Laboratory Scientific Report

Proof of Concept of Visualisation of Dry Matter Content in Yam Tuber during Storage, using Hyperspectral Imaging

High-Throughput Phenotyping Protocols (HTPP), WP3

Montpellier, France, 28/06/2021

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

Fabrice DAVRIEUX, CIRAD, Saint-Pierre, Réunion, France (Validator)

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

To be cited as:

Karima MEGHAR, Fabrice DAVRIEUX (2022). *Proof of Concept of Visualisation of Dry Matter Content in Yam Tuber during Storage, using Hyperspectral Imaging.* Montpellier, France: RTBfoods Laboratory Scientific Report, 16 p. http://doi.org/10.18167/agritrop/00716

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. 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) Karima MEGHAR (CIRAD)	24/01/2022 24/01/2022
Final validation by:	
Fabrice DAVRIEUX (CIRAD)	24/01/2022

CONTENTS

Table of Contents

Table of Figures

ABSTRACT

Fast screening methods are needed for plant breeding. The objective of this research was to evaluate the potential of hypespectral imaging coupled with Near infrared spectroscopy (HSI-NIR) for the analysis of dry matter (DM) in intact discs of fresh yam tubers during storage process (0, 15, 30, 45, 60 and 75 days) at 16 °C and 70 % of relative humidity. Partial least square regression (PLS) was applied to establish calibration. The coefficient of determination for calibration (R^2) was equal to 0.89 and the root mean square of error of cross validation (RMSECV) was 2.85 % with a ratio performance to deviation (RPD) of 3.05. Finally, distribution maps were displayed to visualize DM repartition within tubersat different storage days.

These results suggested that HSI-NIRS can be used to determine dry matter in fresh tuber samples of yam with acceptable accuracy. Further research will have to determine if additional traits can be incorporated into this scheme.

Key Words: Hyperspectral imaging, yam tuber, Dry matter, PLS regression

1 MATERIALS AND METHODS

1.1 Yam samples

Eighteen yam tubers (six tubers per origin) used in this study were purchased from a local market in Montpellier, France. The tubers were from three different origins Ghana (GHA), France (FRA) and Brazil (BRA). After purchase, these were stored in cold room at 16 °C and 71 % of relative humidity. Indoor air flow and daily renewal of the atmosphere are also provided.

1.2 Storage process and samples preparation

For this study, yam tubers from different origins were used in order to have a maximum of inter tuber variability, these tubers were stored under controlled conditions at different storage times 0, 15, 30, 45, 60 and 75 days after purchase. For each storage time, one tuber of each origin was analysed. The tubers must have at least 30 cm length and 7 cm in diameter.

Yam tubers were peeled washed and dried using towel paper. Then, the tuber was cut into six cylindrical slices of equal size (5,7 cm of diameter and 2 cm thick): 2 slices per part of the tuber: proximal, central and distal (fig.1).

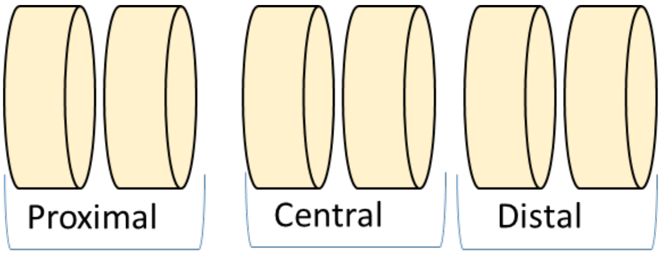


Figure 1 : yam tuber divided in six slices (2 proximal, 2 central and 2 distal).

1.3 Acquisition of hyperspectral images

The hyperspectral measurements were carried out with a hyperspectral camera Specim FX17 (ImSpector, N17E, SPECIM, Finland) with a spectral range of 932.95 - 1721.16 nm. The measurements were carried out at different storage times of 0, 15, 30, 45, 60 and 75 days. For each storage time, one tuber of each origin was used. Two hyperspectral images were acquired per slice immediately after cutting. The interval time between cutting and measuring should be as short as possible to avoid oxidation and water evaporation from the slices. At the end of the storage process 198 images were acquired, table 1 summarises the number of samples (slices) acquired by origin and storage days.

Table 1 : number of images acquired by origin and storage days

Origin/storage days	A00	A15	A30	A45	A60	A75	Total
BRE	12	12	12	8	12	12	68
FRA	12	12	12	10	12	12	70
GHA	12	12	12	12	12	00	60
Total	**36**	**36**	**36**	**30**	**36**	**24**	**198**

1.4 Image correction

The hyperspectral images were firstly corrected with a white and a dark reference. The dark reference was used to remove the effect of the dark current of the thermally sensitive CCD detectors. The corrected image (R) is estimated using Eq. (1):

$$R = \frac{\mathrm{R0} - \mathrm{D}}{W - D}$$

Where R0 is the recorded hyperspectral image, D the dark image and W is the white reference image. The corrected images R will be the basis for the subsequent image analysis to extract information about the spectral properties of fresh yam sample for optimizing surface characteristics identification, selection of effective wavelengths and texture analysis purposes (ElMasry et al., 2007).

1.5 Selection of the region of interest (ROI)

The spectral response of fresh yam samples could be used to characterize and identify the sample uniquely. To collect the spectral response of each sample, a binary mask is first created to produce an image containing only the fresh yam in the image, avoiding any interference from the background. Here, an image at 1325 nm wavelength is taken for this task because the tuber appeared opaque compared with the background and can be segmented easily by simple thresholding at the level of 0.3176. All active pixels in the segmented image are used as a mask to identify all pixels belonging to the yam sample (ROI) and set the others to zero background (fig.4). At each pixel of the region of interest (ROI), the relative reflectance is recorded at each wavelength from 932.95 to 1721.16 nm. Each segmented image contains more than 100000 pixels.

1.6 Unfolding the hyperspectral (hypercube) image (x, y, z) into a 2D matrix (z,x ×y)

Before application of chemometrics methods, it was imperative to unfold the three dimensional hypercube (x, y, z) into a two-dimensional matrix (z,x ×y) where each row represents the spectrum of a pixel and the columns represent one wavelength (fig.2).

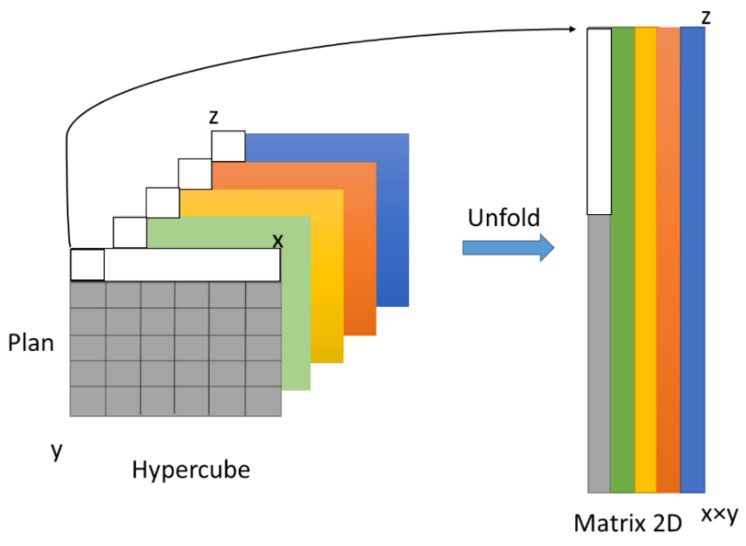


Figure 2 : schema of the unfolding of the hyperspectral image (hypercube) (x y z) to a 2D spectral matrix (z, x×y).

1.7 Calculation of spectral data

The reflectance values of all pixels from one image were averaged to obtain only one mean spectrum for each sample. The same procedure was repeated to obtain the spectrum for all hyper- spectral images of all tested samples. The extracted spectrum of each sample was then arranged together to form a spectral matrix. Background segmentation and extraction of spectral data from hyperspectral images were programmed in MatLab R2018b (The Mathworks Inc., Mass, USA).

1.8 Dry matter measurement

Dry matter (DM) measurements were performed immediately after the acquisition of the HSI images. To do this, each slice is divided into three cylinders A1, A2 and A3 with identical volumes, using dies with diameters of 3.4, 4.9 and 6 cm respectively (fig.3). The DM measurements were performed on the A1, A2 and A3 cylinders of each slice at of 0, 15, 30, 45, 60 and 75 days by freeze drying. In total 297 values of DM were measured, the table 2 summarises the number of DM measurements by origin and storage days. In this work we calculated the mean DM value of the three values of each slice. Therefore, 297/3=99 values of DM were obtained. Each value was associated to 2 images acquired in the same slice.

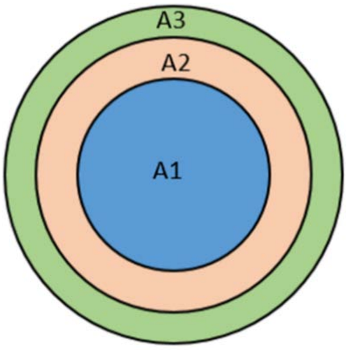


Figure 3 : Fresh yam slice divided into three slices.

Table 2 : number of DM values measured by origin and storage days

Origin/storage days	A00	A15	A30	A45	A60	A75	Total
BRE	18	18	18	12	18	18	102
FRA	18	18	18	15	18	18	105
GHA	18	18	18	18	18	00	90
Total général	**54**	**54**	**54**	**45**	**54**	**36**	**297**

1.9 Multivariate analysis

1.9.1 Principal components analysis (PCA)

Initial investigation was carried out using PCA to differentiate between fresh yam samples from three origins (Ghana, France and Brazil) and also to distinguish between samples from different storage days 0, 15, 30, 45, 60 and 75 days of storage. PCA transforms spectral data into several principal components (PCs), which are linear combinations of the original spectral data. The first few PCs resulting from PCA retain most of the variation present in all of the original variables. Each PC can be interpreted independently, which permits an overview of the data structure by revealing the relationship between the objects.

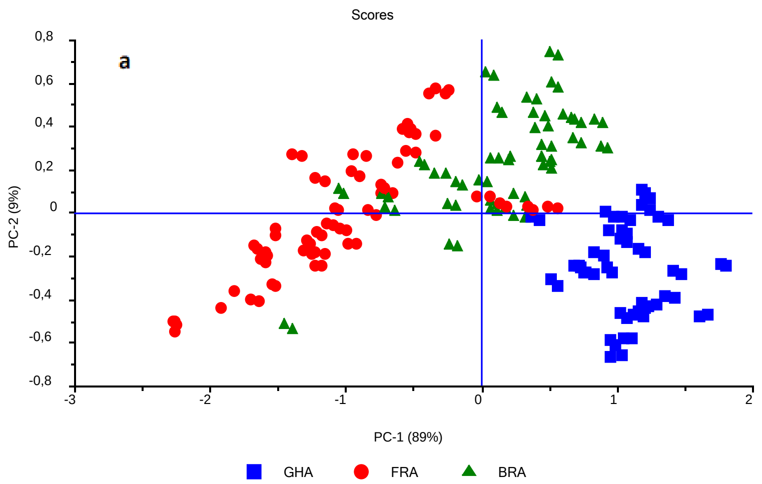
1.9.2 Partial Least Square Regression (PLSR)

PLSR was applied to predict the dry matter content in fresh yam slices during the storage. It is a very popular and effective technique in spectral analysis where predictors (wavelengths) are highly correlated and the number of predictors is much greater than that of observations [52]. The PLSR compresses the spectral data into orthogonal structures called latent variables (LVs) which describe the maximum covariance between X and Y [53]. Ideally; only the first few LVs carry the overwhelming majority of relevant infor mation in the original variables. The optimum number of LVs should be used to obtain efficient and reliable prediction models and in this study, it was determined at the minimum value of the root mean squared error estimated by leave-one-out cross- validation (RMSECV).

2 Results and Discussion

2.1 Primary investigation by PCA

PCA was used to interpret and visualize the spectral data to highlight their properties, groupings, similarities and differences by identifying the most important directions according to spectral features of the samples. PCA was applied to 192 mean spectra of hyperspectral images. The scores plot (fig. 4.a) of the first two PCs (explained 98 % of variation) revealed a clear clustering between yam samples from different origins, where three groups were observed with a slight overlap between samples from France and Brazil. On the other hand no clear clustering between samples from different storage days was observed (Fig.4.b) . These results showed the potential of HIS to differentiate between yam tubers from different origins.



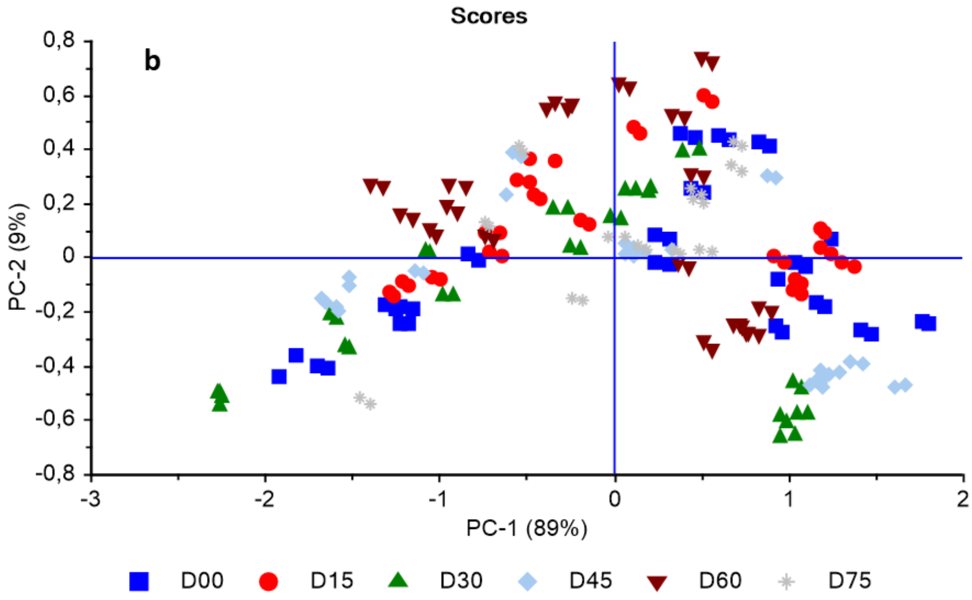


Figure 4 : Scores plot of the first and the second component of PCA for fresh yam slices from three origins, Ghana, France and Brazil at spectral range 968-1685 nm, (a) coloured by origins, (b) coloured by days of storage.

PCA was also applied to 56 Ghana samples from 0 to 60 storage days, the score plot (Fig.5) of the first two PCs (explained 98 % of variation). revealed a clustering between yam samples from different storage times, the samples are grouped as follow: J00 and J15 grouped in one group, J30 an J45 together and J60. These results showed that the storage duration changed the chemical composition of the tubers (especially DM content) and this can be monitoring using HSI.

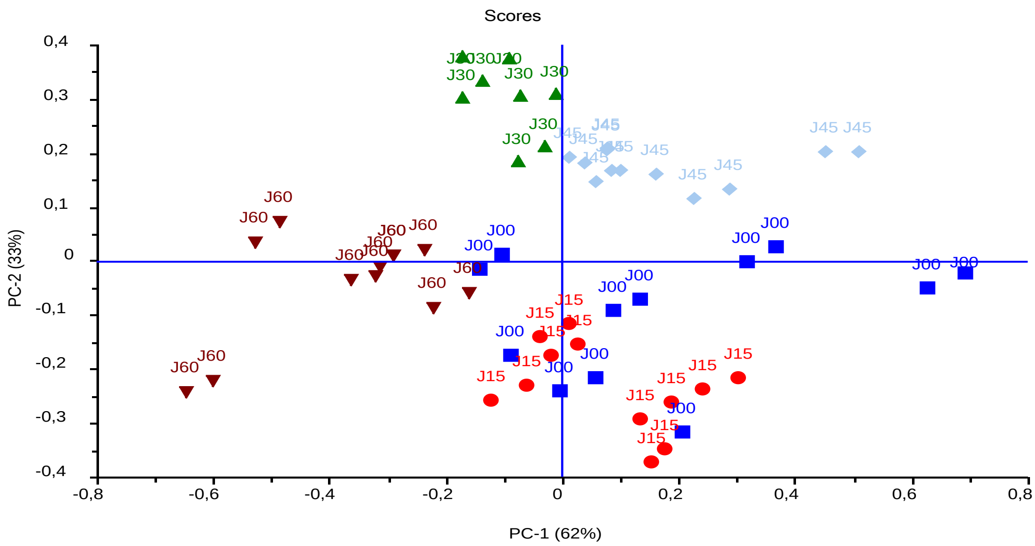


Figure 5 : scores plot of the first and the second component of PCA for fresh yam slices samples for Ghana coloured by storage days.

2.2 Dry matter (DM) analysis

Reference measurements for DM are reported in Table 3. Samples for 0 to 45 days of storage were used as calibration set. DM content ranged from 18.44% to 46.83% with an average of 30.64 % and SD of 8.64 %. samples from 60 to 75 days were used as the validation set, DM content ranged from 16.85% to 35.87 % with an average of 25.05 % and SD of 6.54 %.

Table 3 : Descriptive statistics for dry matter content measured by freeze drying in yam tuber pieces.

DM (%)	N	Min	Max	Mean	SD
Calibration (0 to 45 days)	132	18.44	46.83	30.64	8.64
Validation (60 to 75 days)	59	16.85	35.87	25.04	6.54

The variation of dry matter content in yam tuber of the three origins during the storage is shown in fig.6. DM content from Ghana samples is higher compared to brazil and France samples whatever storage days. The DM content for all origins increase from 0 to 45 days of storage, decrease at 60 of storage days and increase again at 75 days of storage days. Ho 1 demonstrated that the DM content decrease in yam tuber during the storage, the decrease in DM to 60 could be explained by the use of different tubers at each storage days and then the DM content is not known for each tuber upon receipt of yam tubers.

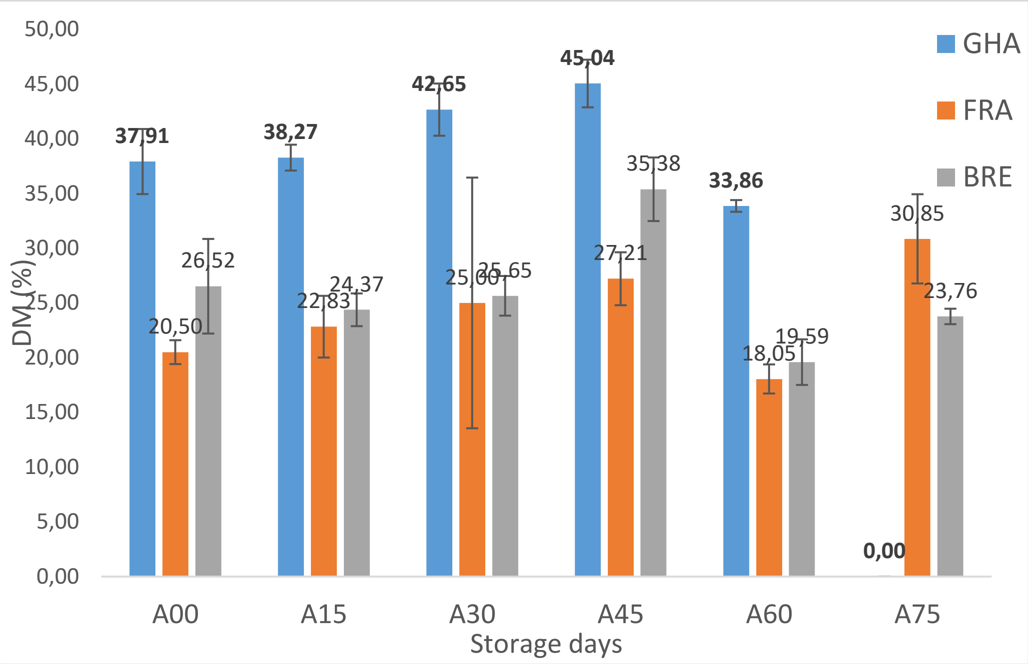


Figure 6 : Evolution of dry matter content in yam tubers from Ghana, France and Brazil from 0 to 75 days of storage (n=192).

2.3 PLS model for the prediction of DM by HSI

The PLS calibration models were established using the average spectra of 132 hyperspectral images from 0 to 45 days of storage in the calibration/training in the spectral range 932-1721 nm. The model was validated using 60 average spectra of 60 samples from 60 to 75 days of storage. The best spectral pre-processing was Savitzky Golay 1st derivative and SNV at 968-1685 nm.

The number of latent factors for PLS model is determined at the lowest value of Root Mean Square Error of Cross Validation (RMSECV).as shown in fig.7. The optimum number of latent variables to predict DM is LV=3.

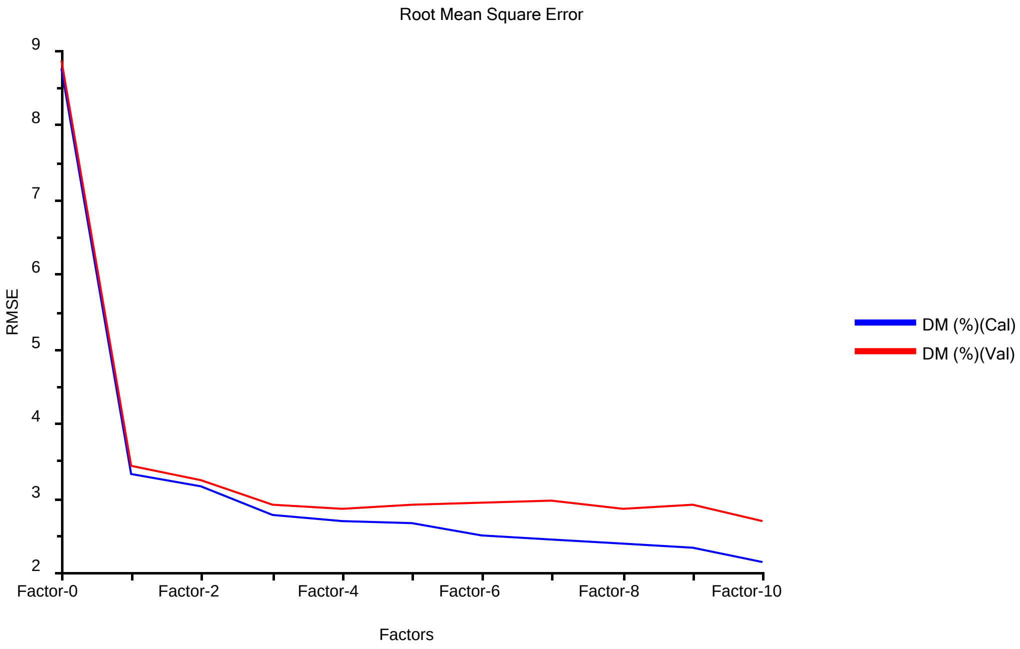


Figure 7 : predict value of Root Mean Square Error of Cross Validation (RMSECV) in red for predicting DM

PLS regression was used to build prediction model. The Performance of the predictive model is shown in Table 4. Good prediction of DM was achieved with $R^2CV=0.89$, RMSECV=2.85 % and in calibration and $R^2P=0.80$ and RMSEP=3.88 % in validation. With an RPD=3.05 the performance of this model is sufficient for the quantification of DM in high throughput phenotyping of yam tuber. This results almost comparable to the work of 2using traditional NIR spectroscopy techniques to predict DM content in fresh yam slices with dry matter content (R2=0.94; standard error of performance, SEP=1.2%). However, as expected for the fresh yam slices the performance of the model are lower than those previously published for ground fresh yam, due to the less uniform homogenous sample presentation.

Table 4 : performance of PLS regression model for DM content in fresh yam tuber using HSI

	Calibration						Validation		
N	**LV**	**RMSEC (%)**	**R^2cal**	**RMSECV(%)**	**R^2 CV**	**RPDcv**	**N**	**R^2p**	**RMSEP (%)**
132	3	2.79	0.90	2.85	0.89	3.05	60	0.80	3.88

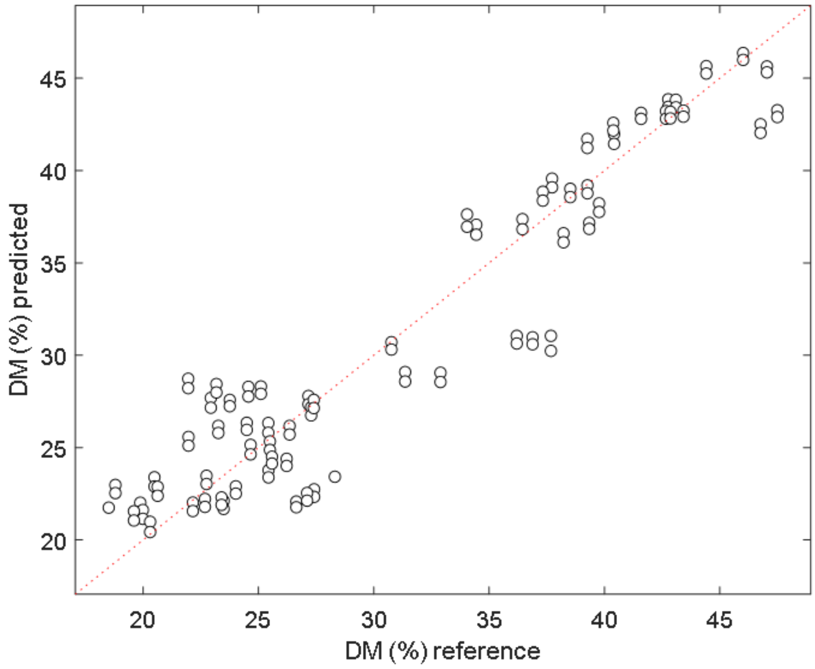


Figure 8 : reference and predicted DM values for calibration set using 3 factors.

2.4 Visualisation of DM content during the storage

Once the calibration established for DM content of fresh yam slices, it is possible to determine DM content for each individual pixel of the original images. Therefore, HSI allows visualisation of the composition of a target compound (here, DM) within slice pixel by pixel. Fig.9 shows a dry matter distribution map of yam slices generated at 0, 30, 45 and 60 days of storage. The dry matter level (red) increase with the storage days. At D00, the DM is high concentrated in the edges and increase with storage days from the edges to the centre of the slices.

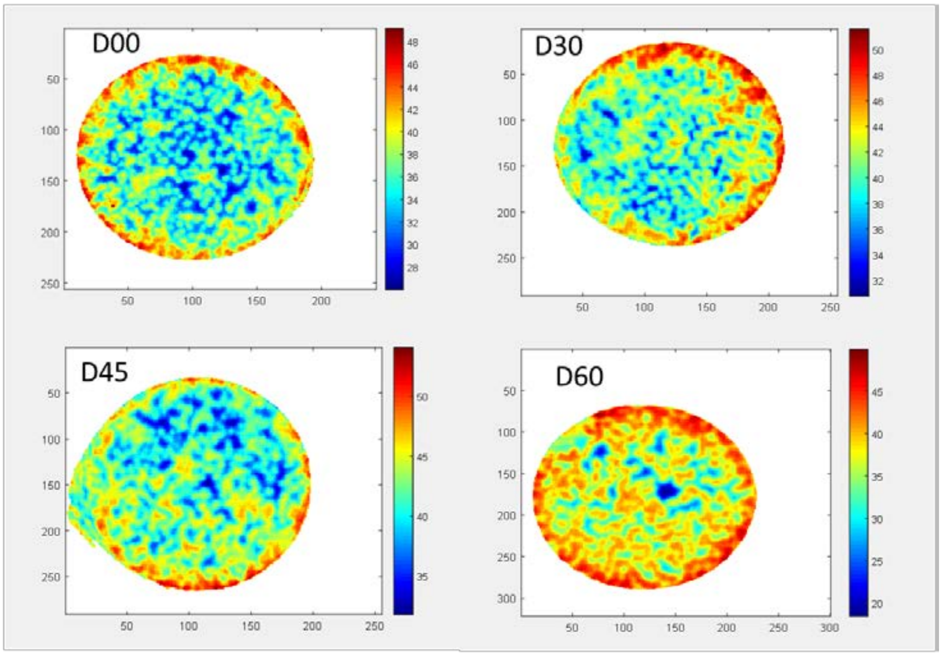


Figure 9 : Dry matter distribution map of yam slices at 0, 30, 45, 60 days of storage generated by using PLS model.

2.5 Estimation of spectral repeatability

The repeatability of HSI measurement was estimated by the calculation of the root mean square error (RMS) between two average spectra of two images acquired on the same slice. This operation was applied on six slices on the same tuber. The RMS was calculated using the equation below:

$$RMS = \sqrt{\frac{\sum_j^p (Xaj - Xbj)}{p}}$$

Where: Xaj, Xbj average of absorbance of wavelength j for the two repliacats a and b, p number of wavelengths (j variate from 1 to p).Therefore, The average and standard deviation (SD) of all RMS were also calculated. A spectral measurement is repeatable if the RMS value is within the interval mean RMS (RMSM) ±3SD. The fig.10 shows the RMS values obtained from samples of the same tuber. The RMS obtained between two images taken on the same slice are between mean RMs + 3SD and mean RMs -3SD compared to the average RMS. In conclusion, only one image per slice is sufficient to catch the spectral variability of the slice.

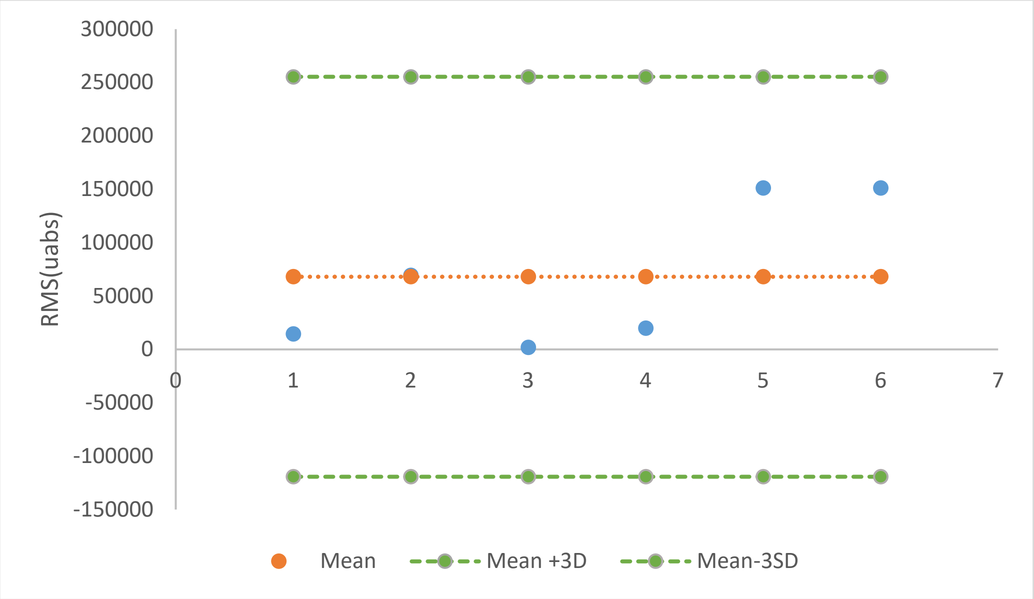


Figure 10 : RMS values between two replicates acquired on the same slice for six slices from one yam tuber

3 CONCLUSION

We report, for the first time, quantitative prediction of DM based on HSI for fresh yam slices during the storage. Our research demonstrated that the estimation or prediction of DM content is possible by NIR-HSI on fresh yam slices. The method developed could be applied at a single pixel level with the potential to provide information on the distribution of composition within slice.

References

(1) Otegbayo, B. O.; Asiedu, R.; Bokanga, M. EFFECTS OF STORAGE ON THE CHEMICAL COMPOSITION AND FOOD QUALITY OF YAM: EFFECT OF STORAGE ON THE FOOD QUALITY OF YAM TUBERS. Journal of Food Processing and Preservation 2012, 36 (5), 438–445. https://doi.org/10.1111/j.1745-4549.2011.00600.x.

(2) Velasco, L.; Grüneberg, W. J. Analysis of Dry Matter and Protein Contents in Fresh Yam Bean Tubers by Near-Infrared Reflectance Spectroscopy. Communications in Soil Science and Plant Analysis 1999, 30 (13–14), 1797–1805.

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/