Calibrattion Report

NIRS Calibration for Dry Matter & Texture on Fresh Matooke at IITA & NaCRRI Uganda

High-Throughput Phenotyping Protocols (HTPP), WP3

Namulonge, 04/12/2021

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

Jennifer ACAYO, International Institute of Tropical Agriculture (IITA), Ibadan Nigeria

Ephraim NUWAMANYA, National Crops Resources Research Institute (NaCRRI), Kampala, Uganda

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, Jennifer ACAYO, Ephraim NUWAMANYA, Fabrice DAVRIEUX, (2022). *NIRS Calibration for Dry Matter & Texture on fresh Matooke at IITA & NaCRRI Uganda. High-Throughput Phenotyping Protocols (HTPP), WP3.* Montpellier, France: RTBfoods Calibration Report, 13 p. http://doi.org/10.18167/agritrop/00714

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:	
Final validation by:	
Fabrice DAVRIEUX (CIRAD)	02/02/2022

CONTENTS

Table of Contents

ABSTRACT

This report concern two data bases related to fresh Matooke: 1) physico-chemicals data and 2) NIRS spectral data. The data were collected on fresh Matooke in NaCRRI and IITA (Uganda). 45 genotypes were analysed, for each genotype one to three bunches were used and each genotype is replicated 3 times (3 bunches are evaluated for each genotype). For each sample, 4 fresh fingers are scanned using NIRS. After scanning, each of these fingers is cut in half. One half of the same 4 fingers are boiled and mashed, then scanned twice for NIRS on mashed sample. The other half is peeled, chopped and used for DM determination. The dried sample (after DM) is then ground into powder/flour, from which we obtain spectra and RVA data.

Partial Least Square regression (PLSR) is used to associate spectral data and physico-chemical parameters. The direct calibrations of physico-chemical parameters (DM, RVA) using different pretreatments (SNV, SNVD, first or second derivative…) failed.

Keywords: Matooke, physico-chemicals data, NIRS, spectral data, dry matter.

1 FRESH MATOOKE

1.1 NIRS spectra

303 NIRS spectra were acquired on fresh Matooke using ASD QualitySpec spectrometer (350-2500 nm). The table below reports the number of spectra acquired for each genotype by fingers and the total number acquired per genotype. From this table, we can see that for the genotypes for 7 genotypes (222K, 660K, N14, N19, N7, NFW) 3 bunches were used. For 17 genotypes 2 bunches were used (254S4, 295S4, 401K, ENT, MBW, N12, N13, N15, N18, N2, N4, N5, N6, N8, NAN, NFU, NKS). One bunch was used for the rest of the genotypes.

Genotype	F1	F2	F3	F4	Bunche
1201K	1	1	1	1	1
1438K	1	1	1	1	1
222K	3	3	3	3	3
249S13	1	1	1	1	1
254S4	2	2	2	2	2
256S11	1	1	1	1	1
277S4	1	1	1	1	4
278S9	1	1	1	1	1
280S23	1	1	1	1	1
295S4	2	2	2	2	2
365K	1	1	1	1	1
401K	1	2	2	2	2
660K	3	3	3	3	3
917K	1	1	1	1	1
ENT	2	2	2	2	2
ENY	1	1	1	1	1
ENZ	1	1	1	1	1
KAZ	1	1	1	1	1
MBW	2	2	2	2	2
N1	1	1	1	1	1
N12	2	2	2	2	2
N13	2	2	2	2	2
N14	3	3	3	3	3
N15	2	2	2	2	2
N18	2	2	2	2	2
N19	3	3	3	3	3
N2	2	2	2	2	2
N20	1	1	1	1	1
N21	1	1	1	1	1
N22	1	1	1	1	1
N23	1	1	1	1	1
N26	1	1	1	1	1
N3	1	1	1	1	1
N4	1	2	2	2	2
N5	2	2	2	2	2

N6	2	2	2	2	2
N7	3	3	3	3	3
N8	3	2	2	2	2
N9	1	1	1	1	1
NAM	1	1	1	1	1
NAN	2	2	2	2	2
NFU	2	2	2	2	2
NKS	2	2	2	2	2
NKW	3	3	3	3	3
TER	3	3	3	3	3
Total	**75**	**76**	**76**	**76**	**79**

1.2 Dry mater content

A total of 55 DM measurements are performed on fresh Matooke by oven drying method. For each bunch, one measurement is performed. the table below reports the statistics descriptive for DM content. The range of variability of DM is 10,06 %.

Variable	N	Min	Max	Mean	Standard deviation
DM (%)	55	19,25	29,31	24,82	2,26

1.3 RVA measurements

RVA measurements are performed on flour samples of each bunch of each genotype with two replications making a total of 88 values of RVA. Four parameters (Peak 1, Final viscosity, peak time and pasting temperature) are measured. The table below contains descriptive statistics for the fourth parameters.

Variable	N	Min	Max	Mean	Standard deviation
Peak 1	88	361,00	7372,00	3921,58	1606,69
Final Visc	88	663,00	5458,00	3797,66	1066,14
Peak Time (min)	88	4,13	7,00	5,18	0,59
Pasting Temp	88	51,50	83,20	79,16	4,45

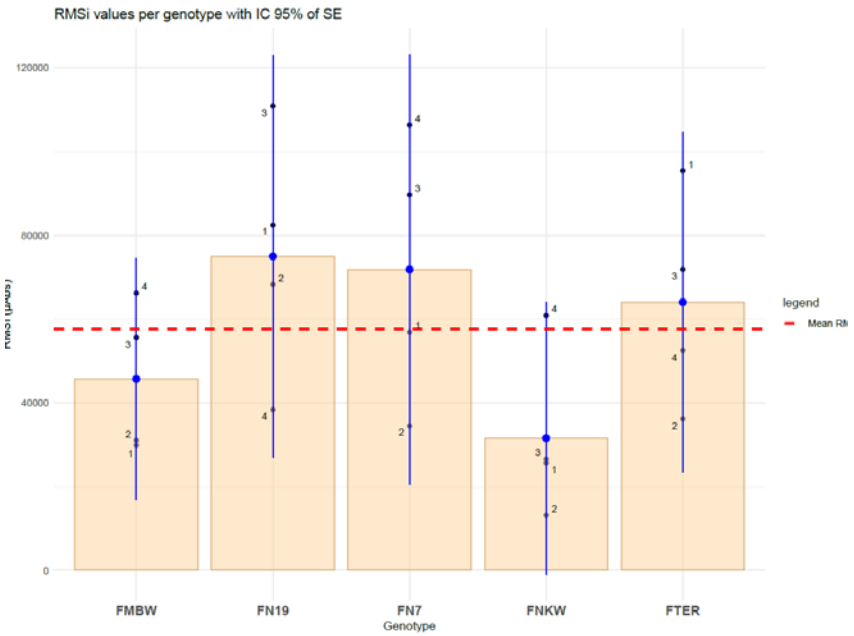
2 RESULTS

2.1 Data exploration

2.1.1 Test of representativeness for NIRS measurement

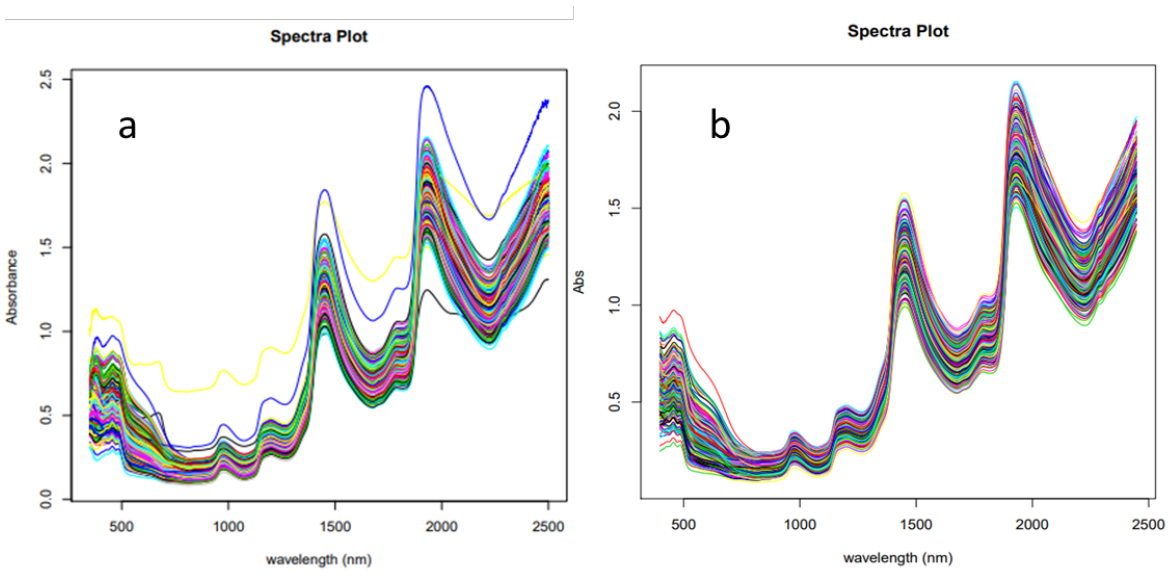
The test of representativeness is important for the determination of the number of fingers per genotype to be scanned in order to capture the sample variability. For this test, the root mean square (RMS) calculation (dispersion of each spectra from the mean spectra) are applied on spectra acquired from 5 genotypes of Matooke. For each genotype, one bunch of four fingers are used, and for each finger, one spectrum is acquired. For each genotype we have 4 spectra, making a total of 20 spectra. The figure below represents the RMS value of all genotype (histogram), mean RMS (dashed red line) and RMS value of each finger (blue vertical dots). From this figure we see two group of genotypes. The first one FMBW and FNKW with low mean RMS values and low dispersion

(spectral variability). The second group with FN19, FN7, FTER with high values of RMS and high spectral variability between fingers of the same genotype. herefore 4 fingers per bunch have to be scanned in order to capture all the variability.



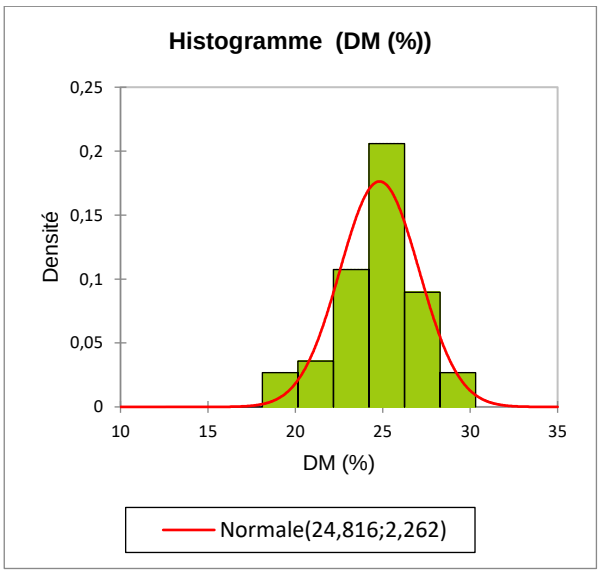
2.1.2 Spectra

The figure below shows the NIRS spectra acquired on fresh matooke, (a) represents raw spectra (350-2500 nm) before removing outliers and (b) represents clean spectra after removing outliers. From the raw spectra, we see 3 spectral outliers (N14B2F1, N9B1F4 and ENTB2F2) to be eliminated and instrumental noise in the extreme of the spectra, additive effect between spectra (difference of intensity of absorbance) and deviation of base line.



2.1.3 Dry matter content

The figure below represents histogram of the distribution of 55 values of DM. the DM content ranges from 19, 25 % to 29,31%with a mean value of 24,82 %. The distribution of the values follows a normal law with no outlier (extreme value).

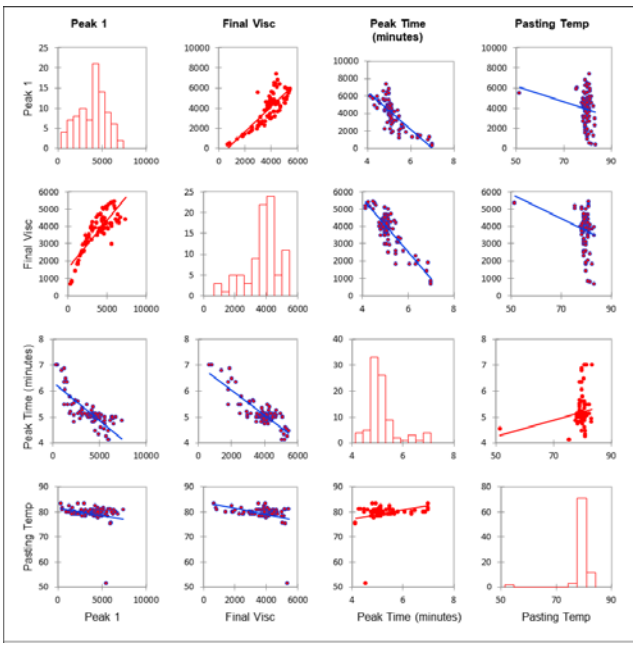


2.1.4 RVA parameters

From the matrix of Pearson correlation and the graphics below we see that there a high correlation between the parameters peak1, Final Visc and peak time (from 0,627 to 0,746). However, there is no correlation between pasting temp and the others parameters. This is due to the low range of variability of this parameter after removing the outlier at 50 °C. The distribution of the parameter peak 1 and peak time follows normal distribution while the distribution of pasting temp is not normal due to the outlier at 50 °C.

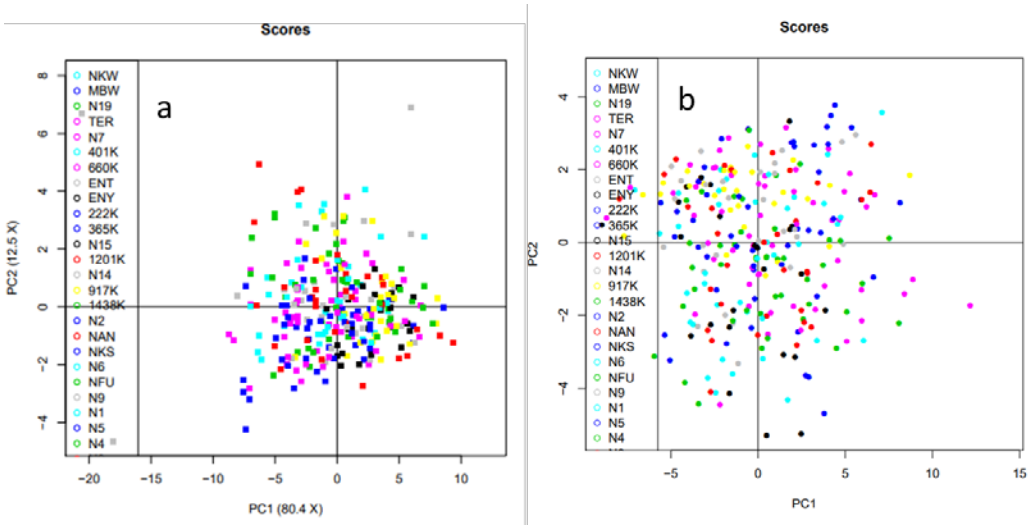
Table1 : Pearson correlation matrix

Paramaters	Peak 1	Final Visc	Peak Time (min)	Pasting Temp
Peak 1	**1**	0,698	0,627	0,045
Final Visc	0,698	**1**	0,746	0,090
Peak Time (min)	0,627	0,746	**1**	0,057
Pasting Temp	0,045	0,090	0,057	**1**



2.1.5 Principal component analyses (PCA)

In order to detect outliers as well as grouping according to genotype PCA is applied to raw spectra (303) and spectra after cleaning and pre-treatments (snvd on segment 426-2401 nm). In the figure below we see score plots PC1-PC1 of PCA applied on raw spectra (a) (with outliers) and after cleaning (b). The samples are colored according to the genotype. from the raw spectra scores plot, we see 3 important outliers which can be removed from the data. When this samples are eliminated we see in plot (b) that the distribution of the samples is more scattered with no discrimination between genotypes.

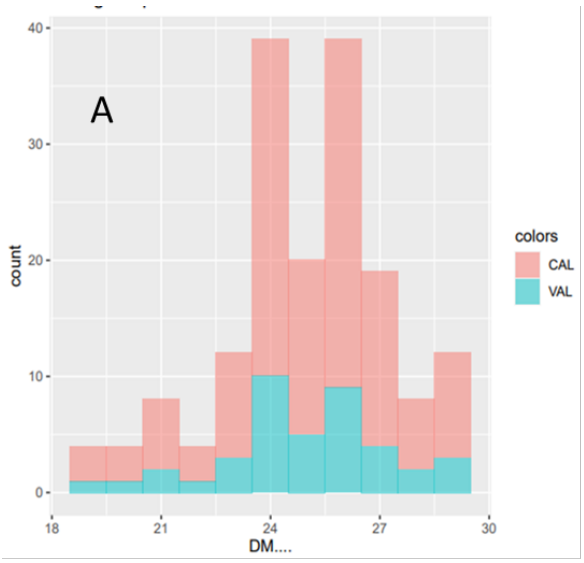


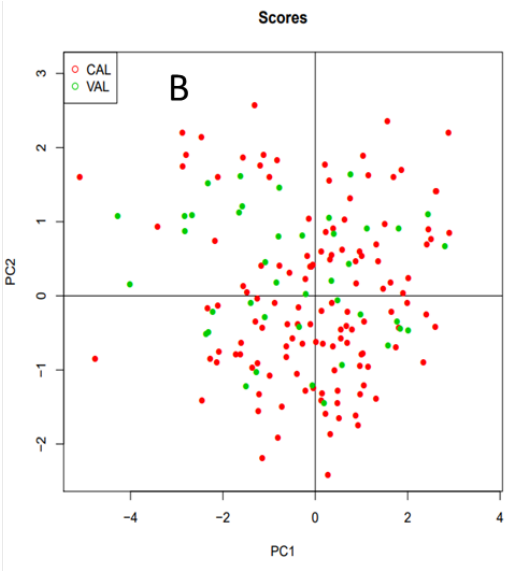
2.2 Prediction of DM

For the development of DM calibration 169 spectra and DM values are used. The DM value measured on each bunch is assigned to spectra measured on this bunch. For example, we have only one value of DM for each bunch while we have 4 spectra (four fingers) for on bunch. To have the same number of DM and spectra for each spectra, we give the same value of DM to the fourth spectra. Then we divided our data set into two subsets: calibration data set which contains 128 samples from fingers 1,2 and 3 and validation dataset which contains 41 spectra from fingers 4. The table below reports the descriptive statistics of calibration and validation datasets.

	n	mean	sd	min	max	range
Xcal	128	25,053	2,285	19,246	29,31	10,069
Xval (F4)	41	24,961	2,320	19,246	29,31	10,064

The figure below shows in A the distribution of DM values for calibration (red) and validation (blue). In B we see the projection of validation spectra on calibration spectra. From these plots we see that the range of variability of validation data (DM, spectra) is similar to the range of variability of calibration data.



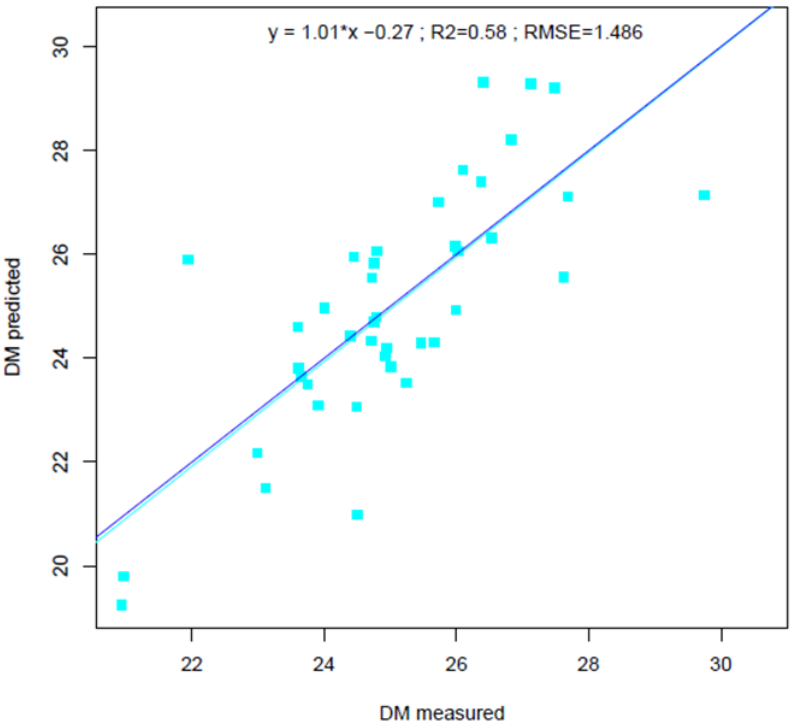


2.2.1 NIRS Calibration

Partial least square regression (PLSR) is applied for the development of the model of calibration of DM. The statistics parameters (table below) show that the calibration for the prediction of DM of fresh matooke presents low performances with RMSEP =1,49 % and R^2p=0,58.

nLV	RMSEC (%)	RMSECV (%)	R^2cv (%)	RMSEP (%)	R^2P	Math
10	1,53	1,67	0,59	1,49	0,58	Snvd +450-2500 nm

The scatter plot of predicted DM using NIRS model versus the measured DM in laboratory (below) confirms the weak performances of the model. model's error expressed as RMSEP should be compared to the laboratory error of DM measurement for a full evaluation of the nirs model

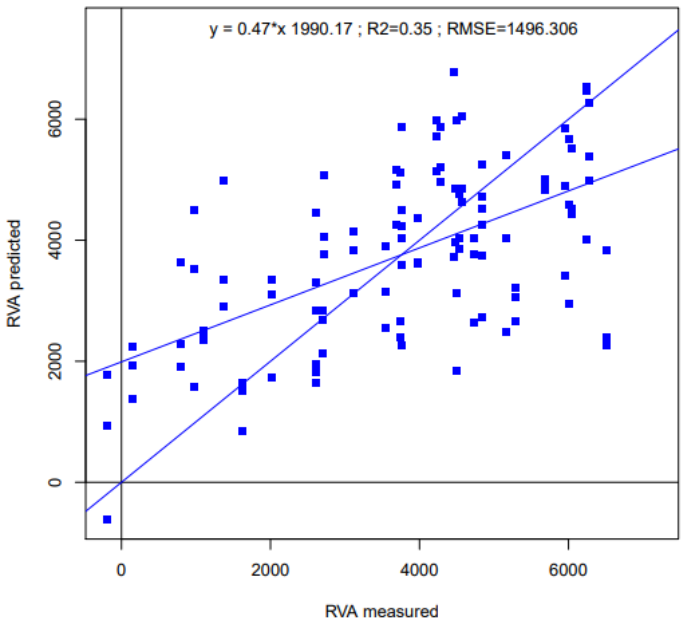


2.3 Prediction of RVA (peak 1)

For the development of the model for RVA, the parameter peak 1 is taken as example. In addition, this parameter is highly correlated to the others parameters. For the calibration, 112 spectra and peak 1 values are used. The statistics parameters (table below) show that the calibration for the prediction of RVA (peak 1) of fresh Matooke has weak performances with RMSECV =1496,31 and R^2cv=0,60.

nLV	RMSEC	RMSECV	R^2cv (%)	Math
13	1211,17	1496,31	0,60	Snvd + 450-2500 nm

The scatter plot of predicted RVA using NIRS versus the measured RVA in laboratory (below) highlights the poor performances of the calibration. This is probably due to that the linear approach is not relevant for the quantification of physical proprieties. There is no linear correlation between texture parameters and NIRS and NIRS is not be able to predict texture parameters that are not linked to the biochemical composition of matooke.



2.4 Conclusion

The calibration models for DM and RVA (peak 1) quantification present low performances.

The solution to improve the performances of these models could be:

- For dry matter, compare the error of the prediction of NIRS model and error of laboratory in order to verify if the error of NIRS model does not come from the laboratory measurements.
-
- For RVA parameters, try non-linear methods (Knnr, SVMR, LWPLSR).
- Increase the range of variability of DM and RVA by adding new samples from other genotypes, year of harvest…etc.
- Check the measurement workflow and accuracy of reference methods

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/