methods.....	7
2.1 Material.....	7
2.2 Experimental design.....	7
2.3 Mid-infrared spectroscopy measurement.....	7
2.4 Pectin content measurement.....	8
2.5 Multivariate analyses.....	8
3 Results.....	8
3.1 Cassava flours MIR spectra.....	8
3.2 Characteristic bands of pectin (galacturonic acid).....	9
3.3 Pectin content.....	10
3.4 Principal component analyses (PCA).....	10
3.5 ANOVA-simultaneous component analysis (ASCA).....	11
3.6 Prediction of pectin M content.....	12
4 Conclusion.....	14
5 References.....	14

LIST OF FIGURES

Figure 1 : Mean MIR spectra of cassava flours obtained by averaging spectra by storage time.....	9
Figure 2 : MIR spectra of cassava flours and pur galacturonic acid (pectin).....	9
Figure 3 : Boxplot of evolution of pectin M (a) and pectin T(b) during the storage.	10
Figure 4 : PCA scores plots PC1-PC2 (a), PC3-PC4 (b) colored by storage time	11
Figure 5 : PCA loadings plot of PC1, PC2, PC3.	11
Figure 6 : ASCA results of MIR spectra. Score plot (a) and loadings plot (b) of the storage submodel and Score plot (c) and loadings plot (d) of genotype submodel.	12

LIST OF TABLES

Table 1 : Statistic descriptive of pectin M and pectin T contents of cassava flours samples.....	10
Table 2 : Calibrations Statistics for the prediction of pectin M in cassava flours samples by MIRS	12

ABSTRACT

Polysaccharides are the main component of cell walls in plants. Among them, pectin, cellulose and hemicellulose are composed of various sugars residues than may affect the structure but also the behaviour and textural properties of roots, bananas and tubers. In raw products but also in cooked products. The main aim of this study was to determine pectin content of different cassava genotypes during storage at different conditions (-20°C and -80°C) and from different drying methods (air drying and freeze-drying) and the acquisition of Mid infrared spectroscopy (MIRS) spectra of these flours in order to evaluate the potential of MIRS for pectin content prediction in cassava flours. Different multivariate analyses such as PLSR, SVMR, CovSel-MLR were applied.

Keywords : Pectin, galacturonic acid, storage, MIRS, prediction, statistics

1 INTRODUCTION

In this study, the first objective was to investigate the effect of storage time, genotype and drying type on total pectins (Pectin T) and methylated pectins (Pectin M) of cassava flour by Mid infrared spectroscopy (MIRS) and pectin measurement. The resulted data was processed by chemometrics methods principal components analyses (PCA) and ANOVA-simultaneous component analysis (ASCA). The second objective was to investigate the potential of MIRS to predict pectin content in cassava flours. Different multivariate analyses were applied PLSR, SVMR, CovSel-MLR were applied.

2 MATERIALS AND METHODS

2.1 Material

Cassava flours from eight genotypes (8HMC, COL1505, COL1516, COL2215, CR63, IND135, PER183, SM1127) with contrasting cooking proprieties were provided in march 2023 from the International Center for Tropical Agriculture (CIAT), Cali (Colombia).

2.2 Experimental design

The samples of cassava flours were prepared in CIAT using two types of drying freeze drying and oven drying. Each sample was subjected in two portion and stored at -20°C , -80°C . MIRS and pectin analyses were performed at 0, 15, 30, 60, 90 and 120 days of storage. Making a total of 8 genotypes \times 2 drying types \times 2 storage temperatures \times 6 storage times samples.

2.3 Mid-infrared spectroscopy measurement

Mid-infrared (MIR) spectra provided higher signal-to-noise ratio, narrower and fewer overlapped absorption bands, minimal baseline changes between samples and provided characteristic fingerprint information improving the ability to resolve and identify specific functional groups as compared to NIR spectra.

MIR spectra are classified into two regions : fingerprint region $800\text{-}1800\text{ cm}^{-1}$ and functional group region $1800\text{-}4000\text{ cm}^{-1}$. The fingerprint ($800\text{-}1800\text{ cm}^{-1}$) region was used to develop calibration model for prediction pectin content, because this region provides unique spectroscopic information due to the vibrations of the C–C, C–H and C–O bonds, while the infrared region $1800\text{-}4000\text{ cm}^{-1}$ contained only broad O–H and aliphatic C–H absorptions which contain little relevant information (1).

In this experiment, Spectra were collected using Thermo Scientific™ Nicolet™ iS50R Research FTIR Spectrometers in the $4000\text{-}400\text{ cm}^{-1}$ region and resolution of 0.09 cm^{-1} . The acquisition system is controlled by Thermo Scientific™ OMNIC™ Series Software. This software provides the acquisition, saving of spectra and setting of parameters such as spectral range, number of scans, offset, gain, integration time, resolution, spectral path. Background spectrum of the air is collected before spectra acquisition. This spectra is used as reference for all measured spectra. Therefore, by using a small spatula, milled flour sample ($\sim 0.1\text{ g}$) is placed directly on the surface of the diamond ATR crystal by pressing the powder sample onto the crystal using a pressure clamp and spectrum is collected.

2.4 Pectin content measurement

Pectin content of cassava flours was measured in Montpellier by following the RTBfoods SOP : SOP for the determination of galacturonic content. Biophysical characterization of quality traits, WP2 (<https://agritrop.cirad.fr/602125/>).

2.5 Multivariate analyses

The multivariate analyses were performed using R (version 4.2.2) and chemflow softwares.

PCA was performed to extract and display the variation in MIR spectra of cassava flours preparing from different genotypes, drying type, storage temperature and storage time. PCA uses orthogonal projections to transform a set of spectra matrix with possibly correlated variables into a set of spatial orthogonal variables called principal components (PCs). In this study, the number of PCs used in the developed PCA models were manually selected.

Furthermore, in order to evaluate which factors are significant in explaining the experimental data, ANOVA-simultaneous component analysis (ASCA), a new approach that can deal with complex multivariate datasets containing an underlying experimental design was used. It is a generalization of analysis of variance (ANOVA) for univariate data to the multivariate case, which allows for easy interpretation of the variation induced by the different design factors. In ASCA analysis, PCA is applied to each ANOVA decomposition matrix, thus there is a PCA submodel for each factor, and the scores for each PCA submodel are scores for factor-level averages (2).

For the prediction of pectin content, different chemometrics methods were tested, partial least square regression (PLSR), K nearest neighbors regression (KNNR)(3), support vector machine regression (SVMR)(4) and CovSel multiple linear regression (CovSel_MLR)(5) in order to obtain the best model of prediction of pectin content in cassava flours.

3 RESULTS

3.1 Cassava flours MIR spectra

The fig1 represents mean spectra at 0, 15, 30, 60, 90 and 120 days of storage. It shows the characteristic MIR bands of cassava flour. Difference of absorbance intensities between spectra. No modification of characteristics bands during the storage. the important band at 995 cm^{-1} correspond to polysaccharids (starch). The band at 1153 cm^{-1} C-O-C stretching from glycosidic. The band 1645 cm^{-1} and symmetric stretching of the carboxylate anion (COO^-) mainly from pectins or due to esterification carboxylic, protein (1). The bands in the region 2300-2380 cm^{-1} is due to atmospheric absorption of N_2 and the wings of distant CO_2 lines (6). The band 2837-2850 cm^{-1} corresponds to symmetric and anti symmetric of lipids. The large band at 3000-3600 cm^{-1} corresponds to OH stretching mode of water.

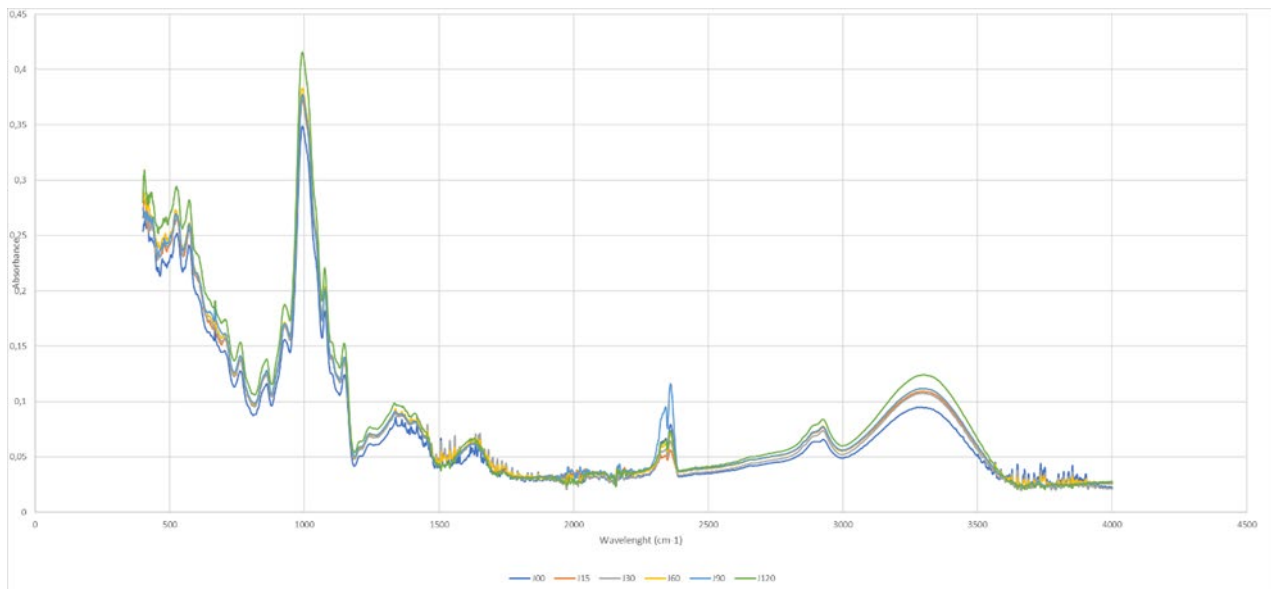


Figure 1 : Mean MIR spectra of cassava flours obtained by averaging spectra by storage time.

3.2 Characteristic bands of pectin (galacturonic acid)

The fig 2 shown, the raw spectra of cassava flour and pur galacturonic acid (pectin). The characteristics bands of pectin are located from 800 to 1800 cm⁻¹ and the 1028, 1094, 1182, 1284 and 1707 cm⁻¹ are the most important. Therefore, this spectral range is relevant for the prediction of pectin content by MIRS.

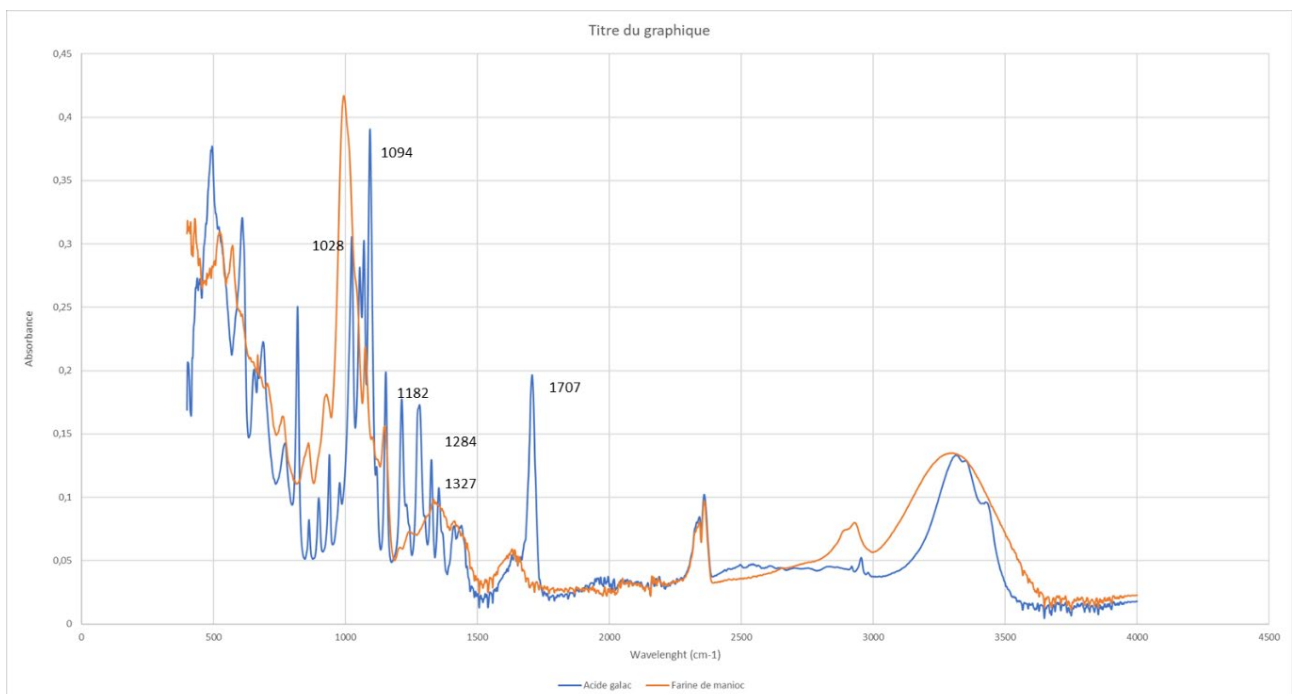


Figure 2 : MIR spectra of cassava flours and pur galacturonic acid (pectin).

3.3 Pectin content

The evolution of pectin content during the storage is shown in fig 3. The pectin M content (a) is stable from 0 to 60 days of storage but increase significantly at 90 days and stay high at 120 days. Regarding pectin T (b) content, the evolution is random and has no trend.

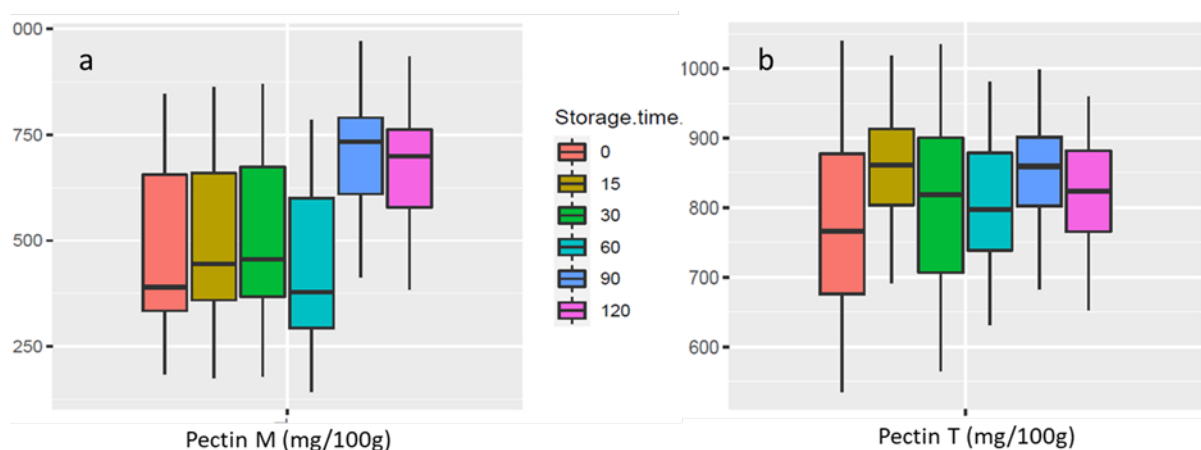


Figure 3 : Boxplot of evolution of pectin M (a) and pectin T(b) during the storage.

The Table 1 below contains the statistics descriptive of pectin M and pectin T content of 172 samples generated in this experiment. This result shows that the variability in pectin M is higher (827.76) than pectin T (504.41) content. This should provide better precision in MIRS calibration model.

Table 1 : Statistic descriptive of pectin M and pectin T contents of cassava flours samples.

Constituants	N	Mean	SD	Min	Max	Range
Pectin M (mg/100g)	172	555.32	199.52	142.29	970.06	827.76
Pectin T (mg/100g)	172	826.08	104.65	535.58	1039.99	504.41

3.4 Principal component analyses (PCA)

PCA was applied to raw MIR spectra (400-4000 cm⁻¹) in order to detect outliers and extract variation during the storage. fig 4 represents the resulting score plots from PCA analysis of raw MIR spectra using all samples and coloured by storage time. The fig 5 represents the PCA loadings plot of PC1,2 and 3 which explain 83.8, 8.9 and 3.5 % of total variance in the data respectively. According to the loading peak (992 cm⁻¹), PC1 explain the variation of starch in the samples, PC2 explain the variation of atmospheric absorption of N₂ and CO₂ with the characteristics band at 2359 cm⁻¹. PC3 seems explain instrumental noise. According to PC1, samples with positive part has a higher starch than those in negative part. A long PC1, the samples in positive side are affected by the atmospheric absorption during the spectra acquisition which are stored during 90 days at -80°C. In fig 4 a.b There are no trend and no clear grouping in function of storage time of samples.

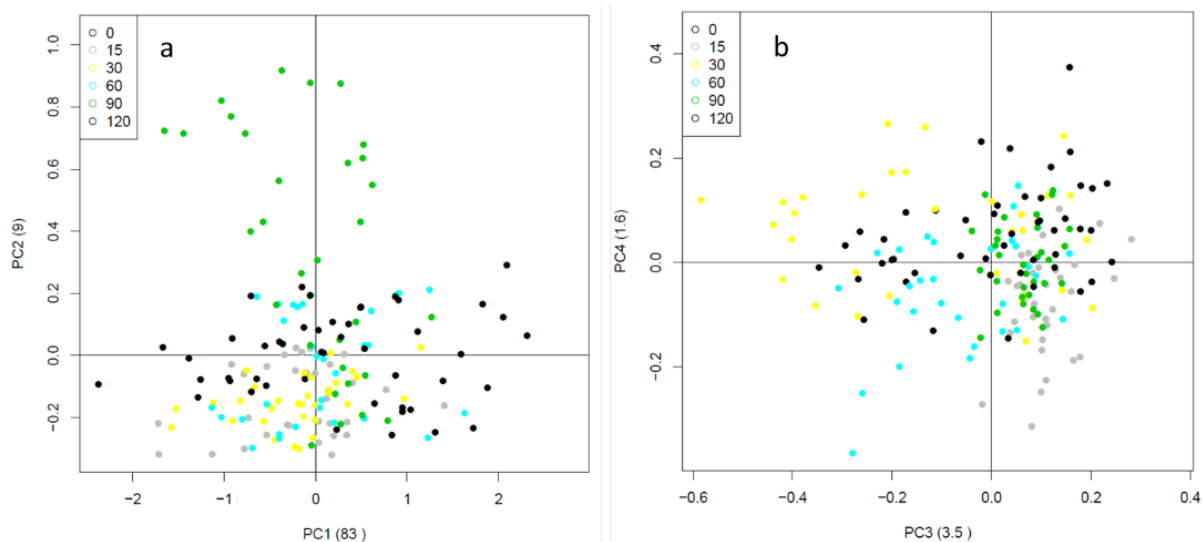


Figure 4 : PCA scores plots PC1-PC2 (a), PC3-PC4 (b) colored by storage time .

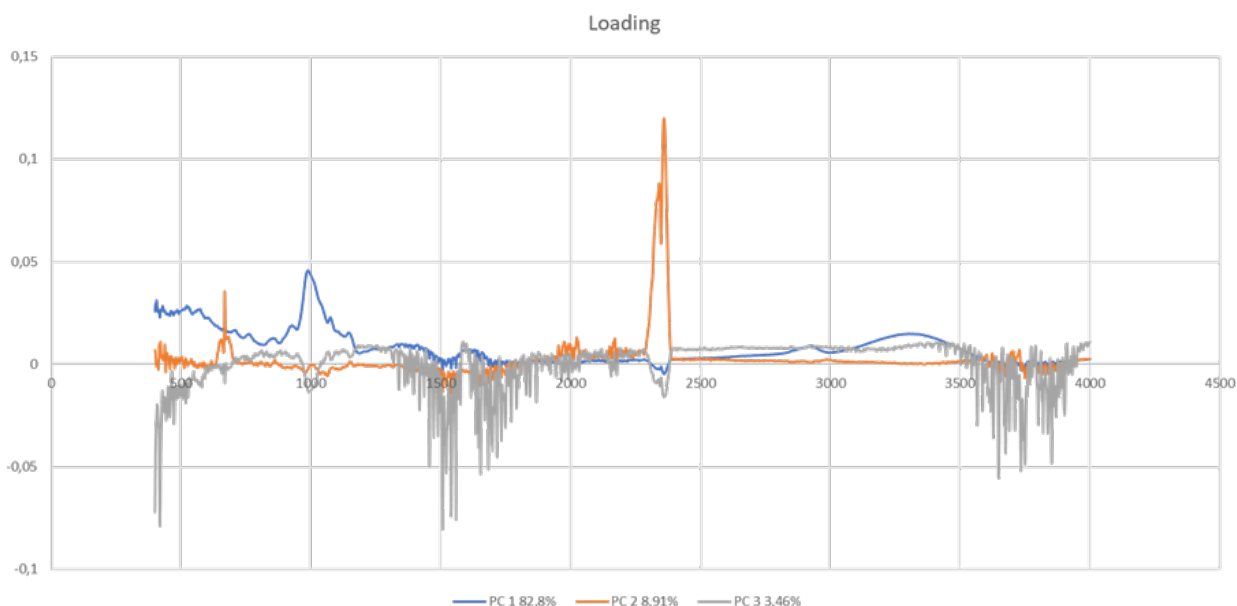


Figure 5 : PCA loadings plot of PC1, PC2, PC3.

3.5 ANOVA-simultaneous component analysis (ASCA)

ASCA analysis was applied to raw spectra on 800-1800 cm^{-1} and experimental design with 3 factors (Genotypes, drying and storage time). The explained variance obtained of genotypes, drying and storage time are 6.78, 10.03 and 28.08 % respectively. Storage time factor has a significant influence on MIR spectra and drying type and genotype are not significant.

The score plots and corresponding loading plot of storage and genotype obtained from ASCA are shown in Figure 6. According to the loadings (pc1, pc2) of storage submodel (Figure 6b), the observed variation (Figure 6a) during the storage seems to be is due to the evolution in amylose/starch content with an important band at 988 cm^{-1} .

The loading PC1 of genotype submodel (Figure 6c) indicate the observed variation between genotypes is maybe due to variation in amylose/starch content also with the characteristic band at 995 cm^{-1} . Moreover,

according to score plot of genotype submodel the starch content of CR63=PER183>COL2215>COL1505>SM1127>COL1516>IND135>.

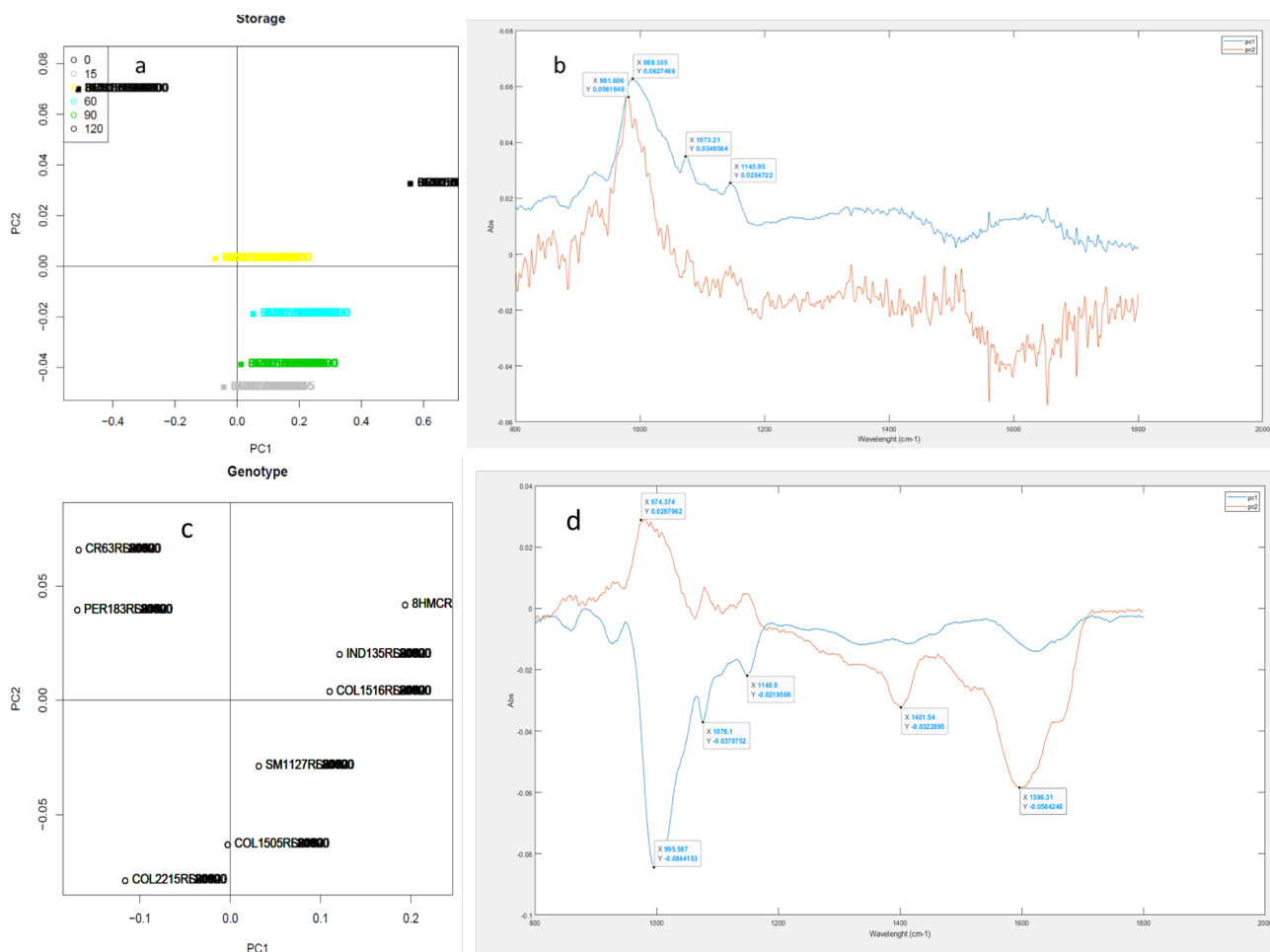


Figure 6 : ASCA results of MIR spectra. Score plot (a) and loadings plot (b) of the storage submodel and Score plot (c) and loadings plot (d) of genotype submodel.

3.6 Prediction of pectin M content

Calibration models (n=116) between pectin M content and MIR spectra in spectral region 800-1800 cm⁻¹ using PLSR, KNNR, SVMR and COVSEL_MLR. The calibration models developed were validated using the remaining samples (n=56). Statistics calculated included the coefficient of determination (R²p), the root mean error of prediction (RMSEP), and RPD. ATR-MIR spectral data was processed using standard normal variate (SNV). The performances of the models are shown in table 2 and the scatter plots of those models are presented in fig 7. The best calibration is developed using PLSR with R²p=0.88, RMSEP=70.46 and RPD=2.95. The RPD values depends on the context and on the purpose of the measurement. For breeding purposes, the RPD=SD ref/RMSEP value obtained (2.95) indicate that the PLSR calibration developed using MIRS spectra can be used only for screening of pectin M content (7).

Table 2 : Calibrations Statistics for the prediction of pectin M in cassava flours samples by MIRS

Model	Calibration				Validation			
	Ncal	Parameters	RMSEC	R ² cal	Nval	R ² p	RMSEP	RPD
PLSR	116	LV=12	36,29	0,96	56	0,88	70,46	2,95
KNNR	116		4,64	0,99	56	0,85	79,06	2,62
SVMR	116		182,86	0,55	56	0,58	194,86	1,06
Covsel_MLR	116	20 variables	62,55	0,89	56	0,79	95,66	2,16

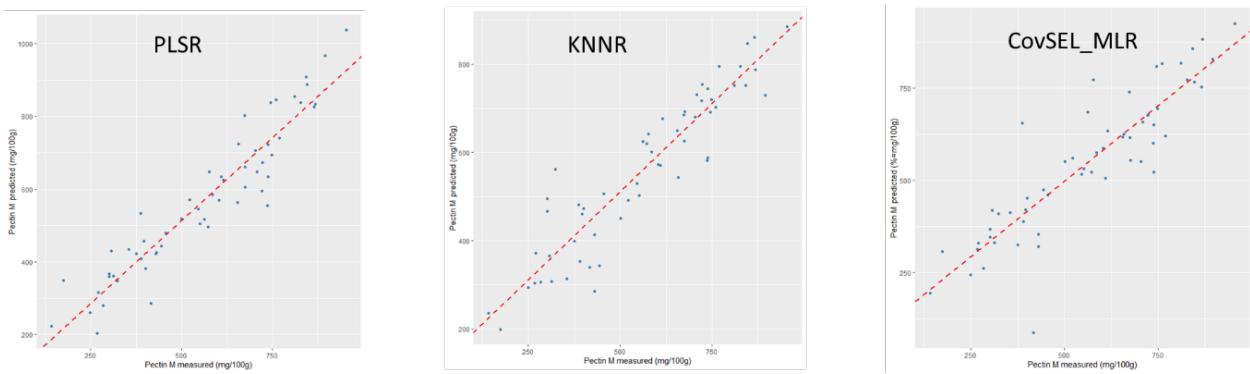


Figure 7 : Scatter plots of measured Pectin M versus predicted pectin M by MIRS using PSR, KNNR and CovSel_MLR.

4 CONCLUSION

The main conclusion of this study indicates that linear approach PLSR is the most relevant method for the pectin content prediction compared to no linear approaches (KNNR, SVMR) as well as variable selection approach (Covsel_MLR). The RPD value obtained (2.95) indicate that the PLSR calibration developed using MIRS spectra could be used only for screening of highly methylated pectin content.

MIRS spectra were not grouped according to storage time. Moreover, the effect of genotype and drying type were not significant in the variation of MIR spectra with thresholds error of 5 %. However, the storage with 28% of explained variability has a significant effect in the variability.

5 REFERENCES

1. Mestres, C., Ollier, L., Domingo, R., Otegbayo, O. B., Rolland-Sabaté, A., & Dutheil de la Rochère, A. (2022). SOP for the Determination of Galacturonic Content. Biophysical Characterization of Quality Traits (pp. 1–13). CIRAD. <https://doi.org/https://doi.org/10.18167/AGRITROP/00605>
2. Ayvaz H, Bozdogan A, Giusti MM, Mortas M, Gomez R, Rodriguez-Saona LE. Improving the screening of potato breeding lines for specific nutritional traits using portable mid-infrared spectroscopy and multivariate analysis. *Food Chem.* nov 2016;211:374-82.
3. Ryckewaert M, Gorretta N, Henriot F, Marini F, Roger JM. Reduction of repeatability error for analysis of variance-Simultaneous Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample. *Anal Chim Acta.* mars 2020;1101:23-31.
4. Lesnoff M, Metz M, Roger JM. Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. *J Chemom.* 2020;34(5):e3209.
5. Teng X, Zhang X, Yu H. Using Least Squares Support Vector Machines for Frequency Estimation. *Int J Commun Netw Syst Sci.* 27 oct 2010;3(10):821-5.
6. Roger JM, Palagos B, Bertrand D, Fernandez-Ahumada E. CovSel: Variable selection for highly multivariate and multi-response calibration: Application to IR spectroscopy. *Chemom Intell Lab Syst.* 15 avr 2011;106(2):216-23.
7. Susskind J, Searl JE. Atmospheric absorption near 2400 cm⁻¹. *J Quant Spectrosc Radiat Transf.* déc 1977;18(6):581-7.
8. Cozzolino D, Roumeliotis S, Eglinton J. Prediction of starch pasting properties in barley flour using ATR-MIR spectroscopy. *Carbohydr Polym.* juin 2013;95(1):509-14.