Calibration Report

Prediction of Yam Cooking Behavior Using Hyperspectral Imaging

High-Throughput Phenotyping Protocols (HTPP), WP3

Montpellier, France, 22/11/2022

Karima MEGHAR, Centre de coopération Internationale en Recherche Agronomique pour le Développement (CIRAD), Montpellier, France

Julien BOYER, CIRAD, Montpellier, France

Fabrice DAVRIEUX, CIRAD, Saint-Pierre, La Réunion, France

Graduation thesis, Master II CACQ-OPEX: Analytical Chemistry, Chemometrics, Quality - Optimization of Experimental Processes Université de Bretagne Occidentale (UBO), Brest, France

This report has been written in the framework of RTBfoods project.

To be cited as:

Karima MEGHAR, Julien BOYER, Fabrice DAVRIEUX (2022). *Prediction of Yam Cooking Behavior Using Hyperspectral Imaging, High-Throughput Phenotyping Protocols (HTPP), WP3.* Montpellier, France: RTBfoods Calibration Report, 25 p. https://doi.org/10.18167/agritrop/00707

Ethics: The activities, which led to the production of this manual, 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. Written consent (signature) was systematically sought from sensory panelists and from consumers participating in activities.

Acknowledgments: This work was supported by the RTBfoods project https://rtbfoods.cirad.fr, through a grant OPP1178942: Breeding RTB products for end user preferences (RTBfoods), to the French Agricultural Research Centre for International Development (CIRAD), Montpellier, France, by the Bill & Melinda Gates Foundation (BMGF).

This document has been reviewed by:	
Fabrice DAVRIEUX (CIRAD) Fabrice DAVRIEUX (CIRAD) Karima MEGHAR (CIRAD)	22/11/2022 24/11/2022 24/11/2022
Final validation by:	
Fabrice DAVRIEUX (CIRAD)	24/11/2022

CONTENTS

List of Figures

ABSTRACT

Context: This scientific report summarizes the Master OPEX internship dissertation (6 months April to October 2022) produced by Julien Boyer. Main results and conclusions are reported here.

Place: Montpellier, France

Date: 22/11/2022

Authors: Julien BOYER (UBO, master OPEX), Karima MEGHAR (CIRAD), Fabrice DAVRIEUX (CIRAD).

This scientific report is based on results obtained within the framework of the Master internship of Julien Boyer (OPEX, University of Western Brittany (UBO), Brest, France). The study aimed to produce classification and quantification models to predict the cooking ability of yams genotypes using hyperspectral images. For this study the yams samples were harvested at CIRAD research station in Guadeloupe. The work was carried out with 10 contrasting varieties of yam presenting cooking ability ranging from very poor to very good. Samples were analyzed for their spectral profiles (HIS) and for their Pectin, starch, dry matter contents and texture properties (hardness) using specific SOPs developed in RTBfoods project.

Different model based on average spectra and physico-chemical traits were developed in order to classify the genotypes according to their cooking ability and to quantify their pectin, starch and DM contents.

The study demonstrated that there is a significant correlation between hardness, measured using a penetrometer and starch content (r = -0,80), pectin content (r = 0.52) and Dry matter content (r = 0.55). no significant correlation exists between starch, pectin and DM and cooking ability except for highly contrasted varieties (very bad vs very good).

The models developed for predicting pectin, starch and texture content do not have sufficient performance for routine analysis, however these models can be improved by increasing the number of samples and the ranges of the constituents.

The local PLS classification model is promising for the yam cooking ability classification with a classification error of 19%. To improve this model, it is recommended to collect more samples with more varied and better distributed cooking qualities. Indeed, there was within the dataset an over representation of middle-class variety and not enough good or bad classes. The model developed for dry matter quantification is efficient with an error of prediction RMSEP = 2.67%.

This study demonstrated the potential of HSI for the selection of Yam genotypes according to some relevant traits such as dry matter content and cooking ability.

Keywords: Hyperspectral imaging, steam cooked yam, cooking ability, texture, dry matter, starch and pectin.

1 OBJECTIVES

1.1 Objective

The objective of this study is to produce classification and quantification models to predict the cooking ability of yams genotypes using hyperspectral images., in order to provide variety breeders with rapid and non-destructive tools for high throughput phenotyping of yam. Therefore, it is important in this project to establish a model from hyperspectral images because it is a fast, simple and non-destructive method that allows it to be used directly on fresh yam tubers in the field with little preparation for the samples and thus carry out a high-throughput varietal selection.

1.2 Specific objectives

The first specific objective focus on developing models, based on HSI data, to quantify the biochemical compounds related to cooking behavior. The targeted compounds are: pectin, starch and dry matter contents

The second specific objective focus on relation between HIS data and texture properties related to cooking behavior. The texture parameters are measured using a texturometer.

2 MATERIALS AND METHODS

2.1 Vegetal material

The yams samples were harvested at a CIRAD research station in Guadeloupe. In order to ensure a high variability, the work was carried out with 10 contrasting varieties of yam presenting cooking ability ranging from very poor to very good (Table 1). All the yams were harvested between April and May 2022. They were stored in a climatic chamber (Firlabo, SP-BVEHF, France) at a temperature of 16°C and with a humidity rate of 76%. For each variety, 6 tubers were sampled, 3 tubers are intended for biochemical analyzes and 3 tubers for texture analysis.

Table 1 : Cooking ability of the 10 Yam varieties

Variété	Qualitée de cuisson
14M	Moyenne
61F	Moyenne
74F	Très mauvaise
A107	Moyenne
A109	Moyenne
A48	Mauvaise
A54	Moyenne
A64	Moyenne
CIR	Bonne
Kabusa	Très bonne

2.2 Hyperspectral imaging

2.2.1 Camera Specim FX17

The FX17 Hyperspectral camera is a complete, multi-purpose, turn-key hyperspectral imaging instrument designed for industrial and laboratory use. It works in what is known as a push-broom

mode, and collects hyperspectral data in the NIR (900 to 1700 nm) region through single fore optics. Each FX17 unit has been factory calibrated for optimum performance (including spectral wavelength calibration and automatic image enhancement operations. The main component (fig .1) of the system are: (1) PGP optical structure spectrograph (ImSpector, N17E, SPECIM, Finland), (2) 12-bit CCD camera (V-light, Lowel Light Inc, USA), (3) 150 W tungsten halogen lamps (Fibre-Lite DC950 Illuminator, Dolan Jenner Industries Inc., Boxborough, MA, USA) and (4) a moving plate LabScanner with dimensions (L × W) of 40 × 20 cm. The operation of the system is ensured by the use (5) of the control software LumoScanner (SPECIM, Finland).

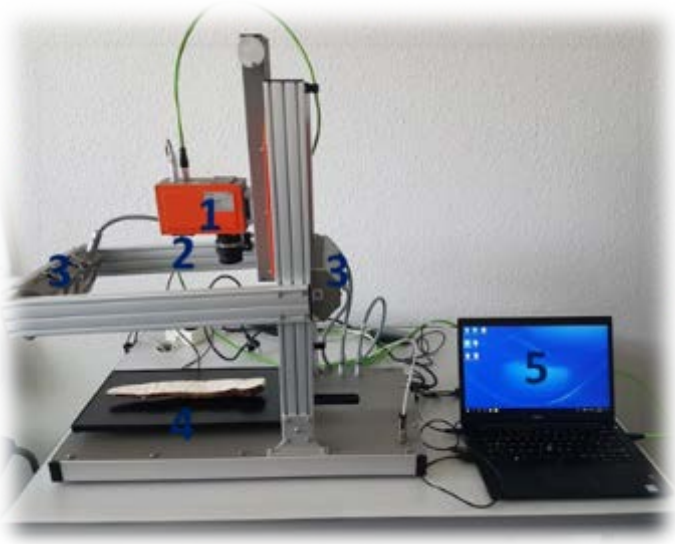


Figure 1 : Camera Specim FX17 system

2.2.2 Images acquisition

Several steps are required for HSI acquisition, the procedure is described in the SOP (https://doi.org/10.18167/agritrop/00667 , K. Meghar and F. Davrieux, 2021). First, the device must be calibrated using the LumoScanner acquisition software (SPECIM, Finland), checking the sharpness of the image using a test image. Then, adjust the acquisition speed to avoid having a distorted image. Finally, the position of the sample and the white reference on the plate must be defined (Annex II).

Once the device is calibrated, image acquisition can begin. A white reference and a "black" image, which corresponds to the noise of the device, are acquired, then the hyperspectral image of the sample is produced. To obtain a correct image of the sample, it must be corrected. The image is corrected with the white reference and the black reference.

The spectrophotometer and camera have a spectral range of 935 to 1720 nm with a spectral resolution of 8 nm, the exposure time is set at 11.5 ms for the entire duration of the test. The distance between the objective and the surface of the yam tuber is fixed at 22 cm, the scanning speed is 8.3 mm/s. After completing the scans on a tuber, a three-dimensional (x,y,z) hypercube is obtained, with the spatial dimensions (x,y) and spectral dimension (z).

The image is processed as following: correction (remove defaults), thresholding, segmentation, unfolding of the hypercube and calculation of the average spectrum. The processing operations are performed using a specific Python script (Spyder, 5.5.2). The average spectra will be used for modelling (Annex III).

2.3 Samples preparation

2.3.1 Samples for biochemical analysis

Tubers intended for biochemical analyzes (pectin and starch dry matter content) were peeled, washed and wiped with paper to remove dirt from the surface. Then, each tuber was cut into three

zones (proximal, central and distal) and then each zone was cut into 2 cylindrical slices 1 cm thick using a large piece (fig. 2). Hyperspectral images are directly acquired on the slices of the fresh samples before laboratory analysis.

Figure 2 : Preparation of the samples for both HSI and Biochemical analysis

The preparation of the yam samples was carried out according to the RTBfoods Standard Operating Procedure (SOP) Sample Preparation and Cooking Time for Texture Analysis of Boiled Yam. This SOP recommends sampling in the form of yam cubes with 2.3 cm edges. This cube format was determined to standardize the cooking and cooling steps necessary for texture analyses. As with tubers for biochemical analyses, tubers for texture analyzes were peeled, washed and wiped with paper to remove dirt from the surface. Then, each tuber was cut into three zones (proximal, central and distal) then, each zone was cut into 3 cubes 2.3 cm thick which were removed using a large piece. A hyperspectral image is acquired on each cube before cooking and texture analysis.

2.3.2 Samples for texture analysis

The preparation of the yam samples was carried out according to the RTBfoods SOP “Sample Preparation and Cooking Time for Texture Analysis of Boiled Yam” (https://doi.org/10.18167/agritrop/00666 , K. Meghar, F. Davrieux, E. Alamu, 2020) . This SOP recommends sampling in the form of yam cubes with 2.3 cm edges. This cube format was determined to standardize the cooking and cooling steps necessary for texture analyses. As with tubers for biochemical analyses, tubers for texture analyzes were peeled, washed and wiped with paper to remove dirt from the surface. Then, each tuber was cut into three zones (proximal, central and distal) then, each zone was cut into 3 cubes 2.3 cm thick using a cookie cutter (fig.3). A hyperspectral image is acquired on each cube before cooking and texture analysis.

Figure 3 : Yam cubes for texture analysis

A general scheme of samples preparation is given in annex I.

2.4 Reference analysis

2.4.1 Texture analysis

Texture analyzes were performed following RTB Foods SOP “Sample Preparation and Cooking Time for Texture Analysis of Boiled Yam” (https://doi.org/10.18167/agritrop/00603, Adinsi L., Honfozo Fifamè L., Akissoé N., Dahdouh L., 2020). The texture analysis was carried out with the Texturometer (TA-XT plus, Stable Micro systems Ltd. Surrey, UK) with a conical probe P/40C on samples at 45°c. During the analysis, the yam samples were placed so that the conical probe

penetrated them in the direction of the fibers (Annex II). The parameters for texture measurements were:

Pré-test Speed	1 mm/s
Test speed	0.5 mm/s
Trigger force	5g
Target distance	10 mm

2.4.2 Dry matter quantification

The sliced yam samples (§ 2.3.1) are freeze-dried after hyperspectral images acquisition. Freeze-drying is a two-step sample drying procedure. The first step is to remove the free water contained in the samples. For this, the samples are placed on trays at -30°C (in small cups) for 24 hours until the samples reach this temperature. Once this temperature is reached, the pressure decreases in the freeze-dryer then the temperature of the samples is gradually raised to 10°C to release the bound water from the samples.

After the freeze-drying step, the samples are weighed again to obtain the water content by weight difference between before and after freeze-drying. About ten witnesses are used to obtain the residual water content after the lyophilization. These witnesses are placed in the oven for 48 hours. Once this residual content has been obtained, the water contents of the samples are corrected.

2.4.3 Pectin quantification

The pectin quantification was carried out according to an adaptation of the method developed by Blumenkrantz and Asboe-Hansen in 1973. Under the action of hot concentrated sulfuric acid, the polysaccharides are hydrolyzed and transformed into furfuralic derivatives, which, put in a solution of NaOH and MHDP can be quantified by an absorbance reading. Extraction makes it possible to obtain acid sugars (galacturonic acids) from the hydrolysis of pectin and neutral sugars, mainly from the hydrolysis of residual starches. An extraction buffer (pH10 + EDTA) is used to extract the pectin from each of the yam flours. Approximately 250 mg (+/- 0,1 mg) of each flour were dissolved in 10 mL of this buffer, then stirred for 1 hour in a water bath at 55°C. Finally, after centrifugation for 10 min at 25° C. at 4000 G, the supernatants were recovered and used for the pectin assay.

The supernatants obtained after extraction were analyzed with a continuous flow analyzer (SKALAR, the Netherlands) to automate the acid hydrolysis of the polysaccharides, the contact with the colored reagent and the absorbance measurements. In order to relate the absorbance data to the concentrations of Ac. Gal., a standard range of galacturonic acid was analyzed under the same conditions. Finally, a reading of the absorbance of the samples in a soda solution without MHDP was carried out to serve as a "blank" and correct the results obtained. This corrects the effect of the natural coloring of the supernatant.

The flow analyzer software translates the measured absorbances into “corrected heights”, taking the form of peaks being calculated relative to the strongest and weakest absorbance signals (baseline). Since other compounds (minority) present in our samples absorb at 520 nm, only the maximum height of the peak (corresponding to the highest absorbance and resulting from the specific reaction between MHDP and galacturonic acids) is considered.

The content of extracted galacturonic acids was expressed in mg per 100g of yam flour, using the following formula:

$$[\textbf{Acid } \textit{Galacturoniques}] \textbf{ (mg/ 100g) } = \alpha V/m * 100 * 1000$$

With:

α: the coefficient allowing the conversion of the corrected heights into [Galacturonic Acids] (mg/L) obtained from the results of the standard range.

V: the extraction volume = $10*10^{-3}$ L

m: the exact mass of sample used for the extraction (mg).

Two replicates were carried out, then the average value for each samples were computed.

2.4.4 Starch content

Starch is a major reserve carbohydrate of higher plants and occurs in the form of water-insoluble granules. As the starch polymer is complex, the combination of enzymes is required to break it down.

The analysis of total starch is carried out on yam flours using a method for determining carbohydrates adapted from the Holm method. (JCS, 1985). This method uses several enzymes that will allow the digestion of starch to transform it into free glucose. Once the glucose is released, it is assayed by absorbance using a reagent that will color the solution.

Approximately 400 mg (+/- 0,1mg) of flour is weighed into an Erlenmeyer flask and then the powder is dispersed in 30 ml of distilled water. Then, 100 µl of a first α-amylase enzyme is added and the suspension is incubated for 25 min at 98°C. After 25 min, the samples is taken out of the water bath and left to cool. The cooled suspension is transfered the contents of the Erlenmeyer flask to a 100 ml volumetric flask and the volume is completed with distillated water. Then, 10 ml of the sample is centrifuged for 5 mins at 3500 rpm.

After centrifugation, 500 µl of the supernatant are transferred in test tubes and added of 1 ml of amylo-glucosidase before incubation during 30 min at 60°C. After incubation, 8.5 ml of distilled water are added to each tube.

Finally, 75 µl of each sample are taken and added of 1.5 ml of an enzymatic solution, GOD-POD (Glucosidase and peroxidase solution) in order to realize the staining of free glucose and allow its quantification by absorbance. Quantification is carried out at 510 nm with a UV/visible spectrometer (Shimadzu, UV-2450, Japan)

Three replicates are made for each sample

A standard range of glucose is produced to measure the concentration of glucose in our samples. The results obtained are expressed in g of starch per 100g of dry matter according to the following equation:

$$\%Amidon = \frac{(Abs * Slope * 0.9 * 10 * 100) * 100}{(0.05 * 0. * 1000 * DM)}$$

With:

- Abs = Absorbance
- Slope = Standard curve slope
- 0.9 = factor of conversion between glucose and starch
- 10 = dilution factor (glass tubes)
- 0.5 = 500 µL sample taken
- 100 = dilution factor (Erlenmeyer flask)
- 0.05 = Dilution factor (75 µL to 1500 µL of GOD-POD)
- 1000: conversion from mg to g
- DM = Dry Matter in mg

2.5 chemometrics

Chemometrics is the science of extracting information from chemical systems by data-driven means. Chemometrics is inherently interdisciplinary, using methods frequently employed in core data-analytic disciplines such as multivariate statistics, applied mathematics, and computer science, in order to address problems in chemistry, biochemistry, and chemical engineering.

Applied to spectral, biochemical and physical data, the chemometric process can be ordered in 4 steps:

1. Data description using statistics and exploration using multivariate methods. The aim is to understand the relation between variables and individuals, to identify clusters and outliers and to know the overall variability.
2. Pretreatments of the data using mathematics procedures such as derivative, normalization, smoothing … The aim is to enhance the quality of the signal (for spectral data), to remove noise, to scale the variable for better comparison…
3. Choose the best model in order to predict the target using the predictors. This aim to define the suitable algorithm and to fix the parameters. To do this a learning set and test set are selected within the data.
4. Develop the final model to be used for routine analysis.

For this study the following software were used for chemometric developments:

R (Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/)

prospectr, tevens A, Ramirez-Lopez L (2022). An introduction to the prospectr package. R package version 0.2.6.

rchemo, Lesnoff, M. 2021. R package rchemo: Dimension reduction, Regression and Discrimination for Chemometrics. https://github.com/mlesnoff/rchemo. CIRAD, UMR SELMET, Montpellier, France

RStudio Team, 2015. RStudio: Integrated Development Environment for R, Boston, MA. Available at: http://www.rstudio.com/

XLSTAT (Addinsoft, 2022, Paris, France).

The R script developed for classification purpose is reported in annex IV

3 RESULTS

3.1 biochemicals analysis

the biochemical analyses were realized on 108 yam samples corresponding to 7 varieties (tab.2) the descriptive statistics for DM, Starch and Pectin are reported in table 3.

Table 2 : Samples used for biochemical characterization

Variable	Modalités	Nbr éch
Variété	14M	12
	61F	18
	74F	18
	A109	18
	A48	12
	A64	18
	Kabusa	12
Tubercule	1	42
	2	42
	3	24
Zone	Centrale	36
	Distale	36
	Proximale	36

Table 3 : Descriptive statistics for Dry matter, Starch and Pectin of the Yam samples

Contituent	N	Minimum	Maximum	Mean	SD
DM (%)	108	19.12	42.9	29.69	5.38
Starch (%/DM)	108	42.39	88.67	74.26	8.98
Pectin (%/DM)	108	483.37	2078.95	932.45	354.77

The only significant correlation (r of Pearson, α = 0.05) is between starch and pectin contents with a r = -0,539.

The varieties with “bad” cooking behavior (long cooking time) present higher pectin content and so lower starch content.

3.1.1 Pectin

The highest pectin contents are in the proximal zones of the tubers (fig. 4). We note that the pectin content of the very poor quality 74F is very different from the Kabusa which is very good. The pectin contents of the bad variety are higher than that of the good variety whatever the zone. The intra-variety variability is not the same according to the cooking quality. Indeed, the Kabusa variety is the most homogeneous over the whole tuber which may justify that it is the variety most appreciated by consumers.

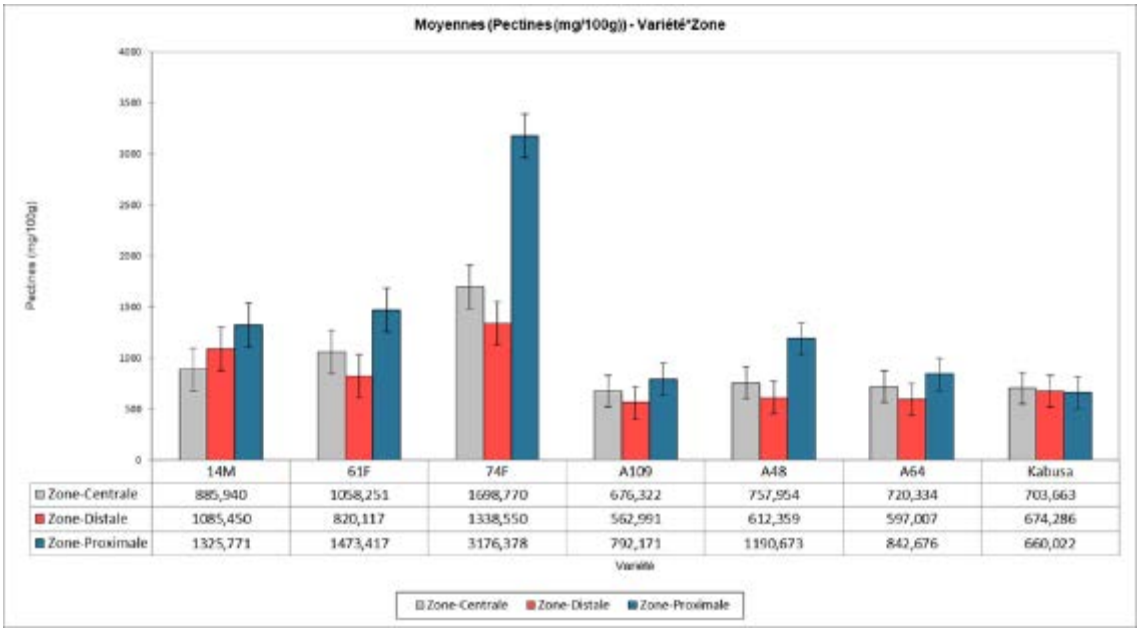


Figure 4 : Pectin contents (mg/100g DM) per variety and zone of the tuber

An analysis of variance was carried out to test the influence of the Variety and Zone of the tuber factors and their interaction on the pectin content. The factors Variety, Zone present significant effect ($\alpha = 0.05$) on pectin content with a significant interaction between Variety*Zone. However, the pectin content does not allow us to establish cooking classes because except for the "very bad" class, the pectin content is similar for the other classes.

3.1.2 Starch

For all varieties, the starch content is homogeneous between the proximal, central and distal zones except for the 74F variety which is the worst (in terms of cooking ability) and the 61F which is an average variety (fig.5). The A48, A64, A109 14M and Kabusa varieties present an average starch content around 80 g/100 g of dry matter, while the 74F and 61F varieties present a lower average content (70 g/100g).

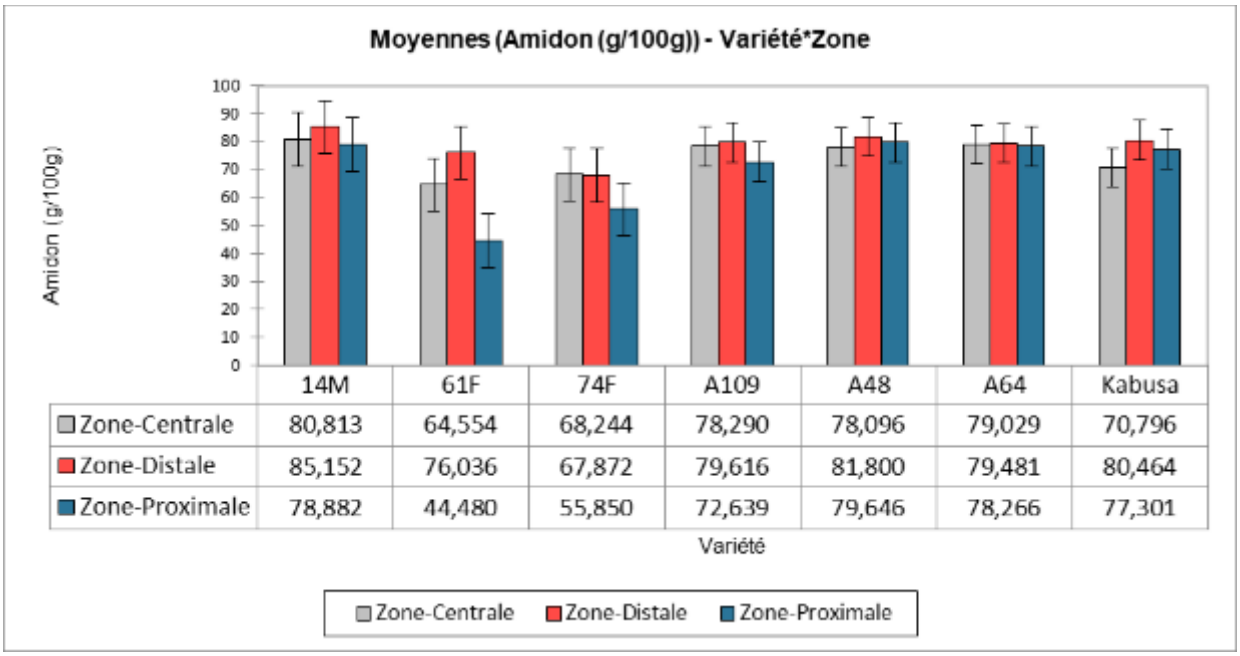


Figure 5 : Starch contents (mg/100g DM) per variety and zone of the tuber

The results of the ANOVA show us that the Variety factor and the Variety*Zone interaction are significant on the starch content because the associated p-value is less than 5%, this confirms our analyzes which show that there is differences in starch content between varieties and for some varieties there are intra-tuber differences such as for 74F and 61F. The Zone factor is not significant on its own because for certain varieties (A64, A48), the starch content is homogeneous. However, cooking classes cannot be created from the starch content.

An analysis of variance was carried out to test the influence of the Variety and Zone of the tuber factors and their interaction on the starch content. The factors Variety significant effect ($\alpha = 0.05$) on starch content with a significant interaction between Variety*Zone. There is no significant effect of zone of the tuber on starch content, starch content is homogenous within the tuber. No clear cooking classes can be defined based on starch content.

3.1.3 Dry Matter

The dry matter content is obtained by the difference of weights (before and after freeze-drying) corrected with the average residual water content after freeze-drying (quantified by oven desiccation). The average DM content ranges between 26% and 39%, with similar contents according to tuber zone: proximal, central and distal (fig. 6). The dry matter content of the 74F varieties is the lowest. There is no relation between DM content and cooking ability of the tubers.

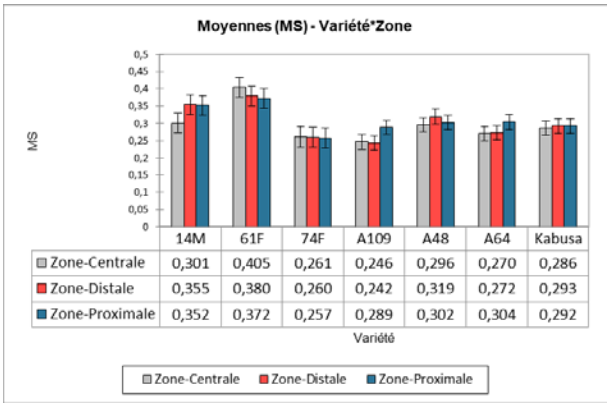


Figure 6 : Dry matter contents (%) per variety and zone of the tuber

3.2 Texture analyses

The texture results were obtained from 172 yam samples comprising 7 different yam varieties (tab.4). Texture parameter corresponds to hardness, the mean value is 4,78 (N) with a minimum value of 0,613 (N) and a maximum of 28,65 (N). The standard deviation of hardness values is 4,23 (N).

Table 4 : Samples used for texture analyses

Variable	Modalités	Nbe éch
Variété	14M	27
	61F	27
	74F	18
	A107	18
	A48	18
	A54	18
	Kabusa	26
Tubercule	1	63
	2	63
	3	26
Zone	Centrale	51
	Distale	51
	Proximale	50

A trend saw observed for hardness according to the tuber zone, the proximal zone is the hardest zone except for kabusa which is very homogeneous for hardness on the 3 zones (fig.7). The KAB variety presents a high ability to cook with a cooking homogeneous throughout the tuber. The 61F variety is the hardest variety of all varieties. The 74F variety is the most heterogeneous in texture when cooked, this variety is not appreciated by consumers. No clear classes for cooking ability can be defined according to hardness

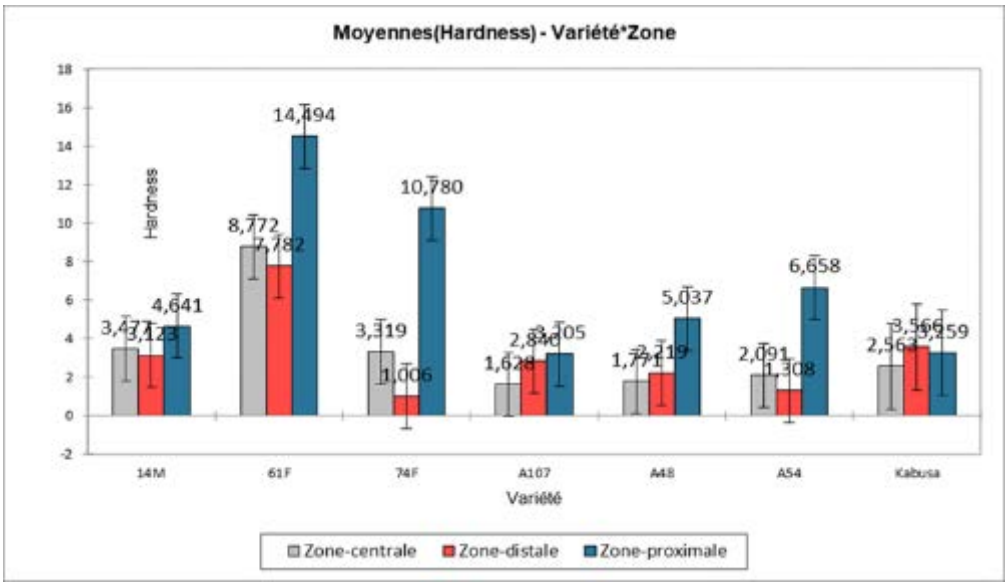


Figure 7 : Hardness values (N) per variety and zone of the tuber

4 MODELLISATION

4.1 Texture

A model to predict hardness was developed based on average spectra of the texture samples. The complete dataset contains 152 spectra (average spectrum of HIS images) and 152 Hardness values. The data set was split into 2 sets: a learning set which corresponds to two thirds of the complete data set, i.e. 101 samples, and a test set which corresponds to the remaining third of the samples (n=51). Before realizing the model, the spectra were transformed. The preprocessing was: a Savitzky-Golay smoothing with the parameter width = 15 and polynomial order = 2 and dorder = 0 ; a baseline correction with the following parameters lambda = 5 and p = 0.001 and finally normalization reduction using SNV.

The chemometric model applied is a PLS model, the number of latent variables was determined by cross validation. The figures 8 and 9 correspond to the scatter plots of predicted hardness (using the PLS model) values versus measured hardness values for both sets, learning (fig.8) and test set (fig.9).

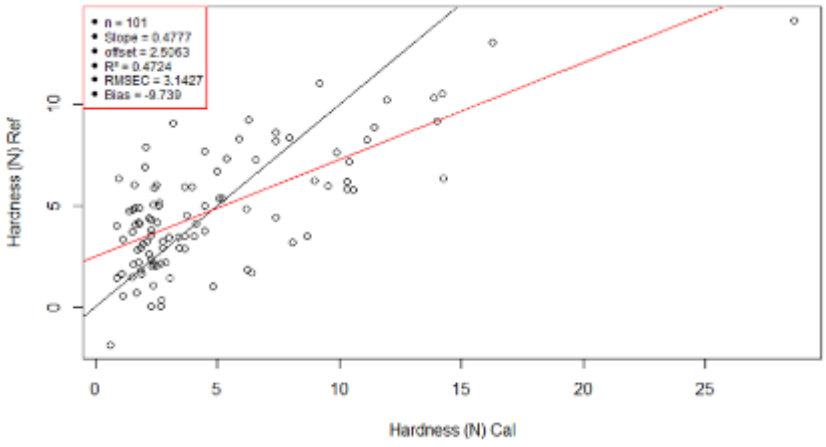


Figure 8 : scatter plot of hardness reference values versus predicted values for learning set

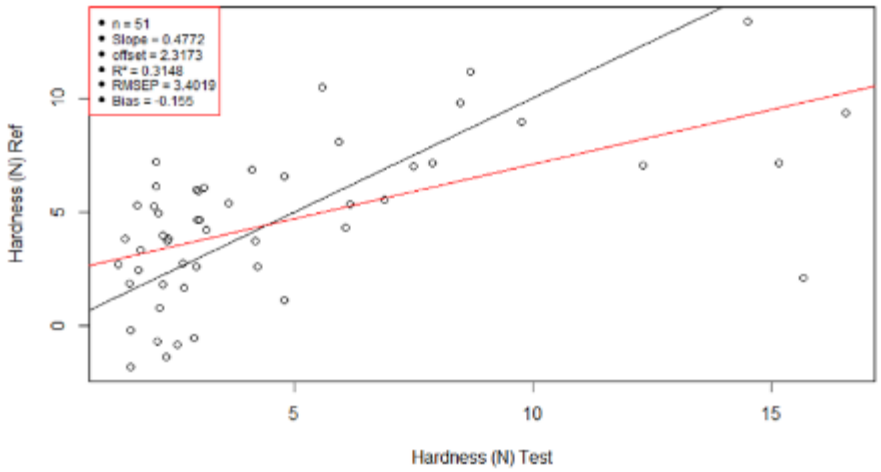


Figure 9: scatter plot of hardness reference values versus predicted values for test set

The model is not sufficiently efficient either in calibration or in validation with an RMSEC of 3.14 N and an RMSEP of 3.40 N knowing that the average hardness over the entire data set is 4.78 N.

4.2 Yam cooking ability

During the internship 306 hyperspectral images of 10 varieties of yam where acquired, for all these varieties the cooking ability are known (breeder's data).

Three classes are defined (tab.5):

• "Good", which includes the "Good and Very good" classes.

• "Average"

• "Bad" which groups together the "Bad and Very Bad" classes.

Table 5 : number of samples per cooking class and varieties

Variété	Classe initiale	Nouvelle classe	Nbr Ech
14M	Moyenne	Moyenne	39
61F	Moyenne	Moyenne	45
74F	Très mauvaise	Mauvaise	36
A107	Moyenne	Moyenne	36
A109	Moyenne	Moyenne	18
A48	Mauvaise	Mauvaise	30
A54	Moyenne	Moyenne	30
A64	Moyenne	Moyenne	18
CIR	Bonne	Bonne	16
Kabusa	Très bonne	Bonne	38

The Duplex algorithm (R and prospect package) was used to create a learning data set (2/3 of the samples, n= 204) and a test set (1/3 of the samples, n = 102). Prior to calibration the spectra are preprocessed as follow: a Savitzky-Golay smoothing with the parameter width = 15 and polynomial order = 2; a baseline correction with the following parameters lambda = 5 and p = 0.001 and finally normalization reduction using SNV.

Three chemometric classification methods were tested on the learning set: a PLSDA model, an LWPLSDA model and a SVMDA model. Each parameter applied for the models was obtained after a cross validation step.

Table 6 : performances of the different model tested for classification of cooking classes

Calibration						Validation	
Ncal	Model	Treatment	Parameters	err C	err CV	Nval	err P
204	LWPLSLDA	smoothing + baseline correction + SNV	nlv = 7, nlvdis = 43, h = 6, k = 38, diss = mahal, prior = prop	0	0,19	102	0,19
204	PLSDA	smoothing + baseline correction + SNV	nlv = 15	0,24	0,31	102	0,31
204	SVMDA	smoothing + baseline correction + SNV	cost = 10^{-3}, epsilon = 0,1, gamma = 10^{-1}	0,12	0,25	102	0,25
err C = erreur de calibration ; err CV = erreur de cross validation ; err P = erreur de prédiction							

The LWPLSLDA model gives the best result with a classification error of 19%. This is an acceptable result for a prescreening of yam tuber according to their cooking ability. Indeed, the prediction of all the pixels of a hyperspectral image of yam tuber using this model will provide a reliable average classification of the tuber.

4.3 Dry matter

The dataset contains 149 samples which has been separated into 2 sets, a learning set of 99 samples and a validation set of 50 samples using duplex algorithm. Prior to calibration the spectra are preprocessed as follow: a Savitzky-Golay smoothing with the parameter width = 15 and polynomial order = 2; a baseline correction with the following parameters lambda = 5 and p = 0.001 and finally normalization reduction using SNV.

Two chemometric methods were tested on the learning data, a PLS model and an LWPLSR model. Each parameter applied for the models was obtained after a cross validation step.

The PLS calibration model is not efficient with an RMSEC of 4.47% and a rather poor linearity with an R^2 of 0.73 which explains only 73% of the total variability. Moreover, the prediction error RMSEP = 6.23 with a the R^2p = 0.56 indicate that the model is not robust.

On the other hand, the LWPLSR model is much better than the PLS model. The RMSEP was divided by 3 (RMSEP of 2.7%) and the linearity of the model increased R^2 = 0.92. The figures 10 corresponds to the scatter plot of predicted hardness (using the LWPPLSR model) values versus measured hardness values for the test set samples.

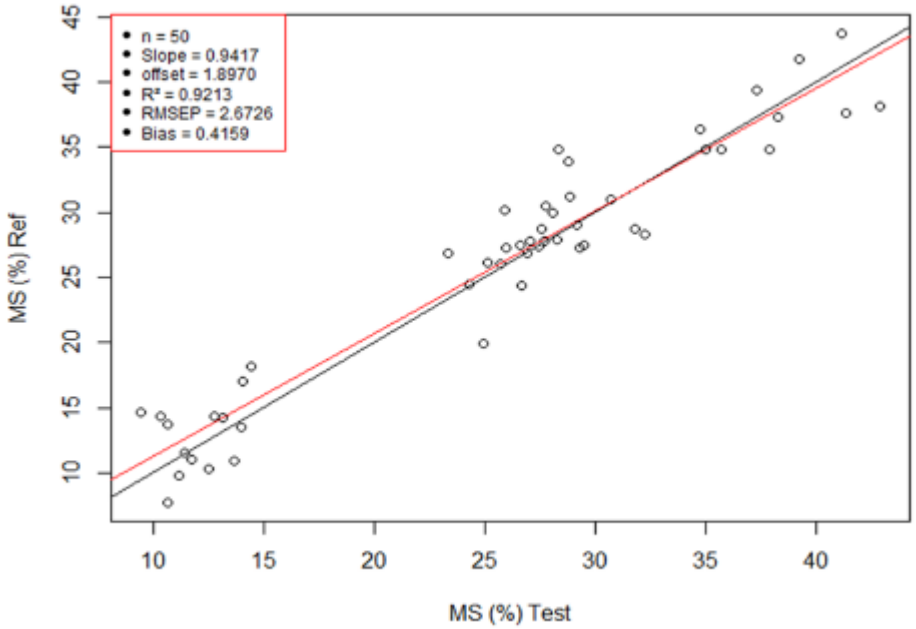


Figure 10 : scatter plot of dry matter content values versus predicted values for test set, model LWPLSR.

Despite the few numbers of samples analyzed, the performances of the LWPLSR model are promising, the HIS method can be applied as routine analysis for DM quantification of fresh yam tubers.

4.4 Pectin and starch

The complete datasets include 108 spectra and 108 values for pectin & starch content. The data set has been separated into 2 sets: a learning set with two thirds of the samples (n = 72) and a test set with samples (n = 36). Prior to calibration the spectra are preprocessed as follow: a Savitzky-Golay smoothing with the parameter width = 15 and polynomial order = 2; a baseline correction with the following parameters lambda = 5 and p = 0.001 and finally a normalization reduction using SNV.

The chemometric models applied are PLS model.

The model for pectin quantification presents poor performances whether in calibration or validation with an RMSEC of 268 mg/100g and an RMSEP of 316.02 mg/100g. Regarding the poor number of samples available no other chemometric approach was investigated, to improve the model, the number of samples has to be increased, with the precaution to select samples that will increase the pectin content range.

As for pectin the PLS model for starch quantification presents poor performances whether in calibration or validation with an RMSEC of 6,83 g/100 g DM and an RMSEP of 9.21 g/100 g /DM. Regarding the poor number of samples available no other chemometric approach was investigated, to improve the model, the number of samples has to be increased, with the precaution to select samples that will increase the starch content range.

5 CONCLUSION

The aim of the study was to develop a predictive model of yam cooking ability using hyperspectral imaging applied to fresh yam tuber.

The study demonstrated that there is a significant correlation between hardness, measured using a penetrometer and starch content (r = -0,80), pectin content (r = 0.52) and Dry matter content (r = 0.55). no significant correlation exists between starch, pectin and DM and cooking ability except for highly contrasted varieties (very bad vs very good).

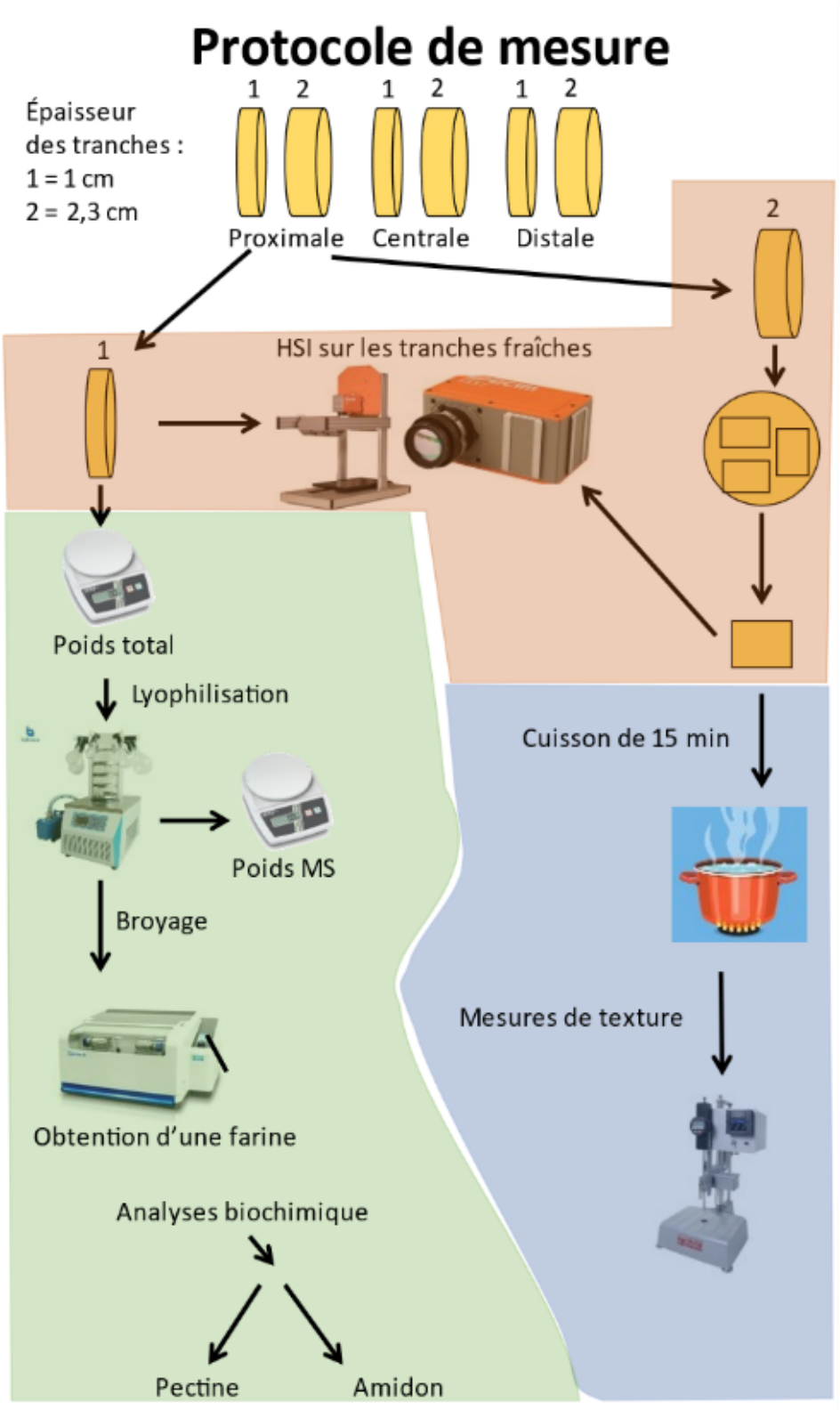
The models developed for predicting pectin, starch and texture content do not have sufficient performance for routine analysis, however these models can be improved by increasing the number of samples and the ranges of the constituents.

The local PLS classification model is promising for the yam cooking ability classification with a classification error of 19%. To improve this model, it is recommended to collect more samples with more varied and better distributed cooking qualities. Indeed, there was within the dataset an over representation of middle-class variety and not enough good or bad classes. The model developed for dry matter quantification is efficient with an error of prediction RMSEP = 2.67%.

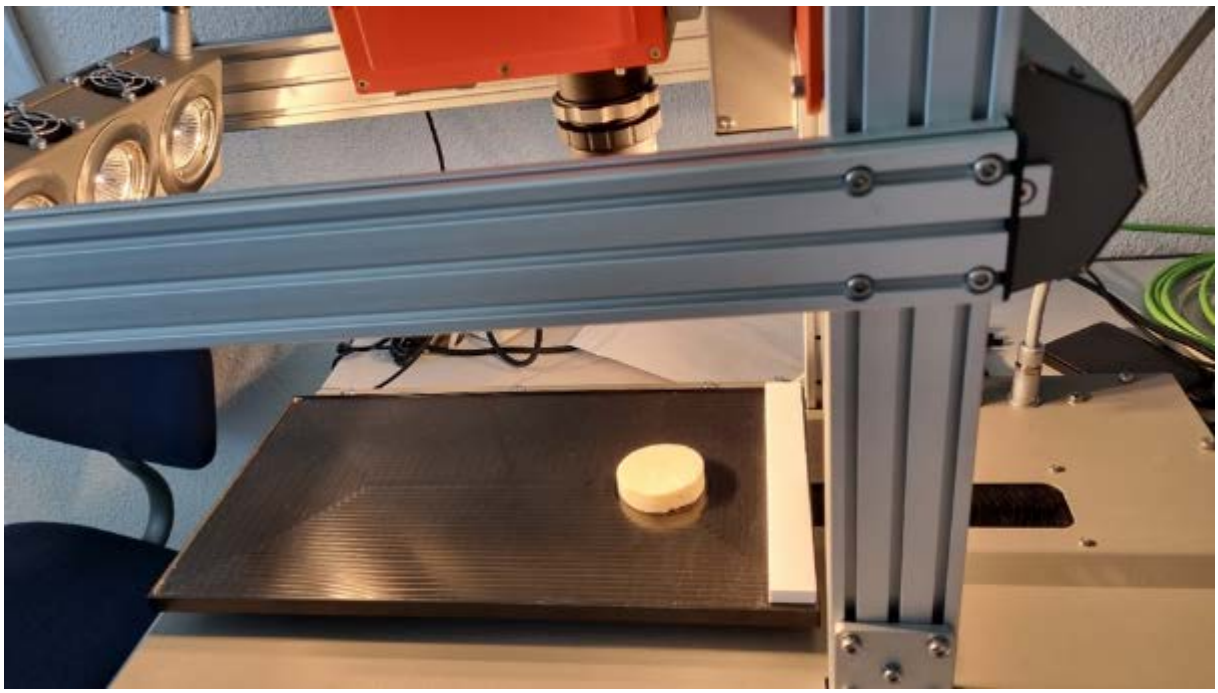
This study demonstrated the potential of HSI for the selection of Yam genotypes according to some relevant traits such as dry matter content and cooking ability. Further investigations have to be done in order to test model's accuracy and robustness and to test new models, for this the number of samples has to be increased. The selection of new samples to be analyzed should be done in such way that the maximum of variability will be integrated in the data set.

6 APPENDICES

6.1 Annex 1: Samples preparation scheme



6.2 Annex 2: HSI and texture measurements



6.3 Annex 3: Python script for images processing

```
# Packages
from spectral import *
import spectral.io.envi as envi
import numpy as np
import matplotlib.pyplot as plt
import matplotlib
import pandas as pd
# Load Parameter
xm = pd.DataFrame()
path = 'D:\Mes Donnees\CIRAD\Texture\HSI Texture\HSI'
dfname = pd.read_excel(r'D:\Mes Donnees\CIRAD\Texture\Données_texture.xlsx',
                       sheet_name='name')
list_name = np.array(dfname)
n = len(list_name)
# Boucle for mean spectra
for i in range(0,n):
    name = np.array(list_name[i,0])
    dark_ref = envi.open(f'{path}\{name}\capture\DARKREF_{name}.hdr',
                         f'{path}\{name}\capture\DARKREF_{name}.raw')
    white_ref = envi.open(f'{path}\{name}\capture\WHITEREF_{name}.hdr',
                          f'{path}\{name}\capture\WHITEREF_{name}.raw')
    data_ref = envi.open(f'{path}\{name}\capture\{name}.hdr',
                         f'{path}\{name}\capture\{name}.raw')
    # Convert to numpy array
    white_nparr = np.array(white_ref.load())
    dark_nparr = np.array(dark_ref.load())
    data_nparr = np.array(data_ref.load())
    # Get matrices right
    dark = np.mean(dark_nparr, axis=0, keepdims=True)
    white = np.mean(white_nparr, axis=0, keepdims=True
    # Calculating reflectance
    corrected_nparr = np.divide(
        np.subtract(data_nparr, dark),
        np.subtract(white, dark))
    # Clustering
    (mask, c) = kmeans(corrected_nparr, nclusters =2, max_iterations =30
    # #Clip the calibrated values to the range 0 - 1
    np.clip(corrected_nparr, a_min=0, a_max=1, out=corrected_nparr)
    # Get 2D dataset
    dim=corrected_nparr.shape
    df=np.reshape(corrected_nparr, (dim[0]*dim[1], dim[2]))
    df=pd.DataFrame(df, index=None, columns=data_ref.bands.centers
    # Mean spectrum of the image
    xm[list_name[i,0],] = np.mean(df,0)
```

```
# save df mean spectra
df_mean = xm.T
df_mean.to_csv('mean_spectra.csv')
# Plot all mean spectra
plt.plot(xm)
plt.title('Mean spectrum of the image')
plt.xlabel('Wavelength (nm)')
plt.ylabel('Reflectance')
plt.grid()
```

6.4 Annex 4: R scripts for classification and quantification models

amidon.pdf DM.pdf Pectin.pdf texture.pdf cooking ability.pdf

Institute: Cirad – UMR QualiSud

Address: C/O Cathy Méjean, TA-B95/15 - 73 rue Jean-François Breton - 34398 Montpellier Cedex 5 - France

Tel: +33 4 67 61 44 31

Email: rtbfoodspmu@cirad.fr

Website: https://rtbfoods.cirad.fr/