Research Article

Received: 20 January 2023 Revised: 18 April 2023 Accepted article published: 22 April 2023 Published online in Wiley Online Library: 26 May 2023

(wileyonlinelibrary.com) DOI 10.1002/jsfa.12654

Hyperspectral imaging for the determination of relevant cooking quality traits of boiled cassava

Karima Meghar,[a,b*] Thierry Tran,[a,b,c,d] Luis Fernando Delgado,[d] Maria Alejandra Ospina,[d] Jhon Larry Moreno,[d] Jorge Luna,[d] Luis Londoño,[d] Dominique Dufour[a,b] and Fabrice Davrieux[a,b,e]

Abstract

BACKGROUND: The purpose of this study was to investigate the potential of hyperspectral imaging for the characterization of cooking quality parameters, dry matter content (DMC), water absorption (WAB), and texture in cassava genotypes contrasting for their cooking quality.

RESULTS: Hyperspectral images were acquired on cooked and fresh intact longitudinal and transversal slices from 31 cassava genotypes harvested in March 2022 in Colombia. Different chemometric methods were tested for the quantification of DMC, WAB, and texture parameters. Data analysis was conducted through partial least squares regression, *K* nearest neighbors regression, support vector machine regression and CovSel multiple linear regression (CovSel_MLR). Efficient performances were obtained for DMC using CovSel_MLR with, coefficient of multiple determination $R_p^2 = 0.94$, root-mean-square error of prediction RMSEP = 0.96 g/100 g, and ratio of the standard deviation values RPD = 3.60. High heterogeneity was observed between contrasting genotypes. The predicted distribution of DMC within the root can be homogeneous or heterogeneous depending on the genotype. Weak predictions were obtained for WAB and texture parameters.

CONCLUSIONS: This study showed that hyperspectral imaging could be used as a high-throughput phenotyping tool for the visualization of DMC in contrasting cooking quality genotypes. Further improvement of protocols and larger datasets are required for WAB and texture quality traits.

Keywords: dry matter content; water absorption; texture; chemometrics; high throughput phenotyping; consumer preferences

INTRODUCTION

Cassava (*Manihot esculenta* Crantz), was first domesticated in South America. It is one of the world's most important root and tuber crops and a fundamental staple of rural and urban households in sub-Saharan Africa.[1] Cassava roots are a major source of calories for ~800 million people in the Americas, Asia, and Africa.[2]

Cassava is consumed in different ways: directly after boiling, steaming, or frying, or after processing into gari, eba, fufu, lafun, attiéké, and agbelima.[3,4] Consumers of boiled cassava prefer genotypes with a short cooking time (CT).[5,6] Dry matter content (DMC), mealiness, water absorption (WAB), and texture were identified as the most important parameters for the characterization of cooking quality of cassava.[5-9]

Currently, the non-destructive techniques of near infrared (NIR) and visible–NIR (VNIR) spectroscopy are used in cassava breeding programs as high-throughput phenotyping methods.[10-18] However, those classical spectral techniques ignore the spatial variations of biochemical and physical properties within a root. Therefore, in recent years, many researchers became interested in the spatial distribution of quality traits in food. Hyperspectral imaging (HSI) has been widely used for the characterization of roots and tubers, taking images either of the inside (sliced root, destructive method) or of the surface (intact root, non-destructive

* Correspondence to: K Meghar, CIRAD Montpellier Centre, CIRAD, UMR Qualisud, F-34398 Montpellier, France, E-mail: karima.meghar@cirad.fr

a UMR Qualisud, CIRAD, Montpellier, France

b QualiSud, Univ Montpellier, Avignon Université, CIRAD, Institut Agro, IRD, Université de La Réunion, Montpellier, France

c UMR QualiSud, CIRAD, Cali, Colombia

d Cassava Program, Alliance of Bioversity–International Center for Tropical Agriculture (CIAT), Cali, Colombia

e UMR QualiSud, CIRAD, Saint-Pierre, France

method). Many applications have been successfully identified, mainly for potatoes. These applications concern quantification of biochemical constituents,[19-25] characterization of physical properties,[19,24,26-29] and detection of diseases and other defects, such as scabs or bruises.[27,30-38]

This study focused on the prediction of the distribution of cooking ability (boiling) of cassava from hyperspectral images acquired on fresh cassava, which to our knowledge has not been investigated on cassava or even on other products.

The objectives of this study were (i) to explore whether HSI can discriminate between roots of contrasting cassava varieties in terms of cooking behaviors, (ii) to explore the potential of HSI to replace physico-chemical methods for the determination of DMC, WAB, and texture, thus saving time for quality traits evaluation of boiled cassava, (iii) to develop models for the prediction of DMC, WAB, and texture parameters, and (iv) to produce DMC, WAB, and texture maps using the best models for the visualization of the spatial distribution of these traits within fresh and boiled cassava slices.

MATERIALS AND METHODS

Plant material

Thirty-one cassava genotypes with contrasting cooking properties from the RTBfoods progenitors collection were grown at the International Center for Tropical Agriculture (CIAT, Cali, Colombia) and harvested at 12 months after planting, from 7 to 11 March 2022. Six genotypes were harvested per day (in the morning) and processed immediately. The RTBfoods progenitors collection was selected and planted for the first time in 2018, and replanted over several years (2018–2022) in order to characterize yearly variability. The collection is described in more detail by Tran *et al.*,[9] with small differences in the number of genotypes, caused by unforeseen situations such as diseases.

The roots used in this study are from good- and poor-cooking quality genotypes. Samples with high values of WAB correspond to genotypes with low CT as well as softer texture and higher mealiness expression. A limit between short and long CT genotypes was set at 12% for WAB (measured after 30 min of boiling), which corresponds to CT = 36 min according to a linear regression between WAB and CT obtained at CIAT using data from several years and harvests.[9] The limit CT = 36 min was based on discussions with cassava breeders, who indicated that they would reject genotypes with CT > 36 min when selecting for cooking quality, whereas genotypes with CT < 25 min and 25 ≤ CT ≤ 36 min would be considered desirable and acceptable respectively. Genotypes with WAB ≤ 12% can therefore be considered as poor cooking (CT ≥ 36 min), and genotypes with WAB > 12% correspond to low CT (CT < 36 min) and are more suitable for end users.[9]

Samples preparation

For each cassava genotype, 12 roots (at least 25 cm long and 5.5 cm diameter in the central part) were selected from at least four plants. For each root, both ends were cut and discarded, and six half-cylinders (6 cm long and 5.5 cm diameter) were obtained from the middle section. The half-cylinders were identified individually by numbering from 01 to 06 for the first root, from 11 to 16 for the second root, from 21 to 26 for the third root, and so on (Fig. 1). Codes ending with 1 or 2, 3 or 4, and 5 or 6 correspond to pieces from, respectively, the proximal, central, and distal parts of the roots. The pieces were allocated semi-randomly for various measurements: WAB, CT, texture, dry matter (of fresh roots), and HSI (Fig. 1). Six half-cylinders from the central parts of different roots were used for HSI measurements (03, 43, 44, 103, 104, and 114). First, a transversal slice (1 cm thick) was cut from each half-cylinder and used for HSI, followed by dry matter analysis. The remaining parts of the pieces were boiled for 18 min (pieces 04, 44, and 104) and 30 min (pieces 03, 43, 103, 114), and then another slice (1 cm thick) was cut for HSI measurements after cooking.

HSI setup

The monitoring mode of the HSI system used in this study and all required steps for the acquisition of HSI data are described in the RTBfoods' standard operating protocol (SOP).[39]

The system is based on pushbroom imaging equipment. The main components include a spectrograph with prism–grating–prism optical structure (ImSpector, N17E; SPECIM, Oulu, Finland), a 12-bit CCD camera (V-light; Lowel Light Inc., New York, NY, USA), 150 W tungsten halogen lamps (Fibre-Lite DC950 Illuminator; Dolan Jenner Industries Inc., Boxborough, MA, USA), and a translation LabScanner with dimension ($L \times W$) of 40 cm × 20 cm by a step motor. The harmonious work of the integral system is assured by using LumoScanner control software (SPECIM).

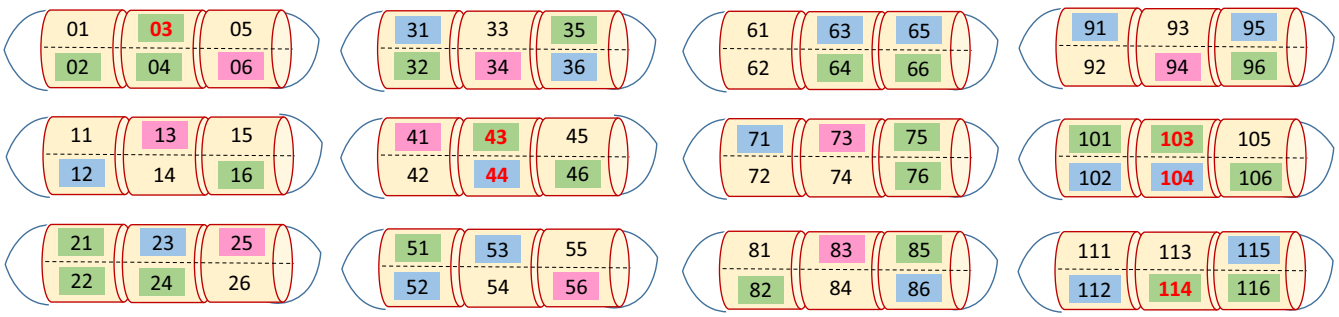

Figure 1. Sample preparation of 12 cassava roots per genotype, for measurements of hyperspectral imaging (HSI), dry matter content, water absorption, cooking time, and texture. Each root was cut in six half-cylinders (6 cm long × 5.5 cm diameter). The half-cylinders in green, pink, and blue were used for water absorption, cooking time, and texture measurements respectively. Half-cylinders numbered in red (03, 43, 44, 103, 104, 114) were also used for hyperspectral image acquisition of transversal slices, before and after cooking. One additional root of each genotype (not shown on the figure) was divided into two longitudinal pieces and scanned by HSI. The remaining half-cylinders without green, pink, or blue color were mashed and used for dry matter content and additional measurements outside the scope of the present HSI study.

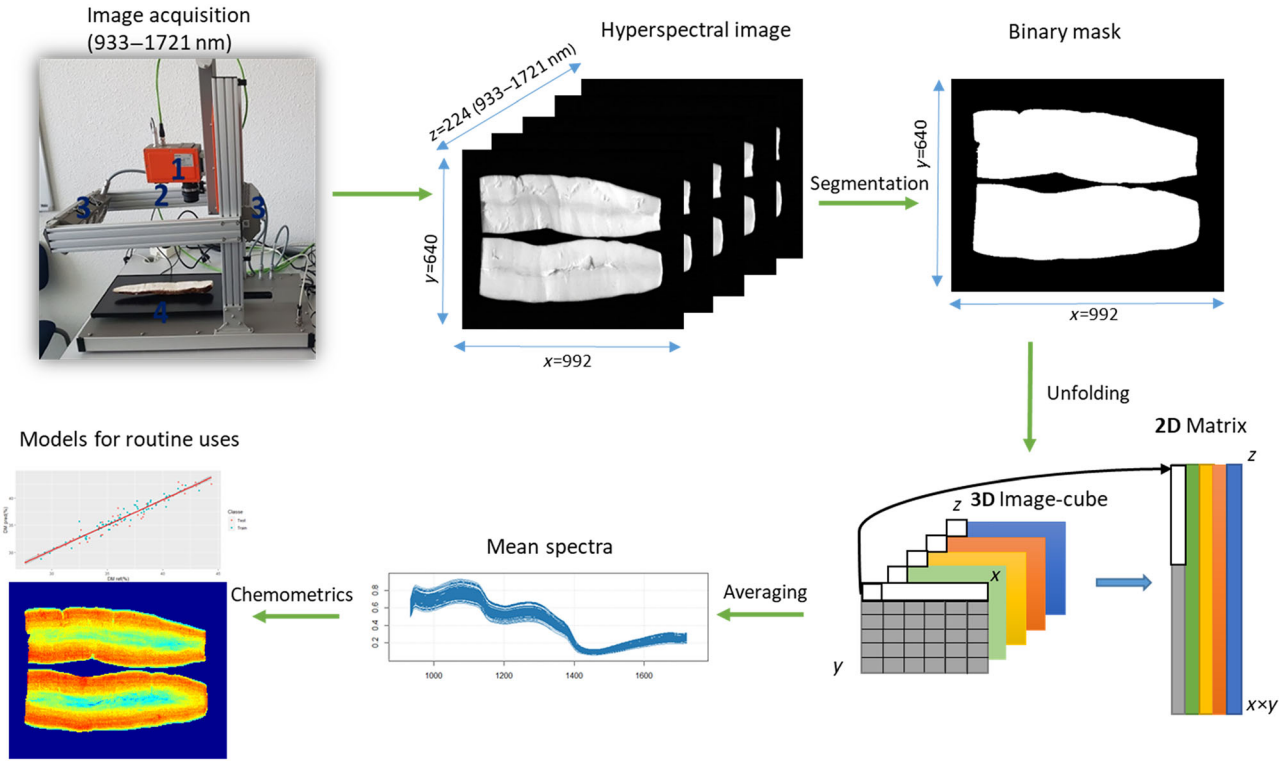

Figure 2. Flowchart of hyperspectral images processing from image acquisition to application of chemometrics methods.

The assembly disperses the incoming line of light into the spectral and spatial matrices and then projects them onto the CCD. The optics, spectrograph, and the camera have high sensitivity from 932 to 1721 nm with a spectral resolution of 8 nm. Hyperspectral images have three dimensions (x, y, z), of which x and y are spatial and z is spectral. The spatial dimension $x \times y$ has dimensions of pixels, and the spectral dimension z is the number of wavelengths (224). The spatial dimensions depend on the size of the sample.

Hyperspectral images measurements procedure

For each half-cylinder allocated to HSI (Fig. 1), one transversal slice was cut (1 cm thick) immediately before measurement. The slices were scanned by HSI at the cut surface and kept for DMC measurement. The remaining parts of the half-cylinders 03, 43, 103, and 114 were cooked for 30 min for WAB measurement and then cooled for 10 min before a second transversal slice was cut (1 cm thick) and immediately scanned by HSI at the cut surface. The remaining parts of the half-cylinders 44 and 104 were cooked for 18 min for texture measurements and then scanned by HSI following the same procedure as for WAB measurement. A total of 186 (6 slices × 31 genotypes) images were acquired for fresh transversal slices and 186 images for cooked transversal slices. Moreover, for each genotype, one hyperspectral image was acquired on a longitudinally cut root, making a total of 31 additional images for longitudinal slices.

Determination of DMC

DMC was measured immediately after HSI analyses by cutting the transversal fresh slices (1 cm thick) from pieces 03, 43, 44, 103, 104, and 114 into smaller pieces and drying them at 105 °C for 24 h.[40] DMC was expressed as the percentage of dry weight relative to fresh weight:

$$\text{DMC (\% wet basis)} = \frac{\text{Weight of the sample after drying}}{\text{Weight of the sample before drying}} \times 100$$

In total, 186 (6 slices × 31 genotypes) DMC values were measured from fresh transversal slices.

Determination of WAB and CT

For each genotype, 24 half-cylinder pieces (shown in green in Fig. 1) were semi-randomly selected among the 72 obtained after sample preparation, distributed among three metal strainers (eight pieces per strainer) and boiled for 30 min. WAB was measured according to Tran *et al.*,[9] and the RTBfoods' SOP.[41] For each genotype, one WAB value was recorded as the average of the three strainers, making a total of 31 values.

WAB was expressed as the change in weight of the cassava pieces in relation to their initial weight:

$$\text{WAB (\%)} = \frac{\text{Weight at } t = 30 \text{ min} - \text{Weight at } t = 0 \text{ min}}{\text{Weight at } t = 0 \text{ min}} \times 100$$

For each genotype, nine half-cylinder pieces (shown in pink in Fig. 1) were semi-randomly selected, placed in a metal strainer, and put in boiling water. CT was determined as described in Tran *et al.*[9] and the RTBfoods' SOP.[41] For each genotype, one CT value was recorded as the time at which six out of the nine pieces became soft, making a total of 31 values.

Texture measurements

For each genotype, 18 half-cylinder pieces (shown in blue in Fig. 1) were semi-randomly selected among the 72 obtained after sample preparation and boiled for 18 min.[42] The pieces were allowed to cool for 10 min in a half-closed plastic container in order to limit

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

water loss. The texture-extrusion test was then carried out according to the RTBfoods' SOP.[42] Several texture parameters were calculated; for example, maximum force, distance at maximum force, area under the compression curve ('Area'). Area was selected in this study as the main texture parameter because it was a better discriminant (lower coefficient of variation). For each genotype, the average of the Area of the 18 pieces was recorded, making a total of 31 values.

Hyperspectral images processing

All the steps of hyperspectral images processing from background segmentation to extraction of spectral data were programmed in Matlab R2018b (The Mathworks Inc., Natick, MA, USA) coupled with PLS Toolbox (version 8.7) and Image Processing Toolbox (version 5.4). The protocol of image processing is described in Fig. 2.

Elimination of the background

The background was removed using a binary mask. In our case, the region of interest was selected from the image at 1325 nm, which provided the best discrimination between the sample surface and the background by thresholding at a reflectance value of 0.318. From these images, the positions of all the pixels on the boundary of the sample were located. All the pixels located inside this boundary were considered as *sample*, and the pixels outside this boundary were classified as *background*. A virtual binary image was then created at this wavelength and used for background removal of all the images at other wavelengths in the hypercube. Each pixel of the three-dimensional (3D) image represents one spectrum from 932 to 1721 nm.

Transformation of the hyperspectral (x, y, z) image into a two-dimensional spectral matrix (z, x × y)

The transformation of a 3D image into a two-dimensional spectral matrix was done by unfolding the 3D hypercube (x, y, z) into a two-dimensional matrix (z, $x \times y$), where each row represents the spectrum of one pixel and the columns represent wavelengths. Each image of one wavelength was infolded in one column in a spectral matrix. The same procedure was repeated to obtain the final spectral matrix (Fig. 2).

Calculation of mean spectra of hyperspectral images

The reflectance values of all pixels from one image were averaged to obtain one mean spectrum for each sample. The same procedure was repeated to obtain an average spectrum for each hyperspectral image. All average spectra of all images were arranged together to form a spectral matrix.

Multivariate data analyses

Multivariate analyses were applied on average spectra using R software (version 4.1.0)[56] coupled with packages such as prospectr[43] and rchemo.[44] The dataset was split into two sets using the duplex algorithm: a learning set, which corresponds to two thirds of the complete dataset, and a test set, which corresponds to the remaining third of the samples. Before the model development, different combinations of preprocessing were applied to the spectra. The best performances were obtained with Savitzky–Golay smoothing with a parameter width of 15 and polynomial order 2, a baseline correction with the parameters $\lambda = 5$ and $P = 0.001$, and, finally, normalization reduction using the standard normal variate method.

Different chemometric methods – partial least-squares regression (PLSR),[45] K nearest neighbors regression,[46] support vector machine regression,[47] and CovSel multiple linear regression (CovSel_MLR)[48] – have been tested in order to obtain the best model for the quantification of quality traits from the spectral information.

The optimum number of latent variables to introduce in the models was determined using the minimum value of the root-mean-squared error estimated by repeated k-folds cross-validation.

The performances of the models developed were evaluated using the coefficient of multiple determination R^2_p for predicted values, the root-mean-square error of prediction (RMSEP),[15] and the ratio of the standard deviation values (RPD).[12] Equations to calculate RMSEP and RPD are provided in the references.[12,15]

Principal components analysis (PCA) was applied on hyperspectral images of transversal slices of cassava after 30 min boiling for the visualization of the variations in water content among cassava genotypes. This variation can be linked to cooking behavior, as evaluated by WAB. PCA was applied to 1000 randomly selected spectra of each image, for a total of 124 000 spectra (31 genotypes × 4 images (slices) per genotype × 1000 spectra per image).

Visualization maps of boiled cassava quality traits

After validation, the best model was applied to predict each pixel of transversal and longitudinal slice images in order to obtain the physical or biochemical composition at pixel level.

Table 1. Calibration and validation performances of chemometrics models applied for the prediction of dry matter content in fresh pieces of cassava by hyperspectral imaging

	Calibration			Validation			
Model	N_{cal}	RMSEC (g/100 g)	R^2_{cal}	N_{val}	R^2_p	RMSEP (g/100 g)	RPD
PLSR	78	0.78	0.94	38	0.93	1.06	3.26
KNNR	78	0.16	0.99	38	0.77	1.94	1.78
SVMR	78	2.85	0.17	38	0.46	3.63	0.95
CovSel_MLR	78	0.88	0.92	38	0.94	0.96	3.60

CovSel_MLR, CovSel multiple linear regression, KNNR, K nearest neighbors regression; PLSR, partial least-squares regression; RMSEC, root-mean-standard error for calibration; RMSEP, root-mean-square error of prediction; RPD, ratio of the standard deviation values; SVMR, support vector machine regression.

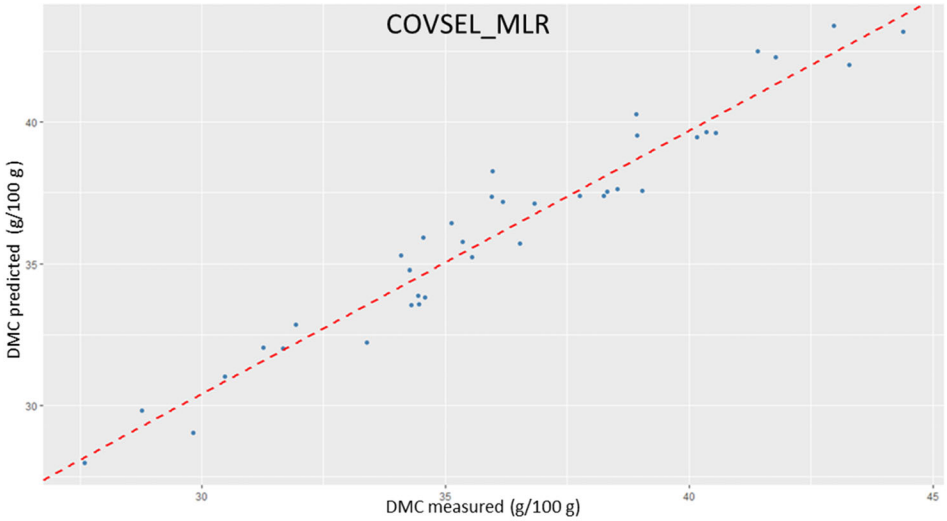

Figure 3. Scatter plot of measured dry matter content (DMC) *versus* predicted DMC by hyperspectral imaging using CovSel_MLR model of validation test set (n = 38).

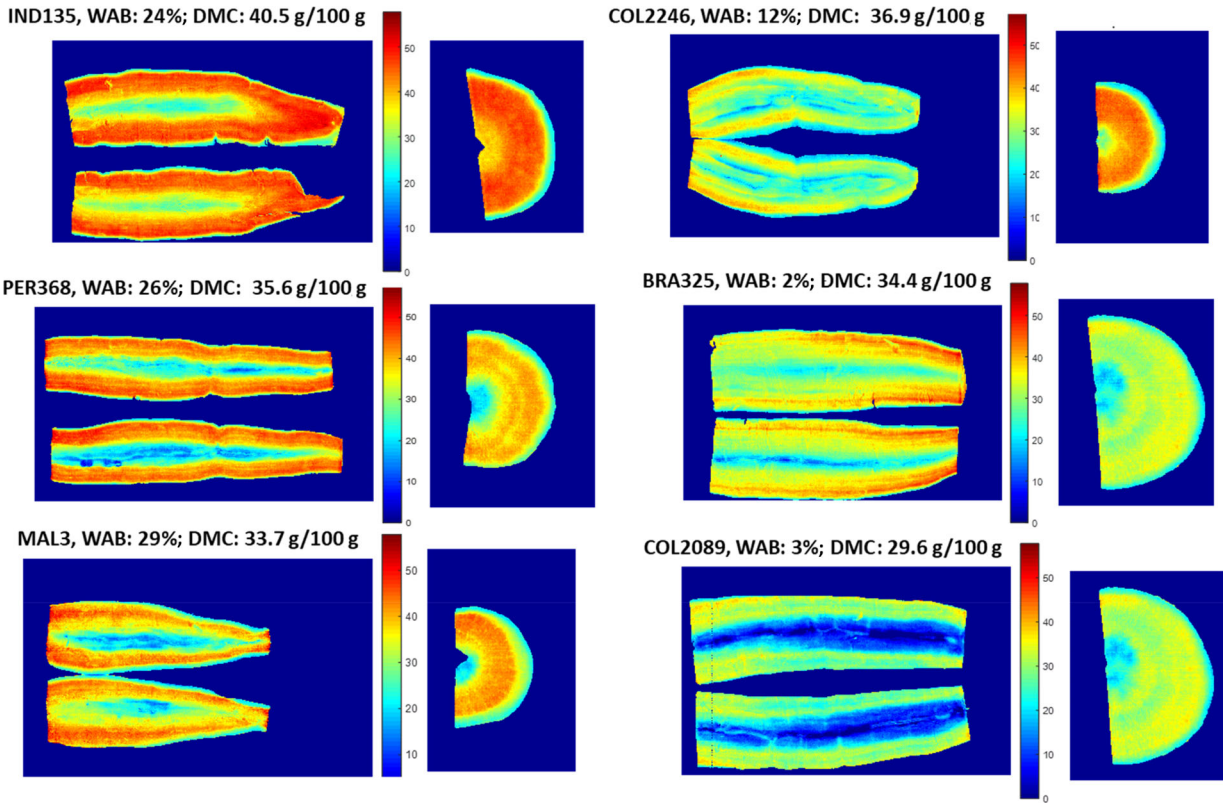

Figure 4. Dry matter content (DMC) distribution map of fresh cassava slices (longitudinal and transversal) generated using the CovSel_MLR model, for good (IND135, PER368, MAL3) and poor (COL2246, BRA325, COL2089) cooking-quality genotypes. WAB, water absorption.

RESULTS

Prediction of DMC

Performance of the models

Among the different models tested, the best prediction of DMC was obtained using the variable selection method (CovSel_MLR; Table 1). This model is based on ten feature wavelengths: 932, 1021, 1130, 1169, 1222, 1385, 1410, 1444, 1484, and 1717 nm.[25] Actually, the goal of the CovSel algorithm is to improve the accuracy of spectral analysis models (such as PLSR) by reducing the number of variables or spectral bands required for the analysis, while preserving the most relevant and discriminative information. This helps to reduce noise, improves separation between classes, and facilitates interpretation of the results. In the present case, the wavelengths selected were related to water content and corresponded to the most discriminative and representative wavelengths in the spectral data according to the criteria to analyze (i.e. DMC). Figure 3 shows the quality of the fitting using this model, where RMSEP = 0.96 g/100 g and R_p^2 = 0.94,

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

Table 2. Cooking time (CT), water absorption (WAB), dry matter content (DMC), and DMC heterogeneity index (HI) values measured for 31 genotypes

Genotype	CT (min)	WAB (%)[a]	DMC (g/100 g fresh root)[b]	DMC HI (g/100 g fresh root)[c]
CM7436-7	24	14.8	35.2	6.19
GUA24	26	9.7	32.3	5.67
PER183	13	21.8	36.0	7.53
PER496	13	21.8	37.6	6.52
VEN208	23	16.1	34.9	11.17
CM5948-1	20	11.5	41.2	9.59
IND135_7	10	34.2	40.9	7.07
COL1505	21	19.2	37.1	9.05
COL2246_9	36	12.4	36.9	5.67
MAL3	21	23.8	33.7	7.89
PER368	15	26.2	35.6	8.92
COL1516	23	16.2	33.8	9.74
VEN77	19	15.3	—	—
COL1722	20	9.6	39.3	5.78
COL2627	18	12.9	—	—
SM593-5	25	9.9	35.8	8.04
VEN117B	17	18.5	33.5	7.50
CM6370-2	19	13.5	36.6	10.16
IND27	20	9.5	33.4	9.01
PER234	20	11.6	36.4	6.50
IND129	21	28.6	39.7	5.88
IND135_22	17	32.3	42.4	5.88
BRA158	36	5.9	37.6	8.61
COL2246_24	17	11.1	34.4	8.44
COL2089	49	2.7	29.6	9.96
COL1736	29	14.9	41.7	5.44
BRA325	60	2.0	34.0	6.54
BRA512	60	3.1	—	—
CUB46	25	13.3	36.0	5.01
VEN25	60	0.4	37.2	7.24
CR63	20	18.3	36.3	6.18

[a] WAB after boiling 30 min; average of 24 pieces of fresh cassava roots per genotype.
[b] DMC determined by oven-drying method; average of six pieces of fresh cassava roots per genotype.
[c] Standard deviation of the dry matter distribution predicted by hyperspectral imaging over a longitudinal slice of fresh cassava root.

with RPD = 3.60. Models with RPD ≥ 3 are considered usable for routine analysis.[12,49] This model was therefore selected for the visualization of DMC distribution in the whole slices of cassava roots (transversal and longitudinal).

Visualization of DMC distribution among cassava genotypes

The model with the best performance (CovSel_MLR) was applied to predict DMC at the pixel level of all hyperspectral images of the 31 cassava genotypes, revealing the distribution of DMC within both longitudinal and transversal fresh cassava slices (Fig. 4). The images are in false colors, with the scale ranging from deep blue (low DMC) to deep red (high DMC). The pixels with the same DMC have the same color. The results highlight a high heterogeneity in DMC distribution: (i) between genotypes with overall higher or lower DMC; and (ii) within a cassava root, with higher DMC values in the peripheral parts where starch reserves are highest, and lower DMC in the internal part, where more fibrous lignocellulosic tissue can be found. The observed distribution of DMC was similar in longitudinal and transversal slices of the same genotype. Although, visually, genotypes with higher WAB appeared to have a higher DMC with a more homogeneous distribution from proximal to distal zone (Fig. 4), no clear correlation was observed between DMC and cooking quality, as indicated by WAB (Table 2) as well as a previous publication.[9]

The heterogeneity index (HI) of the distribution of DMC was evaluated as the standard deviation of the DMC distribution predicted by HSI over the longitudinal slices of fresh cassava roots, for each genotype (Table 2). No correlation between HI and cooking quality of genotypes, as indicated by WAB, was found (Table 2).

Prediction of water absorption and texture (Area)

Performances of models of prediction

PLSR was applied to average spectra (n = 109) of fresh slices of cassava in order to investigate the potential of HSI to predict WAB and texture. Weak prediction performances were obtained for WAB and texture (Area) with, respectively, coefficient of determination for calibration R^2_{cal} = 0.53, root-mean-standard error for calibration

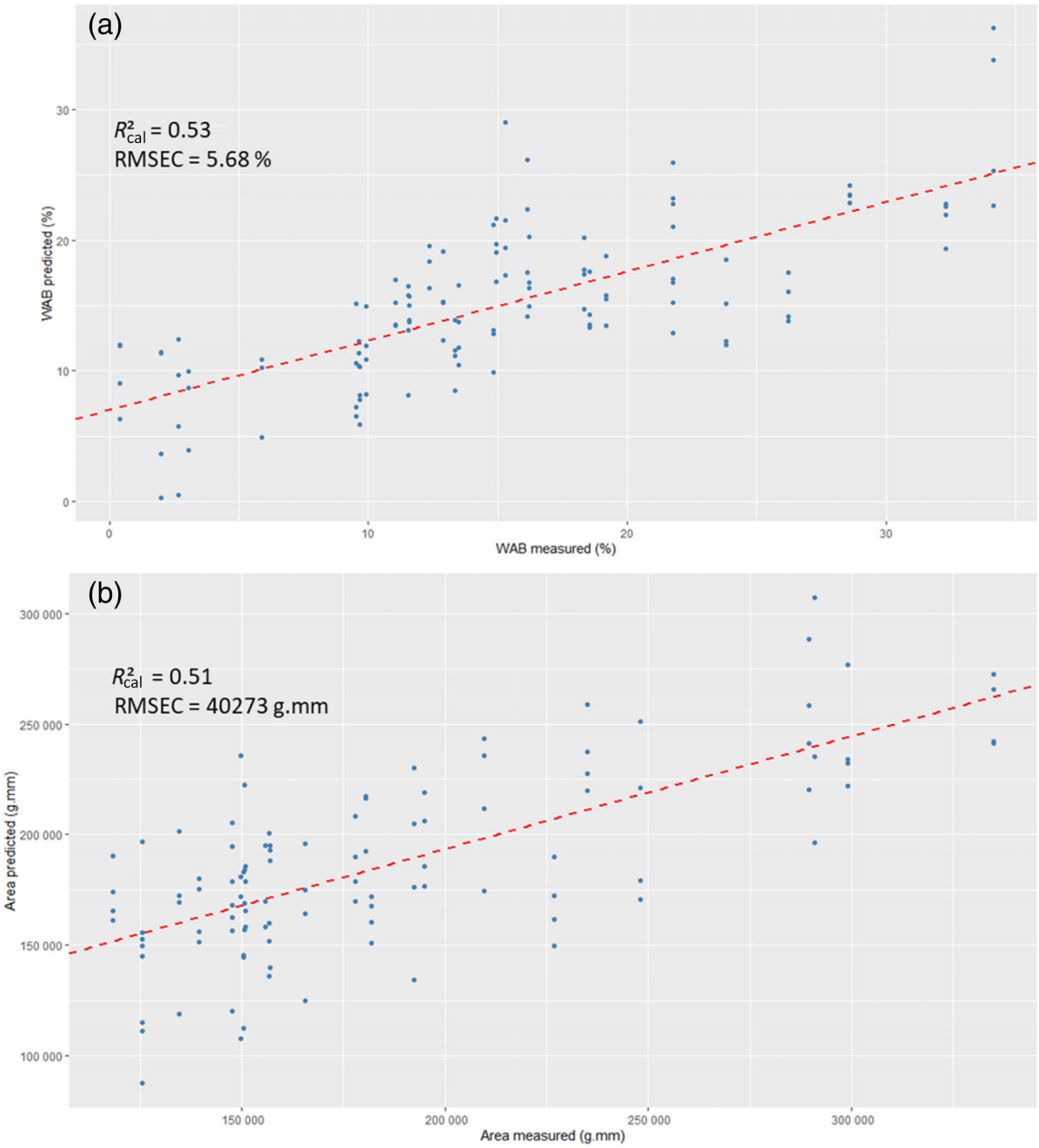

Figure 5. Scatter plot of measured *versus* predicted water absorption (a) and texture (Area) (b) by hyperspectral imaging using partial least-squares regression model performed with training set (n = 109). R^2_{cal}, coefficient of determination for calibration; RMSEC, root-mean-standard error for calibration.

RMSEC = 5.68% and R^2_{cal} = 0.51, RMSEC = 40 273 g mm (Fig. 5). Other texture parameters (initial gradient, maximum force, etc.) gave weak predictions too. Hence, no reliable predictive distribution maps of WAB and texture can be made for the slices of cassava roots, probably due to the limited dataset (n = 109).

Visualization of water absorption distribution in contrasting cooking quality of cassava genotypes

As the previous models did not present satisfactory performances to predict WAB from fresh cassava, a PCA was applied to the HSI data from boiled cassava by following the procedure described in the Materials and methods section. The first component (PC1) loading explained 96% of total variability (Fig. 6). The NIR bands contributing most to PC1 were located at 1131 nm and 1396 nm, corresponding respectively to the first and second overtones of the O—H bond from water.[50] Thus, PC1 was related mostly to water content of the samples after boiling, and by inference to WAB and cooking quality. In order to determine variations in water content within the slices, each pixel of each image was projected onto the PC1 axis. The resulting PC1 score values were arranged together to reconstruct scores-images in false colors, which allow one to observe the variations in water content in boiled cassava slices (Fig. 7). Water content distribution was markedly different among good (CR63, CUB46) and poor (VEN25, BRA512) cooking-quality genotypes, with higher water contents observed in good cooking genotypes (WAB > 12%). Good

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

cooking genotypes appeared to absorb more water in the internal part of the roots, and less in the periphery.

The WAB values measured in the laboratory for the 31 genotypes (Table 2) confirm the high diversity of cooking behaviors among cassava landraces, with variations from 0.4% to 32.6% WAB. The results also confirm the negative correlation between WAB and CT, previously demonstrated by Tran *et al.*[9]

DISCUSSION

The performance of the model developed for DMC quantification (RPD = 3.6) in fresh intact (i.e. not mashed) cassava roots was sufficient for high-throughput phenotyping purposes. These results were better than those obtained by Ikeogu *et al.*[12] and Maraphum *et al.*[15] (RPD = 1.63–2) using classical VNIR spectrophotometry (350–2500 nm) for prediction of DMC in intact cassava roots, possibly because of the much higher number of spectra collected by HSI in one scan. On the other hand, Sánchez *et al.*[51] and Ikeogu *et al.*[12] obtained an even better accuracy (RPD = 4.3) for prediction of DMC in mashed cassava roots, also using VNIR spectrophotometry. This difference in accuracy is due to the higher homogeneity of mashed samples, which results in lower light scattering and thus lower noise in the spectra.

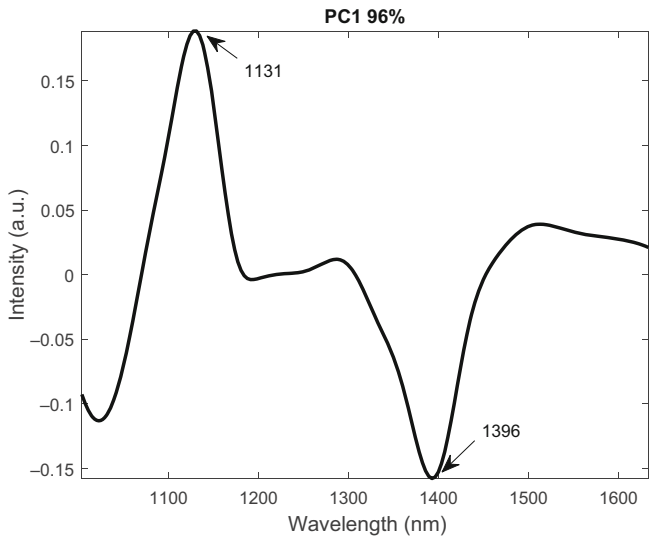

Figure 6. Principal components analysis loading showing the characteristic bands of water, 1131 and 1396 nm (first overtone of O—H), explained by the PC1 axis.

To our knowledge, the present study is the first carried out for the prediction of DMC in cassava roots using HSI, although many studies have been done for DMC prediction in potato tubers.[25] In particular, locally weighted partial least-squares regression (LWPLSR) achieved high accuracy for DMC quantification in potato slices (R^2_p = 0.98, RMSEP = 0.016%). The LWPLSR chemometric method can be used in cases of large databases containing thousands of samples from different locations, years, harvests, and so on.[11] For cassava, further data collection is needed to apply LWPLSR, as the current number of samples is still limited (n = 116).

Using HSI for predicting DMC *per se* may not be practical, as other high-throughput phenotyping approaches, such as portable NIR spectroscopy and the conventional gravimetric method allow for the screening of large numbers of genotypes reliably and inexpensively. However, a unique advantage of HSI is the analysis of the heterogeneity of DMC along and across the sample, which for some products is a key quality trait to take into account (e.g. French fries).

The low performances of the models for prediction of texture (Area) and WAB can be explained by various factors, such as weak correlation between biochemical composition of the samples and physical parameters (cooking quality), defects and diseases of some of the samples, and NIR range. The biochemical constituents related to WAB or texture are in low concentration and so do not impact the spectral fingerprint that much. Also, biochemical reactions and changes in the structure of the root matrix may take place during boiling, which the compositional

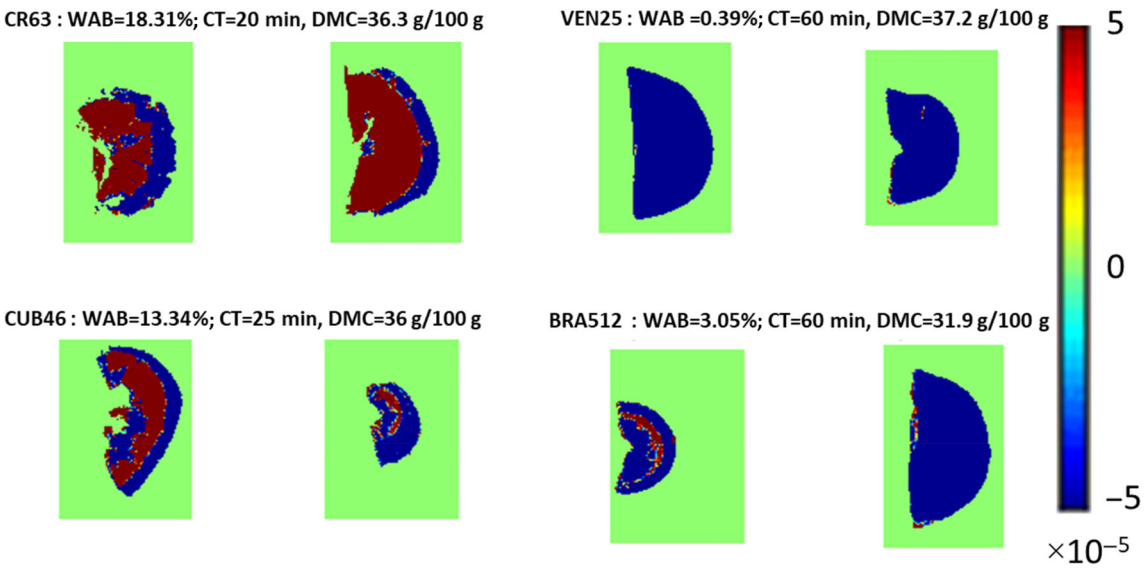

Figure 7. Visualization map of water absorption (WAB) variation of slices (transversal) of boiled cassava generated from PC1 scores of a principal components analysis of hyperspectral imaging data. Four contrasted genotypes are presented, from good (CR63, CUB46; WAB > 12%) to poor (VEN25, BRA512, WAB < 12%) cooking quality. The false colors scale indicates low and high water contents in blue and red hues respectively. DMC, dry matter content.

information contained in the NIR spectra of fresh cassava roots may not be able to fully predict. Nevertheless, high performances were obtained using classification approaches for the prediction of physical parameters by HSI in other crops,[27,28,32,35,52,53] so further data collection to increase the database of HSI, WAB, and texture data may also improve the prediction models for boiled cassava.

The visualization of water content after boiling among contrasting cooking quality cassava genotypes was investigated using the PCA method. The visualization of PC1 scores (Fig. 7) highlighted higher water contents after boiling in high WAB genotypes (WAB > 12%, i.e. good cooking) and that water absorption during boiling spreads from the internal part to the periphery of the half-cylinders of cassava roots. HSI may be used for a more in-depth investigation of water absorption, by taking hyperspectral images of root slices at increasing boiling times (e.g. every 10 min from 10 to 60 min) and analyzing the rate of migration of the water front from the internal part to the periphery for a selection of cassava genotypes with contrasting cooking quality. In this study, the potential of hyperspectral imaging for the visualization of DMC and WAB in fresh and cooked slices of cassava was evaluated. A high-quality model was developed for the prediction of DMC distribution in fresh slices of cassava using the CovSel_MLR model; however, a model for prediction of WAB and texture by HSI may necessitate collecting a larger dataset. In the meantime, PCA coupled with HSI was used for the visualization of WAB variation in cooked slices of boiled cassava, with the potential to investigate the rate of absorption of water during boiling, from the central part towards the periphery of the slices. No correlation was found between cooking quality and DMC of fresh roots, as observed previously.[9] Moreover, HSI revealed that the heterogeneity of DMC distribution within fresh cassava roots had no effect on cooking quality either. Contrary to cassava, in other roots, tubers, and bananas (RTBs), such as yam, dry matter is a key factor determining cooking quality and consumers' preferences.[54] Hence, applying HSI to other RTBs may result in more effective predictions of cooking quality than was found for cassava in the present study. The model to predict and visualize the distribution of DMC and the PCA approach to visualize the distribution of water content after boiling can contribute to a better understanding of cooking behavior of boiled cassava and selection of desirable genotypes. The weak performance of the PLSR models to predict physical parameters (WAB and texture) may be due to the lack of direct correlation between NIR–HSI and physical constituents, and/or the low number of samples in the dataset.

CONCLUSION

The current work demonstrated the potential of HSI as a high-throughput tool, since it can detect differences between genotypes. In particular, the heterogeneity of distribution of DMC between the center and the periphery and between the proximal and distal ends of cassava roots is very useful. This could help in understanding, for example, the variation in concentration of relevant nutritive constituents, such as carotenoids.[55]

Future research on HSI applied to prediction of cooking quality of cassava can investigate several points, including to increase the dataset of HSI, DMC, WAB, and texture to more than 200 samples, in order to confirm the lack of correlation between the heterogeneity index of DMC distribution and cooking quality, and to improve the model of prediction of WAB and/or texture; to develop a dataset of HSI and various components of cassava roots (starch, amylose, pectins, etc.) in order to try to better predict cooking quality; and to apply HSI to predict cooking quality of other RTBs.

ACKNOWLEDGEMENTS

We are grateful to CGIAR Research Program on Roots, Tubers and Bananas (RTB) and the grant opportunity INV-008567 (formerly OPP1178942): Breeding RTB Products for End User Preferences (RTBfoods) from the Bill & Melinda Gates Foundation (BMGF), coordinated by the French Agricultural Research Centre for International Development (CIRAD, Montpellier, France): https://rtbfoods.cirad.fr. Previous versions of this paper and the ideas in it benefited greatly from suggestions and comments by Dr Xiaofei Zhang. We particularly thank Dr Hernan Ceballos for his expert feedback on the paper's content and Clair Hershey for the English editing. Particular thanks also go to the substantial work of the field team of CIAT's Cassava Program for growing and harvesting the cassava genotypes, and to the excellent logistical organization of the Postharvest Quality laboratory team to prepare all the samples and carry out the multiple analyses with accuracy and timeliness.

CREDIT AUTHORSHIP CONTRIBUTIONS STATEMENTS

Conceptualization: Karima Meghar, Fabrice Davrieux, Thierry Tran, and Dominique Dufour.
Data curation: Karima Meghar, Fabrice Davrieux, Thierry Tran, Luis Fernando Delgado, and Jhon Larry Moreno.
Formal analysis: Karima Meghar, Fabrice Davrieux, and Thierry Tran.
Funding acquisition: Dominique Dufour.
Investigation: Meghar Karima, Thierry Tran, Fabrice Davrieux, and Luis Londoño.
Methodology: Thierry Tran, Karima Meghar, Fabrice Davrieux, Luis Londoño, Delgado Fernando Luis, Maria Alejandra Ospina, Jhon Larry Moreno, and Jorge Luna.
Supervision: Fabrice Davrieux and Thierry Tran.
Writing – original draft: Karima Meghar, Fabrice Davrieux, and Thierry Tran.
Writing – review and editing: Karima Meghar, Fabrice Davrieux, and Thierry Tran.

CONFLICT OF INTEREST

All authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data not yet avalable in public.

REFERENCES

1 Petsakos A, Prager SD, Gonzalez CE, Gama AC, Sulser TB, Gbegbelegbe S *et al.*, Understanding the consequences of changes in the production frontiers for roots, tubers and bananas. *Glob Food Secur* **20**:180–188 (2019). https://doi.org/10.1016/j.gfs.2018.12.005.
2 Drapal M, Ovalle Rivera TM, Becerra Lopez-Lavalle LA and Fraser PD, Exploring the chemotypes underlying important agronomic and consumer traits in cassava (*Manihot esculenta* Crantz). *J Plant Physiol* **251**:153206 (2020). https://doi.org/10.1016/j.jplph.2020.153206.
3 Falade KO and Akingbala JO, Utilization of cassava for food. *Food Rev Int* **27**:51–83 (2010). https://doi.org/10.1080/87559129.2010.518296.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.

4 Teeken B, Agbona A, Bello A, Olaosebikan O, Alamu E, Adesokan M *et al.*, Understanding cassava varietal preferences through pairwise ranking of *gari-eba* and *fufu* prepared by local farmer–processors. *Int J Food Sci Technol* **56**:1258–1277 (2021). https://doi.org/10.1111/ijfs.14862.

5 Eggleston G and Asiedu R, Effect of boiling on the texture of cassava clones: a comparison of compressive strength, intercellular adhesion and physio-chemical composition of the tuberous roots. *Trop Sci* **34**:259–273 (1994).

6 Franck H, Christian M, Noël A, Brigitte P, Joseph HD, Cornet D *et al.*, Effects of cultivar and harvesting conditions (age, season) on the texture and taste of boiled cassava roots. *Food Chem* **126**:127–133 (2011). https://doi.org/10.1016/j.foodchem.2010.10.088.

7 Kouadio OK, Nindjin C, Bonfoh B, N'dri D and Amani GN, Water absorption as an evaluation method of cooking quality for yam (*Dioscorea alata*) and cassava (*Manihot esculenta* Crantz). *Proc Food Sci* **1**:153–159 (2011). https://doi.org/10.1016/j.profoo.2011.09.024.

8 Padonou W, Mestres C and Nago MC, The quality of boiled cassava roots: instrumental characterization and relationship with physicochemical properties and sensorial properties. *Food Chem* **89**:261–270 (2005). https://doi.org/10.1016/j.foodchem.2004.02.033.

9 Tran T, Zhang X, Ceballos H, Moreno JL, Luna J, Escobar A *et al.*, Correlation of cooking time with water absorption and changes in relative density during boiling of cassava roots. *Int J Food Sci Technol* **56**:1193–1205 (2021). https://doi.org/10.1111/ijfs.14769.

10 Belalcazar J, Dufour D, Andersson MS, Pizarro M, Luna J, Londoño L *et al.*, High-throughput phenotyping and improvements in breeding cassava for increased carotenoids in the roots. *Crop Sci* **56**:2916–2925 (2016). https://doi.org/10.2135/cropsci2015.11.0701.

11 Davrieux F, Dufour D, Dardenne P, Belalcazar J, Pizarro M, Luna J *et al.*, LOCAL regression algorithm improves near infrared spectroscopy predictions when the target constituent evolves in breeding populations. *J Near Infrared Spectrosc* **24**:109–117 (2016). https://doi.org/10.1255/jnirs.1213.

12 Ikeogu UN, Davrieux F, Dufour D, Ceballos H, Egesi CN and Jannink J-L, Rapid analyses of dry matter content and carotenoids in fresh cassava roots using a portable visible and near infrared spectrometer (Vis/NIRS). *PLoS One* **12**:e0188918 (2017). https://doi.org/10.1371/journal.pone.0188918.

13 Lebot V, Champagne A, Malapa R and Shiley D, NIR determination of major constituents in tropical root and tuber crop flours. *J Agric Food Chem* **57**:10539–10547 (2009). https://doi.org/10.1021/jf902675n.

14 Maraphum K, Saengprachatanarug K, Wongpichet S, Phuphaphud A and Posom J, In-field measurement of starch content of cassava tubers using handheld vis–near infrared spectroscopy implemented for breeding programmes. *Comput Electron Agric* **175**:105607 (2020). https://doi.org/10.1016/j.compag.2020.105607.

15 Maraphum K, Saengprachatanarug K, Wongpichet S, Phuphuphud A and Posom J, Achieving robustness across different ages and cultivars for an NIRS–PLSR model of fresh cassava root starch and dry matter content. *Comput Electron Agric* **196**:106872 (2022). https://doi.org/10.1016/j.compag.2022.106872.

16 Moresco R, Uarrota VG, Pereira A, Tomazzoli M, Nunes EC, Peruch LAM *et al.*, Carotenoid analysis of cassava genotypes roots (*Manihot esculenta* Crantz) cultivated in southern Brazil using chemometric tools, in *9th International Conference on Practical Applications of Computational Biology and Bioinformatics*, ed. by Overbeek R, Rocha MP, Fdez-Riverola F and de Paz JF. Springer International Publishing, Cham, Switzerland, pp. 11–18 (2015). https://doi.org/10.1007/978-3-319-19776-0_2.

17 Phambu N, Meya AS, Djantou EB, Phambu EN, Kita-Phambu P and Anovitz LM, Direct detection of residual cyanide in cassava using spectroscopic techniques. *J Agric Food Chem* **55**:10135–10140 (2007). https://doi.org/10.1021/jf072046i.

18 Sanchez PDC, Hashim N, Shamsudin R and Mohd Nor MZ, Applications of imaging and spectroscopy techniques for non-destructive quality evaluation of potatoes and sweet potatoes: a review. *Trends Food Sci Technol* **96**:208–221 (2020). https://doi.org/10.1016/j.tifs.2019.12.027.

19 Amjad W, Crichton SOJ, Munir A, Hensel O and Sturm B, Hyperspectral imaging for the determination of potato slice moisture content and chromaticity during the convective hot air drying process. *Biosyst Eng* **166**:170–183 (2018). https://doi.org/10.1016/j.biosystemseng.2017.12.001.

20 Kjær A, Nielsen G, Stærke S, Clausen MR, Edelenbos M and Jørgensen B, Prediction of starch, soluble sugars and amino acids in potatoes (Solanum tuberosum L.) using hyperspectral imaging, dielectric and LF-NMR methodologies. *Potato Res* **59**:357–374 (2016). https://doi.org/10.1007/s11540-017-9335-2.

21 Liu H, Pandey MK, Xu Z, Rao D, Huang Z, Chen M *et al.*, Analysis and evaluation of quality traits of peanut varieties with near infra-red spectroscopy technology. *Int J Agric Biol* **21**:491–498 (2019).

22 Nigon TJ, Mulla DJ, Rosen CJ, Cohen Y, Alchanatis V, Knight J *et al.*, Hyperspectral aerial imagery for detecting nitrogen stress in two potato cultivars. *Comput Electron Agric* **112**:36–46 (2015). https://doi.org/10.1016/j.compag.2014.12.018.

23 Rady A, Guyer D and Lu R, Evaluation of sugar content of potatoes using hyperspectral imaging. *Food Bioprocess Technol* **8**:995–1010 (2015). https://doi.org/10.1007/s11947-014-1461-0.

24 Rady AM, Guyer DE, Kirk W and Donis-González IR, The potential use of visible/near infrared spectroscopy and hyperspectral imaging to predict processing-related constituents of potatoes. *J Food Eng* **135**:11–25 (2014). https://doi.org/10.1016/j.jfoodeng.2014.02.021.

25 Su W-H and Sun D-W, Chemical imaging for measuring the time series variations of tuber dry matter and starch concentration. *Comput Electron Agric* **140**:361–373 (2017). https://doi.org/10.1016/j.compag.2017.06.013.

26 Botha EJ, Zebarth BJ and Leblon B, Non-destructive estimation of potato leaf chlorophyll and protein contents from hyperspectral measurements using the PROSPECT radiative transfer model. *Can J Plant Sci* **86**:279–291 (2006). https://doi.org/10.4141/P05-017.

27 Gao H-L, Li X-Y, Xu S-M, Tao H-L, Li X-J and Sun J-F, Comparative study of transmission and reflection hyperspectral imaging technology for potato damage detection. *Spectrosc Spectral Anal* **33**:3366–3371 (2013).

28 Nguyen Do Trong N, Tsuta M, Nicolaï BM, De Baerdemaeker J and Saeys W, Prediction of optimal cooking time for boiled potatoes by hyperspectral imaging. *J Food Eng* **105**:617–624 (2011). https://doi.org/10.1016/j.jfoodeng.2011.03.031.

29 Su W-H and Sun D-W, Comparative assessment of feature-wavelength eligibility for measurement of water binding capacity and specific gravity of tuber using diverse spectral indices stemmed from hyperspectral images. *Comput Electron Agric* **130**:69–82 (2016). https://doi.org/10.1016/j.compag.2016.09.015.

30 Dacal-Nieto A, Formella A, Carrión P, Vazquez-Fernandez E and Fernández-Delgado M, Common scab detection on potatoes using an infrared hyperspectral imaging system, in *Image Analysis and Processing – ICIAP 2011*. Lecture Notes in Computer Science, Vol. **6979**, ed. by Maino G and Foresti GL. Springer, Berlin, Germany, pp. 303–312 (2011).

31 Groinig M, Burgstaller M and Pail M, Industrial application of a new camera system based on hyperspectral imaging for inline quality control of potatoes. In OAGM/AAPR Workshop, Graz (2011), FH Upper Austria, Wels Campus, Vienna, 11–13 May 2011; pp. 1–8.

32 Huang T, Li XY, Jin R, Ku J, Xu SM, Xu ML *et al.*, Multi-target recognition of internal and external defects of potato by semi-transmission hyperspectral imaging and manifold learning algorithm. *Spectrosc Spectral Anal* **35**:992–996 (2015) (in Chinese).

33 Huang T, Li X, Xu M, Jin R, Ku J, Xu S *et al.*, Non-destructive detection research for hollow heart of potato based on semi-transmission hyperspectral imaging and SVM. *Spectrosc Spectral Anal* **35**:198–202 (2015).

34 Islam M, Wahid K and Bhowmik P, Detection of potato diseases using image segmentation and multiclass support vector machine. In 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 1–4 (2017). https://doi.org/10.1109/CCECE.2017.7946594.

35 Lopez-Maestresalas A, Perez C, Tierno R, Arazuri S, Galarreta JIRD and Jaren C, Prediction of main potato compounds by NIRS. *Chem Eng Trans* **58**:385–390 (2017). https://doi.org/10.3303/CET1758065.

36 Noordam JC, van den Broek WH and Buydens LM, Detection and classification of latent defects and diseases on raw French fries with multispectral imaging. *J Sci Food Agric* **85**:2249–2259 (2005). https://doi.org/10.1002/jsfa.2226.

37 Ye D, Sun L, Tan W, Che W and Yang M, Detecting and classifying minor bruised potato based on hyperspectral imaging. *Chemom Intell Lab Syst* **177**:129–139 (2018). https://doi.org/10.1016/j.chemolab.2018.04.002.

38 Zhou Z, Li X, Tao H and Gao H, Detection of potato external defects based on hyperspectral imaging technology. *Trans Chin Soc Agric Eng* **28**:221–228 (2011) (in Chinese with English abstract).
39 Meghar K and Davrieux F, SOP for operating mode and parameters configuration of hyperspectral camera Specim FX17. High-throughput phenotyping protocols (HTPP), WP3. RTBfoods, Montpellier, France (2021). https://doi.org/10.18167/AGRITROP/00667.
40 Sánchez T, Dufour D, Moreno JL, Pizarro M, Aragón IJ, Domínguez M *et al.*, Changes in extended shelf life of cassava roots during storage in ambient conditions. *Postharvest Biol Technol* **86**:520–528 (2013). https://doi.org/10.1016/j.postharvbio.2013.07.014.
41 Escobar Salamanca AF, Tran T and Arufe Vilas S, SOP for characterization of water absorption, cooking time and closing angle of boiled cassava. Biophysical characterization of quality traits, WP2. RTBfoods, Cali, Colombia (2022). https://doi.org/10.18167/AGRITROP/00683.
42 Tran T and Escobar A, SOP for characterization of extrusion-texture of boiled cassava. Biophysical characterization of quality traits, WP2. RTBfoods, Cali, Colombia (2023). https://doi.org/10.18167/AGRITROP/00594.
43 Stevens A and Ramirez-Lopez L, An introduction to the prospectr package. R package Vignette R package version 0.2.6 (2022).
44 Lesnoff, M. (2022). mlesnoff/rchemo: Dimension reduction, Regression and Discrimination for Chemometrics. https://rdrr.io/github/mlesnoff/rchemo/.
45 Wold S, Sjöström M and Eriksson L, PLS-regression: a basic tool of chemometrics. *Chemom Intell Lab Syst* **58**:109–130 (2001). https://doi.org/10.1016/S0169-7439(01)00155-1.
46 Lesnoff M, Metz M and Roger J-M, Comparison of locally weighted PLS strategies for regression and discrimination on agronomic NIR data. *J Chemom* **34**:e3209 (2020). https://doi.org/10.1002/cem.3209.
47 Teng X, Zhang X and Yu H, Using least squares support vector machines for frequency estimation. *Int J Commun Network Syst Sci* **3**:821–825 (2010). https://doi.org/10.4236/ijcns.2010.310111.
48 Roger JM, Palagos B, Bertrand D and Fernandez-Ahumada E, CovSel: variable selection for highly multivariate and multi-response calibration. Application to IR spectroscopy. *Chemom Intell Lab Syst* **106**: 216–223 (2010). https://doi.org/10.1016/j.chemolab.2010.10.003.
49 Nicolaï BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI *et al.*, Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol* **46**: 99–118 (2007). https://doi.org/10.1016/j.postharvbio.2007.06.024.
50 Ignat T, Lurie S, Nyasordzi J, Ostrovsky V, Egozi H, Hoffman A *et al.*, Forecast of apple internal quality indices at harvest and during storage by VIS–NIR spectroscopy. *Food Bioprocess Technol* **7**:2951–2961 (2014). https://doi.org/10.1007/s11947-014-1297-7.
51 Sánchez T, Ceballos H, Dufour D, Ortiz D, Morante N, Calle F *et al.*, Prediction of carotenoids, cyanide and dry matter contents in fresh cassava root using NIRS and Hunter color techniques. *Food Chem* **151**: 444–451 (2014). https://doi.org/10.1016/j.foodchem.2013.11.081.
52 Zhou Z, Zeng S, Li X and Zheng J, Nondestructive detection of blackheart in potato by visible/near infrared transmittance spectroscopy. *J Spectrosc* **2015**:786709 (2015). https://doi.org/10.1155/2015/786709.
53 Su W-H and Sun D-W, Multivariate analysis of hyper/multi-spectra for determining volatile compounds and visualizing cooking degree during low-temperature baking of tubers. *Comput Electron Agric* **127**:561–571 (2016). https://doi.org/10.1016/j.compag.2016.07.007.
54 Otegbayo B, Madu T, Oroniran O, Chijoke U, Fawehinmi O, Okoye B *et al.*, End-user preferences for pounded yam and implications for food product profile development. *Int J Food Sci Technol* **56**:1458–1472 (2021). https://doi.org/10.1111/ijfs.14770.
55 Ceballos H, Luna J, Escobar AF, Ortiz D, Pérez JC, Sánchez T *et al.*, Spatial distribution of dry matter in yellow fleshed cassava roots and its influence on carotenoid retention upon boiling. *Food Res Int* **45**: 52–59 (2012). https://doi.org/10.1016/j.foodres.2011.10.001.
56 R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org.

Journal of The Science of Food and Agriculture published by John Wiley & Sons Ltd on behalf of Society of Chemical Industry.