

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



Institut National Polytechnique
Félix HOUPHOUËT-BOIGNY

Numéro d'ordre : 140 /2023

EDP
École Doctorale Polytechnique

UMRI 78

Laboratoire d'Instrumentation, Image et Spectroscopie (L2IS)

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE

FÉLIX HOUPHOUËT-BOIGNY DE YAMOOUSSOUKRO

Spécialité : Instrumentation optique, Spectroscopie par télédétection

Présentée par

Mlle ASSOI Koko Éliane Nadine

Développement de méthodes de caractérisation de la qualité des fruits tropicaux à l'aide de la spectroscopie par télédétection

Soutenue le 03 Février 2023 devant le jury composé de :

M. SORO Yaya	Professeur Titulaire	Président
	Institut National Polytechnique Félix Houphouët- BOIGNY	
M. ZOUEU Thouakessèh Jérémie	Professeur Titulaire	Directeur
	Institut National Polytechnique Félix Houphouët- BOIGNY	
Mme ZORO Emma Georgina	Maître de conférences	Rapporteur
	Université Félix Houphouët- BOIGNY	
M. MEGNASSAN T. Eugène	Professeur Titulaire	Rapporteur
	Université Nangui ABROGOUA	
M. SORO Pénétiligué Adama	Maître de conférences	Examineur
	Université Félix Houphouët- BOIGNY	

Résumé

Les travaux réalisés durant cette thèse ont porté sur la mesure de la qualité des fruits tropicaux à partir de nouvelles techniques optiques, non invasives et à distance. En effet, la qualité des fruits constitue un facteur important pour le domaine agroalimentaire. Cependant, les méthodes de mesures habituelles de qualité sont parfois subjectives, destructives et lentes, ce qui ne permet pas dans certains cas une analyse rapide, précise, fiable et effective de la qualité des fruits.

La spectroscopie optique permet de caractériser les fruits sans les détruire à partir de leur signature spectrale ou de données spectrales collectées sur ces fruits. Dans ce travail, nous avons utilisé un télescope couplé à un spectromètre pour acquérir les spectres de réflectance et de fluorescence de mangues, de tomates, de pommes de cajou et de papayes. Nous avons développé des modèles mathématiques de prédiction des paramètres de qualité des fruits à partir des données spectrales afin de prédire la qualité de ceux-ci. Nous avons mesuré les paramètres traditionnels de qualité tels que le pH, le taux de sucre, l'eau, l'acidité et la matière sèche. Nous avons cherché à établir des relations entre ces paramètres chimiques et les spectres de fluorescence et de réflectance acquis par télédétection.

Ce travail a contribué au développement de modèles de prédiction des paramètres chimiques à partir des données spectrales obtenues à distance. Les données spectrales contenant du bruit ont nécessité l'application de méthodes de prétraitement avant l'utilisation des méthodes d'analyses statistiques à savoir la méthode de régression PLS et la méthode de réseau de neurones artificiels pour l'établissement des modèles. Les résultats obtenus ont montré l'intérêt de la spectroscopie optique de télédétection comme un moyen d'analyse rapide et efficace de la qualité des fruits.

Mots-clés : spectroscopie optique, télédétection, réflectance et fluorescence optique, fruits tropicaux, régression PLS, réseaux de neurones artificiels.

Abstract

The work carried out during this thesis focused on measuring the quality of tropical fruits using new optical techniques. Indeed, the quality of the fruits is an important factor for the agri-food sector. However, the usual methods for measuring quality are sometimes subjective, destructive and slow, which often does not allow a better analysis of fruit quality. Optical spectroscopy makes it possible to characterize fruits without destroying them based on their spectral signature or on spectral data collected on these fruits. In this work, we used the telescope coupled with a spectrometer to acquire fruit spectra. We have developed mathematical models for predicting fruit quality parameters from spectral data in order to predict fruit quality. The main contribution is the development of models for the prediction of parameters from remotely obtained spectral data. Spectral data containing noise required the application of pre-processing methods prior to the use of statistical analysis methods, namely the PLS regression method and the artificial neural network method for modelling. The results obtained showed the interest of remote sensing spectroscopy as a mean of rapid and efficient analysis of fruit quality.

Keywords: optical spectroscopy, remote sensing, optical reflectance and fluorescence, tropical fruits, PLS regression, artificial neuron networks.

Dédicaces

À l'Éternel, mon Dieu qui a conduit ce travail.

Que toute la Gloire te revienne.

À la mémoire de mon père, qui a œuvré pour ma réussite, aucune dédicace ne saurait exprimer la gratitude, l'amour, l'estime et le respect que j'ai toujours eu pour toi. Puisse Dieu faire en sorte que ce travail porte son fruit pour sa plus grande gloire.

À ma mère, qui de par son amour, son soutien et ses précieux conseils, pour toute son assistance, reçoit à travers ce travail l'expression de mon immuable gratitude. Que Dieu t'accorde une longue vie.

À mon fils TIEU Assena Paul Elijha Yoram

À ma petite-sœur ASSOI Boshoma Edith

À mes frères et sœurs

À tous ceux ou celles qui me sont chers.

Remerciements

Cette thèse a été réalisée à l'Institut National Polytechnique Félix Houphouët BOIGNY de Yamoussoukro (INP-HB) au sein du Laboratoire d'Instrumentation Image et Spectroscopie (L2IS). Ce laboratoire est rattaché à l'Unité Mixte de Recherche et d'Innovation en Génie Electrique et Electronique (UMRI-GEE).

Je tiens tout d'abord à remercier le Dieu Tout Puissant qui m'a donnée la volonté, le courage et la force d'achever cette thèse. Je lui rends toute la gloire.

J'exprime ma plus grande gratitude à mon directeur de thèse, Professeur ZOUEU Thouakessah Jérémie, sans qui jamais cette thèse n'aurait vu le jour. Je lui suis particulièrement reconnaissante pour sa disponibilité et sa grande patience. Professeur, je loue votre ardeur au travail et votre charité. Vos qualités m'ont toujours mise en admiration et je vous remercie de m'avoir aidée à vivre la discipline de la physique comme une expérience palpable et unique. Je voudrais également vous remercier pour la patience dont vous avez fait preuve à mon égard. Je tiens également à vous remercier pour votre apport scientifique et surtout pour le sens aigu de l'encadrement dont j'ai bénéficié. Puisse Dieu vous accompagner, vous éclairer dans votre mission et vous faire accomplir des exploits encore plus grands pour sa plus grande Gloire.

Je remercie Professeur SORO Yaya de l'Institut National Polytechnique Félix Houphouët - BOIGNY, qui, malgré ses multiples occupations, a bien voulu présider le jury de cette thèse ; M. MEGNASSAN Eugène, Professeur titulaire à l'Université Nangui ABROGOUA, Mme ZORO Emma Georgina, Maître de conférences à l'Université Félix Houphouët- BOIGNY, qui ont consacré beaucoup de temps à lire et à rédiger avec intérêt les rapports de thèse, je leur dis merci. Je remercie également M. SORO Pénétiligué Adama, Maître de conférences à l'Université Félix Houphouët-

BOIGNY, qui a accepté de prendre part à ce jury en tant qu'examineur de ce travail de thèse.

Mes remerciements vont aussi à l'endroit du Professeur MEGNASSAN Eugène, directeur du Laboratoire de Physique Fondamentale et Appliquée (LPFA) de l'Université Nangui ABROGOUA. Cher Professeur, vous nous avez séduits par votre simplicité et la qualité de vos enseignements. Ces enseignements nous ont aidés à considérer avec beaucoup d'intérêt les sciences de la matière. Vous nous avez donnés l'envie d'en apprendre toujours davantage.

Je tiens à dire un grand merci à Madame BOSSON Jocelyne, Maître de Conférences, pour son assistance et ses encouragements. Ses remarques et ses conseils avisés m'ont vraiment prêté main forte et m'ont amenée à prendre conscience de l'importance de ce travail.

Je remercie Messieurs KONAN Kouakou, KOUAKOU Abaka et Maîtres de Conférences et Monsieur HABA Cissé, Professeur titulaire.

Je remercie Monsieur SORO Doudjo, Maître de Conférences, pour son assistance, sa disponibilité et ses remarques avisées lors des mesures au laboratoire LAPISEN. Vos conseils, votre aide, votre bonne humeur au quotidien m'ont permise de toujours m'accrocher quelques soient les difficultés.

Je voudrais remercier tous les enseignants de l'Ecole Doctorale, de l'UMRI 78 pour leurs conseils, leur aide et leur disponibilité.

Mes remerciements vont également à l'endroit de Docteur BAGUI Kossan Olivier pour sa contribution technique à toutes les étapes de cette thèse ; merci pour sa présence au quotidien et pour tous les moments de rappels et d'échange. Docteur, merci pour les moments partagés ensembles, pour votre écoute, pour vos conseils et votre bienveillance infaillible qui m'ont donnée la force et la détermination de continuer. Que la faveur de Dieu soit votre partage tout le long de votre carrière.

Cette thèse a été un parcours riche en rencontres humaines. Mes remerciements vont à l'endroit de toute l'équipe du Laboratoire d'Instrumentation Image et Spectroscopie de l'INP-HB, avec qui j'ai passé d'agréables moments d'échange. Merci aux Docteurs ABOU Akoun, TOKOU Zan, WOGNIN Essé Léon, AGNERO Marcel, DIBY wilfried, NIANGORAN Urbain, YALE Pavel, REGNIMA Guy-Oscar, KONIN Edoukoua, YEBOUET Marie Florence, DOUMIN Solange, KOUAKOU Benoît, KOFFI N'guessan Thomas, KOFFI Yao, KOSSONOU Taky Alvarez, à Messieurs KOFFI Kouamé Patrice, AKA Manoua, GBOGBO Yatana Adolphe, DION Sylvère Bienvenu, GNETTO Marius et à tous les autres membres du laboratoire. Votre contribution et soutien ont été très appréciables pour ces travaux de thèse. Mes pensées s'adressent encore à chacun d'entre eux pour leur sympathie, leurs discussions enrichissantes et leur aide au cours de mes travaux.

Au-delà des membres du L2IS, je tiens à remercier les membres et les stagiaires du laboratoire LAPISEN. Votre collaboration au cours de cette thèse a été riche en connaissance et m'a permis de m'approprier des techniques de mesures chimiques.

Merci à tous les collègues et membres des administrations des lycées BAD de Tiénigboué et Mixte 1 de Yamoussoukro pour les conseils et les moments très enrichissants sur le plan professionnel. Je voudrais remercier particulièrement Messieurs KONE Adama et KONE Issiaka, Proviseurs respectifs du Lycée BAD de Tiénigboué et du Lycée Mixte 1 de Yamoussoukro pour leur compréhension et leurs encouragements durant toutes ces années. Je voudrais également remercier tous mes collègues enseignants de Physique-Chimie très particulièrement Messieurs KOUADIO Nicaise et KOUASSI Apaou Prospère pour leur aide, leur collaboration étroite et leur compréhension.

Je remercie Madame GOLY Amenan Armelle, enseignante d'Histoire et Géographie en service au collège Moderne de Port-Bouët pour sa compréhension et sa collaboration étroite. Mes remerciements vont également à l'endroit de Mademoiselle YAO Amani Kra Marcelle, enseignante de Philosophie au lycée BAD de Tienigboué

pour sa compagnie et ses conseils qui ont été d'une grande aide. Mes vifs remerciements vont également à l'endroit de Messieurs N'GORAN Amani, KIMOU Yapou Romaric et CAMARA Yékoman Felix pour leur soutien moral et financier.

Pour terminer, je voudrais dire infiniment merci à mes parents, à mes frères, mes sœurs et toute la famille entière, qui ont toujours su m'apporter leur soutien, leurs encouragements, leur patience et leur affection. Merci pour l'amour et la confiance que vous m'avez toujours accordés.

Table des matières

Résumé.....	I
Abstract	II
Dédicaces.....	III
Remerciements	IV
Table des matières	VIII
Liste des abréviations et sigles	XII
Liste des figures.....	XIV
Liste des tableaux.....	XVI
INTRODUCTION	1
CHAPITRE I : SPECTROSCOPIE OPTIQUE PAR TELEDETECTION.....	7
I.1. Spectroscopie optique.....	8
I.1.1. Spectroscopie de réflectance optique.....	9
I.1.2. Spectroscopie d'absorption.....	10
I.1.3. Spectroscopie de diffusion	11
I.1.4. Spectroscopie de fluorescence.....	13
I.2. Télédétection.....	13
I.2.1. Définition	13
I.2.2. Applications de la télédétection.....	15
I.3. Propriétés optiques des fruits	16
I.3.1. Composition des fruits	16
I.3.2. Propriétés optiques des fruits	17
I.3.2.1. Pigments	17
I.3.2.2. Eau	18
I.3.2.3. Sucres	19
I.4. Paramètres de qualité des fruits	19
I.4.1. Paramètres physiques et chimiques de qualité	20
I.4.1.1. Paramètres physiques	20
I.4.1.2. Paramètres chimiques.....	21
I.4. 2. Paramètres optiques	21
I.4.2.1. Coefficient d'absorption.....	21
I.4.2.2. Coefficient de diffusion	22
I.4.3. Techniques de mesure de qualité	23
I.4.3.1. Techniques destructives.....	23

I.4.3.2. Techniques non destructives.....	24
I.5. Méthodes d'analyses chimiométriques ou d'analyses statistiques multivariées	26
I.5.1. Méthodes statistiques de prétraitement	26
I.5.1.1. Lissage.....	26
I.5.1.2. Dérivée	27
I.5.1.3. Normalisation	27
I.5.2. Méthodes statistiques de traitement	27
I.5.2.1. Analyse en composantes principales	28
I.5.2.2. Régression PLS.....	29
I.5.2.3. Réseaux de neurones artificiels.....	29
CHAPITRE II : MATÉRIEL ET MÉTHODES	32
Introduction.....	33
II.1. Matériel	33
II.1.1. Echantillonnage	33
II.1.2. Préparation des échantillons.....	35
II.1.2.1. Préparation pour les analyses chimiques.....	35
II.1.2.2. Préparation pour les mesures spectroscopiques	35
II.1.3. Dispositif de mesures spectroscopiques	35
II.2. Techniques de mesures.....	38
II.2.1. Techniques de mesures chimiques	38
II.2.1.1. Mesure de la teneur en sucre	38
II.2.1.2. Mesure de l'acidité titrable	39
II.2.1.3. Mesure du Potentiel Hydrogène	40
II. 2.1.4. Mesure de la teneur en eau	40
II.2.1.5. Mesure de la Vitamine C	42
II.2.2. Techniques de mesures optiques.....	42
II.3. Analyses des données	44
II.3.1. Prétraitement des spectres	44
II.3.1.1. Normalisation par les valeurs minimales et maximales.....	45
II.3.1.2. Filtrage gaussien.....	45
II.3.1.3. Dérivation	46
II.3.1.4. Rééchantillonnage par Bootstrap	46
II.3.2. Traitement des données	47
II.3.2.1. Optimisation des données par sélection de variables	47

II. 3.2.1.1. Méthode Monte Carlo	47
II.3.2.1.2. Méthode IRIV de sélection de variables optimales basée sur l'échantillonnage par Matrice Binaire	49
II.3.2.2. Partitionnement des données par la méthode Kennard-Stone	52
II.3.2.3. Modélisation par analyse multivariée	53
II.3.2.3.1. Régression PLS.....	53
II.3.2.3.1.1. Principe.....	53
II.3.2.3.1.2. Algorithme de la régression PLS.....	54
II.3.2.3.2. Réseau de neurones artificiels	55
II.3.3. Evaluation de la performance d'un modèle	58
CHAPITRE III : RÉSULTATS ET DISCUSSION	60
Introduction.....	61
III.1. Estimation de la qualité des mangues par réflectance optique.....	62
III.1.1. Collecte des mangues, mesure et traitement des spectres de réflectances	62
III.1.2. Résultats et discussion	63
III.1.2.1. Photo des échantillons	63
III.1.2.2. Spectres de réflectance optique des échantillons de mangues.....	64
III.1.2.3. Suppression des valeurs aberrantes par Monte Carlo	67
III.1.2.4. Développement et validation des modèles de prédiction	69
III.1.2.4.1. Description générale des modèles	69
III.1.2.4.2. Modèle de prédiction du sucre	70
III.1.2.4.3. Modèle de la teneur en eau	72
III.1.2.4.4. Modèles du Potentiel Hydrogène	74
III.2. Estimation de la qualité des tomates par réflectance optique	76
III.2.1. Méthodologie adoptée.....	77
III.2.2. Résultats et discussion	77
III.2.2.1. Spectres de réflectance	77
III.2.2.2. Modèles de prédiction par régression PLS.....	79
III.2.2.2.1. Présentation des résultats.....	79
III.2.2.2.2. Discussion	84
III.3. Estimation de la qualité des pommes de cajou par réflectance optique.....	85
III.3.1. Méthodologie adoptée.....	85
III.3.2. Spectres de réflectance	88
III.3.2.1. Résultats	88

III.3.2.2. Discussion	89
III. 3.3.1. Modélisation des paramètres de qualité des pommes de cajou	91
III.3.3.2. Discussion	96
III.4. Suivi de la maturité des papayes par la fluorescence	99
III.4.1. Méthodologie adoptée.....	99
III. 4.2. Acquisitions des spectres	99
III. 4.2.1. Mesure de réflectance par télédétection passive	99
III. 4. 2.2. Méthode d'acquisition des spectres de fluorescence	101
III.4. 3. Méthode d'analyse des spectres de fluorescence	102
III. 4.4. Résultats et discussion	104
III. 4.5. Modélisation des paramètres physiques-chimiques à partir du réseau de neurones artificiels	107
III. 4.5.1. Méthode d'analyse prédictive.....	107
III. 4.5.2. Résultats et discussion	108
CONCLUSION	113
RÉFÉRENCES BIBLIOGRAPHIQUES	115

Liste des abréviations et sigles

ACP	: Analyses en Composantes Principales
AFNOR	: Association Française de Normalisation
AOAC	: Association of Official Analytical Chemists
AT	: Acidité Titrable
CARS	: Competitive Adaptive Reweighted Sampling
CCD	: Coupled Charged Devices
Chl	: Chlorophylle
ESS	: Extrait Sec Soluble
FT-NIR	: Fourier Transform- Near Infrared
GA	: Genetic Algorithm
IRIV	: Iteratively Retaining Informative Variables
IRMN	: Imagerie par Résonance Magnétique Nucléaire
IPLS	: Interval Partial Least Squares
KS	: Kennard-Stone
LIDAR	: Light Detecting And Ranging
MC	: Monte Carlo
MCOD	: Monte Carlo Outlier Detection
MC-UVE	: Monte Carlo Uniformative Variable Elimination
MLR	: MultiLinear Regression
MSC	: Multiplicative Scatter Correction
MST	: Matière Sèche Totale
MWPLS	: Moving-Window Partial Least Squares

NIPALS	: Nonlinear Iterative Partial Least Squares
pH	: Potentiel d'Hydrogène
PCR	: Principal Component Regression
PLS	: Partial Least Square
PLSR	: Partial Least Square Regression
RADAR	: Radio Detecting And Ranging
RCP	: Régression sur les Composantes Principales
RMSE:	Root Mean Square Error
RMSEC	: Root Mean Square Error of Calibration
RMSECV	: Root Mean Square Error of Cross-Validation
RMSEP	: Root Mean Square Error of Prediction
RNA	: Réseaux de Neurones Artificiels
RPD	: Ratio Performance Deviation
SNV	: Standard Normal Variate
SONAR	: Sound Navigation And Ranging
TDNN	: Time Dealy Neural Network
Vis/NIR	: Visible and Near Infrared
TH	:Taux d'Humidité
USB	: Universal Serial Bus

Liste des figures

Figure 1: Représentation des phénomènes de réflexion spéculaire (à gauche) et de réflexion diffuse (à droite).....	9
Figure 2: Schéma illustrant la télédétection passive (Gauche) et la télédétection active (Droite).....	14
Figure 3 : Spectre d'absorption des pigments [59]	18
Figure 4 : Spectre d'absorption de l'eau [62].....	19
Figure 5 : Schéma d'un neurone formel.....	30
Figure 6 : Photographie des échantillons de fruits utilisés. (a) Mangues camerounaises ; (b) Tomates ; (c) Pommes de cajou ; (d) Papayes.....	33
Figure 7: Vue d'ensemble du dispositif expérimental.....	36
Figure 8: Image annotée du spectromètre.....	37
Figure 9: Organigramme de l'algorithme MCOV [96].	48
Figure 10: Echantillonnage par matrice binaire. (A), (B) Comparaison des performances de l'inclusion et de l'exclusion d'une variable [97].	50
Figure 11: Organigramme de l'algorithme IRIV [97].	51
Figure 12: Structure du réseau de neurones [102].	55
Figure 13 : Photographie des mangues.	63
Figure 14: Spectres moyens de trois échantillons de mangues non mûres (A) et trois échantillons de mangues mûres (B).....	64
Figure 15: Ecart des erreurs quadratiques moyennes en fonction de la moyenne des erreurs respectivement de a) eau, b) pH, c) en sucre.....	68
Figure 16: Confrontation des concentrations en sucre mesurées par les méthodes de référence avec celles prédites par modélisation PLS : (a) courbe d'étalonnage, (b) Courbe de validation	71
Figure 17: Confrontation des concentrations en eau mesurées par les méthodes de référence avec celles prédites par modélisation PLS : (c) courbe d'étalonnage, (d) Courbe de validation.	73
Figure 18: Confrontation des concentrations en pH mesurées par les méthodes de référence avec celles prédites par modélisation PLS : (e) courbe d'étalonnage, (f) Courbe de validation.	74
Figure 19: Spectres de réflectance des tomates.	78
Figure 20: Graphiques des modèles de prédiction.	83
Figure 21: Spectres de réflectance des pommes de cajou.....	88
Figure 22: Corrélation entre les valeurs expérimentales et prédictives pour (a) la matière sèche, (b) la teneur en eau, (c) le pH, (d) la teneur en sucre et (e) la vitamine C.....	95
Figure 23: Spectres de réflectance des papayes.	100
Figure 24: Spectres de fluorescence normalisés des papayes. (a) Spectres bruts de fluorescence, (b) Spectres de fluorescences filtrés par convolution.	103
Figure 25: Valeur singulière en fonction de la composante spectrale.....	104
Figure 26: Spectres de fluorescence des papayes excitées à 380 nm. (a) Réabsorption de la longueur d'onde émise ; (b) Fluorescence du chromophore.	105
Figure 27: Proportion relative de fluorescence à 685 nm par rapport à la fluorescence totale.....	106
Figure 28: Variation du résidu quadratique en fonction de δ	108
Figure 29: Valeurs expérimentales en fonction des valeurs de sortie des étapes du RNA.	112

Liste des tableaux

Tableau 1: Performances statistiques d'étalonnage et de validation.....	70
Tableau 2: Performances statistiques des modèles de prédiction PLS des paramètres de qualité.....	80
Tableau 3: Résultats des performances statistiques des modèles lors de la validation.....	81
Tableau 4: Résultats statistiques du modèle de prédiction	92
Tableau 5: Valeurs des paramètres statistiques des modèles de prédiction	95
Tableau 6: Récapitulatif des paramètres de configuration du spectromètre pour la mesure du spectre de fluorescence.	102
Tableau 7: Valeurs des paramètres statistiques des modèles de prédiction des papayes.....	109

INTRODUCTION

La mesure de la qualité des fruits est un moyen qui permet de contrôler et d'améliorer leurs propriétés organoleptiques, nutritionnelles et hygiéniques. Cette pratique a suscité au cours du temps, le développement d'instruments et de stratégies pour l'évaluation de la maturité et de la qualité des fruits. Ainsi, une première approche considérée comme intuitive s'appuie essentiellement sur les organes de sens pour évaluer la qualité et la maturité. En effet, cette approche consiste à faire appel à un panel de dégustateurs experts pour effectuer des analyses sensorielles. C'est dans cette optique qu'en 2004, dans une étude expérimentale, Gajewski et Arasimowicz [1] ont évalué la qualité sensorielle des aubergines récoltées à trois stades de maturité. Cette étude leur a permis de montrer que l'espèce végétale et le stade de maturité affectent la qualité sensorielle des aubergines. Outre cela, en 2007, Carbonell et ses collaborateurs [2] ont, par des analyses sensorielles, évalué la qualité des jus de mandarine frais et transformés. Ils ont ainsi pu élaborer une liste de descripteurs permettant d'évaluer la qualité des jus de mandarine et ont sélectionné un groupe d'évaluateurs pour des travaux futurs dans le même domaine. De plus, Christen *et al.* (2011) [3] ont évalué les paramètres sensoriels par analyse sensorielle de nouveaux hybrides d'abricots et de clones de Luizet afin de les comparer à la qualité sensorielle des variétés internationales. Par ailleurs, Populer et Lateur (1990) [4] se sont servis, dans une étude, de ces méthodes d'analyses sensorielles comme outils pour évaluer la qualité organoleptique des pommes. Cependant, des limitations importantes que sont la destruction du fruit et la nécessité d'avoir des experts sont alors énumérées. Pour pallier ces limitations, des méthodologies et des instruments ont été développés pour une évaluation objective de la qualité des fruits et des légumes. Ceux-ci permettent de surmonter la subjectivité des méthodes d'analyses sensorielles par l'utilisation d'outils de mesure de grandeurs physiques et de méthodes d'analyses statistiques pour une meilleure prise de décision.

Les premières méthodes dans cette catégorie sont les méthodes d'analyses chimiques. Ces méthodes permettent de déterminer les composants chimiques tels que le sucre, les lipides, les protéines, le pH, l'acidité titrable, la teneur en eau, en vitamine C, en fibres, en minéraux et en pigments végétaux de manière objective. A cet effet, Serrano *et al.* (2005) [5] ont utilisé les méthodes d'analyses chimiques pour suivre le processus de développement et de maturation des cerises sur l'arbre par la détermination des paramètres de qualité tels que la couleur, la texture, la teneur en sucre, les acides organiques, etc.. Aussi, la détermination de certains de ces paramètres à partir des méthodes d'analyses chimiques a permis à Lingani et Traore (2001) [6] de mettre en évidence les qualités nutritionnelles supérieures de la variété de mangue Amélie en comparaison à la variété mango. De même, Assemand *et al.* (2012) [7] ont pu évaluer les caractéristiques biochimiques des fruits de bananes plantains à partir de ces méthodes d'analyses chimiques. Ceux-ci ont déterminé les différents paramètres que sont les taux de matière sèche, de fibres alimentaires, de glucides totaux, d'amidon, le pH, la valeur énergétique, la teneur en protéine, en lipides et les teneurs en minéraux (Ca, Mg, P, K et Na) afin de montrer la particularité des bananes de la variété "Agnrin" par rapport à la variété de référence Orishele. Par ailleurs, Yuyun *et al.* (2017) [8] ont utilisé les méthodes d'analyses chimiques pour évaluer et comparer les propriétés physico-chimiques de vingt-et-une espèces de canneberges cultivées en Nouvelle-Zélande.

Ces différentes méthodes d'analyses chimiques ont la particularité d'être destructives et d'introduire souvent des erreurs au niveau des résultats du fait des changements des propriétés chimiques des fruits dus à l'oxydation, à la variation de la température et de l'humidité. Par ailleurs, les méthodes d'analyses chimiques utilisent plusieurs réactifs trop coûteux et parfois polluants. Les résultats de ces analyses requièrent un temps d'attente assez long à cause de la préparation des échantillons. Aussi, la réalisation de ces méthodes nécessite, dans de nombreux cas, une main d'œuvre importante. Tous ces facteurs limitent l'utilisation de ces méthodes comme techniques

efficaces d'évaluation de la qualité des fruits. Il est alors nécessaire de recourir à des méthodes ou à des instruments d'analyses rapides et non invasifs d'où la seconde catégorie des méthodologies offrant une évaluation plus objective.

La seconde catégorie regroupe les techniques et instruments de mesure rapides et non destructives. Ces techniques concernent les méthodes optiques, mécaniques, électriques et électromagnétiques. Ce sont entre autres, la technique d'imagerie par résonance magnétique nucléaire (IRMN), la technique d'imagerie Raman, la technique de spectroscopie infrarouge, la technique ultrasonique [9,10]. Celles-ci se sont développées de façon spectaculaire et s'efforcent de surmonter les limites des méthodes classiques. Elles sont rapides et efficaces. De même, elles offrent une certaine flexibilité et mobilité très appréciables. De plus, elles permettent d'obtenir des informations à la fois quantitatives et qualitatives. Ces dernières ne nécessitent pas l'utilisation de réactifs donc sont moins coûteuses en temps et en matériels. Parmi celles-ci, les techniques optiques sont largement utilisées, notamment la technique par spectroscopie dans le proche infrarouge [11,12]. Elle est utilisée pour l'évaluation des paramètres de qualité tels que les solides solubles, le pH, l'acidité titrable, la teneur en eau, la matière sèche, la fermeté [13-15]. C'est dans ce sens que, Kawano *et al.* (1992) [16] ont utilisé la méthode de la spectroscopie proche infrarouge sur des pêches afin de mettre en place des analyseurs de sucre. Egalement, Khodabakhshian *et al.* (2017) [17] ont identifié les différents stades de maturité des grenades à partir de la méthode de spectroscopie visible et proche infrarouge (Vis/NIR). Grâce à des méthodes d'analyses multivariées, ils ont prédit les paramètres de qualité de ces fruits. L'instrumentation utilisée par ce groupe de chercheurs pour atteindre leur objectif est composé d'un spectromètre à double canal AvaSpec-2048TEC équipé d'un logiciel AvaSfot7 pour acquérir les spectres en mode transmission. En 2010, l'équipe de Mape *et al.* (2010) [18] a pu tester le pouvoir discriminant de la spectroscopie Vis/NIR sur deux fruits de natures différentes à savoir les pommes et les raisins. Cette équipe a utilisé un spectromètre Vis/NIR comme instrument. Leur travail a montré l'intérêt

d'utiliser la technique par spectroscopie Vis/NIR comme technique non destructive, applicable *in-situ*. Contrairement aux travaux de [18], Chen *et al.* (2017) [19] ont proposé une technologie de recherche par grille combinée à la spectroscopie par transformée de Fourier proche infrarouge (FT-NIR) afin d'optimiser les résultats de l'analyse quantitative de la teneur en solides solubles des échantillons de fraises. A cet effet, dans le but d'améliorer les performances des modèles de prédiction, les algorithmes iPLS et MWPLS ont été modifiés et associés à la technique de recherche par grille afin d'identifier les bandes spectrales informatives.

L'imagerie hyperspectrale est une autre technique optique couramment utilisée. Elle s'est rapidement développée et est utilisée comme outil scientifique émergent dans l'évaluation non destructive de la qualité des fruits et légumes. Cette technique combine à la fois les techniques d'imagerie et de spectroscopie qui sont deux domaines indépendants. Elle établit un lien entre les informations spatiales et les informations spectrales qui résultent de l'interaction lumière-matière. De nombreuses recherches basées sur l'analyse d'images spatiales ou spectrales ont été publiées proposant l'utilisation de l'imagerie hyperspectrale dans le domaine de l'évaluation de la qualité des fruits et légumes. Ainsi, ElMasry *et al.* (2007) [20] ont testé la possibilité d'utiliser cette technique pour la détermination non destructive des attributs de qualité de la fraise. Baiano *et al.* (2012) [21] ont étudié la possibilité d'utilisation de la technique d'imagerie hyperspectrale pour la prédiction de certains paramètres physico-chimiques et sensoriels des raisins de table. Pour acquérir les spectres de réflectance des baies, un système d'imagerie hyperspectrale a été utilisé. Successivement, les mêmes baies ont été analysées pour la mesure du pH, de l'acidité totale et de la teneur en solides solubles selon des méthodes communes. Toutefois, les informations spectrales n'étaient pas corrélées aux données sensorielles, ce qui a rendu difficile la prédiction de la perception des attributs.

En dépit des nombreux avantages qu'offrent les méthodes optiques de caractérisations que sont l'imagerie hyperspectrale et la spectroscopie Vis/NIR, quelques limitations

sont à relever. Notamment, ces techniques ne permettent pas d'analyser plusieurs échantillons en temps réel. En outre, les acquisitions spectrales sont moins rapides et nécessitent la manipulation des échantillons. En effet, ces techniques utilisent pour la plupart des instruments de mesure à contact direct avec l'échantillon et ils sont coûteux. Celles-ci nécessitent souvent la récolte des fruits entraînant des séances de mesure lentes et destructives. Elles ne peuvent pas être utilisées à distance ne permettent donc pas d'avoir des informations en temps réels.

La méthode de spectroscopie par télédétection semble être une alternative. Cette méthode est couramment utilisée sur l'étude des polluants atmosphériques [22,23], dans le domaine de l'entomologie pour caractériser les espèces d'insectes volants [24,25] et dans le domaine de l'agriculture numérique pour le suivi de la production agricole [26]. Les mesures par télédétection se font *in-situ*. C'est le principal avantage de la méthode. La plupart des méthodes spectroscopiques proposées jusqu'à présent utilisent la région spectrale infrarouge qui exige une instrumentation spécifique et coûteuse. Cette thèse a eu pour objet de développer des méthodes de mesure *in-situ* des propriétés physico-chimiques des fruits par méthodes spectroscopiques de télédétection dans le domaine visible du spectre électromagnétique et de développer des méthodes d'analyses prédictives de la qualité de ces fruits. L'objectif est de fournir une méthode moins coûteuse, peu complexe, rationnelle, rapide et non-destructive de caractérisation des fruits en vue d'éliminer les incertitudes introduites par les impuretés des réactifs chimiques. Il a été spécifiquement question de :

- Effectuer des mesures de réflectance et de fluorescence à distance dans le domaine du visible sur des fruits tropicaux ;
- Effectuer des mesures des paramètres de qualité choisis ;
- Déterminer une corrélation entre les mesures spectrales et les paramètres de qualité.
- Evaluer la capacité de prédiction des paramètres de qualité à partir des mesures spectrales.

Le présent rapport est subdivisé en trois (03) chapitres. Le premier chapitre donne une vue générale sur les notions de spectroscopie, de réflectance optique et de télédétection nécessaires à la compréhension des travaux mentionnés dans ce rapport. La théorie sur les méthodes d'analyse multivariées est abordée et une vue générale sur les propriétés optiques des fruits est également exposée. Le deuxième chapitre est essentiellement consacré au descriptif des techniques expérimentales mises en œuvre au cours de nos travaux. Il présente également les méthodes d'analyses des données de mesures. En ce qui concerne le troisième chapitre, il est consacré aux résultats obtenus ainsi qu'aux discussions qui en découlent. Enfin, nous terminons par une conclusion et des perspectives.

CHAPITRE I : SPECTROSCOPIE OPTIQUE PAR TELEDETECTION

Introduction

L'analyse spectroscopique de la qualité des fruits nécessite dans de nombreux cas, la connaissance de la composition et des propriétés optiques de ceux-ci. L'utilisation d'outils standards de mesure de qualité donne lieu à l'obtention de valeurs moyennes des paramètres de qualité. La mise en œuvre de dispositifs et de méthodes permettant d'accéder plus facilement à la qualité des fruits à partir de données spectrales nécessite la compréhension de la notion de spectroscopie. Ainsi dans cette section, nous aborderons tout d'abord la notion de spectroscopie de réflectance optique et de télédétection. Un accent sera ensuite mis sur la composition des fruits, leurs propriétés optiques et leurs paramètres de qualité ainsi que les méthodes d'analyses statistiques utilisées lors de ces travaux.

I.1. Spectroscopie optique

La spectroscopie optique est l'étude des rayonnements électromagnétiques émis, absorbés, réfléchis, ou diffusés par la matière. Elle étudie les propriétés optiques de la matière. C'est un concept qui a débuté par l'analyse de la lumière blanche par un prisme. Plus tard, cette notion s'est étendue à toute interaction de la lumière avec la matière. Certains phénomènes comme le rayonnement du corps noir, l'effet photoélectrique, la structure, le spectre atomique, et bien d'autres ont été expliqués grâce aux études spectroscopiques. La spectroscopie est devenue ainsi une science incontournable de nos jours. Elle intervient pratiquement dans tous les domaines d'activités tels que la médecine, la sécurité, le contrôle qualité, etc... Ainsi, l'interaction rayonnement – matière donne lieu aux phénomènes de réflexion, d'absorption, de transmission et de diffusion par la matière. L'étude de chaque phénomène correspond à un type de spectroscopie.

- spectroscopie de réflectance lorsque la lumière incidente est réfléchiée dans le même milieu ;
- spectroscopie d'absorption lorsque la lumière incidente est absorbée par la matière ;

- spectroscopie de diffusion lorsque la lumière incidente est redirigée dans toutes les directions à l'issue de son interaction avec la matière ;
- spectroscopie de fluorescence lorsqu'il y a une émission de nouvelles longueurs d'onde après interaction de l'onde incidente avec la matière.

Nous présentons par la suite un aperçu de ces techniques.

I.1.1. Spectroscopie de réflectance optique

La spectroscopie de réflectance optique, cas de notre étude, est une autre forme de spectroscopie de diffusion [27]. Cette technique est basée sur la façon dont la surface d'un objet renvoie un rayonnement électromagnétique. Elle permet de mesurer l'intensité de lumière réfléchi par un objet. La réflexion est définie comme un processus de renvoi du rayonnement électromagnétique quand celui-ci rencontre une surface réfléchissante. En fonction de l'état de la surface de l'échantillon à étudier, nous pouvons définir deux types de réflexions : la réflexion spéculaire et la réflexion diffuse. La figure 1 présente une illustration de ces deux phénomènes.

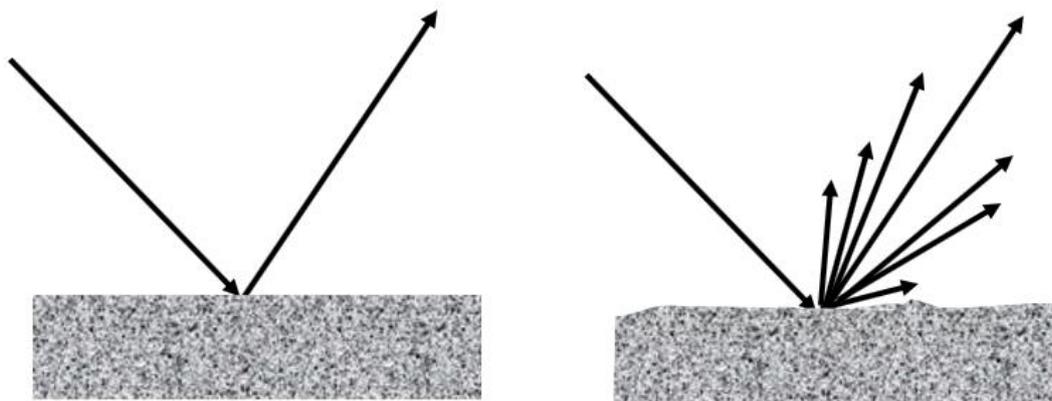


Figure 1: Représentation des phénomènes de réflexion spéculaire (à gauche) et de réflexion diffuse (à droite)

La réflexion est dite spéculaire lorsque le rayonnement incident est réfléchi par la surface de la matière dans le même milieu et dans une seule direction. Elle se produit uniquement sur des surfaces lisses, dont les aspérités ont une taille inférieure à la longueur d'onde du rayonnement incident.

Lorsque les surfaces sont rugueuses et qu'elles présentent des aspérités dont la taille est supérieure à la longueur d'onde du rayonnement incident, la réflexion est diffuse. Dans ce cas, le rayonnement incident est réfléchi par la matière dans toutes les directions. En ce qui concerne les organismes biologiques, ces deux types de réflexion se manifestent [28]. L'intensité du rayonnement réfléchi par l'échantillon renseigne sur sa composition, ses propriétés physico-chimiques et optiques. En effectuant une mesure de réflectance, les deux types de réflexion se produisent simultanément. Cela permet d'avoir un grand nombre d'informations sur l'échantillon à étudier. La contribution relative des deux composantes spéculaires et diffuses varie avec la surface géométrique. Des travaux ont montré que l'expression de la réflexion diffuse est fonction des propriétés géométriques et optiques de l'échantillon [29]. En effet, Tominaga *et al.* (1990) [30] ont montré que l'expression de la réflexion spéculaire met en exergue la géométrie de la substance, par contre, la réflexion diffuse fait ressortir les caractéristiques chimiques et optiques de l'échantillon. En spectroscopie de réflectance diffuse, selon le type d'information recherchée, la réflexion spéculaire peut être considérée comme une source de bruit qu'il faut alors réduire [31]. Pour une matière constituée de différentes couches successives, le coefficient de réflexion $R(\theta_i)$ peut être évalué selon la relation de Fresnel exprimée par l'équation (Eq I.1) [32] :

$$R(\theta_i) = \frac{1}{2} \left[\frac{\sin^2(\theta_i - \theta_t)}{\sin^2(\theta_i + \theta_t)} + \frac{\tan^2(\theta_i - \theta_t)}{\tan^2(\theta_i + \theta_t)} \right] \quad \text{I. 1}$$

Où θ_i est l'angle d'incidence et θ_t l'angle de transmittance, liés par la loi de Snell Descartes, $n_i \sin\theta_i = n_t \sin\theta_t$, où n_i et n_t sont respectivement les indices des milieux de propagation de la lumière.

I.1.2. Spectroscopie d'absorption

La spectroscopie d'absorption est l'étude de l'atténuation de la radiation incidente qui traverse une matière. Elle permet d'identifier les transitions électroniques et d'évaluer les concentrations des substances absorbantes. En effet, lorsque la lumière entre en collision avec les molécules de la matière, celle-ci leur cède son énergie. Si cette énergie

cedée correspond à la séparation entre deux niveaux d'énergie des atomes ou des molécules, alors cette énergie est absorbée. Le phénomène d'absorption est mis en évidence par la loi de Beer-Lambert-Bouguer. Cette loi est basée sur la comparaison de l'intensité du faisceau incident et du faisceau transmis. Elle traduit la proportionnalité entre l'intensité lumineuse absorbée à une longueur d'onde donnée et la quantité de la matière absorbante. L'expression de la loi de Beer-Lambert - Bouguer est donnée par l'équation (Eq I.2) :

$$A_{\lambda} = -\log\left(\frac{I}{I_0}\right) = \varepsilon_{\lambda} \cdot l \cdot c \quad \text{I.2}$$

Avec :

I_0 : intensité de la lumière incidente en Candela (cd)

I : intensité de la lumière transmise en Candela (cd)

ε_{λ} : coefficient d'extinction molaire ; grandeur caractéristique du composé absorbant exprimée en $L \cdot mol^{-1} \cdot cm^{-1}$

C : concentration en mol. L^{-1}

l : longueur du trajet optique en cm

La loi de Beer-Lambert ne s'applique qu'à une lumière monochromatique et aux solutions suffisamment diluées afin d'atténuer le phénomène de diffusion multiple.

I.1.3. Spectroscopie de diffusion

La spectroscopie de diffusion consiste à l'étude de l'interaction de la lumière avec la matière au cours de laquelle la direction de propagation du rayonnement incident et sa polarisation sont changées. La lumière ainsi diffusée permet d'obtenir des informations sur la composition ou les caractéristiques de la matière étudiée et contient l'information sur la dépolarisation [33].

Les phénomènes de diffusion de la lumière interviennent lorsque des hétérogénéités structurelles apparaissent soit dans le volume du milieu de propagation, soit à la

surface de séparation entre deux matériaux de caractéristiques optiques différentes. Les inhomogénéités de volume traduisent des variations plus ou moins locales de l'indice de réfraction et les hétérogénéités en surface correspondent à la rugosité des interfaces. La diffusion renseigne donc sur la rugosité de la matière et son hétérogénéité structurelle.

Il existe deux types de diffusion en fonction de la nature des particules diffusantes : La diffusion élastique et la diffusion inélastique [34].

La diffusion élastique renferme elle-même plusieurs diffusions telles que la diffusion de Mie, la diffusion de Rayleigh et la diffusion géométrique. Les notions de diffusions de Rayleigh et de Mie sont brièvement abordées ci-dessous :

-Diffusion de Rayleigh

La diffusion de Rayleigh correspond à la plus grande partie de l'énergie diffusée lors de l'interaction entre la lumière et la matière. Cette diffusion dite élastique se caractérise par une fréquence identique à celle de l'onde incidente d'où la même énergie. C'est un phénomène qui a lieu lorsque la taille des molécules diffusantes est très inférieure à la longueur d'onde du rayonnement. En effet, lorsqu'une particule est soumise à une onde plane monochromatique, de longueur d'onde λ , le champ électrique de cette onde induit dans la particule un moment dipolaire oscillant [35]. Donc, toutes les molécules oscillent en phase sous l'effet du champ électrique incident. Les hautes couches de l'atmosphère sont affectées par la diffusion Rayleigh. Ce phénomène est à la base de la couleur bleue du ciel pendant la journée.

-Diffusion de Mie

La théorie de la diffusion de Mie concerne la diffusion d'une onde électromagnétique par une particule de forme sphérique homogène dont le rayon est du même ordre de grandeur que l'onde incidente [36]. C'est d'ailleurs, Mie *et al.* (1908) [37] qui furent les premiers à avoir résolu le problème de la diffusion de la lumière par une particule sphérique en utilisant la théorie de l'électromagnétisme de Maxwell. Au niveau de

l'atmosphère, les gouttelettes d'eau, les cristaux de glace, ou les aérosols sont les principaux vecteurs de la diffusion de Mie.

I.1.4. Spectroscopie de fluorescence

La spectroscopie de fluorescence est l'étude de l'émission lumineuse provoquée par l'excitation d'une molécule. En effet, lorsqu'une molécule fluorescente ou un fluorophore absorbe un photon, un électron est transféré vers une orbitale moléculaire d'énergie supérieure. La molécule passe alors de l'état fondamental à l'état excité (instable) d'énergie supérieure. Le retour à l'état de repos stable de cette molécule produit différents phénomènes possibles, dont la fluorescence. Le rendement quantique de la fluorescence d'une molécule détermine son efficacité. Il se définit comme étant le rapport du nombre de photons émis sur le nombre de photons absorbés. Le rendement quantique peut varier en fonction des conditions environnementales des fluorophores. Lors d'une fluorescence, le photon émis a une énergie plus faible que celle du photon d'excitation. De ce fait, les longueurs d'onde d'émission sont supérieures à celles de l'excitation. La spectroscopie de fluorescence est appliquée dans plusieurs domaines, en particulier celui de la biologie où elle est utilisée comme outil d'analyse non invasif depuis la découverte de molécules fluorescentes pouvant être utilisées comme sondes.

I.2. Télédétection

I.2.1. Définition

Le mot télédétection est composé du préfixe « télé » qui signifie à distance et de la racine « détection » qui veut dire découvrir ou déceler. Ainsi, la télédétection peut être définie comme l'ensemble des techniques utilisées pour analyser les caractéristiques physiques et biologiques d'un objet ou d'un environnement à distance. Elle permet d'acquérir des informations sur un objet à partir de l'interaction de celui-ci avec un rayonnement électromagnétique. Il existe plusieurs techniques de télédétection à savoir, le RADAR, le LIDAR, le SONAR, etc. La télédétection intègre non seulement

des traitements numériques (c'est-à-dire tout le processus lié à l'acquisition de données) mais aussi utilise des techniques spectroscopiques. Elle utilise principalement des rayonnements électromagnétiques émis ou réfléchis par les objets d'intérêt afin d'analyser leurs caractéristiques. Il existe deux types de télédétection : la télédétection active et la télédétection passive. Lorsque la source d'illumination de l'échantillon est le soleil, la télédétection est dite passive. On parle de la télédétection active, lorsque qu'une source d'illumination artificielle est émise vers la cible et la réponse de la cible est mesurée par la suite. La figure ci-dessous est une illustration de ces deux types de télédétection.

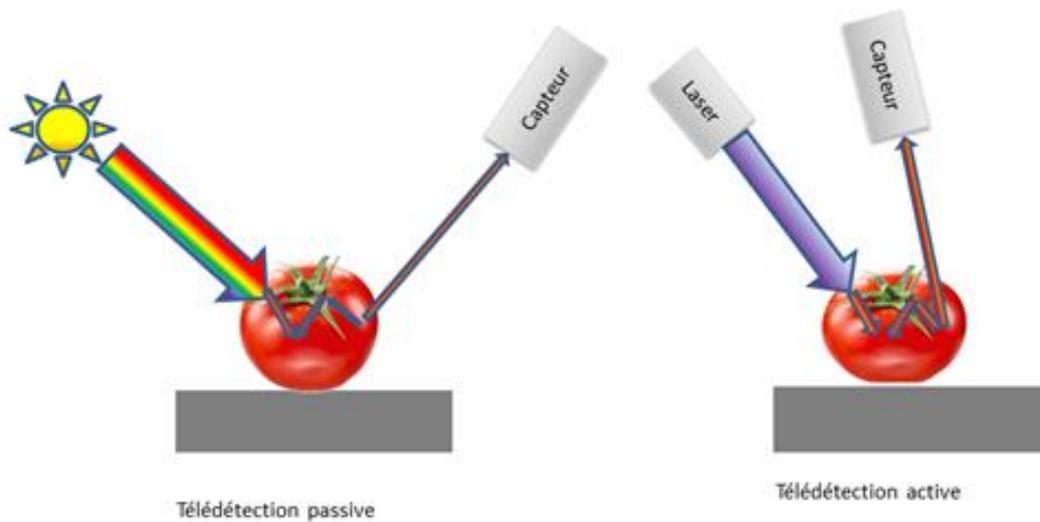


Figure 2: Schéma illustrant la télédétection passive (Gauche) et la télédétection active (Droite).

La source lumineuse utilisée en télédétection passive est une source de lumière naturelle. Cette source n'intègre pas le système d'acquisition : c'est un système de mesure optique passive. La source lumineuse utilisée en télédétection active est une diode laser qui est une source artificielle. Ce type de télédétection active est appelé LIDAR.

I.2.2. Applications de la télédétection

La télédétection est une technique transversale. Elle s'applique dans de nombreux domaines. En géographie, la télédétection intervient dans le suivi de l'évolution des reliefs et des formations végétales. En outre, en climatographie, elle permet d'avoir des images ou des données spectrales pour mettre en évidence certaines propriétés de la matrice d'étude. En météorologie, elle intervient pour l'étude de l'atmosphère. Son intérêt dans le domaine de l'océanographie consiste à offrir une vision synoptique de vastes régions qu'il est impossible d'obtenir par les bateaux. Les applications terrestres de la télédétection sont extrêmement variées. Par ailleurs, ces applications s'étendent également à la prospection minière, à la surveillance des cultures et du couvert forestier et à l'urbanisme. La télédétection s'applique de plus en plus à l'agriculture et à la sylviculture. En agriculture, elle permet de fournir des informations sur l'état des cultures pendant toutes les saisons à différentes échelles. Elle donne également des informations sur l'état du sol et permet d'estimer les superficies agricoles. Egalement, pour le contexte agricole, des techniques de télédétection pertinentes ont été développées [38-40]. Ces techniques permettent d'estimer les variables agronomiques et les caractéristiques végétales comme les caractéristiques morphologiques, biochimiques, physiologiques, structurelles, phénologiques ou comportementales qui influencent la performance ou la forme physique des végétaux. A cet effet, Nock et al. (2016) ont défini ces variables comme des traits ou des caractéristiques qui varient d'une plante, d'une culture ou d'une région à l'autre [41]. De récentes recherches ont permis de renforcer les capacités d'application de la télédétection en fonction des besoins spécifiques. Elles pourront ainsi contribuer à la prévision des rendements des cultures et à la surveillance des terres agricoles. Enfin, la télédétection apparaît comme un outil essentiel pour répondre aux besoins agricoles.

La spectroscopie est une technique non destructive, rapide et fiable permettant d'évaluer de manière précise la composition physico-chimique de la matière [19]. Elle regroupe plusieurs techniques notamment la spectroscopie infrarouge ou ultraviolet,

l'imagerie hyperspectrale, la résonance magnétique nucléaire et la spectroscopie d'émission atomique couramment utilisés en agriculture. La spectroscopie est appliquée en agriculture pour les détections précoces de certaines maladies des plantes [42,43] et des stress hydriques des plantes au niveau des cultures [44]. Aussi, elle est employée pour le suivi de la croissance des plantes agricoles [45]. Les techniques de spectroscopie sont aussi utilisées en agriculture pour le suivi de la maturité, l'analyse de la qualité des fruits et légumes [46-48]. De même, elles interviennent en pédologie pour l'étude de la composition des sols en vue de les cultiver [49,50]. Par ailleurs, les industries agroalimentaires les utilisent pour l'analyse et le contrôle de la qualité des aliments [51].

I.3. Propriétés optiques des fruits

I.3.1. Composition des fruits

Les fruits sont des organes végétaux qui contiennent essentiellement de l'eau, des éléments minéraux (K, P, Ca, Fe, Mg, Mn, Cl, Na, ...), des sucres, des lipides, des acides aminés et des acides organiques. Ils contiennent également des pigments, des protéines, des fibres alimentaires, des composés aromatiques, des polyphénols, etc. Les fruits sont constitués également des vitamines et d'autres métabolites. Ces différents constituants confèrent aux fruits une qualité nutritionnelle exceptionnelle [6]. La grande richesse des fruits en antioxydants tels que les polyphénols, les caroténoïdes et la vitamine C leur confère une grande capacité à neutraliser les excès de radicaux libres dans l'organisme. Les pigments sont des composés chimiques dans les fruits. En général d'excellents antioxydants, ils donnent leurs couleurs aux fruits. L'eau est le constituant le plus important des fruits. Elle représente près de 90% de la masse totale d'un fruit [52, 53]. La quantité maximale d'eau varie d'un fruit à l'autre, eu égard aux différences structurelles. Cette richesse en eau des fruits représente le principal facteur limitant de leur conservation à l'état frais. Les glucides sont à la suite de l'eau, les constituants les plus abondants des fruits. Ils constituent ainsi 50% à 80% de la masse sèche totale. Les principaux glucides présents dans les fruits sont le

fructose, le glucose, le saccharose et l'amidon. Ils contribuent principalement à déterminer la saveur des fruits [54, 55]. Les acides organiques contribuent à la qualité sensorielle des fruits dans le rapport sucre/acide. La quantité des acides dans un fruit n'est pas statique. L'acidité totale des fruits diminue généralement au cours de la maturation. La saveur d'un fruit est liée à la quantité d'acides présents dans le fruit. A l'exception des noix qui sont une source appréciable de protéines, la quantité de protéines dans les fruits est faible. Les tissus sénescents des fruits sur-mûris contiennent des proportions élevées de protéines. En majorité, les produits agricoles sont faibles en lipides. Certains fruits comme les avocats et les olives contiennent assez de lipides. La teneur en matières grasses des fruits et légumes est très faible et varie en fonction de l'espèce. La composition des fruits dépend des facteurs tels que leur variété, leur état de maturité et la lumière. En outre, le climat, la fertilité du sol et les pratiques culturales influencent la composition des fruits. Les concentrations relatives des constituants des fruits sont importantes pour évaluer leur qualité .

I.3.2. Propriétés optiques des fruits

Les propriétés optiques des fruits caractérisent leur capacité à absorber, à réfléchir ou à transmettre la lumière reçue. L'étude de ces propriétés optiques fournit des informations essentielles pour la compréhension des interactions physiologiques et physiques entre une radiation et un tissu biologique. Ces propriétés optiques dépendent des caractéristiques physiques et chimiques des fruits comme la densité, la taille, la teneur en eau, les pigments et les sucres, la présence de graines, etc. A titre d'exemples, nous aborderons le sujet des propriétés optiques des principaux composants que sont l'eau, les pigments et les sucres.

I.3.2.1. Pigments

Les fruits contiennent plusieurs types de pigments. Ces pigments sont responsables de la coloration de leur peau et de leur pulpe. Ils se retrouvent ainsi dans les tissus externes et internes. Parmi ces pigments nous avons la chlorophylle, qui joue le rôle de pigment principal et est responsable de la coloration verte des fruits. La chlorophylle

absorbe préférentiellement les longueurs d'ondes du bleu (400 nm - 500 nm) et celles du rouge (600 nm - 700 nm) dans le spectre du visible [56]. On distingue deux principaux types de chlorophylle au niveau des végétaux : la chlorophylle a (chl a) et la chlorophylle b (chl b) [57]. Les caroténoïdes sont des pigments qui sont responsables de la coloration jaune orangé et absorbent les longueurs d'onde situées entre 400 nm et 500 nm. Il existe plusieurs types de caroténoïdes divisés en deux grands groupes : les carotènes (α -carotène et β -carotène) et les xanthophylles [58]. Ces pigments jouent un rôle protecteur pour les végétaux car ils aident à la régulation de l'excès de lumière absorbée. On a aussi les flavonoïdes dont le plus important, l'anthocyane se retrouve dans les fruits et les fleurs. Ce pigment joue le rôle de pigment secondaire lors de l'activité photosynthétique car il assure le transport de l'énergie vers le centre réactionnel contenant la molécule spéciale de chl a. Les flavonoïdes absorbent dans l'UV et dans le vert.

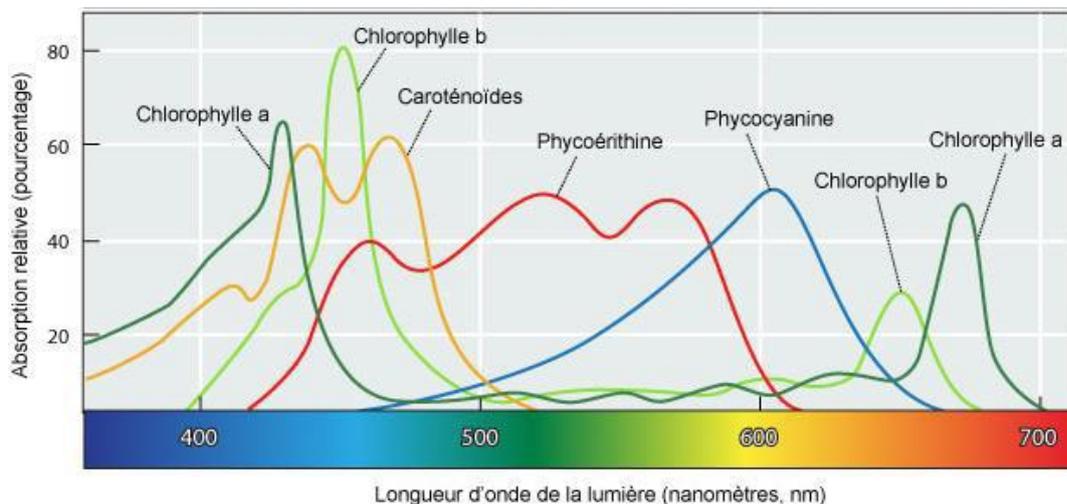


Figure 3 : Spectre d'absorption des pigments [59]

I.3.2.2. Eau

L'eau représente généralement le composé majoritaire des fruits. Elle constitue plus de 80 % de la masse totale. A l'intérieur des fruits, l'eau se trouve sous forme liquide située dans les vacuoles cellulaires. Cette quantité importante d'eau dans les fruits est la raison pour laquelle l'absorption du proche infrarouge rend compte de la présence d'eau. Le spectre d'absorption de l'eau des végétaux et de l'eau pure sont presque

identiques. L'absorption de l'eau est négligeable dans le visible (400 nm - 700 nm), faible dans le proche infrarouge (700 nm -1300 nm) et forte dans l'infrarouge moyen (1300 nm -2500 nm), avec des pics d'absorption observés aux environ de 970 nm, 1450 nm, 1950 nm, 2250 nm [60,61].

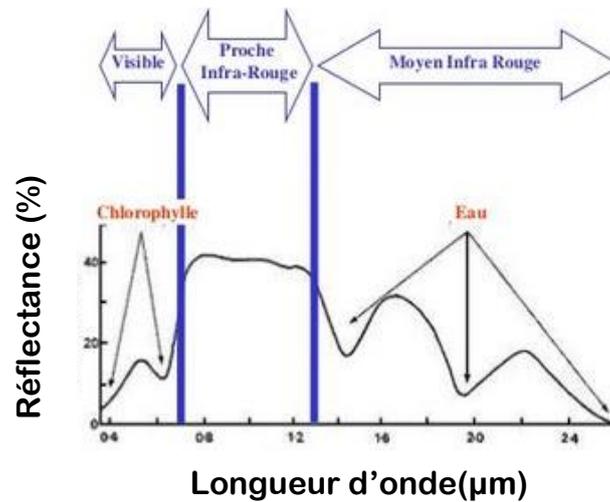


Figure 4 : Spectre d'absorption de l'eau [62]

I.3.2.3. Sucres

Les teneurs en sucres des fruits dont les principaux sont le fructose, le glucose et le saccharose déterminent leurs saveurs [54]. Ces glucides (sucres) contiennent plusieurs groupements CH et OH dans leur structure moléculaire. Ces groupements sont responsables des fortes absorptions de certaines longueurs d'onde du proche infrarouge [63]. Leurs pics d'absorption se situent plus souvent dans les zones de 1160-1200 nm, de 1780 nm et de 1920-1940 nm [64].

I.4. Paramètres de qualité des fruits

La qualité des fruits est une notion complexe qui définit le degré d'excellence ou de supériorité de ceux-ci. Elle renferme en elle-même plusieurs aspects de qualité que sont : la qualité nutritionnelle, la qualité organoleptique, la qualité sanitaire ainsi que la qualité d'usage et de service. La qualité nutritionnelle indique l'aptitude à bien

nourrir au travers des différents composants (antioxydants, fibres, minéraux, sucres, etc.) des fruits. Elle est beaucoup plus liée à la santé. La qualité organoleptique (saveur, parfum, texture, couleur, calibre, défaut, etc.) définie par les différents organes de sens est quant à elle liée à la satisfaction. Elle est subjective car elle varie en fonction du temps et de l'individu. Elle exprime également la perception humaine de la composition d'un fruit. La qualité sanitaire est la qualité qui prend en compte l'absence d'éléments étrangers et toxiques (parasites, résidus de produits phytosanitaires). La qualité d'usage et de service concerne l'aptitude des fruits à la transformation, à la conservation, à leurs commodités d'emploi et à leurs aspects économiques. La qualité globale des fruits est très dépendante de leurs degrés de maturité, des conditions de culture, d'entreposage et de conservation.

I.4.1. Paramètres physiques et chimiques de qualité

Les fruits sont évalués sur la base de leur qualité. Cette qualité se définit suivant plusieurs critères ou paramètres. Les différents critères physico-chimiques principalement utilisés pour évaluer la qualité des fruits se regroupent en des paramètres physiques et des paramètres chimiques.

I.4.1.1. Paramètres physiques

Les paramètres physiques regroupent les critères liés à l'apparence tels que la masse, le volume, la taille, le calibre, la forme et la couleur, à la saveur (astringence, arôme, etc.) et à la texture (fermeté, jutosité, croquant, etc.). L'analyse de ces paramètres requiert l'usage de méthodes d'analyse sensorielle et de méthodes physiques. La masse en g est un paramètre physique qui s'obtient par la pesée. La compressibilité de la peau et les fermetés sont déterminées par l'usage d'un pénétromètre [65]. D'autres caractéristiques physiques telles que l'astringence, l'arôme, etc. sont également mesurées par analyse sensorielle. L'analyse de la couleur s'effectue par la méthode de chromatographie.

I.4.1.2. Paramètres chimiques

Les fruits sont d'importantes sources de paramètres chimiques dont les principales caractéristiques de qualité sont les vitamines (A, B, C, etc.), les fibres et les éléments minéraux (K, P, Ca, Fe, Mg, Mn, Cl, Na, ...). Ils contiennent également des glucides (fructose, glucose, saccharose et l'amidon), des protéines, des acides (acide citrique, acide malique, etc.), des lipides, etc. L'évaluation et l'assurance de la qualité des fruits avant ou après leur récolte représentent des procédures nécessaires pour garantir leur qualité. Elles sont principalement basées sur les caractéristiques de qualité interne telles que le pH, la vitamine C, l'acidité, la teneur en sucre, en eau, etc. La concentration de chaque paramètre chimique dans un fruit est déterminée par une méthode bien spécifique.

I.4. 2. Paramètres optiques

La propagation de la lumière dans les tissus biologiques tels que les fruits et légumes est un mécanisme complexe qui implique principalement des phénomènes de diffusion et d'absorption. L'absorption de la lumière est exclusivement due aux molécules absorbantes (pigments, eau, sucres, etc.) de l'échantillon, tandis que la diffusion de la lumière est davantage liée à ses paramètres structurels (densité, taille des particules et structures cellulaires). Ces deux phénomènes sont caractérisés par le coefficient d'absorption μ_a , le coefficient de diffusion μ_s et l'indice de réfraction n des particules. La connaissance de ces coefficients optiques est nécessaire pour mieux comprendre le processus d'interaction des tissus ou pour améliorer l'optimisation des technologies d'analyse non invasives (non destructives). Diverses méthodes optiques ont été mises au point pour quantifier l'absorption et la diffusion de la lumière dans les échantillons biologiques. Pour ce faire, le coefficient d'absorption et le coefficient de diffusion sont les principaux paramètres optiques considérés dans les études.

I.4.2.1. Coefficient d'absorption

Le phénomène d'absorption est le transfert d'énergie d'un photon à une molécule lors d'une interaction lumière-matière. Le coefficient d'absorption μ_a est décrit comme le

paramètre caractérisant les propriétés absorbantes de la matière ayant une distribution homogène des particules absorbantes. Il représente la probabilité d'absorption d'un photon par la matière pendant son parcours en fonction de l'unité de distance. Son expression est donnée par l'équation I.3 [66] :

$$\mu_a = \rho_a \sigma_a \quad \text{I.3}$$

ρ_a : densité de particules absorbantes par unité de volume (m^{-3})

σ_a : section efficace d'absorption (m^2).

De nombreuses études ont utilisé ce paramètre optique pour l'analyse de plusieurs fruits.

I.4.2.2. Coefficient de diffusion

La diffusion de la lumière est une interaction de la lumière avec la matière au cours de laquelle la direction du rayonnement incident et sa polarisation sont modifiées. La lumière ainsi diffusée porte l'information de dépolarisation [33]. Les phénomènes de diffusion de la lumière interviennent lorsque des hétérogénéités structurelles apparaissent soit dans le volume du milieu de propagation, soit à la surface de séparation entre deux matériaux de caractéristiques optiques différentes. Les propriétés diffusantes du milieu sont caractérisées par le coefficient de diffusion, qui décrit la probabilité de diffusion d'un photon lors de son trajet en fonction de l'unité de distance. Par analogie avec le coefficient d'absorption, il est défini sous la forme suivante :

$$\mu_s = \rho_s \sigma_s \quad \text{I.4}$$

ρ_s : nombre des particules diffusantes par unité de volume du milieu (m^{-3}),

σ_s : section efficace de diffusion (m^2).

I.4.3. Techniques de mesure de qualité

La mauvaise qualité des fruits conduit à une grande perte de production et d'économie au niveau des industries agricoles mondiales. Aussi, elle peut conduire à la réduction de la consommation des fruits et à la cherté. La mesure et le suivi de la qualité des fruits sont importants pour fournir des fruits de qualité supérieure et constante sur le marché et garantir l'acceptation ainsi que la satisfaction des consommateurs. Deux grandes méthodes existent pour la mesure de la qualité des fruits : les méthodes destructives et les méthodes non destructives.

I.4.3.1. Techniques destructives

La plupart des techniques destructives sont des techniques d'analyses chimiques, physiques et sensorielles de qualité des fruits. Ces techniques permettent d'apprécier la qualité des fruits au travers d'un ensemble de paramètres physico-chimiques ou sensoriels. On peut citer :

- les méthodes de l'AOAC (1984,1990, 1995) et la méthode AFNOR qui sont largement utilisées pour déterminer les paramètres tels que le pH, l'acidité titrable, l'extrait sec soluble (sucre), la teneur en cendre, la teneur en acide ascorbique (vitamine C), les protéines, etc. [67] ;

- la méthode de spectroscopie d'absorption atomique qui est une technique de dosage utilisée pour déterminer la concentration des éléments minéraux tels que K, P, Ca, Fe, Mg, Mn, Cl, Na. Cette technique repose sur le principe selon lequel les éléments à analyser peuvent absorber des photons d'une certaine longueur d'onde. Le nombre de photons absorbés étant relié au nombre d'atomes qui les absorbent, on peut ainsi en déduire la concentration de l'élément. C'est une méthode très répandue pour l'analyse minérale des éléments en solution.

- En ce qui concerne les méthodes d'analyses physiques, elles font appel aux méthodes de compression, de pesée, de pénétrométrie et de rhéologie. Ces méthodes sont moins fastidieuses et moins lentes que les méthodes d'analyses chimiques.

-les méthodes d'analyses sensorielles utilisent les différents organes de sens. Elles demandent un panel d'experts entraînés pour les tests de dégustation.

Toutes ces méthodes nécessitent la destruction des fruits, l'extraction des jus pour certaines et des protocoles de mesures différents. Ces différentes méthodes d'analyse ont la particularité d'être destructives et peuvent souvent induire des erreurs au niveau des résultats du fait des changements des propriétés chimiques des fruits par des phénomènes telles que l'oxydation, la variation de la température et de l'humidité. Par ailleurs, les analyses utilisent plusieurs réactifs très coûteux et parfois polluants. Les résultats requièrent un temps d'attente assez long à cause de la préparation des échantillons mais aussi les analyses demandent dans de nombreux cas une main d'œuvre importante. Tous ces facteurs limitent l'utilisation de ces méthodes comme techniques efficaces d'évaluation de la qualité des fruits. Il est alors nécessaire de recourir à des méthodes ou à des instruments d'analyses rapides et non destructifs d'où la seconde catégorie des méthodologies offrant une évaluation plus objective.

I.4.3.2. Techniques non destructives

La mise en place de nouvelles technologies au niveau du secteur agroalimentaire conduit à une nouvelle ère de méthodes automatiques et non destructives dans la mesure de la qualité des fruits. Ces outils de mesure doivent être sensibles pour permettre une évaluation plus efficace de la qualité des fruits. Récemment, un intérêt croissant s'est manifesté pour ces méthodes non destructives d'évaluation de la qualité, et un effort considérable a été fait dans cette direction. Ainsi, différentes techniques non destructives ont été mises au point pour mesurer les composantes de la qualité des fruits. Elles peuvent être classées en techniques mécaniques, optiques, électromagnétiques et dynamiques selon leurs principes de mesure.

Les techniques optiques sont les plus utilisées parmi ces techniques. En effet, l'interaction entre le rayonnement lumineux et la matière fait ressortir les propriétés (absorbance, réflectance et transmittance) de cette matière. Ces propriétés sont alors utilisées pour caractériser la matière. Au niveau des fruits, le domaine visible-

infrarouge du spectre électromagnétique est utilisé. Le spectre obtenu à partir de l'interaction entre le fruit et la lumière est fonction du contenu biochimique du fruit (eau, taux de chlorophylle, de caroténoïdes, etc.) et peut donc être exploité pour la caractérisation du fruit. Il existe plusieurs types de techniques optiques :

-la spectroscopie visible-infra rouge : elle est utilisée comme une méthode rapide, efficace et non destructive pour le suivi de l'évolution et pour la mesure de la qualité des fruits [68, 69, 70]. Cette technique consiste à faire interagir les échantillons avec une source lumineuse et à l'aide d'un détecteur (spectromètre), récupérer le signal réémis par ceux-ci. Certaines longueurs d'onde du visible et de l'infrarouge sont identifiées comme spécifiques pour l'analyse de la qualité, à cause de leurs interactions avec les liaisons C-H, O-H et N-H des principaux composés (eau, sucres, chlorophylles, caroténoïdes, etc.). Ces longueurs d'onde portent un maximum d'informations sur ces composants.

- L'imagerie hyperspectrale : Cette technique consiste en l'acquisition d'images d'un fruit afin d'exploiter au mieux l'information. L'imagerie hyperspectrale permet de reconnaître dans le pixel la signature spectrale des constituants qu'il contient, c'est-à-dire de les identifier. La position en longueur d'onde des pics d'absorption spécifiques dépend de leurs compositions chimiques spectrales. Le principal avantage de l'imagerie hyperspectrale est qu'elle facilite la visualisation de la distribution des différents composants chimiques des échantillons. La combinaison de l'imagerie et de la spectroscopie offre également un grand avantage à cette technique, car elle permet d'avoir à la fois des informations spatiales et spectrales dans l'intervalle de longueur d'onde d'étude.

D'autres techniques optiques sont aussi utilisées pour la mesure de la qualité des fruits. On peut citer la technique de vision par machine [71], la spectroscopie de réflectance à résolution temporelle, etc.

Une importante partie des techniques décrites plus haut utilise la chimiométrie comme moyen de visualisation et d'analyses des données acquises.

I.5. Méthodes d'analyses chimiométriques ou d'analyses statistiques multivariées

La chimiométrie ou l'analyse statistique multivariée est la science de l'analyse de données. Elle permet d'extraire les informations pertinentes et utiles à partir de données chimiques et physiques mesurées ou connues brutes. C'est aussi une discipline qui utilise des méthodes mathématiques et statistiques pour analyser des données de façon optimale ou pour synthétiser l'information issue de plusieurs variables afin de mieux l'expliquer. Elle trouve son intérêt dans le traitement des systèmes complexes dits multi variés. A l'origine, l'analyse multivariée se limitait à l'analyse des données en chimie analytique. Cela a conduit à l'appellation « chimiométrie » proposée par Wold et Kowalski. De nos jours, elle comprend toutes les applications de méthodes statistiques et mathématiques. En outre, elle se sert de l'outil informatique pour résoudre de nombreux problèmes dans divers domaines tels la chimie, la physique, l'économie et la sociologie. La chimiométrie regroupe de nombreuses méthodes d'analyse de données. Celles-ci permettent de répondre à des problématiques variées. Cependant, la technique d'analyse la plus appropriée pour une étude donnée est fonction du type d'étude et des principales questions de recherche. Les méthodes d'analyse chimiométrique renferment les méthodes statistiques de prétraitement et de traitement adaptées aux données spectrales.

I.5.1. Méthodes statistiques de prétraitement

I.5.1.1. Lissage

Le lissage est une méthode de prétraitement des données qui permet de diminuer les bruits de haute fréquence contenus dans les données. L'application de cette méthode aux spectres permet également d'améliorer le rapport signal/bruit. Le lissage utilise comme idée de base la recherche d'une valeur d'estimation optimale par l'ajustement de points dans une fenêtre. La résolution spectrale est fonction du choix de la largeur de la fenêtre. En effet, un lissage trop important peut entraîner un risque de perte d'informations pertinentes. Il existe plusieurs méthodes de lissage à savoir le lissage par moyenne mobile, le lissage par le filtre gaussien, le lissage par le filtre médian, le

lissage Savitzky-Golay et bien d'autres. La méthode de lissage utilisée dans cette étude est le filtrage gaussien. C'est une méthode de lissage rapide et simple à calculer, car le temps de calcul est indépendant de la largeur de la fenêtre.

I.5.1.2. Dérivée

La dérivée est une méthode de prétraitement utilisée pour réduire l'écart par rapport à la ligne de base et de séparer plus clairement les bandes spectrales. Elle permet de faire ressortir certaines informations spectrales. La méthode de dérivée permet également de mettre en exergue certaines informations cachées des spectres. Les dérivées furent les premières méthodes de prétraitement, elles sont utilisées en spectroscopie afin d'éliminer les effets additifs et multiplicatifs dans les spectres. Les dérivées premières et secondes ont été appliquées aux spectres dans cette étude. Elles ont permis de réduire les dérives et les dispersions dans les spectres.

I.5.1.3. Normalisation

La normalisation est une technique de prétraitement de données spectrales souvent utilisée en spectroscopie en amont des analyses statistiques. L'objectif de cette technique est de corriger les spectres par l'élimination de toute variation non désirée du signal spectral. La normalisation élimine ainsi les différences d'intensité dues à la variation de l'intensité lumineuse et à la différence de volume ou de taille des échantillons. De nombreuses techniques de normalisation sont développées et permettent d'explorer les données à savoir la normalisation SNV (Standard Normal Variate), la normalisation MSC (Multiplicative Scatter Correction), la normalisation Min-Max, etc. [72]. La normalisation Min-Max a été appliquée aux spectres dans cette étude.

I.5.2. Méthodes statistiques de traitement

Les méthodes prédictives permettent d'estimer les valeurs d'une ou de plusieurs variables, à partir de données de base. L'intérêt de ces méthodes en spectroscopie est de remplacer les analyses de référence, généralement longues et coûteuses, par des

techniques plus efficaces et rapides. Il s'agit de l'ACP, de la PLS, du réseau de neurones et bien d'autres.

I.5.2.1. Analyse en composantes principales

L'analyse en composantes principales est une méthode d'analyse statistique exploratoire. Elle a pour but de rechercher les liaisons entre les variables afin de faire ressortir les variables portant la même information (corrélées). Cela amène à construire des groupes d'individus homogènes ou la partition des individus. Etudier les individus revient à considérer un nuage de points dans lequel un point correspond à un individu. Ainsi de nouvelles variables appelées composantes principales ou axes principaux sont obtenues. L'ACP est donc utilisée lorsque plusieurs variables quantitatives sont utilisées simultanément et qu'on veut en faire un graphique global. La difficulté vient de ce que les variables étudiées ne sont plus représentées dans un plan (espace de dimension 2) mais dans un espace de dimension plus importante. L'analyse en composantes principales permet ainsi de revenir à un espace de dimension réduite en déformant le moins possible la réalité. Il s'agit donc d'obtenir le résumé le plus pertinent possible des données initiales. Pour ce faire, la matrice des variances-covariances (ou celle des corrélations) obtenue permet de réaliser ce résumé pertinent par analyse de la dispersion des données considérées. Alors, par un procédé mathématique adéquat, les facteurs recherchés sont extraits, en petit nombre. Ces facteurs permettent de réaliser les graphiques désirés dans cet espace de petite dimension en déformant le moins possible la configuration globale des variables initiales. L'interprétation de ces graphiques permet de comprendre la structure des données analysées. Cette interprétation peut être guidée par un certain nombre d'indicateurs numériques et graphiques. Ces indicateurs aident l'utilisateur à interpréter de la façon la plus juste et la plus objective [73]. En effet, l'ACP crée la première composante principale de façon à expliquer le maximum de variance, la seconde composante est contrainte à être orthogonale à la première et d'expliquer la variance résiduelle non prise en compte par la composante précédente et ainsi de suite.

I.5.2.2. Régression PLS

La méthode de régression PLS (Partial Least Square ou Projection to Latent Structure) est une technique statistique essentiellement utilisée dans plusieurs domaines (chimie, pharmacie, alimentation, ...) pour décrire de nombreux problèmes. C'est une technique qui utilise l'algorithme NIPALS (Nonlinear Iterative Partial Least Squares) développé par Herman Wold [74]. Cet algorithme permet de réduire le nombre de variables des prédicteurs en utilisant une méthode semblable à l'analyse en composantes principales pour extraire l'ensemble des composantes qui décrit la corrélation maximale entre les variables de prédiction et de réponse. La PLS est alors utilisée lorsque les variables explicatives présentent une forte corrélation ou lorsque le nombre de variables explicatives dépasse largement le nombre d'observations. La régression PLS met l'accent sur le développement de modèles de prédiction. C'est une technique itérative qui utilise à la fois les caractéristiques de l'analyse en composantes principales et celles de la régression multiple [75, 76, 77]. L'objectif principal de la régression PLS linéaire est de construire un modèle linéaire. Pour faire la régression d'une variable Y sur un ensemble de variables explicatives, la méthode PLS détermine un ensemble de nouveaux facteurs appelés variables latentes qui jouent le même rôle que les variables explicatives initiales. Elle est à présent la méthode la plus utilisée pour les analyses spectroscopiques. Un des avantages de la régression PLS est que l'algorithme d'origine ne nécessite pas l'inversion de matrice. Selon l'algorithme PLS NIPALS, les variables X et les variables Y sont décomposées

I.5.2.3. Réseaux de neurones artificiels

Les réseaux de neurones artificiels (RNA) sont des outils de modélisation de données statistiques non linéaire ou des modèles informatiques d'inspiration biologique. Ils ont été conçus pour simuler la façon dont le cerveau humain traite l'information. C'est une discipline qui a été mise au point dans les années 1890 par le psychologue américain W. James [78] sur le concept de la mémoire associative. Elle s'applique de nos jours dans plusieurs domaines de recherche étant donné son efficacité à résoudre des

problèmes réalistes et complexes [79, 80, 81]. A l'instar du cerveau humain, les réseaux de neurones artificiels sont des systèmes complexes composés d'unités de calcul (ou neurones) regroupées en couches et reliées les unes aux autres par des connexions (synapses) d'intensité variable. Chaque unité reçoit de l'information des unités auxquelles elle est connectée. Aussi, elle calcule son niveau d'activité en intégrant ces informations puis, si son niveau d'activité dépasse un certain seuil, elle transmet l'information aux unités auxquelles elle est connectée. Par conséquent, l'on dispose d'une couche d'entrée recevant les informations externes, d'une ou de plusieurs couche(s) intermédiaire(s) appelées aussi couches cachées car elles ne sont pas en contact avec l'extérieur et enfin d'une couche de sortie. Rappelons qu'il n'y a pas d'interconnexions au niveau d'une couche, car chaque neurone reçoit ses entrées de la couche inférieure.

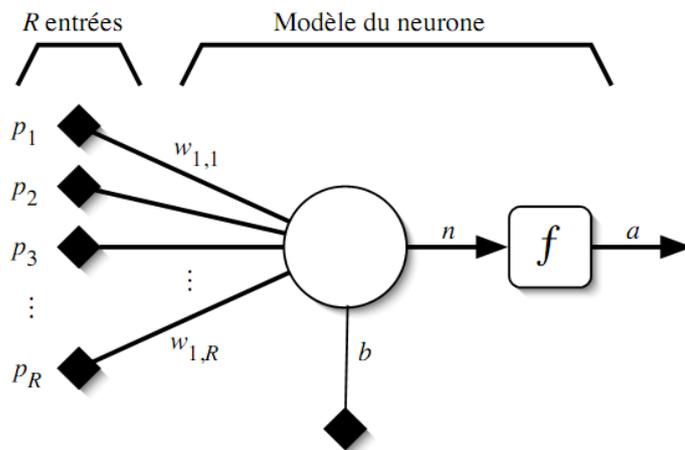


Figure 5 : Schéma d'un neurone formel

La qualité est une notion multicritère et importante pour tous les acteurs du secteur agricole mais aussi pour le consommateur. Elle peut être considérée comme un concept multiple englobant des paramètres physiques, physiologiques, nutritionnels et pathologiques qui affectent la durée de conservation des fruits. La compréhension des notions relatives à la qualité des fruits et à la spectroscopie optique est nécessaire à la compréhension du processus d'analyse de la qualité des fruits. Ce chapitre nous a permis de comprendre la notion de spectroscopie de réflectance optique, la notion de

téledétection et la composition des fruits. Cela est nécessaire pour interpréter les résultats relatifs à la mesure de la qualité des fruits.

CHAPITRE II : MATÉRIEL ET MÉTHODES

Introduction

L'analyse spectrale par télédétection fournit des moyens puissants pour caractériser des échantillons dans une grande variété d'applications. Ce deuxième chapitre va consister à présenter le matériel qui a servi comme outil d'étude pour la détermination des propriétés optiques des fruits tropicaux. Nous avons utilisé plusieurs échantillons dans cette étude à savoir : la mangue, la pomme de cajou, la tomate et la papaye. Nous présenterons donc dans un premier temps les échantillons utilisés pour notre expérimentation. Par la suite, le dispositif expérimental ayant servi de base d'étude pour l'acquisition des spectres de réflectance et de fluorescence sera décrit. La méthode d'acquisition des spectres y sera également décrite. Les données spectrales obtenues à partir de ce dispositif expérimental contiennent des informations qui sont bruitées et cachées, nous avons recours aux méthodes de prétraitement, de traitement de données et d'analyse pour l'extraction des paramètres chimiques.

II.1. Matériel

II.1.1. Echantillonnage

Les échantillons qui ont servi à cette étude sont : les mangues, les tomates, les pommes de cajou et les papayes.



Figure 6 : Photographie des échantillons de fruits utilisés. (a) Mangues camerounaises ; (b) Tomates ; (c) Pommes de cajou ; (d) Papayes.

-Les mangues analysées proviennent de la ville de Yamoussoukro plus précisément du site correspondant aux coordonnées GPS 6°52'57.7"N 5°13'51.3"W. Nous avons utilisé les mangues de la variété *Mangifera indica* appelées communément mangues camerounaises. Le choix de cette variété de mangue a été guidé par le fait qu'elle est de haute valeur nutritive et commerciale. Elle est beaucoup appréciée par la population car elle est très sucrée et contient moins de fibres. Cette variété contient également une importante quantité de chlorophylle au niveau de la peau. Cette quantité peut être considérée comme un indicateur de maturité qui varie très peu au cours de la maturation.

-Les tomates utilisées pour l'analyse proviennent également de la ville de Yamoussoukro. Elles ont été achetées au marché local situé à la Gare INP. Pour cette expérimentation, la variété de tomate *Solanum Lycopersicon L.* du genre Rio Grande de la famille des Solanacées a été utilisée. Cette variété est très cultivée et très appréciée par les consommateurs en raison de la taille des fruits et de leurs caractéristiques organoleptiques. Solide et épaisse, la tomate de la variété Rio Grande peut être stockée pendant une semaine environ. C'est aussi une variété adaptée à toutes les saisons de l'année.

- Les pommes de cajou analysées proviennent d'une plantation située à 15 km de Yamoussoukro aux coordonnées GPS 6°52'14,6N 5°12'14,2W. Pour cette expérimentation, les pommes de cajou ont été choisies car la Côte d'Ivoire est le premier pays producteur d'anacarde au monde, principalement pour la noix de cajou. La pomme de cajou est un coproduit de la noix de cajou. Cependant, la quasi-totalité des pommes de cajou est délaissée à la récolte car elle n'est pas exploitée industriellement et elle est peu consommée à cause de son astringence. Pourtant, elle présente un fort potentiel nutritionnel car elle est très riche en vitamine C, en polyphénols et en caroténoïdes. L'analyse de la qualité nutritionnelle de la pomme de cajou contribuera à la valorisation ce fruit.

-les papayes utilisées dans cette étude ont été collectées dans une plantation située à (6°49'39.8"N 5°15'01.5"W). Les fruits matures et non mûrs ont été récoltés et exposés au laboratoire. Pour l'expérimentation, nous avons choisi les papayes comme échantillons à cause de leur grande disponibilité sur le marché local, de leur importante consommation et de l'extrême finesse de sa membrane externe. De par leurs propriétés (riches en eau et en éléments minéraux), les papayes sont incontournables dans le cadre du suivi d'un régime alimentaire. De plus, elles contribuent à l'économie et est une source de revenus de la grande partie des populations vulnérables à travers le pays [82].

II.1.2. Préparation des échantillons

II.1.2.1. Préparation pour les analyses chimiques

Pour la mesure de la matière sèche ou du taux d'humidité d'un fruit, nous prélevons 5 g de cet échantillon, le plaçons dans une coupelle en papier aluminium. Avant toute mesure, nous étiquetons la coupelle en papier aluminium à vide et la pesons afin d'en connaître la masse à vide m_A .

Concernant les mesures de pH, de l'acidité, de la vitamine C et de la teneur en sucre, nous avons extrait une quantité suffisante de jus du fruit. Afin d'éliminer les éventuelles impuretés qui auraient pu s'y ajouter, les différents jus obtenus ont été filtrés.

II.1.2.2. Préparation pour les mesures spectroscopiques

Avant de réaliser les mesures optiques, nous avons lavé et nettoyé les échantillons afin d'éviter la contribution spectrale d'autres paramètres externes. Ceux-ci ont par la suite été étiquetés et exposés à l'air libre.

II.1.3. Dispositif de mesures spectroscopiques

La figure 7 présente le dispositif de mesure spectrale. Elle montre une vue d'ensemble des différents éléments qui composent le dispositif excepté les sources lumineuses.

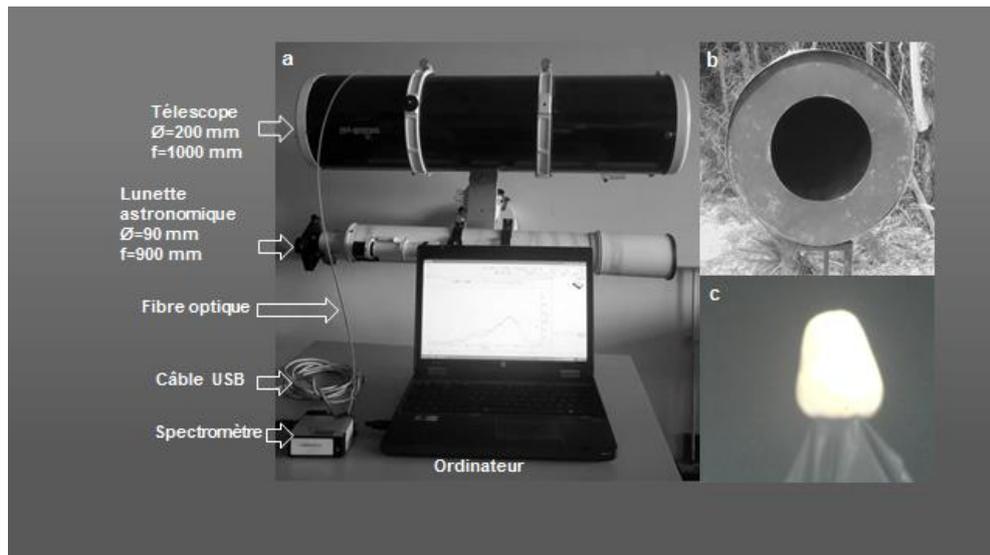


Figure 7: Vue d'ensemble du dispositif expérimental

Le dispositif de mesure des spectres de réflectance et de fluorescence est constitué de deux télescopes (le télescope de Newton et la lunette astronomique), d'un ordinateur portable et d'un spectromètre à réseau Ocean Optics USB4000 FL. Il comporte également une caméra CCD et une fibre optique. Un filtre optique et une source laser sont en plus utilisés pour les mesures de fluorescence.

Le télescope de Newton (Sky-Watcher 200/1000, AZEQ6 GOTO) a pour rôle principal de collecter la lumière réfléchiée ou émise par les échantillons. La quantité de lumière collectée est dépendante du diamètre de son ouverture. En effet, plus le diamètre de l'ouverture est grand, plus il collecte de la lumière pour avoir plus d'informations spectrales. Le télescope de Newton utilisé dans cette étude est un système réflecteur constitué d'un miroir concave (miroir principal) placé au fond du tube, d'un miroir plan (miroir secondaire) permettant de dévier les rayons vers l'oculaire. Le miroir principal collecte la lumière provenant de l'échantillon à analyser et la réfléchit vers le miroir plan incliné à 45° par rapport à l'axe optique du miroir principal. Ce dernier renvoie cette lumière vers l'oculaire pour visualiser l'image.

La lunette astronomique (Bresser Messier AR-90/900 EXOS1/EQ4) couplée au télescope a servi à envoyer le faisceau laser sur les échantillons lors des mesures de

fluorescence. De plus, équipée d'une caméra CCD, elle a contribué à l'alignement du dispositif expérimental et à acquérir les images de chaque échantillon.

L'ordinateur (HP, corei3) a pour rôle de contrôler le processus d'acquisition à l'aide du logiciel SpectraSuite disposant des fonctionnalités de calibrage et d'enregistrement des spectres. Il permet également le stockage des données spectrales et chimiques. Par ailleurs, à l'aide du logiciel Matlab, ces données subiront diverses opérations de prétraitement selon le type d'analyse multivariée pour lequel elles ont été prévues.

Le spectromètre à réseau *USB4000 FL* (figure 8) de la compagnie *Ocean Optics* ayant une réponse spectrale de 345-1042 nm et une résolution optique de 0.22 nm a été utilisé. Il est principalement composé de fentes et d'un système dispersif. Ce système dispersif est formé par un réseau plan et permet de répartir l'intensité de la lumière étudiée en fonction de la fréquence ou de la longueur d'onde. Le spectromètre se compose également de deux miroirs et est contrôlé par le logiciel *SPECTRA SUITE* installé sur l'ordinateur.

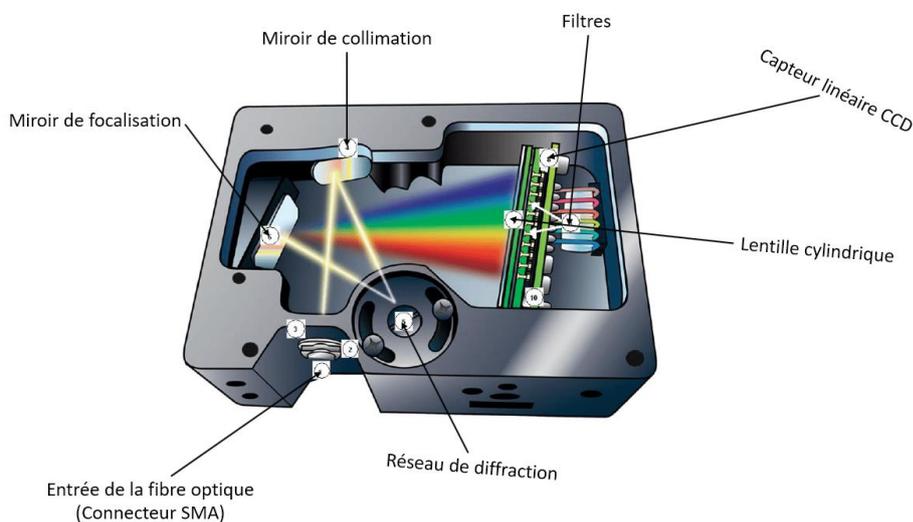


Figure 8: Image annotée du spectromètre

La caméra CCD est utilisée pour visualiser et enregistrer les images de l'échantillon à analyser. Il comporte un capteur CCD qui permet de convertir les photons du signal lumineux qu'il reçoit en signal électrique.

La fibre optique connectée au télescope reçoit le signal collecté (les données de fluorescence et de réflectance) par celui-ci et le transmet au spectromètre. C'est une fibre optique monomode de 50 μm de diamètre.

Le filtre optique passe-bas de longueur d'onde de coupure à 650 nm placé devant la fibre optique permet d'isoler un certain intervalle spectral. Les signaux de fluorescence seront ainsi recueillis. Le laser de longueur d'onde d'excitation 408 nm disposé sur la lunette astronomique est utilisé comme la source lumineuse d'excitation pour obtenir le spectre de fluorescence des échantillons.

II.2. Techniques de mesures

II.2.1. Techniques de mesures chimiques

Des analyses physiques et chimiques ont été réalisées sur les différents échantillons présentés plus haut. Ces analyses concernent : le potentiel hydrogène (pH) ; l'acidité titrable (AT), la matière sèche totale (MST), la teneur en eau et l'extrait sec soluble (ESS) ou teneur en sucre. La teneur en vitamine C a également été déterminée. Ces analyses ont porté aussi bien sur le jus brut filtré que sur la pulpe des fruits.

II.2.1.1. Mesure de la teneur en sucre

La détermination du taux de sucre ou de la teneur en sucre des échantillons a été effectuée au moyen d'un réfractomètre d'Abbe (Novex, Indice de réfraction 0,0002 nD, dispersion 0,0005 nD et échelle BRIX 0,5 %). En effet, une seule goutte du jus de l'échantillon est déposée sur le prisme primaire du réfractomètre (mesure en degrés Brix). On oriente ensuite le réfractomètre en direction d'une source lumineuse pour lire la valeur affichée sur la partie supérieure de l'échelle. Le taux de sucre exprimé en degré Brix n'est rien d'autre que la valeur affichée par cette échelle. L'opération est répétée trois fois sur le même échantillon et une moyenne est calculée par la suite.

La mesure de la teneur en sucre des fruits est fondamentale car elle permet de donner une indication sur l'état de maturité du fruit. Pour une même espèce de fruit, cette teneur en sucre varie d'un fruit à un autre. Du point de vue gustatif, une teneur

minimale en sucre est nécessaire. En ce sens que plus le fruit est sucré, plus il est apprécié par les consommateurs. Le sucre est un paramètre important dans l'analyse de la qualité des fruits du fait qu'il joue un rôle important dans les industries alimentaires et dans l'organisme humain. L'organisation mondiale de la santé (OMS) recommande qu'au maximum 10 % de calories apportées soit sous formes de sucre. Cela équivaut à 25 g de sucre comme apport journalier. Cette dose recommandée exclut le fructose qui est le sucre majoritaire dans les fruits. En effet, le fructose n'est pas nocif pour l'organisme et il a en outre un faible index glycémique.

II.2.1.2. Mesure de l'acidité titrable

L'acidité titrable est la quantité d'acide dans une prise d'essais. Sa détermination consiste à réaliser le dosage de 10 mL de la prise d'essais avec du NaOH 0,1N jusqu'au point équivalent (pH=8,2) en présence de phénolphthaléine. L'acidité titrable des échantillons se détermine par titrage. En effet, il faut préalablement prélever 10 mL du jus filtré de l'échantillon. Ensuite, il faut y ajouter 2 gouttes de phénolphthaléine qui est un indicateur coloré. Puis, il faut remplir une burette de 25 mL avec de l'hydroxyde de sodium (NaOH) 0,1 N. Enfin, il faut verser la soude 0,1 N dans le filtrat prélevé jusqu'au virage au rose de la solution contenant de la phénolphthaléine. Le volume versé de NaOH est par la suite enregistré pour calculer l'acidité titrable comme suit :

$$AT = \frac{N.V}{V_0} \quad \text{II.1}$$

N: normalité de la soude (NaOH) utilisée (mol/L)

V: volume, en millilitres (mL), de la solution d'hydroxyde de sodium versée

V₀: volume, en millilitres (mL), de filtrat prélevé

La méthode utilisée pour la détermination de l'acidité est la méthode de titration AOAC. Au total, trois essais de mesure sont effectués sur chaque échantillon pour ensuite calculer la moyenne des valeurs obtenues qui sera retenue. L'acidité titrable indique la quantité d'acides présents dans le fruit.

L'acidité titrable des fruits est un paramètre important qui permet de déterminer la maturité et le goût des fruits. Elle contribue à la qualité sensorielle des fruits dans le rapport sucre/acide. Le rapport sucre/acide permet d'évaluer l'état de maturité des fruits. Ce rapport varie au cours du développement des fruits et d'un fruit à un autre. CHRISTEN et al. (2011) [3] ont rapporté pour les abricots des valeurs se situant dans la fourchette de 0,37 à 1,24.

II.2.1.3. Mesure du Potentiel Hydrogène

Le principe de la détermination du pH consiste à créer une différence de potentiel dans la solution à analyser au niveau de la double électrode. Avant la détermination du pH des solutions, le pH-mètre est étalonné à la température de lecture avec deux solutions tampons de pH = 4 et pH = 7. Le pH des solutions liquides est déterminé en introduisant l'électrode dans 20 ml de la solution prélevée dans un bécher de 50 ml. Au bout de 5s, la valeur est affichée par le pH-mètre de marque HANNA, HI 8424.

La détermination du pH et de l'acidité des fruits permet d'estimer la conformabilité et les propriétés qui n'apparaissent pas à l'œil nu. On peut les considérer comme des indicateurs de la maturité ou du stade de mûrissement du fruit. Les acides jouent un rôle important dans la qualité du fruit, car la saveur est essentiellement liée à un équilibre entre la teneur en sucre et en acide. L'acidité indique l'état de maturité du fruit. Elle décroît pendant la maturation. Pour une même espèce, les niveaux d'acidité varient en fonction de la variété et de la climatologie de l'année. Le rapport des taux de sucres et d'acidité détermine le caractère doux, équilibré ou acidulé du fruit [83]. Un organisme humain en bonne santé doit présenter un pH équilibré qui se situe autour de 7 la neutralité.

II. 2.1.4. Mesure de la teneur en eau

La mesure du taux d'humidité ou de la teneur en eau des échantillons consiste à déterminer la quantité d'eau contenue dans ces échantillons. Cette mesure implique également la détermination de la teneur en matière sèche du fruit. Elle est effectuée à partir d'une balance numérique et d'une étuve selon la méthode décrite par l'AFNOR

(1981) [84]. Avant toute mesure, il faut étiqueter un récipient vide (une coupelle en papier aluminium), le peser afin de connaître la masse à vide m_A . Il faut ensuite placer 5g d'échantillon dans le récipient et enregistrer la nouvelle masse m_B . L'échantillon est automatiquement conservé dans un dessiccateur pour empêcher qu'il réagisse avec le milieu extérieur qui modifierait ses propriétés. On répète ce mode opératoire sur les autres échantillons soumis à notre analyse. Une fois ces opérations terminées, les échantillons pesés sont portés à l'étuve qui est ventilée à 105°C pendant 24 heures afin de les sécher. Après 24 heures, les échantillons sont sortis de l'étuve et transférés à nouveau dans le dessiccateur. Puis, ils sont refroidis à la température ambiante. A la fin du séchage, chaque échantillon m_C est pesé à nouveau. Le taux d'humidité et la teneur en matière sèche des échantillons sont calculés comme suit :

Masse de l'échantillon humide, m_D :

$$m_D = m_A - m_B \quad \text{II. 2}$$

Masse de l'échantillon sec, m_E :

$$m_E = m_C - m_A \quad \text{II. 3}$$

Taux d'humidité TH :

$$TH = 100 \frac{m_D - m_E}{m_D} \quad \text{II. 4}$$

Teneur en matière sèche, MS :

$$MS = 100 - TH \quad \text{II. 5}$$

Le taux d'humidité et la teneur en matière sèche sont des critères de qualité importants. Ils permettent notamment de déterminer si un accroissement de rendement à la récolte est dû à l'augmentation de la teneur en eau ou à une augmentation véritable de la

masse des fruits récoltés. Les besoins en eau diffèrent en fonction des périodes de la vie. Ainsi, les apports en eau au quotidien pour l'homme sont estimés à environ 2 L par jour. Ces apports incluent l'eau provenant à la fois de la consommation d'aliments et de boissons. En général, la teneur minimale en eau des fruits est estimée à environ 90 % de leur masse totale [52,53].

II.2.1.5. Mesure de la Vitamine C

La vitamine C est un antiscorbutique et un anti-infectieux qui joue un rôle important dans la synthèse du collagène. Elle est synthétisée par de nombreux êtres vivants mais pas par l'homme. Ainsi, il doit le trouver par l'alimentation. Les besoins journaliers de vitamine C de l'homme sont de l'ordre de 75 mg. La vitamine C possède également des propriétés réductrices, c'est-à-dire qu'elle peut réduire le dioxygène. La vitamine C est extraite en présence d'une solution d'acide métaphosphorique/ acide acétique dosée par le 2,6-dichlorophénol indophénol. Pour le dosage, 1 mL du filtrat de l'échantillon est prélevé et dosé avec une solution de 2,6-dichlorophénol indophénol jusqu'à l'apparition d'une coloration rose champagne persistante et un calcul s'en suit en fonction du volume versé.

Les méthodes d'analyses chimiques ont montré leur importance à travers plusieurs applications dans le domaine d'analyse de la qualité des fruits [85, 86]. Toutefois, ces techniques sont destructives et requièrent assez de temps. De plus, elles nécessitent une méthodologie bien élaborée et spécifique dans la préparation de l'échantillon.

II.2.2. Techniques de mesures optiques

Les spectres de réflectance ont été acquis à l'aide du spectromètre USB 4000 Ocean Optics connecté au télescope de Newton. Le spectromètre est piloté par le logiciel Spectrasuite. Ce logiciel présente dans son interface plusieurs modes de mesures à savoir l'absorption, la réflexion, la diffusion et la transmission. Il permet de faire le choix du mode de mesure et de contrôler le flux de photons qui arrivent au détecteur du spectromètre. Le contrôle du flux de photons se fait à travers le temps d'intégration. Plus le temps d'intégration est important, plus le nombre de photons qui arrive au

détecteur est important. Cependant, il faut éviter d'avoir une intensité de photons trop importante. Cela entraînerait une saturation de la caméra et une perte d'informations. Aussi, un temps d'intégration trop faible entraîne également une perte d'informations. Il convient donc d'avoir un temps d'intégration adéquat (environ 07 millisecondes) de sorte à avoir une bonne bande dynamique pour permettre une bonne détection par la caméra. Dans le mode réflexion, cas de notre mesure, on active le mode « Réflexion » indiqué dans l'interface donné par le logiciel.

L'obtention du spectre de réflectance nécessite trois types de mesures :

- Mesure de la référence R_b ;
- Mesure du bruit R_d ;
- Mesure brute de la réflectance de l'échantillon R_s contenant du bruit dû aux réflexions de l'environnement et aux systèmes électroniques.

Ces trois signaux R_b , R_d et R_s sont acquis dans les mêmes conditions :

- Il faut pour la caméra un temps d'intégration identique pour tous les signaux ;
- L'échantillon et le réflecteur doivent être placés à la même position par rapport au dispositif expérimental ;
- L'illumination provenant de l'environnement doit être la même.

Le spectre corrigé R de l'échantillon est alors donné par la relation [87] :

$$R = \frac{R_s - R_d}{R_b - R_d} \quad \text{II. 6}$$

Avant d'effectuer toute mesure, il faut s'assurer de l'alignement du dispositif expérimental. L'alignement du système optique consiste à orienter l'ouverture du télescope vers l'échantillon. Il s'agit de focaliser l'ouverture sur notre objet d'intérêt pour observer au travers de l'oculaire du télescope l'image de l'échantillon. L'observation des images à l'ordinateur se fait au moyen de la caméra. L'alignement

permet de montrer que l'image affichée par l'ordinateur est celle de l'échantillon analysé par le télescope. Les rayons lumineux arrivant à l'oculaire sont transmis au spectromètre par une fibre optique. Le spectromètre permet d'obtenir le spectre de l'échantillon. L'alignement est une étape primordiale de l'acquisition des données, car un mauvais alignement implique immédiatement une mauvaise acquisition des données spectrales.

La mesure de référence R_b est effectuée en mettant un réflecteur à la place de l'échantillon, car le spectre de réflectance se mesure par rapport à un réflecteur parfait. L'utilisation du réflecteur permet une distribution régulière d'intensité lumineuse dans le champ entier de vision de la caméra ou du télescope, de sorte qu'une illumination légèrement irrégulière dans ce champ soit automatiquement corrigée [88]. Le signal R_d de bruit traduit les réflexions dues à l'environnement, au système optique et aux bruits électroniques. Le signal R_s bruité de l'échantillon est mesuré après avoir placé l'échantillon sur le site de mesure.

II.3. Analyses des données

II.3.1. Prétraitement des spectres

Les données spectrales acquises sur les fruits contiennent des bruits. Elles doivent subir certaines corrections ou transformations afin de pouvoir être analysées. Alors l'étape du prétraitement est nécessaire pour éliminer les effets incontrôlés d'intensités, corriger les décalages et faire ressortir certains pics masqués par l'absorption de l'eau [89,90]. L'objectif du prétraitement est donc de réduire ou d'éliminer les informations non pertinentes, tels que les bruits, les incertitudes, la variabilité, les interactions afin d'améliorer le rapport signal/bruit et les résultats de la modélisation (régression PLS, réseaux de neurones, etc.) [91]. Il permet également d'améliorer la qualité visuelle des spectres. Ainsi, à l'aide du logiciel Matlab, nous avons appliqué la normalisation par les valeurs minimales et maximales, le filtre gaussien et les dérivées premières et secondes sur nos spectres.

II.3.1.1. Normalisation par les valeurs minimales et maximales

La normalisation est une méthode de prétraitement couramment utilisée en spectroscopie pour corriger les problèmes de dispersion dans les spectres. Cette méthode consiste à corriger chaque spectre de façon individuelle en supprimant ou en réduisant l'effet lié à la variation incontrôlée de l'intensité spectrale. L'intégralité du spectre est centrée sur sa moyenne et l'ensemble de l'intensité est réajusté grâce à l'écart type du spectre. La normalisation par les valeurs minimales et maximales obéit à l'hypothèse selon laquelle le nombre maximum de photons à chaque longueur d'onde demeure constant tout au long de l'acquisition. Ainsi, chaque spectre est normalisé pour avoir une valeur minimale de zéro et une valeur maximale de un. Dans cette étude, la normalisation utilisée est la normalisation Min/Max décrite par la formule [92] :

$$x_{norm} = \frac{(x - \min x_i)}{(\max x_i - \min x_i)} \quad \text{II. 7}$$

Où $\min x_i$ et $\max x_i$ sont respectivement l'intensité minimale et maximale du spectre x .

II.3.1.2. Filtrage gaussien

Les spectres acquis sont toujours perturbés par du bruit. L'opération de filtrage ou de lissage des spectres a pour but de réduire ce bruit en vue de faciliter une analyse ultérieure. Le filtrage permet également une amélioration de la qualité visuelle des spectres par la suppression des informations inutiles. La méthode de filtrage utilisée dans notre étude est le filtrage gaussien. C'est un filtre passe-bas qui utilise la fonction de densité de probabilité normale implantée dans le logiciel Matlab pour lisser les données spectrales suivant la loi normale ou la fonction de Gauss. Cette fonction a pour expression :

$$f(X, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X-\mu)^2}{2\sigma^2}} \quad \text{II. 8}$$

Où, X , μ et, σ sont respectivement la taille de la fenêtre, la moyenne et l'écart quadratique moyen du spectre.

II.3.1.3. Dérivation

La dérivation est une méthode de prétraitement qui consiste à séparer plus clairement les pics qui se chevauchent. Elle agit comme un filtre passe-haut. La spectrométrie dérivée permet d'améliorer la précision des dosages lorsqu'on est en présence de mélanges de composés dont les spectres sont peu différents. La dérivation des spectres, au sens mathématique du terme, amplifie les faibles variations de pente du spectre de départ (spectre d'ordre zéro). De ce fait, elle permet d'accroître la résolution des données spectrales en faisant ressortir certaines informations spectrales cachées des spectres. En effet, Hruschka (1987) [93] ont montré que les méthodes utilisant la dérivée pourraient améliorer le rapport signal sur bruit et résoudre les chevauchements caractéristiques des spectres. De plus, Rieppo et al. (2012) [94] ont noté que la spectroscopie dérivée au second ordre permet une identification plus précise des faibles pics d'absorption voisins qui ne sont pas résolus dans le spectre d'origine. Cette opération offre ainsi des moyens d'augmenter la spécificité des pics d'absorption pour certaines molécules du tissu.

$$f'(\lambda_i) = \frac{f(\lambda_{i+1}) - f(\lambda_{i-1})}{2\Delta\lambda} \quad \text{II. 9}$$

Où $f(\lambda_i)$ est la réflectance à la longueur d'onde λ_i et $\Delta\lambda$ le pas d'échantillonnage.

II.3.1.4. Rééchantillonnage par Bootstrap

Le bootstrap est une technique statistique de rééchantillonnage qui permet de créer de « nouveaux échantillons » statistiques. Les nouveaux échantillons sont uniquement créés par tirage avec remise à partir du jeu de données initial. Cette technique est donc utilisée pour simuler des échantillons à partir d'un ensemble de données en utilisant la technique de remplacement. Elle permet d'augmenter le nombre d'échantillons d'un jeu de données. Sa désignation anglaise bootstrap est inspirée de l'aspect autocentré et

itératif de la technique. Cette désignation signifie « se hisser en tirant sur ses propres lacets ou plus précisément sur ses « bootstraps ».

II.3.2. Traitement des données

Le traitement des données spectrales a consisté à construire des modèles de prédiction de chaque paramètre de qualité en utilisant la méthode régression PLS et la méthode de réseaux de neurones artificiels. Cela a nécessité l'utilisation du logiciel Matlab pour le développement des modèles de prédiction et leur évaluation.

II.3.2.1. Optimisation des données par sélection de variables

Le développement de modèles robustes et performants nécessite au préalable une optimisation des données. L'optimisation consiste à rechercher dans un jeu de données les informations les plus pertinentes avant la construction de ces modèles. Les informations non pertinentes sont retirées de la base de données. En effet, l'optimisation regroupe toutes les techniques permettant de faire ressortir les données utiles afin d'améliorer la qualité de prédiction. Dans cette étude, nous avons eu recours à trois méthodes d'optimisation des données, tant au niveau des variables spectrales que des échantillons de fruits.

L'ensemble de données spectrales acquises contient de nombreux échantillons avec des milliers de variables. Afin de simplifier la complexité du calcul et d'augmenter la robustesse des modèles de prédiction, la sélection des variables constitue une étape nécessaire et importante. Cette étape permet de sélectionner les variables optimales (informatives) et supprimer les variables non informatives [95]. De nombreuses méthodes basées sur différents critères ont été développées à cet effet mais, pour cette étude, nous avons utilisé les méthodes Monte Carlo et IRIV pour la sélection des variables.

II. 3.2.1.1. Méthode Monte Carlo

La modélisation est une étape clé en analyse spectroscopique. Cette opération a pour but de développer des modèles de prédiction des paramètres de qualité recherchée.

Afin d'élaborer de bons modèles, il est primordial de détecter parmi les valeurs d'une base de données toutes celles qui sont aberrantes et de les supprimer par la suite de la base de données avant d'établir les modèles. Cela permettra l'amélioration des performances prédictives des modèles. La méthode de détection Monte Carlo utilise le jeu de données initial afin d'étudier la distribution des erreurs de prédiction de chaque échantillon à partir de la validation croisée de Monte Carlo [96]. La figure 9 permet de présenter les différentes étapes suivies par de la méthode MC.

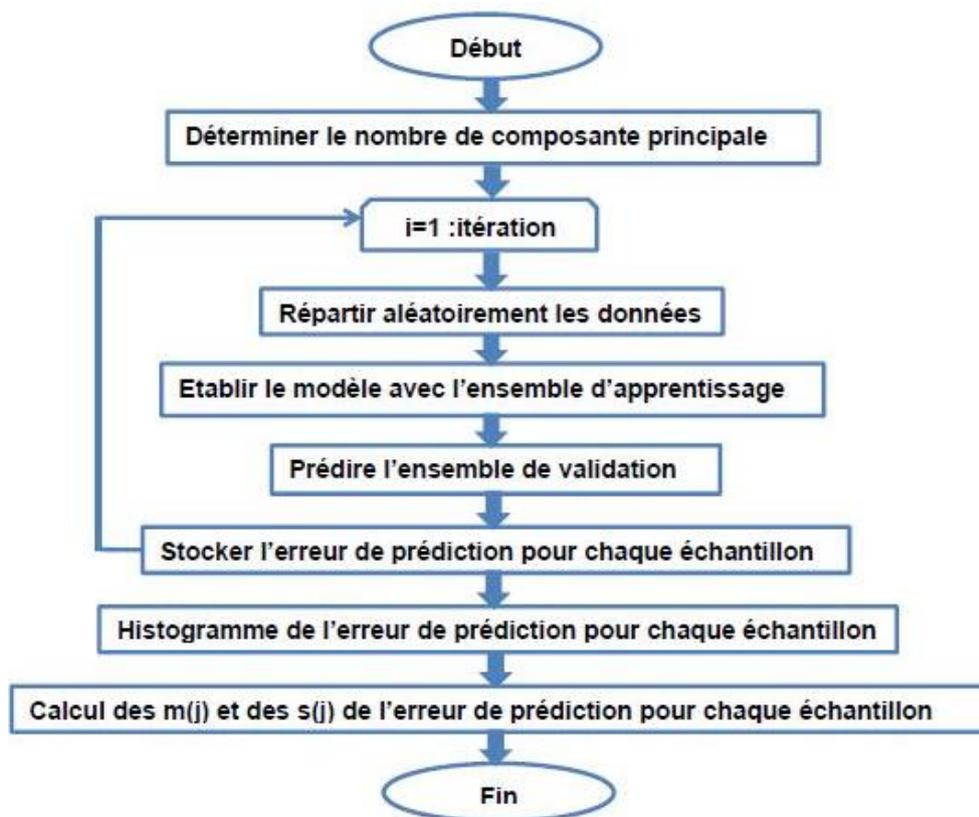


Figure 9: Organigramme de l'algorithme MCO [96].

L'algorithme détermine d'abord le nombre de composants principaux en utilisant la validation croisée dans les méthodes PLS et PCR. Ensuite, l'algorithme divise aléatoirement les données entières en deux parties qui sont : ensemble d'entraînement et ensemble de validation, respectivement. D'une manière générale, les données de l'apprentissage constituent les 2/3 de l'ensemble des données. Comme pour les autres modèles, l'ensemble d'apprentissage est utilisé pour établir le modèle en utilisant la

composante principale déterminée. L'ensemble de validation est utilisé pour la prédiction et une erreur de prédiction est obtenue pour chaque échantillon de validation ($error(i)$). L'algorithme exécute ce cycle autant de fois selon un nombre prédéterminé égal à 1000, de façon à obtenir la répartition des erreurs de prédiction pour chaque échantillon [96]. Les histogrammes de ces distributions ont été tracés et leurs caractéristiques statistiques ont été utilisées pour détecter les valeurs aberrantes. Ensuite, la moyenne $m(j)$ et l'écart-type $s(j)$ de la distribution des erreurs sont employés pour caractériser chaque échantillon.

$$m(j) = \frac{1}{k} \sum_{i=1}^k error(i) \quad \text{II. 10}$$

$$s(j) = \left(\frac{1}{k-1} \sum_{i=1}^k (error(i) - m(j))^2 \right)^{1/2} \quad \text{II. 11}$$

II.3.2.1.2. Méthode IRIV de sélection de variables optimales basée sur l'échantillonnage par Matrice Binaire

La méthode IRIV est une technique utilisée en vue de sélectionner les variables optimales de la base de données spectrales dans cette étude. Pour ce faire, la méthode considère une matrice X de données contenant N échantillons en ligne et P variables en colonne et une autre matrice y de dimension $N \times 1$ désignant la grandeur d'intérêt. Ainsi, l'échantillonnage par matrice binaire consiste à générer une matrice binaire M de dimension $K \times P$. A cette matrice binaire, on attribue de manière aléatoire le nombre '1' ou '0' à chaque colonne. Chaque ligne de la matrice binaire correspond à un échantillonnage individuel, où '1' indique les échantillons sélectionnés pour la modélisation et '0' pour les échantillons non sélectionnés. On obtient ainsi le même nombre de "1" et de "0" dans chaque colonne.

La permutation sur la matrice M permet de supprimer les éventuelles correspondances linéaires entre les réponses et les observations (Figure 10 A). Par le processus d'inclusion et d'exclusion des valeurs binaires '1' ou '0', la matrice M_1 est obtenue en changeant pour la première colonne tous les '0' en '1' et inversement (Figure 10 B).

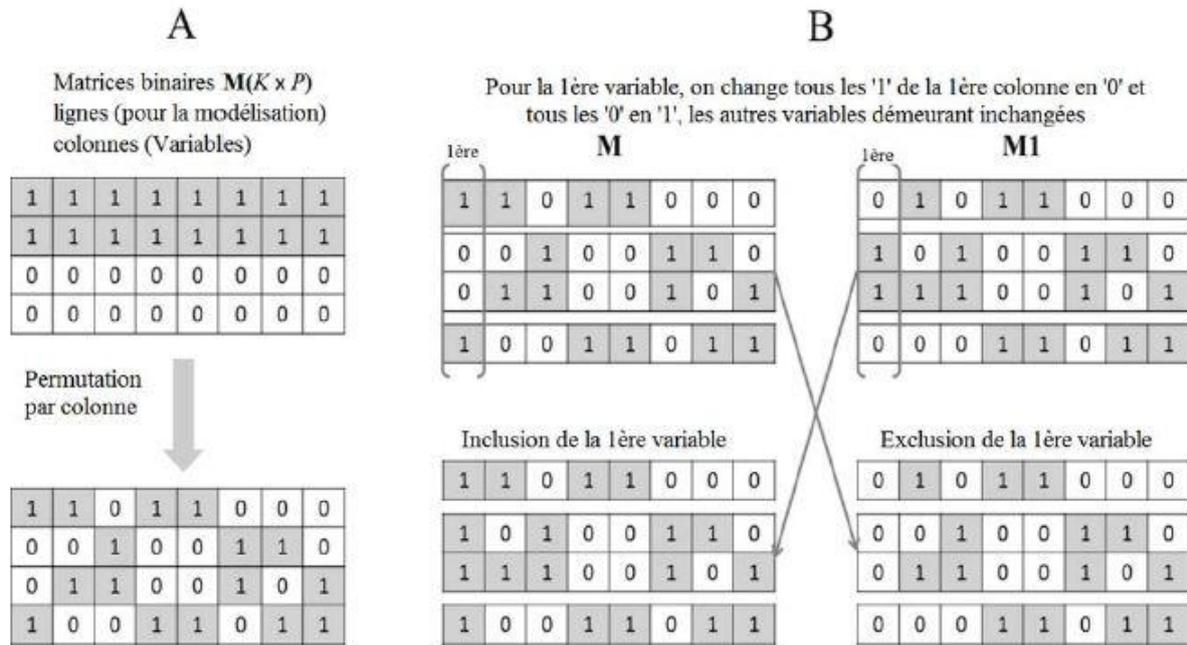


Figure 10: Echantillonnage par matrice binaire. (A), (B) Comparaison des performances de l'inclusion et de l'exclusion d'une variable [97].

Un sous-ensemble (variable i ème colonne) est ainsi évalué par PLS et une distribution statistique des RMSECV à l'inclusion et à l'exclusion est obtenue. Cette distribution statistique des RMSECV pour une variable est testée avec l'hypothèse selon laquelle la distribution des données est la même pour deux groupes afin d'identifier l'importance de cette variable. Le test non paramétrique U de Mann-Whitney est utilisé à cet effet [98]. Selon la valeur de la probabilité critique obtenue, les variables sont classées en quatre catégories : fortement informatives, faiblement informatives, non informatives et interférentes. Après le retrait des deux derniers types de variables, un nouveau sous-ensemble est généré et le même cycle de calcul est réitéré. A la fin de l'itération, lorsqu'aucune variable non informative, ni interférente n'existe, une procédure d'élimination en arrière-plan est conduite pour donner en sortie les variables optimales retenues. L'organigramme de l'algorithme IRIV est décrit à la figure 11.

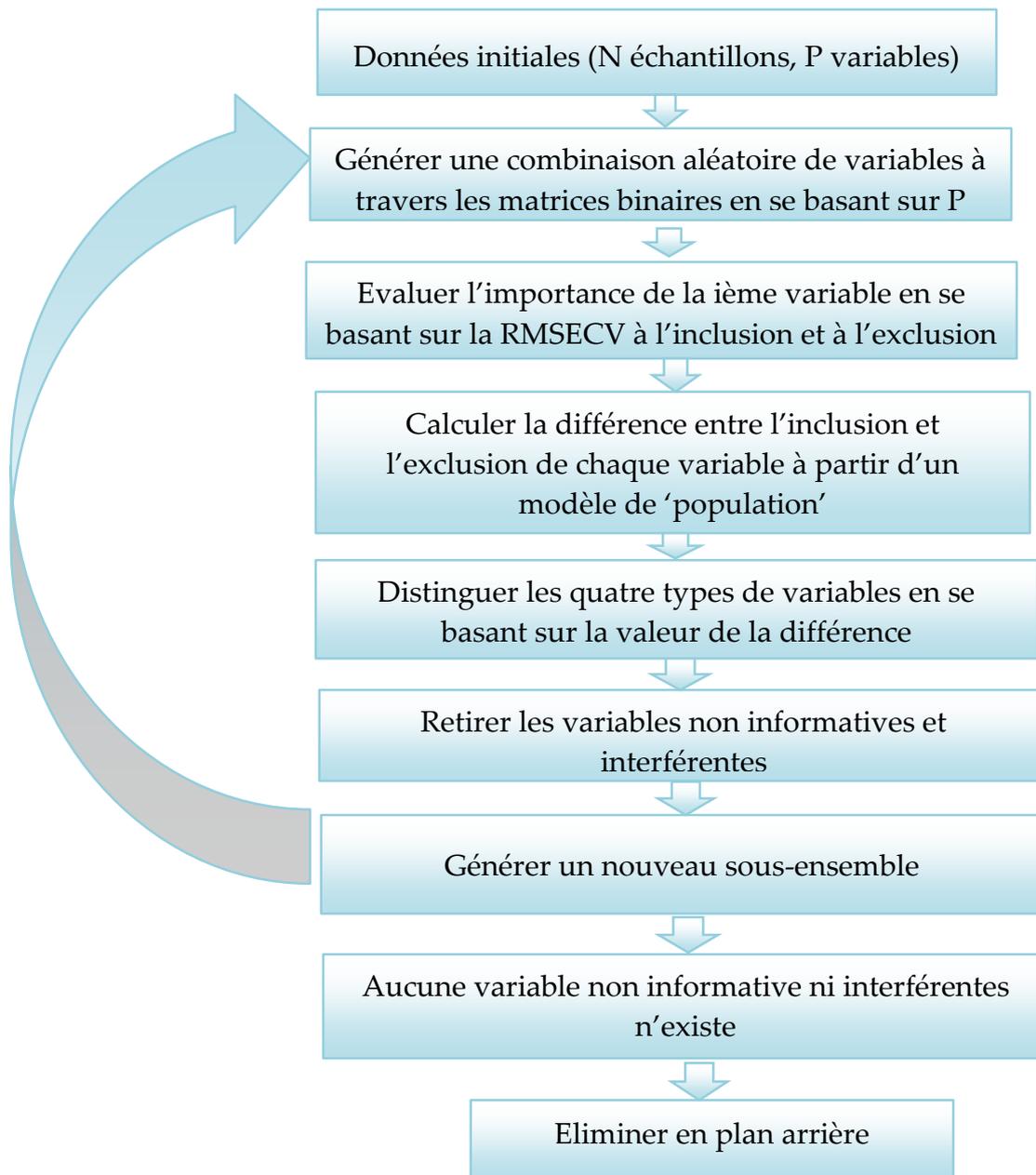


Figure 11: Organigramme de l'algorithme IRIV [97].

L'étude des méthodes multivariées de sélection de variables, selon les travaux de Yuna et al. (2014) [97] a démontré que l'algorithme IRIV possède les meilleures performances devant trois (03) autres méthodes de sélection à succès utilisant les algorithmes génétiques (GA), l'élimination des variables non informatives par la méthode Monte Carlo (UVE-MC) et l'échantillonnage adaptatif compétitif repondéré (CARS). Cependant, son usage dans nos travaux nous a permis d'observer un long temps

d'exécution (3 à 4 jours) sur un ordinateur de bureau intel®Core (TM) i3-4160 3.6 GHz CPU et 16 GB de RAM avec le système d'exploitation Microsoft Windows 10.

II.3.2.2. Partitionnement des données par la méthode Kennard-Stone

La méthode Kennard et Stone (KS) est un algorithme de partitionnement de données. Elle offre une solution pratique pour diviser nos données en un ensemble d'entraînement et en un ensemble de test. Le principe de l'algorithme consiste à extraire un sous-ensemble de N points d'un ensemble de N_c représentatif uniformément réparti dans l'espace expérimental [99]. A chaque itération de l'algorithme, il retient le point le plus distant de l'ensemble des points déjà sélectionnés.

1. Considérer un ensemble de N_c points candidats dans l'espace à D dimensions ;
2. Calculer la matrice des distances euclidiennes de l'ensemble des points candidats comme suit :

$$d_{ij} = \|x_i - x_j\| = \sqrt{\sum_{r=1}^D (x_{ir} - x_{jr})^2} \quad \text{II. 12}$$

d_{ij} : distance euclidienne entre les points i et j ;

3. Choisir les points I et J tels que :

$d_{ij} = \max(d_{ij})$ jusqu'à ce que $N = N_c$, $N=N_c$ avec le nombre de points souhaités

- (a) Calculer les distances des $(N_c - N)$ points restants par rapport aux N points choisis et retenir la valeur minimale :

$$\Delta_i(N) = \min\{d_{i1}, d_{i2}, \dots, d_{iN}\}$$

Avec $\Delta_i(N)$ la distance du point candidat i non encore dans la matrice des points sélectionnés au point le plus proche dans les points sélectionnés.

(b) Pour le $(N + 1)^{ème}$ point de la matrice, choisir parmi les $(N_c - N)$ points candidats restants celui pour lequel :

$$\Delta(N + 1) = \max\{\Delta_i(N)\}$$

pour que le nouveau point soit le plus éloigné des points déjà sélectionnés.

II.3.2.3. Modélisation par analyse multivariée

En générale, la modélisation vise à extraire les informations dans un jeu de données en construisant des modèles de prédiction reliant les données spectrales aux données chimiques. Il s'agit donc dans le cadre de notre travail d'élaborer des modèles de prédiction des paramètres de qualité. Ces paramètres concernent le sucre, le pH, l'acidité titrable, la matière sèche, l'eau et la vitamine C. Les méthodes de régression des moindres carrés partiels (PLSR) et réseau de neurones artificiels sont les techniques utilisées pour réaliser cette modélisation.

II.3.2.3.1. Régression PLS

II.3.2.3.1.1. Principe

Dans de nombreuses situations où l'on possède des variables X sur lesquelles on peut plus ou moins agir et des variables Y que l'on ne peut qu'observer, la régression PLS intervient. Comme les autres méthodes d'analyse, l'objectif de la PLS est alors de décrire les relations entre Y et X , en l'absence de modèle théorique. Le problème est que le nombre de variables X est souvent très important par rapport au nombre d'observations. La régression PLS se présente comme une méthode adaptée pour ce type de problème. C'est une méthode itérative qui généralise et combine les caractéristiques de l'analyse en composantes principales et la régression multiple [75, 76, 77]. Elle met l'accent sur la prédiction de la réponse et pas nécessairement sur la mise en évidence d'une relation entre variables. Pour faire la régression d'une variable y sur un ensemble de variables explicatives, la méthode PLS détermine un ensemble de nouveaux facteurs qui jouent le même rôle que les variables explicatives initiales. Ces nouveaux facteurs sont nommés variables latentes. Chaque variable latente est une combinaison des variables explicatives de départ. La particularité de cette méthode

par rapport à la méthode de régression sur les composantes principales (RCP) est qu'elle utilise non seulement les corrélations entre les variables prédictives mais aussi les corrélations entre les variables de réponse. Le modèle linéaire général construit par la régression PLS est :

$$Y = BX + E \quad \text{II.13}$$

où Y est une matrice de n observations par m variables de réponse, X est une matrice de n observations par p variables prédictives (plan), B est une matrice de coefficients de régression p par m, et E est le terme d'erreur du modèle de même dimension que Y. D'une manière générale, les variables de X et Y sont centrées en retranchant leurs moyennes, et réduites en les divisant par leurs écarts-types.

II.3.2.3.1.2. Algorithme de la régression PLS

L'algorithme recherche de façon successive les i composantes latentes, combinaisons linéaires des variables de départ $C_i = Xw_i$, liées avec y, et les utilise ensuite pour modéliser 'y' linéairement. On recherche ainsi le maximum des covariances entre les nouvelles composantes et les variables à observer :

$$\max w_i \{cov(y, C_i = Xw_i)\} \quad \text{II.14}$$

Sous les contraintes :

$$w_i^t w_i = \| w_i \|^2 = 1$$

$$C_i^t w_l = 0, \text{ pour } l \in \{1, 2, \dots, (i-1)\}$$

Pour un ensemble de variables, $X = (x_1, x_2, \dots, x_p)$ la première composante latente se présente comme suit:

$$C_1 = Xw = w_{11}x_1 + w_{12}x_2 + \dots + w_{1p}x_p = \sum_{j=1}^p w_{1j} x_j \quad \text{II.15}$$

$$\text{Avec } w_{1j} = \frac{cov(x_j, y)}{\sqrt{\sum_{k=1}^p cov^2(x_k, y)}}$$

A partir de cette première composante latente (C_1), la régression simple pour y est de la forme :

$$y = \alpha_1 c_1 = c_1 w_{11} x_1 + \dots + c_1 w_{1p} x_p + y^{(1)} \quad \text{II. 16}$$

La construction de la composante suivante, c'est-à-dire la deuxième, se fait par la détermination de c_2 , qui est la combinaison linéaire des variables de X non corrélées à c_1 , expliquant au mieux le résidu de $y, y^{(1)} = y - \alpha_1 c_1$. On répète cette procédure jusqu'à l'obtention de toutes les composantes (variables latentes).

II.3.2.3.2. Réseau de neurones artificiels

On peut définir le réseau de neurones artificiels comme un système de calcul inspiré du fonctionnement des neurones biologiques. Il consiste en un système composé de plusieurs unités de calcul simple bien organisées fonctionnant en parallèle [100,101]. Il est beaucoup plus proche des méthodes statistiques. La figure 12 présente la structure du réseau de neurones de notre étude.

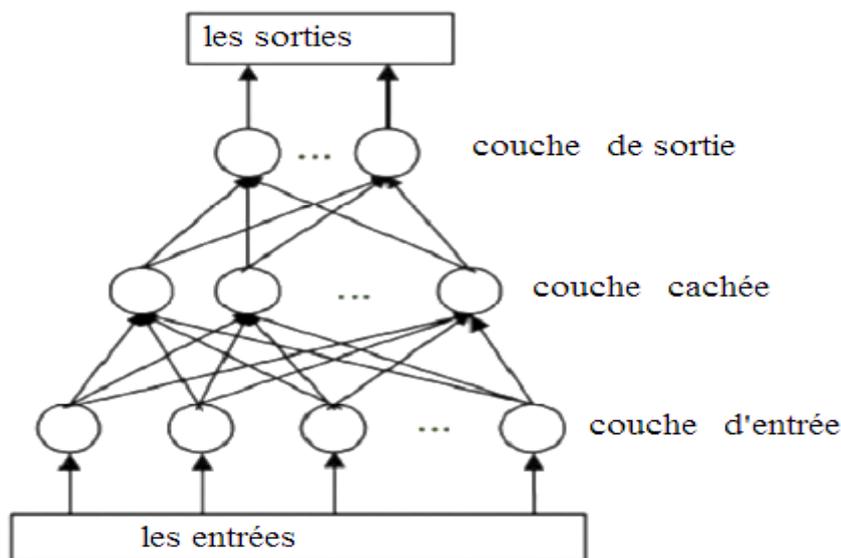


Figure 12: Structure du réseau de neurones [102].

Il est composé de trois couches de cellules ou neurones que sont :

- la couche d'entrée : en tant que première couche du réseau, la couche d'entrée a une propriété de réception des données soumises à l'analyse. Elle reçoit ce que le chercheur veut analyser.

- la couche cachée : la seconde couche dite couche cachée constitue la couche intermédiaire entre la couche d'entrée et la couche de sortie. Elle n'a pas de contact direct avec l'extérieur.

- la couche de sortie : elle donne le résultat obtenu après compilation par le réseau, des données d'entrée dans la première couche. Le nombre de réponses que l'on souhaite en sortie définit la taille de cette couche.

Ces différents neurones sont interconnectés. Ces interconnexions sont traduites par des fonctions de transfert élémentaires appelées fonctions d'activation. Cela permet aux neurones de disposer d'un canal pour recevoir et envoyer des signaux en provenance d'autres cellules du réseau [102]. Quatre principales étapes résument le principe de fonctionnement du RNA : la préparation des échantillons, l'élaboration de la structure du RNA, l'apprentissage et la validation suivie du test. La première étape, c'est-à-dire la préparation des échantillons consiste dans le choix du type de réseau approprié pour résoudre son problème. Elle est très importante car le succès de toutes les autres étapes en dépend. En clair, le type de réseau, le nombre de cellules d'entrée tout comme le nombre de cellules de sortie et la façon dont il faut mener l'apprentissage, les tests et la validation sont tributaires de la présentation même de l'échantillon.

L'étape de l'élaboration de la structure du réseau, comme nous l'avons signifié plus haut, est strictement dépendante du type des échantillons. En effet, il existe plusieurs types de modèle de réseau : le réseau de Hopfield, le perceptron standard, le réseau de Kohonen et le réseau à décalage temporel (TDNN). Il convient d'en choisir un modèle qui permettra une maîtrise de son sujet d'étude. Dans le modèle du perceptron (exemple de notre étude), le choix du perceptron nous impose de déterminer le nombre de neurones dans la couche cachée. Du coup nous atteignons l'étape de l'apprentissage.

La phase d'apprentissage est une étape importante. Elle consiste à la construction même du réseau de neurones. A cette étape, le réseau est ajusté par les valeurs appelées

les poids jusqu'à obtenir un réseau stable. Les algorithmes d'apprentissage se regroupent en deux types d'apprentissage ou grandes familles : l'apprentissage supervisé et l'apprentissage non supervisé. L'apprentissage supervisé consiste à estimer une fonction qui traduit la relation entre les éléments et leurs classes. Dans l'apprentissage non supervisé, les données sont regroupées par rapport à des éléments ayant des propriétés identiques. Dans le cadre de notre travail, nous avons utilisé l'apprentissage supervisé. Il estime premièrement les poids optimaux des différentes liaisons, en utilisant un échantillon. Lors de l'apprentissage du réseau, des valeurs dites observations sont présentées aux neurones d'entrée. Ces valeurs sont compilées pour obtenir une ou des nouvelles valeurs « S » en sortie. Une comparaison est faite entre S et Y la réponse désirée. En fonction de l'écart entre S et Y (erreur), à la fin de l'apprentissage, on ajuste le réseau en corrigeant les poids accordés aux pondérations. On répète ce processus, jusqu'à obtenir une erreur plus faible. Un réseau de neurones surentraîné devient moins performant. C'est pourquoi, il faut faire attention à ne pas le surentraîné. Ce processus est basé sur la minimisation d'une fonction de coût par un algorithme adaptatif de type gradient. La fonction de coût souvent utilisée est l'erreur moyenne quadratique.

Lors de l'apprentissage du réseau, l'ajustement des poids permet d'augmenter le coefficient de corrélation R tout en minimisant les erreurs quadratiques moyennes MSE et standard $RMSE$ [103]. Ces paramètres évaluent la performance du réseau de neurones et se définissent ainsi :

$$R = \frac{\sqrt{\sum_{i=1}^N (V_{\text{exp},i} - \bar{V}_{\text{exp}})^2 - \sum_{i=1}^N (V_{\text{cal},i} - V_{\text{exp},i})^2}}{\sqrt{\sum_{i=1}^N (V_{\text{exp},i} - \bar{V}_{\text{exp}})^2}} \quad \text{II.17}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (V_{\text{exp},i} - \bar{V}_{\text{exp}})^2} \quad \text{II.18}$$

$$MSE = \sqrt{\frac{\sum_{i=1}^N (V_{\text{exp},i} - \bar{V}_{\text{exp}})^2}{N-1}} \quad \text{II.19}$$

Où N est le nombre d'observations (d'expériences), $V_{\text{exp},i}$, $V_{\text{cal},i}$ et \bar{V}_{exp} sont respectivement la valeur expérimentale, la valeur calculée par le réseau et la valeur expérimentale moyenne [103]. Dans cette étude, les fonctions d'activation tangente hyperbolique et fonction linéaire sont utilisées respectivement pour la couche cachée et pour la couche de sortie. Nous arrivons à l'étape de la validation. En effet, après l'apprentissage du réseau ou l'élaboration d'un modèle par le réseau, il doit être testé afin de vérifier son bon fonctionnement ou ses performances et ainsi procéder à sa généralisation (c'est le test). Cette dernière étape va consister à vérifier les performances ou la capacité de généralisation du réseau ou du modèle en introduisant d'autres échantillons dans le modèle établi afin de vérifier ses performances de prédiction.

II.3.3. Evaluation de la performance d'un modèle

L'évaluation de la performance d'un modèle consiste à tester la capacité du modèle à restituer les données utilisées pour son estimation mais aussi son aptitude à prédire de manière précise de nouvelles observations, non utilisées lors de la construction du modèle. Les procédures de validation de modèle sont cruciales pour évaluer l'exactitude de l'étalonnage et pour éviter tout surajustement. Afin d'évaluer quantitativement les modèles développés, les différents paramètres statistiques ci-dessous ont été calculés à savoir :

- le coefficient de détermination

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad \text{II.20}$$

- la racine carrée de l'erreur quadratique moyenne

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad \text{II. 21}$$

- le ratio de performance de déviation (RPD)

$$RPD = \frac{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}_i)^2}}{RMSE} \quad \text{II. 22}$$

où y_i est la valeur observée d'une série à prédire, \hat{y}_i , la valeur prédite par le modèle pour une valeur observée et \bar{y}_i , la moyenne de la valeur observée.

Les critères de Malley *et al.* (2004) [104] basés sur le coefficient de corrélation et le RPD des données de validation, nous ont permis de classer les modèles que nous avons développés. En général, selon ces critères, un coefficient de corrélation et un RPD élevés, ainsi que des valeurs de RMSEC, de RMSEP et de biais faibles indiquent que le modèle est bon [104].

Cette partie de notre rédaction nous a permis de présenter les échantillons utilisés pour notre étude à travers leur mise en place et leur préparation avant l'acquisition des données. Le dispositif expérimental utilisé dans ce travail a par la suite été décrit et les protocoles d'acquisition des données ont été présentés. La résolution du problème du bruit et d'effets indésirables dans les données a été faite grâce aux techniques de prétraitement de données. Les différentes techniques d'analyse de données et de modélisation ont été passées en revue. En outre, un accent a été particulièrement mis sur celles qui interviennent dans ce travail. Les méthodes statistiques multivariées utiles dans la réduction de la redondance d'informations ont été décrites. L'usage de tous ces outils nous a permis d'aboutir aux résultats que nous présentons dans le chapitre suivant.

CHAPITRE III : RÉSULTATS ET DISCUSSION

Introduction

Ce chapitre est consacré à la présentation et à la discussion des résultats des analyses effectuées avec notre dispositif. Les spectres des échantillons présentés dans cette partie correspondent aux résultats des analyses réalisées. Ils comprennent les spectres des mangues, les spectres des tomates, les spectres des pommes de cajou et les spectres des papayes. Les informations à l'intérieur de ces spectres sont celles des composants chimiques et sont caractérisées par une répartition des intensités des longueurs d'ondes qui les représentent. L'objectif de ce travail est de développer des méthodes d'analyse optique de la qualité des fruits utilisant la spectroscopie de réflectance et de fluorescence par télédétection dans le domaine du visible.

Les études consisteront dans un premier temps à effectuer des mesures spectrales sur les échantillons et dans un second temps à effectuer des analyses chimiques sur ces mêmes échantillons pour déterminer les valeurs de paramètres chimiques. Nous avons utilisé le dispositif de mesure des spectres décrit dans le chapitre II pour acquérir nos données spectrales. Les données spectrales et chimiques acquises sont affectées de bruits qu'il faut obligatoirement réduire ou éliminer afin d'améliorer les résultats de prédiction. L'analyse des données passera donc au préalable par une phase de prétraitement. Pour ce faire, les méthodes de prétraitement que nous avons appliquées aux données spectrales sont : la normalisation, le filtrage et la dérivation. Nous avons eu recours à deux méthodes de sélection de variables telles que la méthode Monte Carlo et la méthode IRIV pour éliminer certaines variables non informatives. La méthode Kennard-Stone (KS) a été utilisée pour partitionner nos données en deux groupes. Un groupe servira à la construction des modèles de prédiction par la régression PLS et l'autre à la validation de ces modèles. La méthode de rééchantillonnage bootstrap a été utilisée pour augmenter le nombre d'échantillons d'un jeu de données.

Nous présenterons et discuterons également les résultats obtenus.

III.1. Estimation de la qualité des mangues par réflectance optique

III.1.1. Collecte des mangues, mesure et traitement des spectres de réflectances

Pour cette étude, soixante (60) échantillons de mangues ont été directement récoltés sur un manguier situé à l'INP-HB centre. Pendant la récolte, des précautions ont été prises afin d'éviter l'endommagement de ces fruits. Une fois les mangues transportées au laboratoire, celles-ci ont été triées, lavées et nettoyées pour éviter toute contribution externe dans les résultats. Sur chaque fruit, un chiffre identifiant a été marqué afin de les distinguer. Les mesures de réflectance et les analyses chimiques ont été effectuées sur la portion non marquée de chaque échantillon pour éviter la présence d'artefacts dus à la contribution des marqueurs.

Nous avons acquis les spectres optiques de réflectance de mangues non mûres et mûres à l'aide du dispositif décrit au chapitre II. Les mesures de réflectance ont été effectuées sur chaque mangue à raison de deux mesures sur une même surface. Nous avons réalisé ces mesures sur trois (3) jours successifs. La méthodologie de mesure du spectre de réflectance est décrite au chapitre II.

Une fois la séance de mesures spectroscopiques terminée, les échantillons ont été transportés au laboratoire de chimie (LAPISEN) pour des analyses chimiques (la détermination du pH, de la teneur en eau et de la teneur en sucre).

Les spectres optiques acquis sont affectés par des bruits provenant des systèmes optiques utilisés et des variations incontrôlées de l'intensité lumineuse. Ces bruits se superposent aux spectres attendus. Pour réduire ou éliminer la contribution spectrale de ces bruits, nous avons eu recours à un filtrage gaussien.

Pour rehausser la qualité des pics noyés dans le bruit, nous avons utilisé les dérivées premières et secondes du spectre de réflectance. Les valeurs aberrantes qui se sont introduites dans les mesures chimiques ont été identifiées par la méthode Monte Carlo afin d'améliorer la performance des modèles de prédiction. La méthode Monte Carlo pour la recherche des valeurs aberrantes consiste à calculer les erreurs de prédiction

des paramètres de qualité des échantillons et à étudier la distribution des erreurs de prédiction de chaque échantillon de la base de données. Après le prétraitement des données, nous avons appliqué la méthode de régression PLS pour établir une correspondance entre les mesures chimiques et les mesures physiques.

III.1.2. Résultats et discussion

III.1.2.1. Photo des échantillons

Nous avons pris en photo les échantillons en vue de faire une analyse visuelle.

La figure 13 présente la photo d'un lot d'échantillons de mangues non mûres et mûres disposées ensemble. Nous pouvons observer des mangues qui paraissent visuellement toutes vertes. Il est difficile de distinguer le stade de maturité des mangues (mûres et non mûres) par une simple analyse visuelle. La dégradation de la teneur en chlorophylle de la peau de ces fruits est très faible pour refléter le stade de maturité. Cela signifie qu'une analyse sensorielle simple ne peut pas nous permettre d'évaluer la qualité de ces mangues d'où la nécessité d'employer des méthodes de mesures non destructives.



Figure 13 : Photographie des mangues.

III.1.2.2. Spectres de réflectance optique des échantillons de mangues

La figure 14 présente les spectres de réflectance de trois échantillons de mangues non mûres et de trois échantillons de mangues mûres du lot de fruits.

L'observation de la figure 14 montre des spectres qui présentent les mêmes profils et les mêmes allures sur toute la plage de longueurs d'ondes quel que soit le stade de maturité. Cette similarité au niveau du profil spectral des spectres révèle donc la présence de constituants communs aux fruits en dépit de la différence liée à leur stade de maturité. Toutefois, une différence est enregistrée au niveau des intensités relatives des pics de réflectance et d'absorption à certaines longueurs d'ondes pour chaque spectre. En effet, l'intensité relative des pics de réflectance est maximale dans certaines régions spectrales et minimales dans d'autres régions. Cela s'explique par les concentrations des constituants qui diffèrent et qui varient au cours de la maturation des fruits [105]. La maturation du fruit conduit à des transformations physiques et chimiques au niveau du fruit. Aussi, cette différence au niveau des intensités peut se justifier par le fait que les constituants (sucre, pH, eau, chlorophylles, caroténoïdes, etc..) sont séparément identifiables par leur propre signature spectrale et selon la concentration.

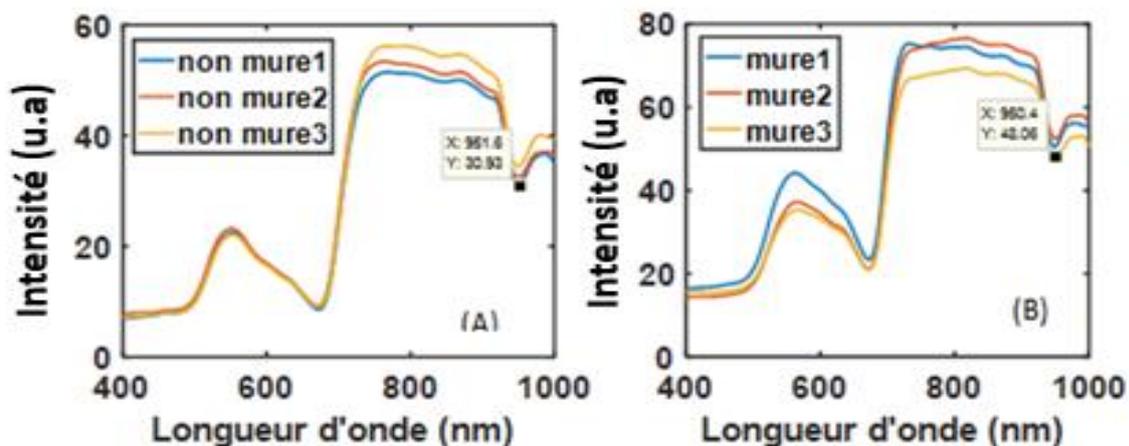


Figure 14: Spectres moyens de trois échantillons de mangues non mûres (A) et trois échantillons de mangues mûres (B).

Nous observons également sur les courbes des réflectances et des absorptions à certaines zones spectrales bien spécifiques (500-650 nm, 650-680 nm et 900-952 nm). En ce qui concerne les réflectances, un pic maximal et deux plateaux sont observés. Le pic maximal de réflectance est observable autour de 550 nm. Le premier plateau est situé entre 780 nm et 950 nm. Le second plateau à peine visible s'étend de 952 nm à 1000 nm. En général, un pic de réflectance élevé traduit un maximum de réflectance à cette longueur d'onde par un composant organique spécifique. Ainsi, le pic de réflectance visible autour de 550 nm s'explique par la réflexion de la lumière par les pigments en particulier les caroténoïdes et les pigments chlorophylliens [106]. Les caroténoïdes sont dominés par les pigments chlorophylliens au début de la maturation. Les longueurs d'ondes du vert expliquant la coloration verte de ces fruits sont réfléchies par les pigments chlorophylliens. Au cours de la maturation, la diminution du pic de réflectance dans la région de 550 nm est liée aux variations des quantités de caroténoïdes et de chlorophylle. La concentration en caroténoïdes des fruits augmente avec leur mûrissement (stade de maturité). Par contre, on assiste à une légère dégradation de teneur en chlorophylle des mangues [107].

Le plateau de réflectance observé entre 780 nm et 900 nm traduit une contribution à quantité presque égale de nombreux composants organiques. Ces composants responsables du plateau de réflectance pourraient être les éléments constituant de la structure cellulaire du fruit. L'observation des figures A et B montre que la réflectance dans cette région augmente avec la maturation des fruits.

L'autre plateau situé entre 952 nm et 1000 nm est moins important. Ce plateau pourrait être lié à la structure cellulaire du fruit. L'eau étant le composant le plus abondant dans les fruits, les spectres des fruits dans cette région sont dominés par son absorption [108].

Quant aux absorptions observées, elles se situent dans trois régions spectrales distinctes. La première est forte et se situe entre 400-500 nm. Cette forte absorption montre une importante interaction de cette bande spectrale avec des composants

chimiques des mangues. Les caroténoïdes sont les pigments responsables de l'absorption dans cette région spectrale. Les travaux de Governici *et al.* (2017) [109] menés sur des tomates ont identifié des bandes d'absorption similaires. Le comportement spectral dans cette région est fortement influencé par la présence de chlorophylles et de caroténoïdes qui varient avec la maturation des fruits.

La seconde absorption identifiée autour de 675 nm est due à la coloration verte des fruits. Cette forte absorption est due à l'influence de teneur en chlorophylle sur ces échantillons. En effet, la forte absorption dans cette région du spectre traduit une forte teneur en pigments chlorophylliens des mangues à ce stade de maturité [110]. La réponse spectrale d'un fruit dépend de ses caractéristiques chimiques, principalement de la teneur en chlorophylle et en caroténoïdes, qui donnent respectivement une couleur verte et rouge aux fruits.

Enfin, la troisième absorption située entre 920 nm et 970 nm. Elle est moins forte en raison de la faible teneur en eau des mangues. Les spectres dans cette gamme sont influencés par les larges bandes d'absorption d'eau (liaison O-H) autour de 970 nm [111].

Un examen des spectres des mangues mûres et des mangues non mûres révèle un léger décalage du pic de réflectance vers l'infrarouge lorsque les fruits mûrissent. Cela pourrait être dû aux variations des teneurs en caroténoïdes et en chlorophylles. En effet, la maturation des mangues entraîne l'apparition d'autres pigments tels que les caroténoïdes. Ceux-ci apparaissent lors du mûrissement des fruits. Nous observons aussi une augmentation du plateau de réflectance situé entre 780 nm et 950 nm. Cette augmentation de la réflectance pourrait s'expliquer par la transformation de la structure cellulaire des mangues pendant la maturation. Le second plateau observé entre 952 nm et 1000 nm a également augmenté. Le changement de la structure du fruit lié à la maturation pourrait justifier cette augmentation. Sur la figure, une faible absorption est observée autour de 675 nm. La faible absorption montre que la teneur

en pigments responsables de l'absorption a diminué à cours de la maturation des mangues.

On observe également trois zones d'absorption. D'abord, une première absorption est observée entre 400 nm et 500 nm, ensuite une deuxième observée autour de 675 nm et enfin la troisième se localisant aux alentours de 950 nm. Les deux premières absorptions ont légèrement diminué. Cette diminution est due à la réduction des quantités des pigments absorbants dans ces zones. La dernière liée à la présence des molécules d'eau a connu une légère variation. Ces légers changements au niveau des spectres traduisent alors une faible variation de l'ensemble des paramètres de qualité surtout la coloration.

III.1.2.3. Suppression des valeurs aberrantes par Monte Carlo

Dans cette partie de l'étude, les valeurs aberrantes (échantillons aux données chimiques ou spectrales aberrantes) ont été recherchées dans toute la base de données pour chaque paramètre chimique de qualité (la teneur en eau, le pH et la teneur en sucre). Pour rechercher les valeurs aberrantes, nous avons calculé les erreurs de prédiction des paramètres de qualité de chaque échantillon de mangue. A cet effet, l'algorithme de Monte Carlo a été utilisé. L'algorithme consiste à établir des modèles de prédiction des paramètres de qualité de chaque échantillon à l'aide de la méthode de régression. Une fois les modèles établis, leurs erreurs de prédiction sont calculées et sont représentées sur des diagrammes. Les lignes horizontales et verticales permettent de délimiter la zone d'intérêt (retenir les erreurs les plus faibles). Les échantillons aux erreurs de prédictions trop élevées sont retirés de la base de données en vue d'améliorer les performances statistiques des modèles de prédiction. Ces valeurs qui s'écartent significativement des autres valeurs sont des valeurs aberrantes. Ainsi, ces échantillons ont été repérés et retirés de la base de données. La figure 15 présente la distribution des erreurs de prédiction des échantillons dans les diagrammes respectifs après l'application de la méthode Monte Carlo. Nous pouvons observer sur la figure 15 les distributions des erreurs de prédiction des différents

paramètres de qualité. Les erreurs de prédictions élevées sont observées pour certains échantillons. Les résultats de l'application de la méthode Monte Carlo montrent l'existence de différents points aberrants en nombre et en type de paramètres. L'examen de la figure révèle que certains échantillons aux valeurs aberrantes se retrouvent aussi bien dans les jeux de données du pH, de la teneur en sucre et en eau. Ces valeurs aberrantes retrouvées sont dues aux erreurs intervenues lors des analyses chimiques ou lors des acquisitions spectrales. Ces erreurs élevées pourraient se justifier par les conditions d'analyse chimique incontrôlées (variations de température, matériel de mesure pas précis, etc.) et spectrales [112].

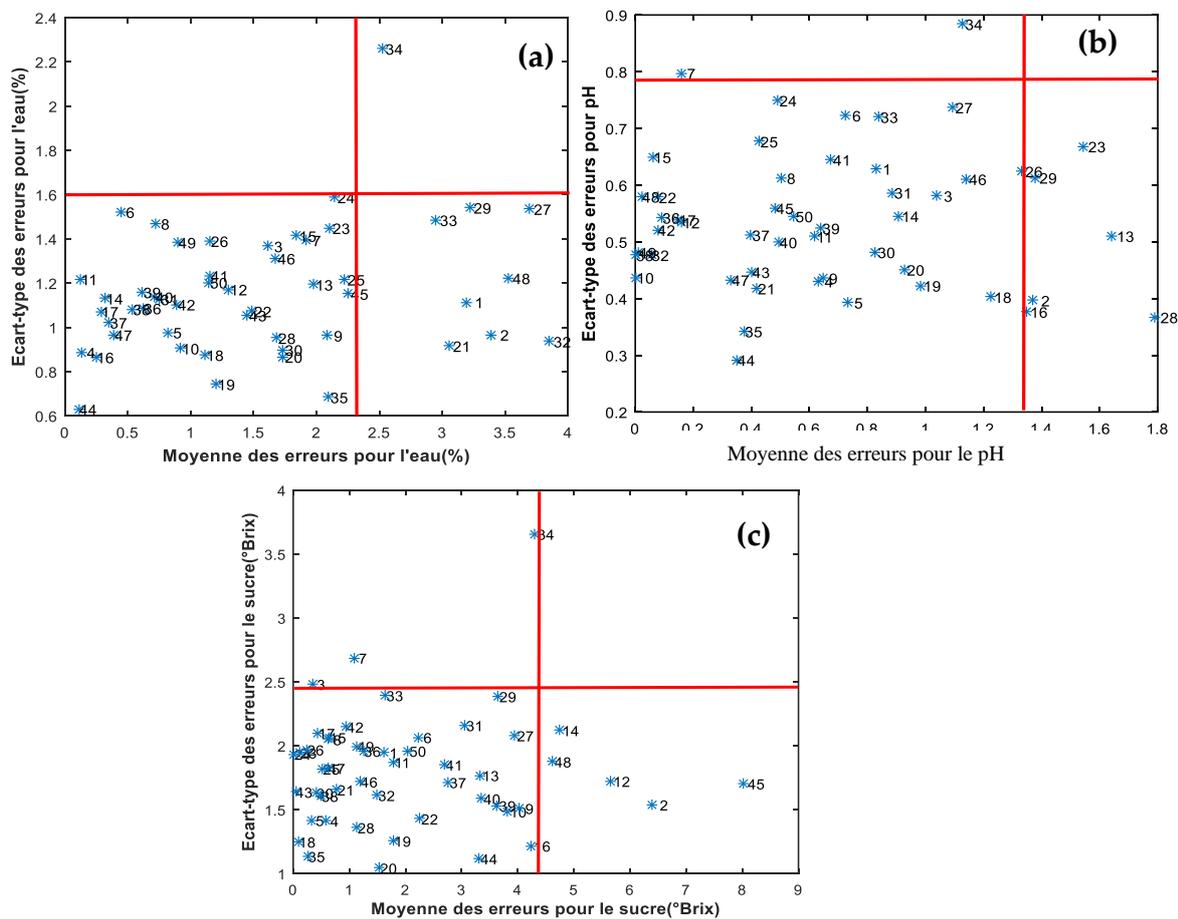


Figure 15: Ecart des erreurs quadratiques moyennes en fonction de la moyenne des erreurs respectivement de a) eau, b) pH, c) en sucre.

Elles peuvent également être dues aux conditions d'acquisition des données spectrales (variations incontrôlées de l'intensité lumineuse et de la température). Ainsi, à l'aide de la méthode Monte Carlo, nous avons pu retirer 8 échantillons dans le jeu de données de teneurs en sucre et du pH, et 9 échantillons dans le jeu de données des teneurs en eau. Certains échantillons aux valeurs aberrantes se retrouvent aussi bien dans le jeu de données de la teneur en sucre, du pH et de la teneur en eau. Il s'agit des échantillons 2, 29, 34 et 48. L'application de cette méthode sur notre base de données avait pour objectif d'améliorer la performance du modèle qui sera élaboré par la régression PLS.

III.1.2.4. Développement et validation des modèles de prédiction

Le développement des modèles constitue une étape primordiale de l'analyse des données. Elle consiste à établir des modèles de prédiction de chaque paramètre de qualité (le pH, la teneur en eau et la teneur en sucre). La validation des modèles permet de tester la performance de prédiction des modèles établis sur des données différentes de celles utilisées lors de l'apprentissage. Pour développer et valider les modèles de prédiction des paramètres de qualité des mangues, nous avons utilisé la méthode de régression PLS. De cette méthode, des modèles ont été établis à partir des données spectrales. Ces modèles ont été établis séparément pour chaque paramètre. Avant, l'établissement des modèles, les données ont été partitionnées en deux groupes pour former les ensembles d'apprentissage et de test ou validation, de sorte que 70% des données ont été utilisées pour construire les modèles de prédiction tandis que le reste des données a été utilisé pour tester les modèles développés. Ce partitionnement a été réalisé par la méthode de Kennard-Stone qui est un algorithme appliqué aux données en vue de les ranger en deux groupes.

III.1.2.4.1. Description générale des modèles

Selon des études déjà menées, un modèle de prédiction établi est considéré comme idéal lorsque le coefficient de corrélation entre les valeurs prédites par le modèle et les valeurs calculées tend vers 1 et l'erreur de prédiction vers 0 [113].

Le tableau 1 résume les performances statistiques des modèles d'étalonnage et de validation de prédiction des différents paramètres de qualité (la teneur en sucre, la teneur en eau et le pH).

Tableau 1: Performances statistiques d'étalonnage et de validation

Paramètres	Calibration				Test		
	R_C^2	R_{CV}^2	RMSEC	RPD	R_P^2	RMSEP	RPD
Teneur en sucre	0,96	0,93	1,28	5,06	0,919	1,69	3,58
Teneur en eau	0,71	0,62	0,9	1,9	0,60	1,56	1,52
pH	0,77	0,70	0,58	2,14	0,61	0,8	1,446

Au vu dudit tableau, nous pouvons dire que pour l'ensemble des modèles établis, les valeurs minimales en calibrage sont de $R_c^2 = 0,71$ et de $RPD_{cal} = 1,9$. Ces valeurs indiquent que les modèles ont été bien entraînés avec des performances de prédiction satisfaisantes. Par ailleurs, l'examen général des valeurs des résultats de la validation montre que les modèles sont efficaces et robustes dans la mesure où les valeurs des coefficients de validation croisés sont comprises entre 0,91 et 0,60.

III.1.2.4.2. Modèle de prédiction du sucre

La figure 16 met en évidence les droites de régression du modèle de calibrage (a) et de validation (b) de la teneur en sucre. Ainsi, nous aurons à partir de sa configuration à visualiser les valeurs de teneur en sucre réellement mesurées par les méthodes de référence par rapport aux valeurs prédites par modélisation PLS. L'observation de la figure montre des nuages de points qui sont proches et allongés autour de la droite de régression. Cette observation révèle un modèle de prédiction très précis pour la teneur en sucre [113]. Cette précision se traduit par le coefficient de détermination élevé entre les valeurs de référence et les valeurs prédites par le modèle. Les valeurs prédites sont très proches des valeurs mesurées.

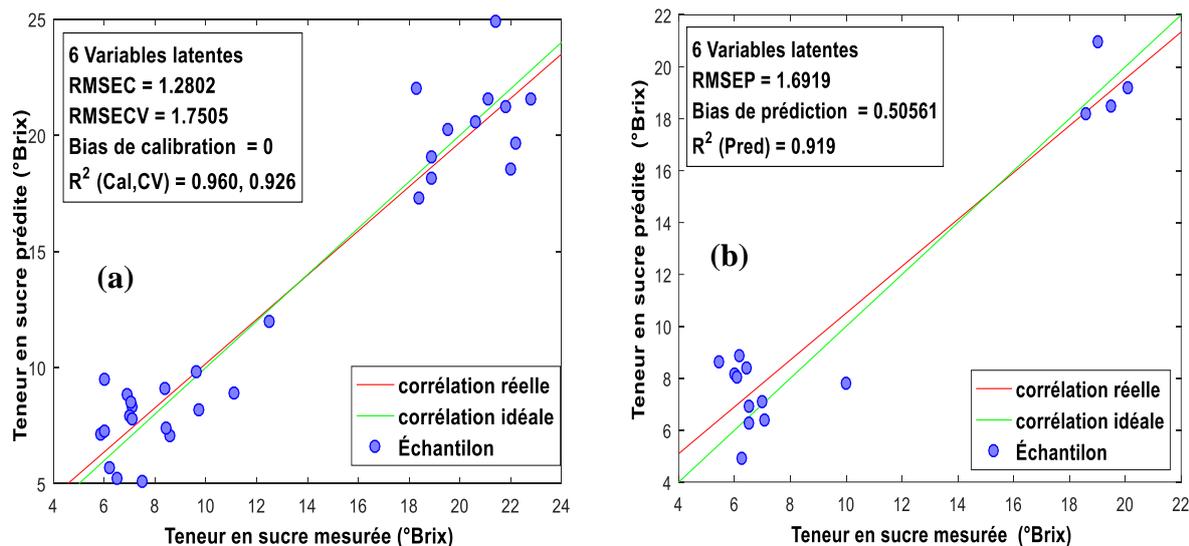


Figure 16: Confrontation des concentrations en sucre mesurées par les méthodes de référence avec celles prédites par modélisation PLS : (a) courbe d'étalonnage, (b) Courbe de validation

Pour le calibrage du modèle de prédiction de la teneur en sucre, nous avons enregistré un coefficient de détermination R^2 de 0,96 avec une RMSEC de 1,28. Le coefficient de détermination proche de 1 (0,96) traduit une relation linéaire entre les valeurs prédites et les valeurs mesurées. Ces valeurs indiquent que le modèle ainsi établi a une excellente performance prédictive bien que l'erreur de prédiction soit trop éloignée de 0. Une affirmation que soutient les travaux de Williams et Norris (2001) [114] qui stipulent que lorsque le R^2 d'un modèle est supérieur à 0,70, la performance du modèle est alors excellente. Ainsi, le modèle a été bien entraîné à la prédiction des teneurs en sucre. La validation du modèle avec des échantillons autres que ceux utilisés lors de son calibrage a donné quant à elle une valeur de coefficient de détermination R^2 de 0,91 avec une RMSEP de 1,69. Ce qui vient confirmer l'excellente performance du modèle de prédiction de la teneur en sucre. Les valeurs de RPD pour le calibrage et pour la validation sont respectivement 5,06 et 3,58. Or, selon les critères de Malley *et al.* (2004) [104], un modèle est considéré comme excellent lorsque la RPD de celui-ci est supérieur à 2. D'où, le modèle ainsi construit est excellent et précis. D'autres travaux comme celui d'ElMasry *et al.* (2007) [20] confirment également cette performance. La prédiction de la teneur en sucre est donc très précise car les valeurs prédites ont une aussi bonne corrélation avec les valeurs chimiques calculées tant pour le calibrage que

pour la validation. Les modèles de prédiction de la teneur en sucre établis dans d'autres études ont obtenu des coefficients de détermination élevés que ceux des travaux de Phuangsoombut *et al.* (2020) et de Lu (2001) [115,116]. Par ailleurs, dans notre étude, la valeur du coefficient de détermination de 0,96 obtenue est supérieure à celle observée dans les travaux d'Elmasry *et al.* (2007) [20], lorsqu'ils ont utilisé l'imagerie hyperspectrale sur des fraises dans la zone du visible-proche infrarouge (400 nm-1000 nm). Ils ont mené des études sur des fraises et ont obtenu une valeur de coefficient de corrélation de 0,85. Aussi, Leiva-Valenzuela *et al.* (2013) [117] ont réalisé des travaux sur des bleuets et ont rapporté des valeurs de R^2 situées entre 0,78 et 0,82 pour l'élaboration du modèle de la teneur en sucre. Tous ces résultats montrent que la teneur en sucre peut être déterminée avec plus de précision par les mesures de télédétection. Par conséquent, la méthode de mesure par télédétection peut servir à la détermination de la teneur en sucre des mangues. Des performances statistiques similaires ont été obtenues par Slaughter *et al.* (1998) [118].

III.1.2.4. 3. Modèle de la teneur en eau

La figure 17 présente les droites de régression des valeurs mesurées en fonction des valeurs prédites des modèles de calibrage (c) et de validation (d) pour la teneur en eau. L'observation de ces deux graphes montre des nuages de points légèrement dispersés autour de la droite de régression. Ces légères dispersions signifient que les valeurs prédites par le modèle ne sont pas assez proches des valeurs expérimentales. Cela est traduit par les valeurs des coefficients de corrélation qui sont moyennement élevées pour le calibrage et la validation. Les valeurs des erreurs sur le calibrage et des erreurs sur la validation sont aussi élevées. Elles indiquent une faible performance de prédiction du modèle de la teneur en eau.

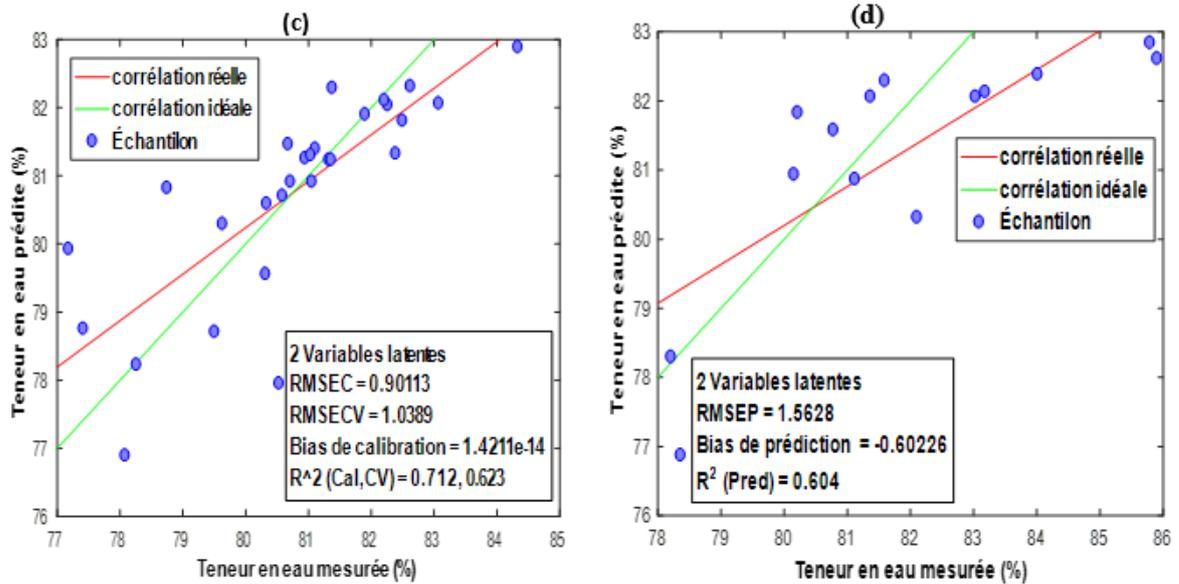


Figure 17: Confrontation des concentrations en eau mesurées par les méthodes de référence avec celles prédites par modélisation PLS : (c) courbe d'étalonnage, (d) Courbe de validation.

Le calibrage du modèle de prédiction de l'eau a donné des valeurs de coefficient de détermination R^2 moyennement élevées de 0,71 avec une RMSEC assez faible de 0,9. Tenant compte de ces valeurs, il convient d'affirmer qu'à la différence du modèle de prédiction de la teneur en sucre, le modèle de prédiction de l'eau se référant aux travaux de Zhang *et al.* (2013) [119] a une performance assez acceptable. En ce qui concerne la validation du modèle, les valeurs du coefficient de corrélation et de REMEP obtenues sont respectivement 0,60 et 1,56. Ces valeurs traduisent un modèle de prédiction moyennement satisfaisant. Le modèle de prédiction de la teneur en eau est moins performant du fait que ses valeurs de RPD de 1,9 pour le calibrage et 1,52 pour la validation soient faibles. Encore, par rapport au modèle de prédiction des teneurs en sucre, le modèle de prédiction de la teneur en eau est moins précis. Cela s'explique par la faible variabilité des valeurs chimiques de teneur en eau des mangues [113]. En comparaison à d'autres travaux, l'équipe de recherche de Seyed Ahmad *et al.* (2010) [120] a réalisé une étude sur des dattes. Elle a construit des modèles de prédiction de teneur en eau et obtenu une valeur de coefficient de détermination $R^2 = 0,98$ supérieure à celle de notre modèle. Cependant, selon la classification des performances de modèles faite par Wang *et al.* (2013) [121], les valeurs des ratios RPD

et des coefficients de détermination respectivement dans les plages 1,4-2 et 0,5-0,75 sont acceptables. Nous pouvons expliquer la faible performance du modèle de l'eau par le faible nombre d'échantillons utilisés pour son calibrage. En se servant de notre technique spectroscopique de télédétection, nous avons développé un modèle fiable pour déterminer la teneur en eau des mangues. Toutefois, un réglage fin du modèle d'étalonnage est nécessaire afin d'augmenter la performance de prédiction et d'améliorer la précision de la prédiction de la teneur en eau des mangues.

III.1.2.4.4. Modèles du Potentiel Hydrogène

La figure 18 présente les droites de régression des valeurs mesurées en fonction des valeurs prédites par le modèle lors du calibrage (e) et de la validation (f) pour le pH. On observe une dispersion considérable du nuage de points autour de la droite de régression. Cela indique qu'il existe un écart considérable entre les valeurs prédites par le modèle et les valeurs calculées par les méthodes chimiques. Cette dispersion traduit également que le modèle n'a pas été suffisamment entraîné pour la prédiction. Cela peut être dû au nombre d'échantillons utilisés pour entraîner le modèle mais aussi pour le tester.

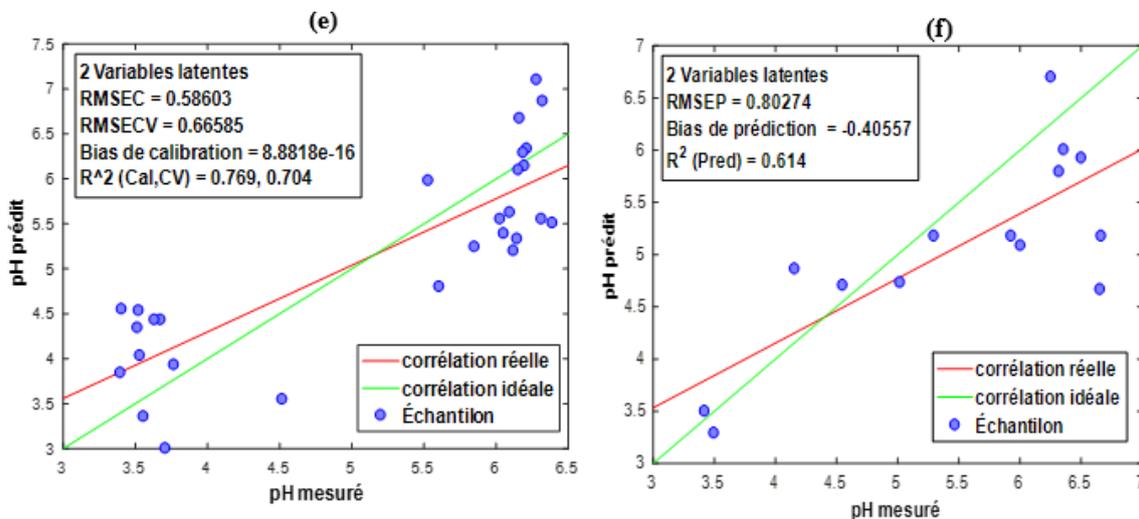


Figure 18: Confrontation des concentrations en pH mesurées par les méthodes de référence avec celles prédites par modélisation PLS : (e) courbe d'étalonnage, (f) Courbe de validation.

Selon l'étalonnage du modèle, le coefficient de corrélation obtenu est de 0,77 avec une erreur RMSEC de 0,58. Soit un coefficient moyennement élevé et une RMSEC

relativement faible. Ces valeurs montrent que le modèle établi est acceptable. En effet, lorsque nous avons utilisé le même modèle pour prédire les pH d'autres échantillons c'est-à-dire pour valider le modèle, nous avons obtenu un coefficient de détermination de 0,61 et un *RMSEC* de 0,8. Le coefficient de détermination est peu élevé et la *RMSEC* est assez faible. Quoique ces valeurs soient différentes à quelques chiffres près, elles viennent confirmer notre première interprétation. Le modèle est acceptable. Ainsi ces résultats traduisent une performance assez faible dans la prédiction de ce paramètre. D'après le *RPD* de 2,14 obtenu pour le pH, le modèle est considéré comme utile selon les travaux de Wang et *al.* (2013) [121]. La prédiction du pH par le modèle est alors satisfaisante et précise avec des *RMSEP* faibles. Ce modèle a également pu être prédit avec de faibles variables latentes (2 variables latentes). Ces faibles variables latentes ont permis d'améliorer les résultats des erreurs de calibrage des modèles du pH. Nous avons pu trouver une confirmation de nos résultats dans ceux de Pu et *al.* (2015) [122] dont les travaux ont porté sur les fruits de litchi. En effet, leurs résultats ont pu montrer de fortes corrélations entre les valeurs de pH et les valeurs prédites. Notre résultat corrobore aussi l'étude de Shyam et *al.* (2012) [123] qui a montré que la spectroscopie proche infrarouge est une technique capable de prédire la teneur en sucre et le pH des mangues avec deux méthodes de régressions. Notre étude a donc obtenu des résultats satisfaisants. Toutefois, ces résultats sont moins satisfaisants en comparaison à ceux de El Masry et *al.* (2007) [20], qui ont mené une étude sur des fraises et ont estimé à 0,87 le coefficient de corrélation et à 0,129 le *RMSEP*; ceux de Baiano et *al.* (2012) sur les raisins rouges et noirs dont les résultats sont $R=0,893$ et 0,947 [124] ; et ceux de Moghimi et *al.* (2010) sur les fruits de kiwis soit un R de 0,943 avec une *RMSEP* de 0,076 très faible [125]. Cependant ces différences pourraient être justifiées par la structure des fruits ayant servi aux différentes études. Etant donné que la mangue a une membrane externe plus épaisse que les fruits cités précédemment, la pénétration de la lumière dans les tissus de celle-ci pour aller rechercher les informations est assez faible. L'étude de Lu Xu et *al.* (2018) [126] a montré qu'il était possible de fournir une méthode innovante de calcul de profondeur de pénétration de la lumière chez le

melon. Aussi comme le soutient Pu et *al.* (2015) [122], la mangue a un niveau de sucre élevé par rapport aux autres fruits et elle a des niveaux d'acides organiques faibles.

Dans cette étude, une technique spectroscopique par télédétection a été développée afin de déterminer des paramètres de qualité des mangues. L'étude a démontré qu'il est possible d'accéder à distance aux paramètres tels que le taux de sucre, le pH et la teneur en eau des fruits de mangue. La régression PLS que nous avons utilisée, nous a permis d'établir des modèles de prédiction de la teneur en sucre, en eau et le pH. Cette technique a ainsi montré son intérêt compte tenu de la robustesse des modèles établis. Cela se caractérise par une excellente performance de prédiction de la teneur en sucre ($R^2 = 0.96$) et une performance moyenne de prédiction du pH ainsi que la teneur en eau. Ces résultats obtenus confirment qu'il est possible d'évaluer les paramètres de qualité de la mangue de manière non destructive par télédétection. La spectroscopie par télédétection apparaît ainsi comme une nouvelle voie de détection des paramètres de qualité des fruits à grande échelle.

III.2. Estimation de la qualité des tomates par réflectance optique

La mesure optique de la qualité des fruits consiste à faire une analyse de ces fruits à partir de leurs propriétés optiques (réflectance, transmittance, absorbance ou diffusion). Elle est de plus en plus utilisée comme une technique non destructive pour évaluer la qualité des fruits ou légumes de manière rapide et efficace. Cette partie du travail étudie alors la capacité et la potentialité de la spectroscopie par télédétection à évaluer la qualité des tomates. Le choix de la tomate comme échantillon dans l'étude est dû au fait qu'elle possède une membrane externe fine pouvant faciliter la pénétration des rayons lumineux afin d'obtenir plus efficacement les informations recherchées. Ce choix s'explique également par le fait que les tomates possèdent de grandes richesses en antioxydants (principalement le lycopène et la vitamine C), en eau et en éléments minéraux offrant ainsi à ces derniers une place importante dans le régime alimentaire des populations.

Dans cette étude, nous avons utilisé soixante-dix (70) spectres de tomates. Ces spectres ont été acquis en mode réflectance. Nous avons utilisé les longueurs d'onde de la gamme de 400 à 800 nm correspondant à la zone du visible. Les analyses chimiques ont été réalisées sur les mêmes échantillons utilisés lors des mesures spectrales. Pour estimer les valeurs des paramètres de qualité, des modèles de prédiction ont été établis et évalués à l'aide de la méthode de régression PLS.

III.2.1. Méthodologie adoptée

Des échantillons de tomate (70 échantillons) ont été achetés sur le marché local de Yamoussoukro. Ils ont été transportés au laboratoire de physique pour les mesures spectroscopiques. Les spectres des échantillons ont été obtenus à partir des instruments spectroscopiques décrits au chapitre II. Les analyses chimiques portant sur le pH, l'acidité, la matière sèche, la teneur en eau et la teneur en sucre ont été effectuées. Toutes les analyses ont été réalisées dans le laboratoire LAPISEN. Le logiciel libPLS installé sur matlab a été utilisé pour le prétraitement et le traitement de l'ensemble des données [127].

III.2.2. Résultats et discussion

III.2.2.1. Spectres de réflectance

Les spectres de réflectance des tomates acquis à l'aide du spectromètre USB 4000 couplé à un télescope de Newton sont présentés dans cette section. Ces spectres ont subi au préalable un lissage en vue de réduire le bruit et d'améliorer la qualité visuelle des spectres. La figure 19 présente les spectres des tomates acquis en mode réflexion.

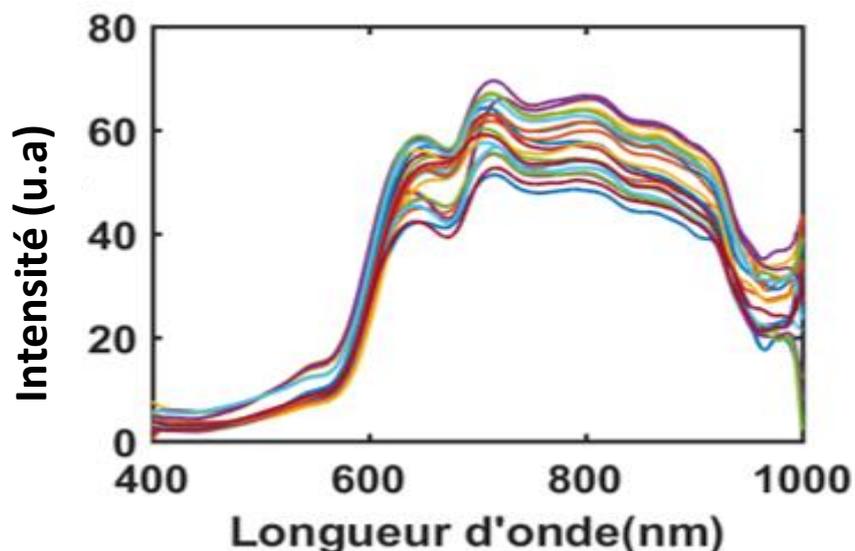


Figure 19: Spectres de réflectance des tomates.

Ils ont été construits dans la gamme spectrale de 400 à 1000 nm pour l'observation visuelle. L'examen de ces spectres indique une similarité au niveau du profil de tous les spectres. Aussi, les spectres présentent tous la même allure car les échantillons sont de la même espèce. Par contre, une observation attentive de ces spectres laisse apparaître une différence visuelle au niveau des intensités des pics de réflectance. En effet, deux pics principaux de réflectance et trois régions principales d'absorptions sont observés sur ces spectres.

Les deux principaux pics de réflectances avec les intensités maximales sont respectivement observables à 644 nm et à 718 nm. Ces deux pics sont liés à la présence des caroténoïdes plus précisément des lycopènes [128]. Ainsi, l'important pic de réflectance dans ce domaine traduit une augmentation de ces molécules du fait de la maturation des tomates. Les lycopènes sont des pigments qui réfléchissent énormément les longueurs d'ondes du proche infrarouge. Ils sont les principaux pigments responsables de la coloration rouge que prend la tomate pendant le mûrissement. Ils appartiennent à la famille des caroténoïdes, plus précisément des carotènes. Ces pigments deviennent plus abondants lorsque la tomate mûrit. Lu et *al.* (2015) [129] ont également expliqué que le comportement spectral des tomates dans ces régions était lié aux variations des concentrations en lycopène et en chlorophylle. Aussi, un dernier pic sous la forme d'un plateau de réflectance est situé dans la gamme spectrale de 800 nm à 900 nm. Le maximum de réflectance pourrait être attribué à la structure de la tomate. Il traduit alors une contribution égale à la réflexion de plusieurs autres composants chimiques des échantillons.

La région de 400 à 550 nm se caractérise par une forte absorption. Les principales longueurs d'ondes correspondant à cette région sont celle du violet, bleu et vert. Dans cette gamme, l'absorption est attribuée à la forte présence de caroténoïdes plus précisément le lycopène [128]. Lorsque les tomates sont complètement mûres, la teneur en chlorophylle est fortement réduite et le lycopène devient le pigment dominant de

son tissu. Par conséquent, on observe un pic d'absorption important à la zone du 400 à 550 nm pour les tomates rouges en raison de leur importante quantité de lycopène, ainsi qu'à la disparition de l'absorption de la chlorophylle à 675 nm. La nature des courbes spectrales de tomates est en étroite association avec la maturité physiologique du fruit de la tomate.

La région autour de 675 nm est aussi marquée par une autre absorption. Cette absorption est faible par rapport à celle de la bande de 400 à 550 nm. Elle est due aux pigments chlorophylliens. La faible absorption dans cette zone renseigne sur la quantité de chlorophylle dans les tomates qui diminue avec la maturation au profit du lycopène. En effet, en raison de la dégradation de la teneur en chlorophylle des tomates à la maturité, les spectres dans cet intervalle ont été faiblement influencés.

Par ailleurs, une forte absorption est observée dans la zone de 960 nm. Elle rend compte de la présence des molécules d'eau [114]. Le pic d'absorption est assez considérable dans cette région, ce qui suggère que la teneur en eau des tomates était élevée et traduit clairement la variation de la teneur en eau au cours de la maturation. L'absorption dans cette zone est sensiblement plus forte justifiant ainsi l'abondance de l'eau dans la tomate [130]. En effet, l'eau représente le composant le plus dominant dans les fruits.

Toutes ces observations et interprétations des spectres confirment l'existence d'une relation entre les valeurs chimiques et les longueurs d'onde. D'où la nécessité de faire de la modélisation.

III.2.2.2. Modèles de prédiction par régression PLS

III. 2.2.2.1. Présentation des résultats

Les mesures spectrales effectuées ne permettent pas d'obtenir directement les valeurs des différents paramètres de qualité. Pour accéder aux informations recherchées, nous devons passer par une phase de modélisation. Pour ce faire, nous avons développé des modèles de prédiction à partir de la méthode de régression PLS. Pour établir ces modèles de prédiction, nous avons utilisé le logiciel MATLAB 2016a qui est un logiciel

d'analyses statistiques libres. Sur ce logiciel, nous avons installé un autre logiciel libPLS [127]. Le logiciel libPLS a été modifié afin d'être adapté au type de calcul que nous souhaitons. Pour le prétraitement des données spectrales, nous avons ajouté une méthode de lissage. Ce prétraitement a impliqué la suppression de certains bruits générés par les variations incontrôlées de l'intensité lumineuse. Il a également contribué à améliorer les résultats de la prédiction. A l'aide du même logiciel libPLS qui contient plusieurs algorithmes dont Monte Carlo, Kennard Stones et la régression PLS, les différents modèles de prédiction ont été élaborés. L'algorithme de Monte Carlo a été utilisé dans cette étude pour la suppression de données aberrantes et celui de Kennard Stone pour le partitionnement de l'ensemble des données en deux sous-groupes. Les deux sous-groupes obtenus ont respectivement servi à l'apprentissage (calibrage) et à la validation des modèles. Les éléments du groupe d'apprentissage ont servi à la construction des modèles de prédiction et ceux du groupe de validation ont servi à tester ces modèles. Le but de cette approche était de montrer la possibilité d'estimer plus efficacement les paramètres de qualité des tomates à partir des données spectrales obtenues à distance. Les résultats du calibrage et de la validation des modèles de prédiction sont rapportés dans le tableau 2. Chaque modèle établi sera évalué à partir de ses performances statistiques (R^2 , RMSEC et RMSEP).

Tableau 2: Performances statistiques des modèles de prédiction PLS des paramètres de qualité

Paramètres de qualité	Paramètres statistiques	
	R^2	RMSE
pH	0,982	0,991
Acidité	0,978	0,989
Teneur en sucre	0,983	0,991
Teneur en eau	0,978	0,153
Matière sèche	0,988	0,993

Le tableau 2 regroupe les résultats de l'établissement des modèles de prédiction des différents paramètres de qualité par la régression PLS. Les résultats dans ce tableau

nous renseignent sur les performances statistiques obtenues lors des entraînements des modèles établis. L'examen des résultats du tableau montre des valeurs élevées des coefficients de corrélation (R^2 supérieurs à 0,97) traduisant d'excellentes capacités prédictives des modèles au cours de leur entraînement. Ces valeurs sont sensiblement égales à 1. Aussi, l'on observe des faibles valeurs des erreurs. Ces valeurs indiquent des modèles très précis pendant leur entraînement. Ces modèles prédisent ainsi des valeurs très proches des valeurs expérimentales. Tous ces résultats prouvent que les modèles ont été très bien entraînés et peuvent servir à la prédiction des paramètres recherchés.

Le tableau 3 montre les résultats des performances statistiques des modèles lors de leur validation.

Tableau 3: Résultats des performances statistiques des modèles lors de la validation

Paramètres de qualité	Paramètres statistiques	
	R^2	RMSE
pH	0,97	0,055
Acidité	0,956	0,005
Teneur en sucre	0,969	0,189
Teneur en eau	0,978	0,993
Matière sèche	0,978	0,153

Les résultats du tableau montrent que pour l'ensemble des paramètres, les valeurs des coefficients de détermination R^2 sont très élevées. Ces valeurs sont toutes comprises entre 0,95 et 0,98. Elles se rapprochent beaucoup plus de 1 montrant ainsi de très fortes corrélations entre les valeurs prédites par les modèles et les valeurs expérimentales. Ces fortes corrélations nous renseignent sur les capacités prédictives des modèles. Aussi, ces capacités des modèles se confirment à travers de faibles valeurs d'erreurs de prédiction (RMSEV) de chaque modèle. Les erreurs de prédiction sont toutes comprises entre 0,005 et 0,99. Elles tendent en général vers zéro (0). En effet, une faible

erreur de prédiction traduit une bonne précision du modèle de prédiction. Les coefficients de détermination élevés et les erreurs de prédiction faibles traduisent alors des modèles linéaires et précis donc d'excellents modèles. Quel que soit le paramètre, le modèle établi fournit une bonne précision lors de la validation.

Les résultats des apprentissages et des validations des modèles sont présentés au travers de la figure 20.

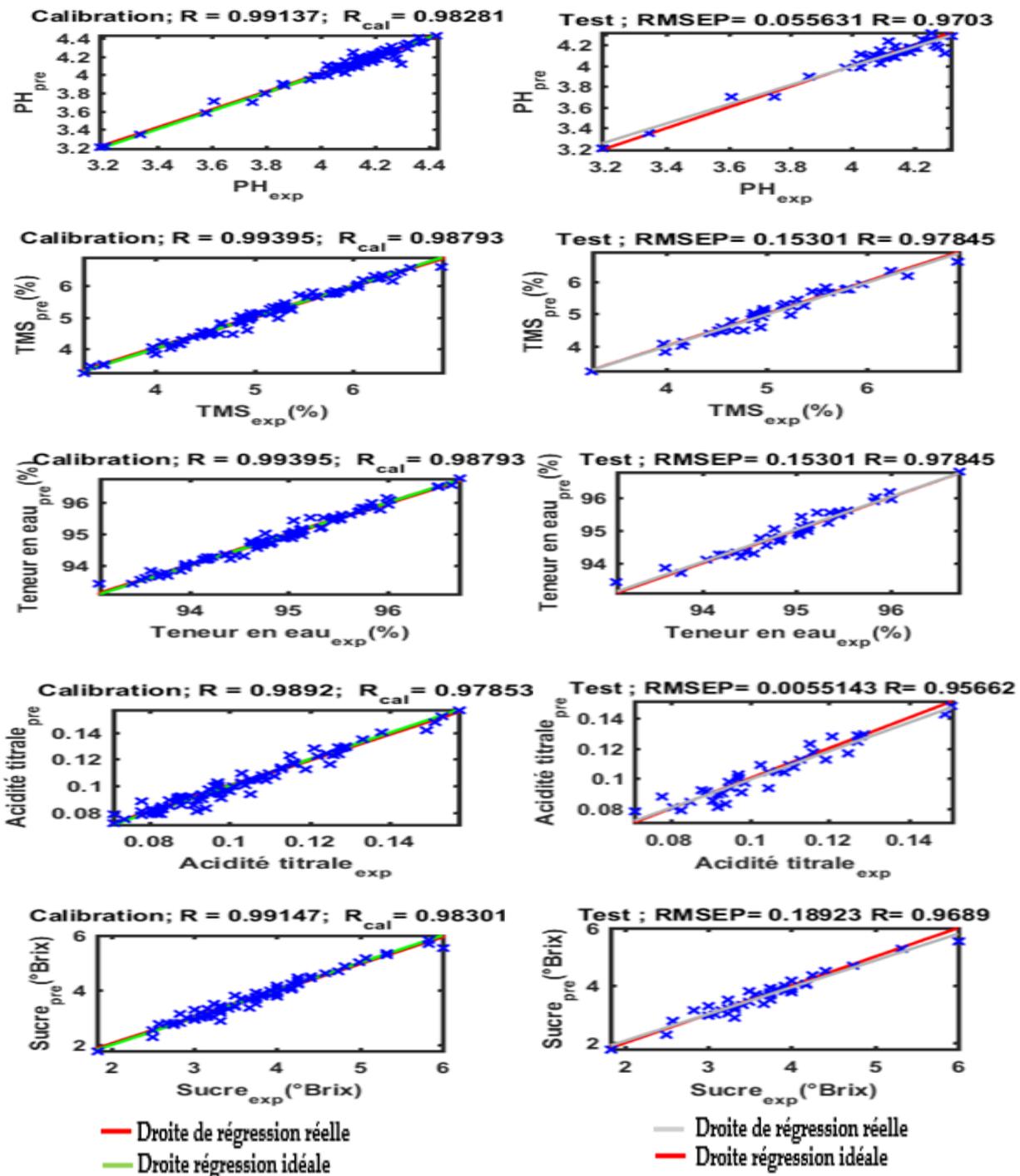


Figure 20: Graphiques des modèles de prédiction.

La figure 20 présente les résultats de chaque étape de la régression partielle par moindres carrés (PLS) des différents modèles de prédiction. Les colonnes 1 à 2 de la figure 20 représentent respectivement l'étape de calibration et de test pour chaque paramètre physico-chimique que sont l'acidité, la matière sèche (MS), le pH, la teneur en sucre (TS) et la teneur en eau (TE). Les valeurs en ordonnée représentent les valeurs physico-chimiques prédites par le modèle généré par PLS et en abscisse, les valeurs expérimentales des paramètres physico-chimiques mesurées au laboratoire. Au-dessus de chaque sous-figure de l'étape de calibration, on distingue l'erreur quadratique moyenne de calibration (R) et le coefficient de régression lors de la calibration (R_{cal}). Au-dessus de chaque sous-figure de l'étape de test, on a l'erreur quadratique moyenne de prédiction ($RMSEP$) et le coefficient de détermination de la prédiction (R) lors du test. Les erreurs quadratiques moyennes de calibration des modèles générés par PLS sont autour de 0,98 pour l'acidité, la MS, le pH et la TE. Le coefficient de régression lors de la calibration est approximativement égal à 0,98 pour le pH, 0,99 pour la TMS, l'acidité et le sucre. Les différentes valeurs des coefficients sont en adéquation avec la distribution des données autour de la droite d'équation $y=x$. Cela montre qu'il y a une excellente corrélation entre les valeurs expérimentales et les valeurs prédites lors de la calibration.

Au-dessus de chaque sous-figure de l'étape de test, les erreurs quadratiques moyennes de prédiction sont respectivement de 0,06 pour le pH, 0,15 pour le TMS et la teneur en eau, 0,006 pour l'acidité et 0,19 pour le sucre avec des coefficients de détermination de prédiction autour de 0,98. L'acidité présente une erreur quadratique moyenne plus faible par rapport aux autres modèles mais un coefficient de prédiction de 0,95 inférieur aux coefficients de prédiction des autres paramètres. Les erreurs quadratiques de prédiction du pH et de l'acidité sont les plus faibles, cela justifierait le caractère acide de la tomate.

III.1.3.2. Discussion

Tous les échantillons présentent le même comportement spectral. Cependant, nous pouvons noter une différence au niveau des intensités à certaines longueurs d'ondes.

L'examen des résultats de l'entraînement (calibrage) de chaque modèle révèle un excellent entraînement des différents modèles par la régression PLS. Les coefficients de corrélation R sont tous supérieurs à 0,97 et sont dans l'ensemble très proches de 1. Ces valeurs de coefficients de détermination en calibrage très proches de 1 représentent des résultats très satisfaisants. Elles traduisent des modèles idéaux et très bien entraînés pour la prédiction des paramètres de qualité. En effet, selon les critères de performance de calibrage donnés par Malley *et al.* (2004) [104], les valeurs des coefficients de corrélation supérieurs à 0,95 sont excellentes. Donc les modèles seront capables de prédire les paramètres de qualité de la tomate. Aussi, les valeurs des erreurs de calibrage renseignent sur la qualité de la précision lors de l'entraînement. L'analyse des résultats lors des calibrages révèle une parfaite corrélation entre les valeurs prédites par les modèles et les valeurs expérimentales. Cela traduit donc de très bons apprentissages des modèles.

En ce qui concerne l'étape de la validation, l'examen de tous les résultats révèle donc que l'ensemble des modèles de prédiction établis par la méthode de régression PLS ont une excellente capacité de prédiction et sont tous très robustes. Tous ces résultats donnent de meilleures performances pour tous les constituants. Les meilleures performances des modèles peuvent s'expliquer par les méthodes statistiques (prétraitement et traitement) appliquées aux données. En effet, certains auteurs dans leurs travaux ont montré que l'utilisation des méthodes de prétraitement pendant la modélisation pouvait améliorer considérablement les résultats des modèles de prédiction. A cet effet, nous pouvons citer DEÁK *et al.* (2015) [131] qui ont utilisé les méthodes de dérivée première et de dérivée seconde pour prétraiter leurs données spectrales. Ils ont pu obtenir des équations optimales pour la modélisation des

paramètres. Gomez *et al.* (2006) [132] ont eux aussi montré que les prétraitements par moyenne mobile et par correction des dispersions multiplicatives peuvent influencer les performances des modèles de prédiction. Nous rappelons aussi que les meilleures performances des modèles de cette étude pourraient s'expliquer par les structures externes des échantillons utilisés car les fines couches externes ont facilité la pénétration des rayons lumineux. Ainsi les rayons lumineux ont pu facilement interagir avec les paramètres recherchés.

Par ailleurs, la plage spectrale sélectionnée pour l'étude serait un facteur majeur d'amélioration des résultats de la modélisation. Ainsi, la sélection a permis de réduire les bruits par élimination des régions assez bruyantes des différents spectres.

A partir de l'analyse des modèles, nous pouvons affirmer que la capacité prédictive générale donnée par les résultats de la validation est très satisfaisante et reflète les valeurs de prédiction pour chaque paramètre de qualité.

Les modèles développés pour chaque paramètre de qualité sont acceptables et très encourageants. Cela exprime que notre technique est capable de déterminer à distance les paramètres de qualité de la tomate.

Dans ce chapitre, nous avons appliqué notre technique de mesure par télédétection sur des tomates. Cette application repose sur l'acquisition de données spectrales suivie de la modélisation des paramètres de qualité des tomates. L'ensemble des résultats présentés montre une importante performance de prédiction des différents modèles. Les valeurs des paramètres de prédiction sont très satisfaisantes. Ces résultats mettent alors en évidence l'aptitude de notre dispositif à évaluer la qualité des tomates. Ceci ouvre une voie d'analyse à distance de la qualité des fruits.

III.3. Estimation de la qualité des pommes de cajou par réflectance optique

III.3.1. Méthodologie adoptée

Pendant la période de mesure qui a duré une semaine, trente (30) plantes d'une plantation d'anacarde ($6^{\circ}52'14.6''N$ $5^{\circ}12'14.2''W$) ont été sélectionnées. Dix (10)

échantillons ont été prélevés quotidiennement sur le lot de plantes sélectionnées à raison de deux pommes par plante. Les fruits prélevés ont été immédiatement transportés au laboratoire après la récolte en vue d'effectuer des mesures spectrales et des analyses chimiques. Au total, soixante-dix (70) pommes de cajou ont été utilisées comme échantillons dans cette étude. Les spectres de réflectance ont été acquis sur chaque pomme de cajou intact à l'aide du dispositif décrit au chapitre II. Les mesures chimiques ont été réalisées sur la même partie que la partie ayant servi aux mesures optiques. Chaque paramètre chimique (pH, teneur en sucre, teneur en eau, matière sèche et Vitamine C) a été déterminé selon une méthode bien appropriée. Après les séances de mesure, une étape de prétraitement a eu lieu. Elle a consisté à la sélection de spectres ou de variables afin d'améliorer la précision et la performance prédictive des modèles de prédiction lors de la modélisation. Dans le cas notre étude, cette démarche a consisté à sélectionner les régions spectrales et les échantillons les plus prédictifs de la base de données avant d'établir les modèles de prédiction des paramètres de qualité. Pour cela, nous avons sélectionné les longueurs d'ondes correspondant aux régions de variables informatives. Ainsi, nous avons exclu les régions spectrales inférieures à 400 nm et celles supérieures à 800 nm qui sont des régions dans lesquelles les bruits sont très élevés. Cela nous a permis de limiter la région spectrale utile pour l'analyse statistique de 400 nm à 800 nm. Les échantillons dont les spectres présentaient de grandes variabilités par rapport aux autres ont été également supprimés de la base de données. Ainsi, sur 70 échantillons, nous avons retenus 52 échantillons avec 2004 variables pour chaque échantillon. Notre spectromètre attribue à chaque longueur d'onde une valeur d'intensité et celle-ci correspond à une variable. Ainsi, pour la pomme de cajou, nous avons 3648 valeurs d'intensité correspondant à 3648 variables qui sont passées à 2004 variables après sélection. Il faut noter qu'il a été question d'identifier et d'éliminer par la suite les spectres aberrants où les gammes spectrales contenant un maximum de bruits. En plus d'améliorer la précision et la robustesse des modèles, la sélection permet d'accélérer le processus d'analyses des données en réduisant le temps de calcul.

Le processus d'analyse des données a consisté à établir des modèles mathématiques afin de prédire les paramètres de qualité à partir des données spectrales de réflectance. Cependant, du fait de la réduction du nombre d'échantillons, nous avons procédé à un rééchantillonnage par la méthode bootstrap. Cette méthode consistait à créer des « nouveaux échantillons » statistiques à partir des échantillons initiaux à l'aide de l'algorithme de Wong et Easton [133]. Les nouveaux échantillons créés ont alors servi à établir les différents modèles mathématiques par la méthode des réseaux de neurones artificiels. Les spectres de réflectance visibles sont utilisés comme entrées dans l'application de réseau neuronal. Avant de traiter les données par application de réseau de neurones, elles ont d'abord été divisées en deux groupes : Un groupe d'apprentissage et un groupe de validation. Par la suite, le groupe d'apprentissage obtenu a été subdivisé en groupe d'apprentissage, en groupe de test et en groupe de validation. Les pourcentages relatifs sont respectivement de 70% pour l'apprentissage, 15% pour le test et 15% pour la validation [134]. L'évaluation de la performance d'un modèle de réseau de neurone dépend fortement du partitionnement de l'ensemble de données. Les données d'apprentissage sont présentées au réseau pendant l'entraînement et le modèle est ajusté en fonction de l'erreur de prédiction. Les données de validation sont utilisées pour évaluer la généralisation du réseau et pour arrêter l'apprentissage lorsque la généralisation cesse de s'améliorer. Les données tests n'ont aucun effet sur l'apprentissage et fournissent donc une mesure indépendante de la performance du réseau pendant et après l'apprentissage. Le réseau utilise comme fonction d'activation au niveau des couches cachées, la fonction de tangente hyperbolique. Il utilise une fonction de transfert linéaire dans la couche de sortie et un algorithme Levenber-Marquart pour réduire le temps de calcul. Par défaut, le nombre de neurones pour les couches cachées est dix (10).

III.3.2. Spectres de réflectance

III.3.2.1. Résultats

Nous avons mesuré les spectres de réflectance de soixante-dix (70) pommes de cajou à maturité à l'aide du spectromètre USB couplé au télescope. Les spectres nous permettent d'avoir l'information spectrale globale sur la qualité des pommes. La figure 21 présente les spectres de réflectance des pommes de cajou.

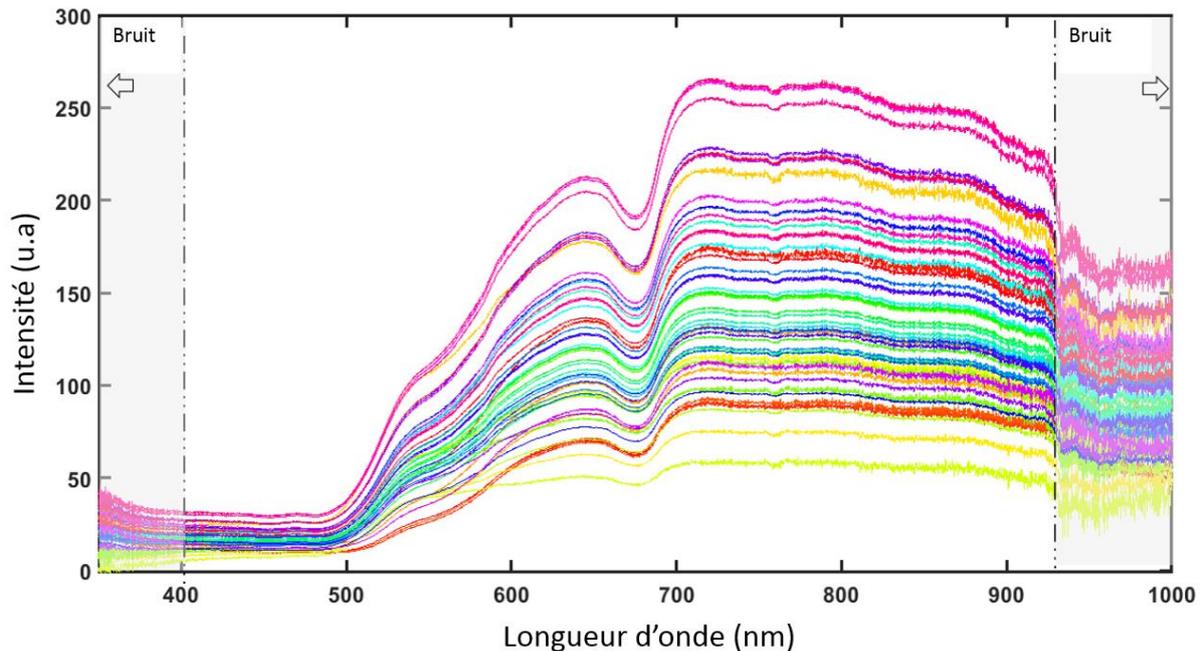


Figure 21: Spectres de réflectance des pommes de cajou.

Nous observons une variation des spectres de réflectance sur toute la plage spectrale. Nous pouvons distinguer différentes zones de réflectance :

- 400-500 nm : cette région présente une très faible intensité de réflectance. Les spectres sont quasiment constants et présentent tous les mêmes aspects avec une petite inflexion vers 500 nm. L'intensité lumineuse ne varie quasiment pas dans cette zone et les spectres sont tous concentrés. Ce qui traduit une forte absorption des constituants des fruits dans le bleu.

- 500-520 nm : Le spectre est croissant dans cette région. La réflectance ici est faible avec un pic observé à 520 nm. Cette région voit une brusque augmentation de la réflectance dont le maximum de la dérivée permet de déterminer un point d'inflexion.

Des variations sont observées au niveau des spectres. Il y a un décalage des pics pour certains spectres.

- 520-675 nm : Le spectre continue de croître dans cette région et l'intensité de réflectance est assez élevée. La réflectance dans cette zone est caractérisée par un maximum au voisinage de 650 nm. La région présente des spectres qui sont dispersés et les pics des spectres n'ont pas le même niveau d'intensité de réflectance. Certains spectres présentent des pics d'intensités faibles.

- 675-685 nm : Le spectre connaît une décroissance dans cette région. La décroissance est importante pour certains spectres et pour d'autres, elle est moins importante. L'intensité de réflectance est basse dans cette zone et elle est marquée par un minimum au voisinage de 680 nm. La région présente des spectres qui sont dispersés et qui n'ont pas le même niveau de réflectance. Certains spectres présentent des pics avec des intensités de réflectance quasiment faibles.

- 685-920 nm : La figure montre une croissance des spectres dans cette région. L'intensité de réflectance est quasiment constante sur toute la zone. Cependant, nous remarquons des spectres avec des intensités de réflectance plus élevées que d'autres. Tous les spectres ne coïncident pas.

- 920-980 nm : Cette zone présente une faible réflectance. Elle est essentiellement influencée par du bruit. Il est difficile de distinguer les spectres. La réflectance dans cette zone est caractérisée par une constance.

III.3.2.2. Discussion

Nous remarquons ici un comportement similaire entre les spectres des pommes de cajou. En effet, cette similarité révèle une même composition chimique pour toutes les pommes de cajou. La différence se trouve encore au niveau des intensités à certaines longueurs d'onde.

Dans la région de 400-500 nm, les spectres des pommes présentent une intensité plus faible sur toute la plage des longueurs d'onde d'étude. Cette faible réflexion traduit une forte absorption par ces fruits, signe d'une importante concentration en molécules

absorbantes des longueurs d'onde de cette région. Les molécules responsables ici concernent les caroténoïdes qui influencent énormément les longueurs d'onde de la région de 400-500 nm. Ainsi la forte absorption observée dans cette région est une indication de l'abondance des caroténoïdes. Le comportement spectral est relié à l'état de maturation des fruits. L'importante absorption des longueurs d'ondes de cette zone est due à une augmentation de la concentration en caroténoïdes des fruits. Ils sont responsables de la coloration à la maturation.

Quant à la région de 500 à 520 nm, nous observons une faible croissance. Cette faible croissance traduit une faible intensité de réflectance par ces fruits et une absorption. La réflectance est faible du fait de la faible teneur en chlorophylle des pommes de cajou mûres. Les caroténoïdes sont les pigments responsables de la couleur rouge et jaune des pommes de cajou. En effet, à la maturation la concentration en caroténoïdes augmente et celle de la chlorophylle diminue [135].

Dans le domaine de 520-675 nm, le spectre continue de croître dans cette région et l'intensité de réflectance est assez élevée. L'augmentation de l'intensité de réflectance suggère une accumulation de substances organiques liée au processus de maturation. Le maximum observé au voisinage de 650 nm s'explique par le décalage du pic de réflectance de la chlorophylle autour de 550 nm. Ce décalage de pic vers les autres longueurs d'onde montre qu'une dégradation du taux de chlorophylle laissant place à d'autres pigments (caroténoïdes, xanthophylles). La dégradation du taux de chlorophylle est responsable du changement de coloration des fruits au cours de la maturation. Ainsi les caroténoïdes sont à l'origine de la coloration jaune pâle ou jaune orangé tandis que la coloration rouge est attribuée aux anthocyanes. Aussi, la différence du niveau de réflectance des spectres dans cette zone peut être justifiée par la différence des niveaux de maturité des fruits [136].

Au niveau de 675 à 685 nm, la décroissance observée indique une faible réflexion et une importante absorption par les fruits. La faible réflexion est liée à une faible contribution des pigments responsables de la coloration à la maturation. Le minimum

observé au voisinage de 680 nm est généralement dû aux absorptions des pigments chlorophylliens [137, 138]. Ce minimum est moins important. Cela suggère une baisse d'absorption qui indique qu'à cette étape de la maturation, les constituants photosynthétiques (pigments chlorophylliens) des pommes de cajou sont en quantité insuffisante pour assurer la synthèse de leur matière organique. La séparation des spectres et des niveaux d'intensité de réflectance dans cette région peut s'expliquer par la différence des états de maturité.

A la zone de 685-920 nm, l'intensité de réflectance élevée est une indication de l'abondance des composants chimiques liés à la réflexion de l'intensité lumineuse. Ici, cette réflectance est liée à la structure interne du fruit. La forte intensité de réflexion des fruits cette zone traduit la croissance des constituants de la structure interne des fruits au cours de la maturation.

De 920-980 nm, la faible réflectance observée est liée aux absorptions des molécules d'eau [139] et les autres constituants (sucre) contenant des liaisons O-H et C-H dans leur structure. Cette région est dominée par le spectre d'absorption de l'eau car l'eau est le constituant le plus abondant des fruits. Le bruit qui influence le spectre dans ce domaine peut s'expliquer par le fait que dans cette région (infrarouge), les spectres sont sensibles aux variations de température des échantillons. Une augmentation de la température de la surface de l'échantillon, en raison de la chaleur de la source lumineuse, affecte leur spectre. Les travaux de Xie et *al.* (2016) [140] ont rapporté l'effet de la variation de la température sur les spectres visibles et proches infrarouges.

La redondance des informations au niveau de ces spectres ne permet pas une analyse directe des paramètres recherchés à partir des données spectrales obtenues de nos échantillons. D'où la nécessité d'utiliser des méthodes d'analyse statistique.

III. 3.3.1. Modélisation des paramètres de qualité des pommes de cajou

Nous avons utilisé le réseau de neurones artificiels dans cette étude dans le but de construire des modèles mathématiques de prédiction des paramètres de qualité des pommes de cajou. Ces modèles permettront de prédire les cinq valeurs chimiques que

sont le pH, la teneur en sucre, la teneur en eau, la matière sèche et la vitamine C. Pour ce faire, nous avons constitué une base de données de 52 échantillons avec 2004 variables pour chaque échantillon. Avant l'application du réseau de neurones artificiels, la base de données a été subdivisée en deux groupes : un groupe pour l'entraînement et un autre pour la validation avec les proportions de 70% pour l'entraînement et 30% pour la validation. Les données d'entraînement ont été présentées au réseau lors de l'entraînement (apprentissage) du réseau et le réseau est ajusté en fonction des erreurs. La performance d'un modèle de prédiction établi par le réseau neuronal dépend fortement du partitionnement de l'ensemble des données. Il fallait donc un nombre important d'échantillons afin d'avoir de bons modèles de prédiction.

Les modèles de prédiction que nous avons mis au point utilisent un réseau à deux couches en aval avec une fonction de transfert hyperbolique tangente (tanh) dans la couche cachée, une fonction de transfert linéaire dans la couche de sortie et un algorithme de Levenberg-Marquardt pour réduire le temps de calcul. Le nombre de neurones sur la couche cachée est dix (10) par défaut.

Ainsi, le réseau de neurones a permis d'élaborer cinq modèles de prédiction pour les cinq paramètres de qualité.

Nous présentons ainsi les résultats statistiques obtenus après les entraînements des différents modèles de prédiction dans le tableau 4.

Tableau 4: Résultats statistiques du modèle de prédiction

Paramètres de qualité	Paramètres statistiques	
	R ²	RMSE
pH	0,988	0,025
Teneur en sucre	0,989	0,13
Teneur en eau	0,976	0,314
Matière sèche	0,974	0,32
Vitamine C	0,993	4,72

Le tableau 4 présente les résultats de l'apprentissage du réseau à construire les modèles de prédiction des paramètres de qualité des pommes de cajou. L'observation du tableau révèle que tous les modèles de prédiction construits par le réseau ont obtenu des valeurs de coefficient de détermination très élevées et supérieures à 0,97. Les coefficients de détermination des modèles sont beaucoup proches de 1. Nous remarquons que chaque modèle présente une faible valeur d'erreur de prédiction lors de l'entraînement. Ces valeurs tendent vers 0 pour la plupart des modèles. L'erreur de prédiction du modèle de pH est très faible par rapport aux autres (0,025). L'erreur de prédiction du modèle de la vitamine C est élevée (4,72). Rappelons que la fonction d'activation utilisée pour les couches cachées est la fonction tangente hyperbolique (tanh). Ainsi, la relation entre la sortie et l'entrée du réseau peut se mettre sous la forme :

$$S = (LW \times \text{Tanh}(IW \times \text{Input} + B1)) + B2 \quad \text{III.1}$$

Avec :

LW : le poids des connexions entre les couches d'entrée et les couches cachées ;

IW : poids des connexions entre les couches cachées et la couche de sortie ;

Input : paramètres d'entrée (données spectrales retenues) ;

B1 : biais des couches d'entrées ;

B2 : biais des couches de sortie ;

S : paramètres de sortie.

Les poids et les biais contribuent à fixer les relations entre les variables d'entrée. Les poids et les biais d'entrées permettent le calcul des entrées de l'ensemble des neurones des couches cachées. Ainsi, l'entrée des N_i neurones est calculée par la formule suivante :

$$N_i = (\sum_{i=1}^{13} IW_i \times \text{input}_i) + B_1 \quad \text{III.2}$$

La fonction d'activation est appliquée, afin d'avoir la valeur de sortie des neurones.

La valeur de sortie du neurone est obtenue par application de la fonction d'activation. Dans ce cas, la sortie du neurone S a été calculée selon la formule qui suit :

$$S = \left(\sum_{i=1}^{10} LW_i \times \tanh(N_i) \right) + B2 \quad \text{III.3}$$

Nous avons utilisé la deuxième partie du jeu de données pour tester les différents modèles construits à partir du premier ensemble de données lors de l'entraînement. Ce deuxième groupe de données fut introduit dans chaque équation afin de prédire les différents paramètres. La comparaison des valeurs prédites aux valeurs expérimentales fournit des informations sur le pouvoir prédictif des modèles et la performance de chaque modèle. La figure 22 montre les corrélations entre les valeurs prédites et les valeurs expérimentales pour chaque paramètre de qualité.

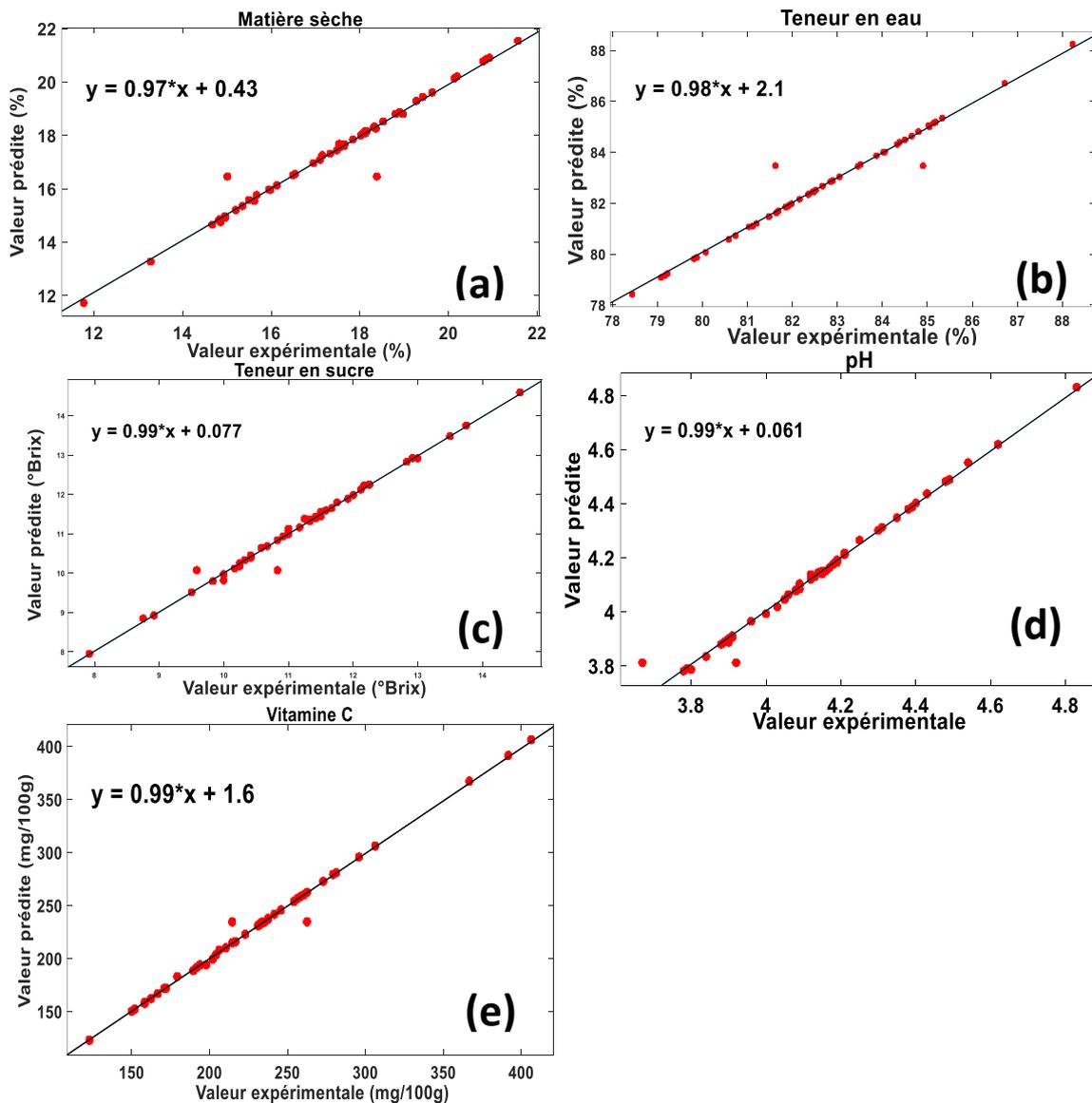


Figure 22: Corrélation entre les valeurs expérimentales et prédictives pour (a) la matière sèche, (b) la teneur en eau, (c) le pH, (d) la teneur en sucre et (e) la vitamine C.

Les valeurs prédites sont obtenues avec la formule III.1 et comparées aux valeurs expérimentales. L'objectif de ces comparaisons est de valider les modèles. En général, les modèles de prédiction sont dits précis lorsque l'ensemble des points autour de la droite de régression (la diagonale $y=x$) sont étirés et resserrés. Dans le cas de cette étude, la figure 22 présente des points qui sont très resserrés autour de la droite de régression pour chaque modèle de prédiction. En plus, la figure révèle une bonne régression entre les valeurs prédites et les valeurs expérimentales car les coefficients de détermination sont tous supérieurs à 0,97 et les erreurs sont faibles (voir le tableau 5). Cela signifie que les valeurs prédites par les modèles sont très proches des valeurs expérimentales et que les corrélations entre ces valeurs sont très élevées. Cela reflète une très bonne prédiction des différents paramètres de qualité par les modèles établis par le réseau de neurones artificiels. Le tableau 5 ci-dessous résume les résultats de la validation des modèles.

Tableau 5: Valeurs des paramètres statistiques des modèles de prédiction

Paramètres de qualité	Critères statistiques	
	R ²	RMSE
pH	0,987	0,026
Teneur en sucre	0,987	0,139
Teneur en eau	0,972	0,333
Matière sèche	0,970	0,346
Vitamine C	0,992	5,03

L'examen de ce tableau montre que les coefficients de détermination R² sont beaucoup élevés et tous supérieurs à 0,97. Ils sont sensiblement égaux à 1. De plus, les erreurs de validation (RMSE) sont très faibles. Ces erreurs se rapprochent de 0. Les résultats de ce tableau prouvent que les modèles sont excellents et peuvent être utilisés pour faire de la prédiction.

III.3.3.2. Discussion

Les analyses spectrales des pommes de cajou montrent que les spectres présentent tous un comportement spectral similaire. Toutefois, la différence est à noter au niveau des intensités à certaines longueurs d'onde. Cette différence au niveau des intensités peut s'expliquer par le fait que certains constituants biochimiques des pommes de cajou interagissent avec le rayonnement lumineux en absorbant ou réfléchissant à certaines longueurs d'onde.

Les performances des modèles de prédiction ont été évaluées à l'aide des indicateurs statistiques suivants : le coefficient de détermination (R^2) qui représente essentiellement la proportion de variance expliquée de la réponse dans l'ensemble des données, l'erreur de prédiction définie comme l'erreur quadratique moyenne (RMSE) et l'erreur de validation croisée ou la performance prédictive (RMSECV ou RMSEP). Plus la valeur du coefficient de détermination d'un modèle est élevée, plus ce modèle a de chances de prédire le paramètre chimique de l'échantillon avec plus de précision. De plus, les faibles valeurs de RMSE ($RMSE < 1$) et la valeur élevée de R^2 indiquent un modèle performant [134].

Au niveau de l'entraînement des modèles de prédiction, dans l'ensemble, les coefficients de détermination sont très élevés et les erreurs de prédiction sont beaucoup faibles. Ces résultats révèlent que les modèles ont été bien entraînés. Comme le montre le tableau 4, le réseau de neurones a entraîné les modèles avec de très grandes précisions. Les coefficients de détermination sont supérieurs à 0,97 pour l'ensemble des paramètres. Plusieurs raisons principales expliquent les meilleures performances des modèles. En effet, l'étape de prétraitement des modèles constitue une étape très importante de l'analyse des données. Ainsi, le fait de réduire la gamme spectrale en supprimant les zones contenant des informations non utiles pourraient contribuer à améliorer la performance des modèles de prédiction. Notre intervalle spectral sélectionné pourrait expliquer les résultats satisfaisants que les modèles ont obtenus. Clément et *al.* (2008) [141] ont démontré que la région 360-750 nm convient à

la prédiction de la teneur en lycopène de la tomate. Aussi, les régions de courtes longueurs d'ondes ont la meilleure pénétration dans les tissus biologiques pour fournir des informations pertinentes du fait de leur forte énergie. Les travaux de Lammertyn et *al.* (2000) [142] ont montré que la profondeur de pénétration de la radiation lumineuse dépend de la longueur d'onde. De plus, la base de données a été augmentée afin de permettre aux modèles de mieux faire l'apprentissage. L'augmentation du nombre d'échantillons par l'algorithme de Wong et Easton représente une des raisons de l'amélioration des performances des différents modèles de prédiction [133].

La méthode proposée et utilisée dans la présente étude est simple sur le plan conceptuel, facile à comprendre et à utiliser. Les résultats indiquent un grand potentiel d'utilisation des données spectrales en combinaison avec la méthode de réseau de neurones artificiels. La méthode proposée comprend deux phases : une phase d'entraînement et une phase de validation. Mais, l'étape la plus importante ici consiste d'abord à prétraiter efficacement les données afin d'éliminer les spectres bruyants qui ne viennent pas des échantillons. De plus, Cette méthode utilise l'acquisition à distance de spectres optiques contrairement aux techniques spectroscopiques usuelles utilisées, comme celles des travaux de Shyam et *al.* (2014) [143] et de Zhang et *al.* (2019) [144]. Cette procédure rend alors la mesure moins fastidieuse. Le principal avantage de cette étude est que notre méthode exploite le potentiel de l'intelligence artificielle qui est en pleine expansion, notamment la technique du réseau de neurones artificiels pour établir les modèles de prédiction basés sur les spectres de réflectance du visible uniquement. Ce n'est pas pratique, car la plupart des techniques développées jusqu'à présent exploitent les données spectrales de la gamme de l'infrarouge et du visible, ou du proche infrarouge comme le font Olarewaju et *al.* (2012) [145] et Kusumiyati et *al.* (2019) [146]. Les résultats obtenus et évalués sur la base des paramètres statistiques tels que le coefficient de détermination R^2 et l'erreur quadratique moyenne (RMSE) montrent un grand potentiel pour l'utilisation des données spectrales combinés à

l'intelligence artificielle. À partir de l'analyse des résultats statistiques, nous pouvons affirmer que tous les modèles ont été bien entraînés. Aussi, tous les modèles ont obtenu de bonnes prédictions avec des résultats de la validation très satisfaisants. Mais dans la validation du modèle de la vitamine C, la valeur d'erreur de prédiction reste élevée du fait de la variation de la vitamine C au contact de l'oxygène de l'air. Au total, les modèles développés pour chaque paramètre sont acceptables et très encourageants. L'analyse spectroscopique de télédétection de la qualité des pommes de cajou associée aux réseaux de neurones artificiels donne un moyen efficace pour la détermination de la qualité des pommes. Cette analyse est faite à travers le modèle établi par le réseau qui permet d'obtenir la valeur des différents paramètres de qualité des échantillons à partir de leur signature spectrale. Le réseau de neurones se présente ainsi comme un outil plus facile d'utilisation et efficace de modélisation des paramètres de qualité.

Ce travail avait pour objectif de développer une technique optique pour estimer la qualité des pommes de cajou à partir de la spectroscopie de télédétection. La technique de modélisation non linéaire (le réseau de neurones artificiels) appliquée à la base de données spectrale nous a permis de développer des modèles de prédiction des paramètres de qualité. Cette étude a montré l'efficacité des modèles à prédire les paramètres de qualité des pommes de cajou avec des valeurs de coefficient de détermination R^2 très élevées pour l'entraînement et pour la validation. Les erreurs de prédiction des différents modèles confirment la précision de chaque modèle. Les résultats de cette étude apportent des éléments nouveaux en termes de prédiction de la qualité par la mesure spectroscopique de télédétection et permettent d'envisager cette mesure comme un outil complémentaire d'aide au suivi et à l'analyse de la maturité ou de la qualité de ces fruits. La spectroscopie de télédétection se révèle être un bon outil de mesure rapide et non destructive permettant d'accéder à la qualité globale des pommes de cajou.

III.4. Suivi de la maturité des papayes par la fluorescence

III.4.1. Méthodologie adoptée

Pour cette expérimentation, soixante-six (66) papayes visuellement non mûres ont été récoltées à la plantation. Les échantillons ont été sélectionnés selon leur calibre et l'absence de défaut, puis stockés au laboratoire durant la période des mesures (07 jours). Les différentes mesures sont effectuées sur des échantillons (en moyenne 09) prélevés de façon journalière dans le lot de 66 fruits. Les spectres de réflectance et de fluorescence ont été acquis sur chaque échantillon intact à l'aide du même dispositif expérimental complété d'un faisceau laser et d'un filtre. Les mesures de réflectance ont été effectuées dans la journée lorsque le soleil se trouvait au zénith afin d'avoir le maximum d'intensité. Les mesures de fluorescence ont été effectuées la nuit à partir de 19 h. Les analyses chimiques ont été réalisées sur les mêmes parties que celles qui ont servi aux mesures optiques. Chaque paramètre chimique (pH, teneur en sucre, teneur en eau et matière sèche) a été déterminé selon la méthode précédemment utilisée.

III. 4.2. Acquisitions des spectres

III. 4.2.1. Mesure de réflectance par télédétection passive

Les spectres de réflectance des papayes ont été acquis en mode réflectance, en plaçant les échantillons sur le porte-échantillon. La source lumineuse de cette mesure est le soleil. Les mesures sont effectuées à partir de midi afin d'avoir une intensité maximale pour éclairer directement la surface du fruit. Les spectres sont enregistrés à l'ordinateur connecté au spectromètre de longueur d'ondes allant de 345 à 1042 nm. Trois acquisitions spectrales sont effectuées par fruit.

L'objectif initial de notre projet était d'utiliser la spectroscopie de réflectance et de fluorescence pour analyser et prédire la qualité des fruits. Mais, nous avons dû avoir recours à la fluorescence pour la papaye, car nous avons été confrontés à une difficulté : Le rendement quantique de la fluorescence est en moyenne cent pour cent

(100%) plus petit que la réflectance et l'intensité de la réflectance est proportionnelle à la section efficace de l'objet qui réfléchit l'onde lumineuse.

Les premiers fruits analysés avaient une section efficace moins importante que les papayes. L'intensité lumineuse du soleil pouvait facilement pénétrer dans ces fruits, et, après plusieurs parcours aléatoires à l'intérieur du fruit, ressortir en collectant des informations sur la composition chimique interne de ces derniers. Cela pouvait être observé par les pics d'absorption sur les spectres de réflectances mesurés. Malheureusement, les pics d'absorption ont été quasiment absents sur les spectres de réflectance de la papaye présentés à la figure 23. Ces spectres ne pouvaient donc pas nous renseigner sur la composition chimique de la papaye. Ils étaient donc inexploitable.

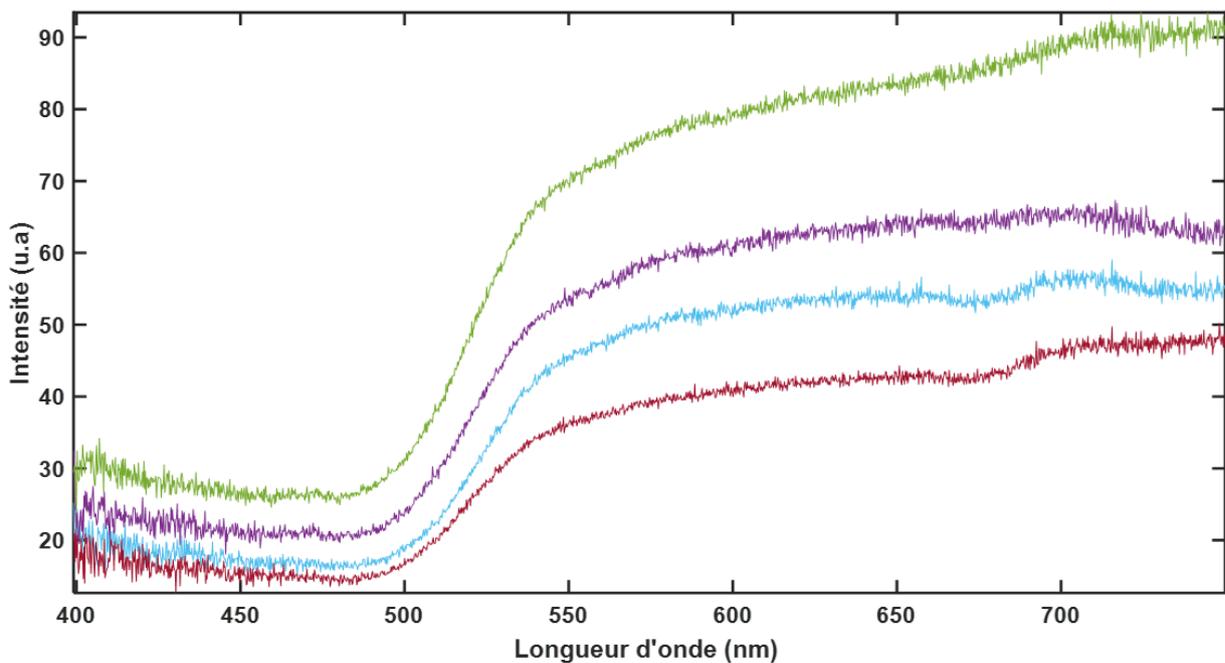


Figure 23: Spectres de réflectance des papayes.

Nous avons alors tiré profit d'une section efficace plus importante de la papaye pour enfin obtenir des spectres de fluorescence exploitables.

Cette situation n'a pas permis d'utiliser les données de la réflectance. Nous avons donc fait le choix d'utiliser les mesures de fluorescence pour démontrer les capacités de cette

technique en spectroscopie par télédétection lorsque la valeur de la section efficace permet d'obtenir des informations utiles sur la maturité.

III. 4. 2.2. Méthode d'acquisition des spectres de fluorescence

Les mesures de télédétection passives utilisent le soleil comme source d'illumination des échantillons. L'intensité de cette source varie au cours du temps en fonction de la position du soleil et de son intensité. Cela rend contraignant l'utilisation de cette technique. Dans ce paragraphe, nous présentons une alternative de mesure et de caractérisation des fruits de papayer par télédétection active. Il sera question de :

- mesurer les spectres de fluorescence ;
- développer une méthode d'analyse ;
- modéliser les paramètres de qualité des fruits.

Le dispositif d'acquisition est similaire au dispositif décrit au chapitre II. Toutefois, la source lumineuse naturelle a été remplacée par une diode laser de longueur d'onde 380 nm et de puissance 8 mW. La diode laser a été montée sur la lunette astronomique et le spectromètre sur le télescope. Un filtre passe-bas circulaire en verre de couleur rouge a été placé devant l'orifice de la fibre optique à l'aide d'un réducteur de diamètre. Le rôle du filtre est de supprimer les longueurs d'ondes en dessous de 650 nm, la longueur d'onde d'excitation. Les mesures ont été effectuées à partir de 19h00 pour réduire la contribution des sources parasites et augmenter le rapport signal/bruit. Le spectromètre est piloté à partir de l'ordinateur, par un logiciel d'acquisition, *SpectraSuite*, développé par la firme *Ocean Optics*. Ce logiciel est fourni lors de l'achat du spectromètre. L'ensemble télescope-lunette astronomique est aligné sur l'orifice d'une boîte noire placée à 50 m du dispositif expérimental. Pour éviter tout risque d'irradiation au faisceau laser, les échantillons de papaye sont placés devant l'orifice de la boîte noire avant d'allumer le laser une fois que personne ne se trouve dans le champ du faisceau laser. La configuration de l'acquisition a été effectuée à partir du *SpectraSuite*. Cette configuration concerne le temps d'intégration (*integration time*), le

nombre de spectre acquis (*scan to average*) et la largeur de la fenêtre du filtre (*boxcar*). Le tableau 6 présente les paramètres de configuration.

Tableau 6: Récapitulatif des paramètres de configuration du spectromètre pour la mesure du spectre de fluorescence.

Paramètre	Unité	Valeur
<i>Integration time</i>	s	3
<i>Scan to average</i>	Échantillon	5
<i>Boxcar</i>	-	2

III.4. 3. Méthode d'analyse des spectres de fluorescence

Les données de fluorescence acquises se présentent sous forme d'un fichier codé sur 16 bits. Chaque fichier contient 3648 valeurs d'intensités et à chaque valeur est associée une longueur d'onde allant de l'ultra-violet (345 nm) au proche infra-rouge (1042 nm). Les pics d'intensités de fluorescence apparaissent dans la bande spectrale 620 nm à 800 nm après excitation avec un laser de longueur d'onde 380 nm. La figure 23 (a) présente les intensités en fonction des longueurs d'ondes de fluorescence dans la bande spectrale d'émission de 44 échantillons de papaye. Ces intensités comprennent les intensités de fluorescence de la molécule de chromophore excitée, des intensités de bruit de fond dues aux diffusions des particules atmosphériques dans la direction de détection du système d'acquisition et aux bruits électroniques inhérents à tout dispositif électronique. Pour extraire l'intensité de fluorescence du chromophore qui a absorbé à la longueur d'onde de 380 nm, les intensités de fluorescence, I_{λ} , ont été filtrées en utilisant un filtre de convolution. Ensuite, ces intensités ont été normalisées, $I_{norm,\lambda}$, en utilisant l'intensité de fluorescence minimale, $I_{min,\lambda}$, l'intensité de fluorescence maximale, $I_{max,\lambda}$, selon l'équation III.4

$$I_{norm,\lambda} = \frac{I_{\lambda} - I_{min,\lambda}}{I_{max,\lambda} - I_{min,\lambda}} \quad \text{III.4}$$

La figure 24-(b) présente les spectres normalisés après filtrage.

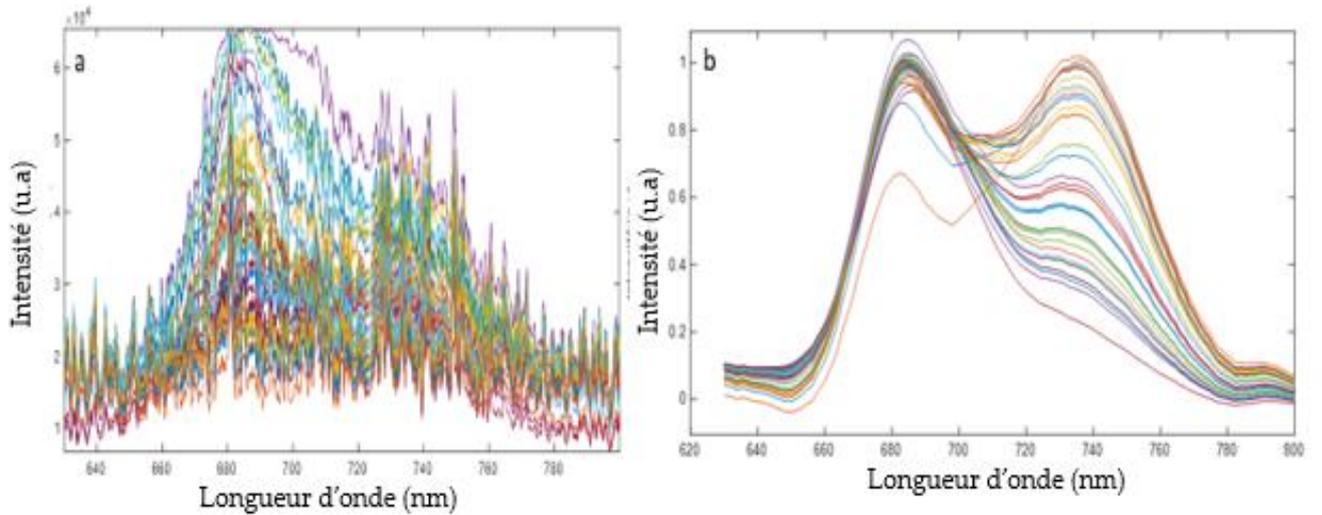


Figure 24: Spectres de fluorescence normalisés des papayes. (a) Spectres bruts de fluorescence, (b) Spectres de fluorescences filtrés par convolution.

Au cours de l'interaction laser-papaye, les intensités de fluorescence émises peuvent être réabsorbées par la papaye et réémettre une longueur d'onde en compétition d'intensité avec la première émission. Pour identifier ces deux types de spectres de fluorescence, une analyse non-supervisée a été effectuée. La démarche a consisté, d'abord, à la décomposition en valeurs singulières de la matrice I_f obtenue en regroupant en colonne les différents spectres de fluorescence, $I_{norm,\lambda}$. Cette décomposition a été effectuée en utilisant l'équation III.5.

$$I_{f,(n,p)} = U_{(n,n)} * S_{(n,p)} * V_{(p,p)}^T \quad \text{III. 5}$$

Dans cette équation, p représente le nombre de spectres de fluorescence des papayes, n représente le nombre de longueurs d'onde dans la bande spectrale d'émission de fluorescence du chromophore. La multiplication ordinaire est indiquée par le signe « * » et la transposée de matrice par « T ». Les valeurs singulières sont contenues dans la diagonale principale de S ; U et V représentent respectivement les matrices des vecteurs propres à gauche et à droite de I_f .

Une fois le nombre de composantes spectrales trouvées, nous avons procédé à l'analyse en classe. Puis, nous avons évalué la proportion relative de fluorescence dans chaque classe.

III. 4.4. Résultats et discussion

La figure 25 présente le graphique de décroissance de la valeur singulière en fonction du nombre de composantes spectrales. La valeur singulière décroît lorsque le nombre de composantes spectrales augmente. A partir des deux premières composantes spectrales, la courbe de décroissance tend vers une asymptote. Cela est dû aux spectres présentant les mêmes profils d'intensités dans le jeu de spectres de fluorescence mesurés et aux valeurs redondantes.

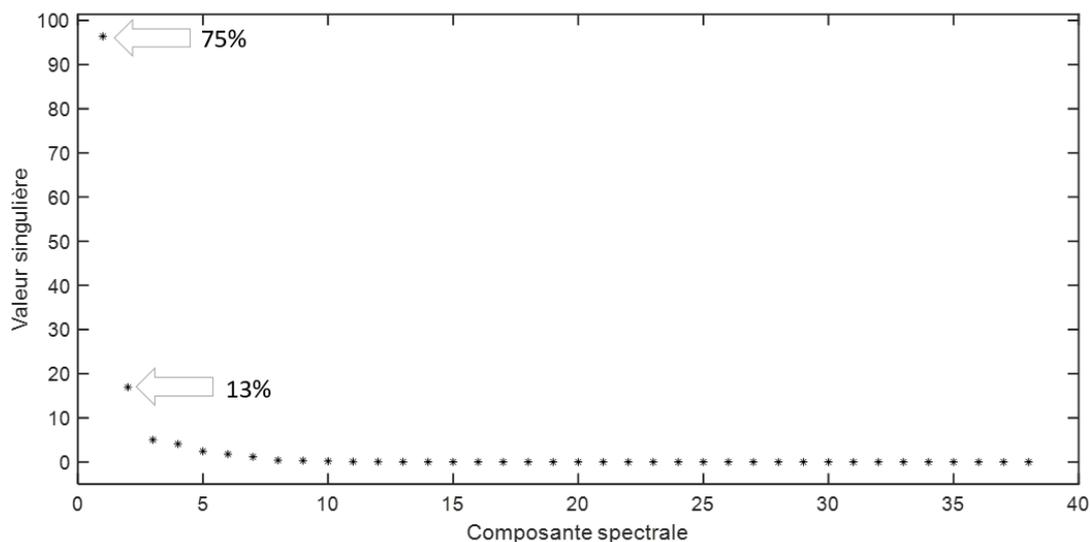


Figure 25: Valeur singulière en fonction de la composante spectrale.

La part d'inertie expliquée des deux premières composantes spectrales est indiquée à l'extrémité des flèches.

Les travaux de Kouakou et *al.* (2021) [147] ont montré que le nombre de composantes spectrales est étroitement lié au nombre optimal de classe. Pour obtenir le nombre optimal de classes, la part d'inertie de chaque valeur singulière a été calculée. Les parts d'inertie des deux premières composantes spectrales contribuent à 88%. Elles sont nécessaires pour expliquer la variabilité entre les différents spectres. Le nombre

optimal de classes est donc égal à 2. Par la suite, les distances euclidiennes entre les différents spectres de fluorescence ont été calculées et les spectres de fluorescence ont été regroupés en fonction de la similarité/dissimilarité de la distance euclidienne. La figure 26 présente les deux classes de spectres de fluorescence.

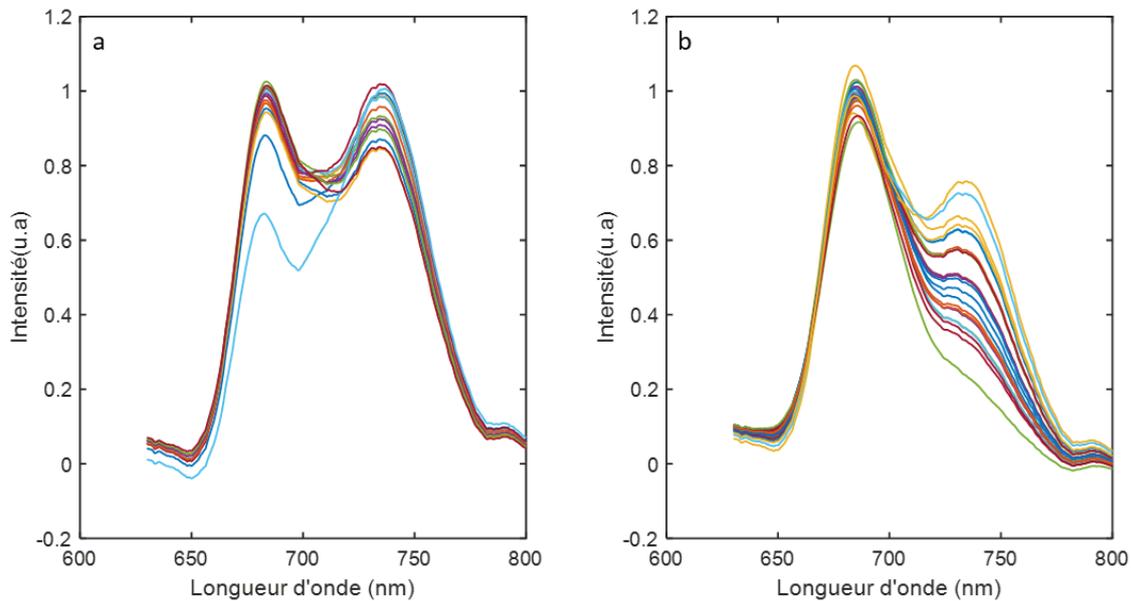


Figure 26: Spectres de fluorescence des papayes excitées à 380 nm. (a) Réabsorption de la longueur d'onde émise ; (b) Fluorescence du chromophore.

Le profil d'intensité de chaque classe spectrale présente deux gaussiennes d'intensités maximales aux longueurs d'ondes 683 nm et 735 nm. Les intensités maximales de la classe spectrale de la figure 26 (a) ont pratiquement la même valeur. Cela serait dû au phénomène de réabsorption de la longueur d'onde émise à 683 nm et à une réémission à la longueur d'onde de 735 nm qui s'additionne à l'émission précédente. Le spectre en vert de la figure 26 (a) a un pic à 683 nm plus faible que le pic à 735 nm. Cela pourrait être dû à l'immaturité du fruit de la papaye ou à une faible concentration de chromophore absorbant à 380 nm.

Les spectres de fluorescence de la figure 26 (b) présentent un pic à 683 nm plus important que le pic à 735 nm. Les spectres de fluorescence obtenus avec une longueur d'onde d'excitation de 380 nm présentent, en général, deux pics. Le premier pic se situe entre 685 nm et 690 nm et le second, moins important que le premier, dans le proche infrarouge, entre 735 nm et 740 nm.

La forme des spectres de fluorescence diffère très peu selon le stade de maturité. De même, pour un spectre de fluorescence donné, les variations des intensités sont très faibles d'un fruit à un autre en fonction de la teneur en chlorophylle. En effet, la bande de longueur d'onde 685-690 nm correspond à la fluorescence chlorophyllienne émise par le PS II tandis que la bande de longueur d'onde 715-740 nm est liée à la fluorescence chlorophyllienne dans le PS I et le PS II [148,149]. Le mécanisme de la photosynthèse consiste en un transfert d'énergie suite à l'absorption de la lumière par les pigments photosynthétiques et met en œuvre des réactions chimiques au niveau des photosystèmes. Ainsi, si l'activité photosynthétique est perturbée en cas de stress nutritionnel ou hydrique, par exemple, l'émission de photons est perturbée, la chaîne de transport est détériorée et la dissipation d'énergie (chaleur ou fluorescence) est augmentée.

La figure 27 présente la proportion de fluorescence dans chaque classe spectrale. Sur la figure 27(a), l'évolution du ratio $I_{\lambda 685}/I_{\lambda 730}$ est aléatoire. Cela montre que le phénomène de réabsorption est un phénomène aléatoire et n'est pas lié à l'état de maturité ou de dégradation du fruit. Par contre, le ratio $I_{\lambda 685}/I_{\lambda 730}$ de la figure 27(b) suit une loi linéaire décroissante. Cela traduit une diminution du taux de chlorophylle dans le fruit au cours du temps.

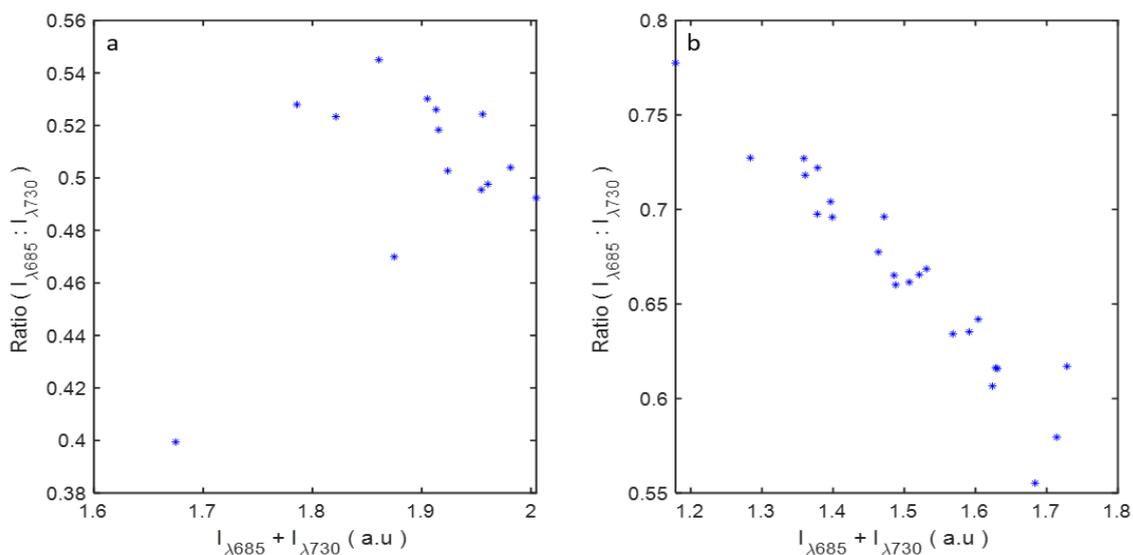


Figure 27: Proportion relative de fluorescence à 685 nm par rapport à la fluorescence totale.

Les variations des teneurs en chlorophylle étant associées au stade de maturité et de dégradation des fruits, ces ratios peuvent être des indicateurs des paramètres physico-chimiques mesurés. Ces ratios ont été utilisés pour construire un modèle de prédiction de la maturité des fruits et cela est l'objet du paragraphe suivant.

III. 4.5. Modélisation des paramètres physiques-chimiques à partir du réseau de neurones artificiels

Dans les phénomènes physiques, lorsque la loi de distribution n'est pas à priori connue ou dans le cas de phénomènes complexes, les réseaux de neurones artificiels sont, dans la plupart des cas, utilisés pour modéliser ce phénomène physique. Dans ce paragraphe, il est question de modéliser les paramètres physico-chimiques des échantillons de papaye à partir des ratios de fluorescence en utilisant les réseaux de neurones artificiels (RNA). En effet, le rapport des intensités de fluorescence présente une relation quasi linéaire avec l'activité photosynthétique du fruit. Cette activité étant elle-même corrélée à l'évolution du stade de maturité du fruit, nous avons donc mesuré les valeurs de fluorescence des papayes puis, nous avons calculé les rapports des intensités de fluorescence à 685 nm et à 730 nm pour modéliser les critères de suivi de maturité.

III. 4.5.1. Méthode d'analyse prédictive

Un RNA de régression multilinéaire a été développé à partir de l'algorithme de Levenberg-Marquardt. Le nombre de nœuds de couches cachées a un impact sur la taille et la capacité de généralisation du RNA. Plus le nombre de nœuds cachés est faible, plus la taille du réseau et la capacité de généralisation sont faibles. On parle de sous-ajustement. Si le nombre de nœuds cachés est trop important, les informations de non-régularité des échantillons d'apprentissage peuvent être mémorisées avec précision. C'est le phénomène de surajustement [150]. Par conséquent, le nombre de nœuds cachés a été déterminé à partir de l'équation III.6 :

$$\text{Noeuds} = \sqrt{p} + p_0 + \delta$$

III.6

p est le nombre de variables d'entrée, p_0 est le nombre de variables de sortie et δ est une constante comprise entre 1 et 10. Pour déterminer le nombre optimal de nœuds, nous avons fait varier δ de 1 à 10 et calculé l'erreur quadratique moyenne. La valeur minimale de l'erreur quadratique moyenne donne la valeur optimale de δ .

A l'entrée du réseau, nous avons les ratios de fluorescence à 685 nm et à la sortie du réseau les paramètres physico-chimiques mesurés et normalisés (PH, MS, TE, TS, Acidité). Ces deux variables ont été rééchantillonnées à 240 pour avoir un jeu de données statistiquement suffisant pour une modélisation. Pour l'entraînement du RNA, 80% des données en entrée ont été utilisées et 10% pour le test et 10% pour la validation. La fonction de transfert du RNA est la fonction tangente-sigmoïde.

III. 4.5.2. Résultats et discussion

La figure 28 présente l'évolution du résidu quadratique moyen (RMSE) en fonction de δ . La valeur optimale de δ est donnée par la valeur minimale de RMSE. Les valeurs optimales de nœuds obtenues pour l'acidité, MS, pH, TS et TE sont respectivement 9, 8, 9 et 9. Chaque point sur chacune des courbes indique la valeur optimale de δ .

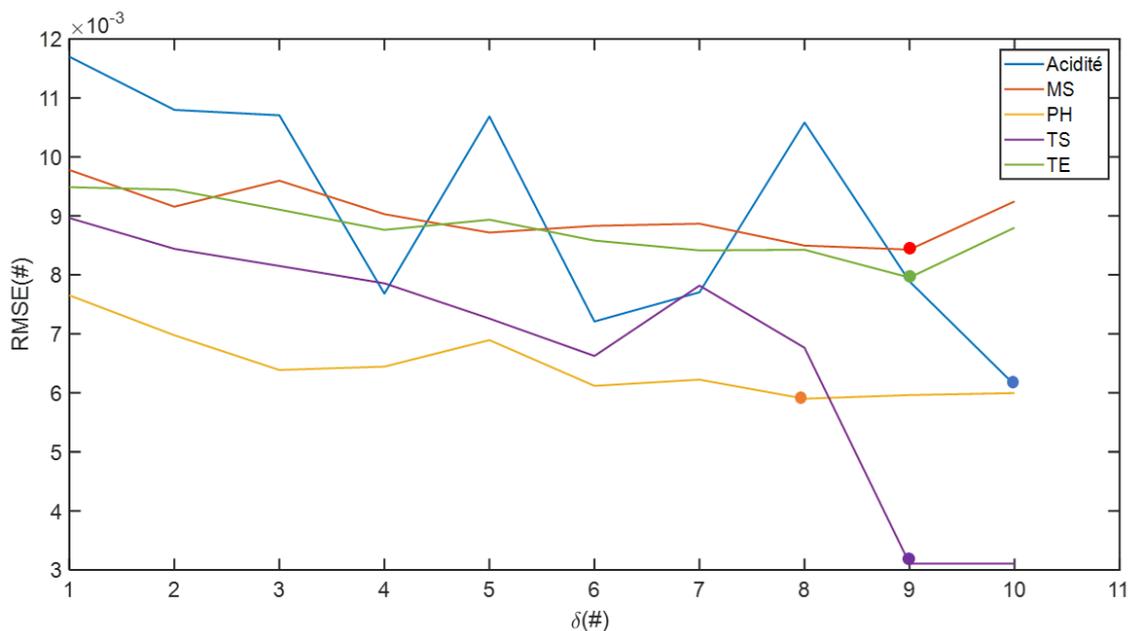


Figure 28: Variation du résidu quadratique en fonction de δ .

Le nombre de nœuds des couches cachées obtenu pour chaque paramètre est compris entre 8 et 10. Cela montre un réseau adapté avec une bonne capacité de généralisation. Le réseau est bien ajusté pour les calculs.

Le tableau 7 résume les résultats des modélisations RNA des différents paramètres de qualité.

Tableau 7: Valeurs des paramètres statistiques des modèles de prédiction des papayes

Paramètres de qualité	Paramètres statistiques	
	R	RMSE
Acidité	0,691	0,222
Matière sèche	0,696	0,192
pH	0,716	0,134
Teneur en sucre	0,956	0,095
Teneur en eau	0,729	0,196

Concernant l'acidité, la matière sèche, le pH et la teneur en eau, les prédictions ont donné des valeurs de R comprises entre 0,69 et 0,72. Ces valeurs montrent des linéarités acceptables des modèles établis. Pour le sucre, la prédiction a donné une valeur R de 0,95 très élevée. Cette valeur confirme que le modèle du sucre est excellent. L'erreur quadratique moyenne obtenue par le modèle de sucre est de 0,095. Ce modèle est très précis avec une valeur d'erreur très faible. Le modèles du sucre est alors plus précis que les modèles de l'acidité, de la matière sèche, du pH et de l'eau qui ont obtenues des erreurs respectives de 0,22 ; 0,19 ; 0,34 et 0,19.

La figure 29 présente les résultats de chaque étape du RNA des différents modèles de prédiction. Les colonnes 1 à 4 de la figure 29 représentent respectivement l'étape d'apprentissage, de validation, de test et de sortie du RNA pour chaque paramètre physico-chimique que sont l'acidité, la matière sèche (MS), le pH, la teneur en sucre (TS) et la teneur en eau (TE). Les valeurs en ordonnée représentent les valeurs physico-chimiques prédites par le modèle généré par le RNA et en abscisse, les valeurs expérimentales des paramètres physico-chimiques. Au-dessus de chaque sous-figure,

le coefficient de régression entre les valeurs prédites et les valeurs expérimentales est indiqué et l'erreur de prédiction à la sortie est indiqué au bas de ces droites. Les coefficients de régression des modèles d'apprentissage sont autour de 0,70 pour l'acidité, la MS, le pH et la TE. Cela montre qu'il y a une bonne corrélation entre les valeurs expérimentales et les valeurs prédites. Toutefois, les erreurs de prédiction sont toutes faibles. Cela signifie que les modèles ont de bonnes précisions. Cette apparente contraction met en lumière l'influence de la gamme spectrale sur la précision du modèle. En effet, un modèle de prédiction est fonction du traitement statistique appliqué et de la gamme spectrale sélectionnée. Pour la TS, nous constatons une corrélation forte avec un coefficient de régression égal à 0,95 et une grande précision avec une erreur de 0,095. Les observations sont similaires au cours de la validation, du test et de la sortie du RNA. Le modèle de la TS présente les coefficients de corrélation les plus élevés dans chacune des étapes du RNA. La corrélation forte observée dans le cas du modèle de la TS indique une forte concentration en caroténoïdes et une destruction de la chlorophylle dans le tissu de l'épiderme. Durant cette phase, la couleur de la papaye passe du vert au jaune et le saccharose est transformé en sucre réducteur. Aussi, l'acidité est réduite pendant la maturation des fruits, cela agit sur le pH des fruits qui va réagir faiblement avec les longueurs d'onde du visible. Ainsi, ces paramètres sont difficilement identifiés par les longueurs d'onde de la zone du visible. Les différentes transformations sont liées à l'état de maturité du fruit.

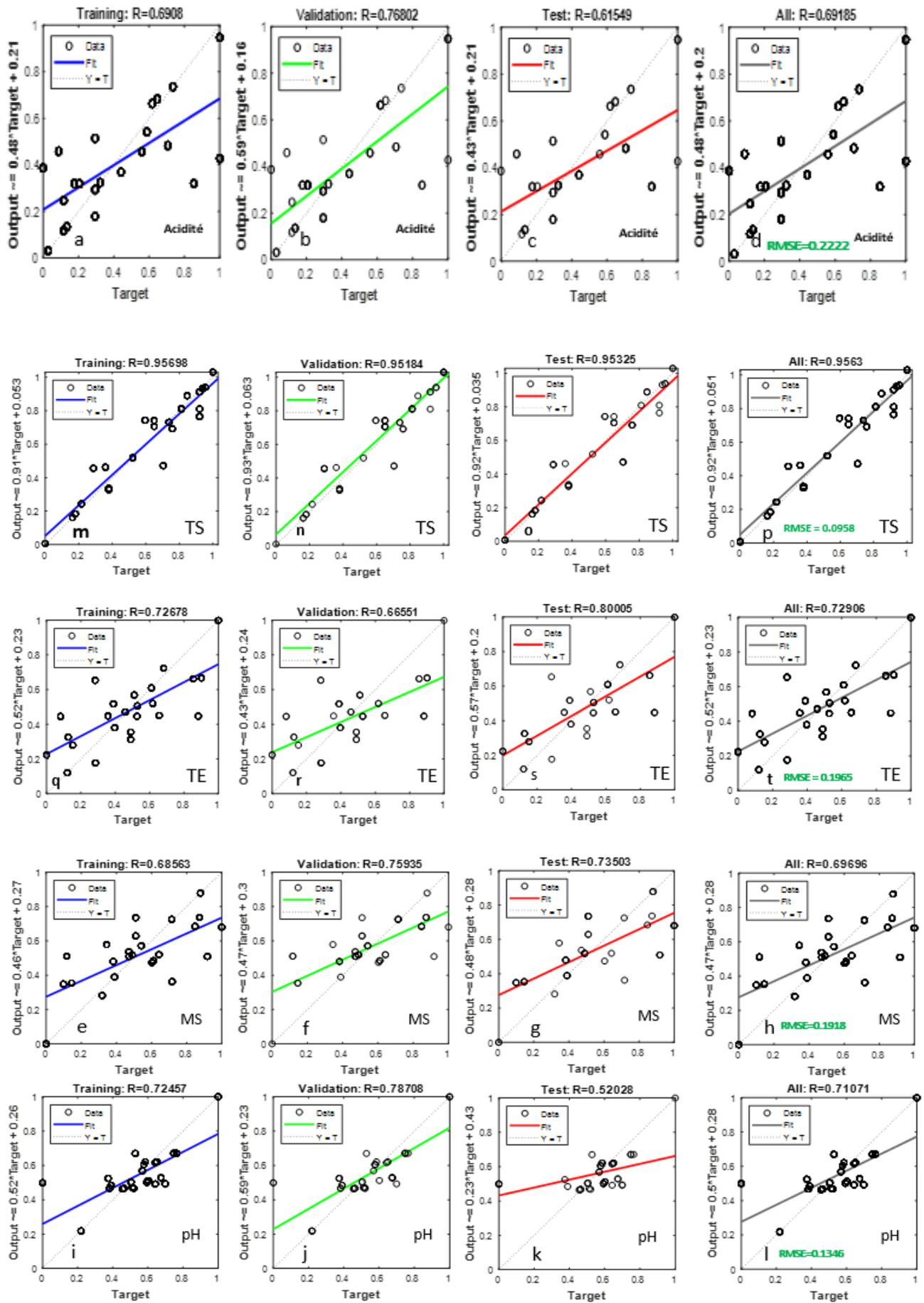


Figure 29: Valeurs expérimentales en fonction des valeurs de sortie des étapes du RNA.

Cette étude a permis de mettre en œuvre une nouvelle technique pour étudier l'état de maturité de la papaye. Nous avons mesuré la fluorescence de la papaye à la longueur d'onde d'excitation de 380 nm. A partir des données de fluorescence, nous avons classé les spectres en fonction de leur stade de maturité et nous avons établi des modèles de prédiction en utilisant le réseau de neurone artificiel. Les corrélations fortes observées entre les valeurs expérimentales et les valeurs prédites montrent que cette technique peut être utilisée pour suivre *in-situ* l'état de maturité des fruits et servir comme outil d'aide à la décision pour la récolte.

CONCLUSION

Dans ce travail de thèse, nous nous sommes attelés à explorer la faisabilité du développement de techniques optiques d'estimation de la qualité des fruits afin de les mettre à la disposition du domaine de l'agroalimentaire. Le développement de ces techniques est fondé sur les informations de fluorescence et de réflectance. Ces travaux ont mis en évidence la pertinence et l'efficacité de l'utilisation des mesures de fluorescence et de réflectance à travers l'atteinte de quatre objectifs spécifiques s'y rapportant. La stratégie mise en œuvre dans notre approche expérimentale a essentiellement consisté à l'utilisation des données spectrales de réflectance et de fluorescence pour estimer les paramètres de qualité des fruits. Pour exploiter les données, nous avons eu recours à deux techniques de modélisation : la méthode de régression PLS et le réseau de neurones artificiels. Cependant, les spectres acquis sur les fruits étant confrontés à la présence de plusieurs données bruitées et aberrantes, des méthodes statistiques de prétraitement ont été sélectionnées. Les résultats en sortie des modèles établis par la régression PLS et le réseau de neurones ont montré que l'exploitation des spectres de réflectance acquis à distance permet d'estimer efficacement les paramètres de qualité des fruits.

Dans le cas du suivi de la maturité des fruits, nous avons aussi prouvé qu'il est possible de suivre et d'estimer de façon non destructive par la spectroscopie de fluorescence la qualité des fruits à distance. L'analyse des résultats issus de ces travaux a montré que l'association de la technique de réseau de neurones artificiels et des données spectrales acquises est d'un grand apport dans l'évaluation de la qualité des fruits. Ainsi, l'observation des spectres optiques des fruits et des résultats des modèles a conduit à la conclusion que la spectroscopie de télédétection est une technique efficace de mesure de la qualité des fruits.

L'utilisation du réseau de neurones artificiels comme méthode d'analyse statistique constitue donc une méthode efficace de modélisation. Cependant, s'il est vrai que l'ensemble des travaux de cette étude suggère un remarquable potentiel de l'utilisation

des données de réflectance et de fluorescence acquises à distance, il n'en demeure pas moins que des travaux complémentaires soient entrepris afin d'améliorer les résultats.

Ceux-ci concernent particulièrement :

- La constitution d'une large banque de données spectrales acquises à différentes saisons pour plusieurs variétés de fruits ;
- l'étude comparative des deux méthodes d'analyse statistique utilisées (PLS, réseau de neurones) ;
- l'étude comparative des deux méthodes spectroscopiques (fluorescence et réflectance),
- l'augmentation du nombre de paramètres de qualité (la quantité de chlorophylle et d'autres pigments).

RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] GAJEWSKI, M. and ARASIMOWICZ D. (2004). Sensory quality of eggplant fruits (*solanum melongena* l.) as affected by cultivar and maturity stage. *polish journal of food and nutrition sciences*, 13/54 (3), 249–254.
- [2] CARBONELL, L., IZQUIERDO L. and CARBONELL I. (2007). Sensory analysis of Spanish mandarin juices. Selection of attributes and panel performance. *Food Quality and Preference*, 18, 329–341. Doi : 10.1016/j.foodqual.2006.02.008
- [3] CHRISTEN, DANILO, ROSSIER J., ÉVÉQUOZ C., NICOLLIER G., REY-SIGGEN J. et MORIN-AUDEBRAND L., (2011). Analyses sensorielles de variétés d'abricots. *Revue suisse Viticulture, Arboriculture, Horticulture*, 43 (3), 152–159.
- [4] POPULER, C. et LATEUR M. (1990). Variétés fruitières anciennes et résistantes aux maladies : comportement des variétés de pommier et de prunier RGF recommandées aux pépiniéristes. *Meded. Fac. Landbouwwet. Rijksuniv. Gent*, 55 (3a), 851–858.
- [5] SERRANO, M., GUILLEN F., MARTINEZ-ROMERO D., CASTILLO S. and VALERO D. (2005). Chemical constituents and antioxidant activity of sweet cherry at different ripening stages. *Journal of Agricultural and Food Chemistry*, 53, 2741–2745.
Doi: 10.1021/jf0479160
- [6] SAWADOGO-LINGANI, H. et TRAORE A. S. (2001). Composition chimique et valeur nutritive de la mangue Amelie. *Journal des Sciences*, 2, 1, 35 – 39
- [7] ASSEMAND, E., CAMARA F., KOUAME F., KONAN V. et KOUAME L. P. (2012). Caractérisation biochimique et évaluation sensorielle des fruits de plantain variété Agrin. *Journal of Applied Biosciences*, 60: 4438– 4447 ISSN 1997–5902
- [8] LU, Y., PEKERTI B. N., TOH Z. S., BROOM F., SAVAGE G., LIU S. Q. and HUANG D. (2017). Physico-chemical parameters and proanthocyanidin profiles of cranberries cultivated in New Zealand. *Journal of Food Composition and Analysis*, 63, 1–7, Doi.org/10.1016/j.jfca/2017.07.002

[9] JAMSHIDI, B. (2020). Ability of near-infrared spectroscopy for non-destructive detection of internal insect infestation in fruits: Meta-analysis of spectral ranges and optical measurement modes. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 225, 117479. Doi: 10.1016/j.saa.2019.117479.

[10] ZHANG, H., WANG J., YE S. and CHANG M. (2009). Application of Electronic Nose and Statistical Analysis to Predict Quality Indices of Peach. *Food and Bioprocess Technology*, 5(1), 65–72. Doi:10.1007/s11947-009-0295-7.

[11] ZHANG, W., PAN L., ZHAO X. and TU K. (2014). A study on soluble solids content assessment using electronic nose: persimmon fruit picked on different dates. *International Journal of Food Properties*, 19(1), 53–62.

Doi: 10.1080/10942912.2014.940535

[12] CLARK, C. J., MCGLONE V. A., REQUEJO C., WHITE A. and WOOLF A. B. (2003). Dry matter determination in “Hass” avocado by NIR spectroscopy. *Postharvest Biology and Technology*, 29(3), 301–308. Doi:10.1016/s0925-5214(03)00046-2.

[13] JHA, S. N., JAISWAL P., NARSAIAH K., GUPTA M., BHARDWAJ R. and SINGH A. K. (2012). Non-destructive prediction of sweetness of intact mango using near infrared spectroscopy. *Scientia Horticulturae*, 138, 171–175.

Doi: 10.1016/j.scienta.2012.02.031.

[14] ZHANG, Y., NOCK J. F., AL SHOFFE Y. and WATKINS C. B. (2019). Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy. *Postharvest Biology and Technology*, 151, 111–118.

Doi: 10.1016/j.postharvbio.2019.01.009.

[15] HASNAH AR, N., ARIS PURWANTO Y., BUDIASTRA I. W. and SOBIR S. (2019). Prediction of soluble solid content, vitamin C, total acid and firmness in astringent persimmon (*Diospyros kaki* L.) cv. Rendeu using NIR spectroscopy. *IOP Conference*

Series: *Materials Science and Engineering*, 557, 012086. Doi:10.1088/1757-899x/557/1/012086.

[16] KAWANO, S., WATANABE H. and IWAMOTO M. (1992). Determination of Sugar Content in Intact Peaches by Near Infrared Spectroscopy with Fiber Optics in Interactance Mode. *Journal Japan of Society Hortical Science*, 61(2):445-451.

[17] KHODABAKHSHIAN, R., EMADI B., KHOJASTEHPUR M., GOLZARIAN M. R. and SAZGARNIA A. (2016). Non-destructive evaluation of maturity and quality parameters of pomegranate fruit by visible/near infrared spectroscopy. *International Journal of Food Properties*, 20(1), 41–52. Doi :10.1080/10942912.2015.1126725.

[18] MAPPE, I., SIRET R., JOURJON F., BLIN M., TURBILLON C. et MEHINAGIC E. (2010). Spectroscopie visible proche infrarouge pour décrire les fruits et prédire leurs paramètres sensoriels. *Revue suisse Viticulture, Arboriculture, Horticulture*, 42 (4) : 248–255.

[19] CHEN, H., LIU Z., CAI K., XU L. and CHEN A. (2018). Grid search parametric optimization for FT-NIR quantitative analysis of solid soluble content in strawberry samples. *Vibrational Spectroscopy*, 94, 7–15. Doi: 10.1016/j.vibspec.2017.10.006.

[20] ELMASRY, G., WANG N., ELSAYED A. and NGADI M. (2007). Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering*, 81(1), 98–107. Doi: 10.1016/j.jfoodeng.2006.10.016

[21] BAIANO, A., TERRACONE C., PERI G. and ROMANIELLO R. (2012). Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. *Computers and Electronics in Agriculture*, 87, 142–151. Doi: 10.1016/j.compag.2012.06.002.

[22] MEI, L., KONG Z. and GUAN P. (2018). Implementation of a violet Scheimpflug lidar system for atmospheric aerosol studies. *Optics Express*, 26(6), A260. Doi:10.1364/oe.26.00a260.

- [23] KOUAKOU, B. K. JANSSON S., BRYDEGAARD M. and ZOUEU J. T. (2020). Entomological Scheimpflug lidar for estimating unique insect classes *in-situ* field test from Ivory Coast. *OSA Continuum*, 3, 2362-2371.
- [24] ATZBERGER, C. (2013). Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs. *Remote Sensing*, 5(2), 949-981.
- [25] ORYNBAIKYZY, A., GESSNER U. and CONRAD C. (2019). Crop type classification using a combination of optical and radar remote sensing data: a review. *International journal of remote sensing*, 40, 6553-6595.
- [26] DEEPA U PUJAR, UDACHAPPA U PUJAR, SHRUTHI CR. ANURADHA WADAGAVE and MADHU CHULAKI (2017). Remote sensing in fruit crops. *Journal of Pharmacognosy and Phytochemistry*, 6, 2479-2484.
- [27] WALLACE, M. B., WAX A., ROBERTS D. N. and GRAF R. N. (2009). Reflectance Spectroscopy. *Gastrointestinal Endoscopy Clinics of North America*, 19 (2): 233–242.
Doi: 10.1016/j.giec.2009.02.008.
- [28] SEREBRENNIKOVA, Y. M., SMITH J. M., HUFFMAN D. E., LEPARC G. F. and GARCIA-RUBIO L. H. (2008). Quantitative interpretations of Visible-NIR reflectance spectra of blood. *Optics Express*, 16(22), 18215. doi:10.1364/oe.16.018215.
- [29] HUFFMAN, D. E., SEREBRENNIKOVA Y. M., SMITH J. M., LEPARC, G. F. and GARCÍA-RUBIO L. H. (2009). New method for the detection of micro-organisms in blood: application of quantitative interpretation model to aerobic blood cultures. *Journal of Biomedical Optics*, 14(3), 034043. Doi:10.1117/1.3156837.
- [30] TOMINAGA, S. and WANDELL B. A. (1990). Component estimation of surface spectral reflectance. *Journal of the Optical Society of America A*, 7(2), 312-317.
Doi :10.1364/josaa.7.000312.

- [31] GERAUDIE, V., ROGER J. M. et OJEDA H. (2010). Développement d'un appareil permettant de prédire la maturité du raisin par Spectroscopie proche infra-rouge. *Revue française d'Oenologie*, 240 p. 2 – 8.
- [32] CHICEA, D. and TURCU I. (2005). A random walk Monte Carlo approach to simulate multiple light scattering on biological suspensions. *Romanian Reports in physics*, 57,3, 418-425.
- [33] ATKINS, P. W., MONNET M. et DE PAULA J. (2004). Chimie Physique. 2e édition. De boeck, 794p.
- [34] SAIDI, I. S., JACQUES S. L. and TITTEL F. K. (1995). Mie and Rayleigh modeling of visible-light scattering in neonatal skin. *Applied Optics*, 34(31) : 7410-7418.
Doi :10.1364/ao.34.007410
- [35] LAUG, M. (1980). Traitement optique du signal et des images : bases théoriques et applications. Cepaduès édition, 208p
- [36] UNGUT, A., GREHAN G., and GOUESBET G. (1981). Comparisons between geometrical optics and Lorenz-Mie theory. *Applied Optics*, 20(17), 2911.
Doi :10.1364/ao.20.002911.
- [37] MIE, G. (1908). Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Annalen Der Physik*, 330(3), 377–445. Doi :10.1002/andp.19083300302.
- [38] MONDELAIN, D. (2001). Lidar linéaire et non-linéaire dans l'infrarouge moyen. *Thèse de doctorat d'état de l'Université Claude Bernard*. 154p.
- [39] PANDA, S. S., HOOGENBOOM G. and PAZ J. O. (2010). Remote Sensing and Geospatial Technological Applications for Site-specific Management of Fruit and Nut Crops: A Review. *Remote Sensing*, 2(8), 1973–1997. Doi:10.3390/rs2081973.
- [40] XIE, H., TIAN Y. Q., GRANILLO J. A. and KELLER G. R. (2007). Suitable remote sensing method and data for mapping and measuring active crop fields. *International Journal of Remote Sensing*, 28(2), 395–411. Doi:10.1080/01431160600702673.

- n [41] NOCK, C. A., VOGT R. J. and BEISNER B. E. (2016). Functional Traits. *eLS*, 1–8. Doi :10.1002/9780470015902.a0026282.
- [42] KOUAKOU, A. K., BAGUI O. K., AGNEROH T. A., SORO A. P. and ZOUEU J. T. (2016). Cucumber mosaic virus detection by artificial neural network using multispectral and multimodal imagery. *Optik International Journal for Light and Electron Optics*, 127(23), 11250–11257. Doi: 10.1016/j.ijleo.2016.09.035.
- [43] SANGARE, M., BAGUI O.K., TRAORE I., BABANA A. H., BA A. and ZOUEU J.T. (2015). Discrimination de différentes sous Phénotypes du *Ralstonia Solanacearum* dans une feuille de tabac par imagerie multi-spectrale. *Afrique Science*. 11, 95–103.
- [44] OKAMURA, N. K., SHIMOMACHI T., TAKEMASA T. and TAKAKURA T. (2001). Nondestructive Detection of Water Stress in Tomato Plants by NIR Spectroscopy. *Environment Control in Biology*, 39(2), 75–85. Doi:10.2525/ecb1963.39.75.
- [45] DIBI, W. G., FOTSO B., BROU C. Y., ZOUEU J. T., ZEZE A. and BOSSON J. (2016). Fluorescence and Reflectance Spectroscopy for Early Detection of Different Mycorrhized Plantain Plants. *Applied Physics Research*, 8(3), 17. Doi:10.5539/apr.v8n3p17.
- [46] CORTES, V., BLASCO J., ALEIXOS N., CUBERO S. and TALENS P. (2019). Monitoring strategies for quality control of agricultural products using visible and near-infrared spectroscopy: A review. *Trends in Food Science & Technology*, 85, 138–148. Doi: 10.1016/j.tifs.2019.01.015.
- [47] NORDEY, T., DAVRIEUX F. and LECHAUDEL M. (2019). Predictions of fruit shelf life and quality after ripening: Are quality traits measured at harvest reliable indicators? *Postharvest Biology and Technology*, 153, 52–60. Doi: 10.1016/j.postharvbio.2019.03.011.

- [48] SARANWONG, I., SORNSRIVICHAI J. and KAWANO S. (2003). On-Tree Evaluation of Harvesting Quality of Mango Fruit Using a Hand-Held NIR Instrument. *Journal of Near Infrared Spectroscopy*, 11(4), 283–293. Doi:10.1255/jnirs.374.
- [49] JIN, X., LI S., ZHANG W., ZHU J. and SUN J. (2020). Prediction of Soil-Available Potassium Content with Visible Near-Infrared Ray Spectroscopy of Different Pretreatment Transformations by the Boosting Algorithms. *Applied Sciences*, 10(4), 1520. Doi:10.3390/app10041520.
- [50] YANG, M., XU D., CHEN S., LI H. and SHI Z. (2019). Evaluation of Machine Learning Approaches to Predict Soil Organic Matter and pH Using vis-NIR Spectra. *Sensors*, 19(2), 263. Doi:10.3390/s19020263.
- [51] ZHU, H., CHU B., FAN Y., TAO X., YIN W. and HE Y. (2017). Hyperspectral Imaging for Predicting the Internal Quality of Kiwifruits Based on Variable Selection Algorithms and Chemometric Models. *Scientific Reports*, 7(1) : 7845. Doi :10.1038/s41598-017-08509-6.
- [52] AGASSOUNON, D. T. M., GOMEZ S., TCHOBO F., SOUMANOU M. and TOUKOUROU F. (2012). Essai de conservation de la tomate par la technique de la déshydratation imprégnation par immersion (DII). *International Journal of Biological and Chemical Sciences*, 6(2). Doi:10.4314/ijbcs. v6i2.10.
- [53] ABDUL AZIZ, N.A., WONG L.M., BHAT R. and CHENG L.H. (2011). Evaluation of processed green and ripe mango peel and pulp flours (*Mangifera indica* var. Chokanan) in terms of chemical composition, antioxidant compounds and functional properties. *Journal of the Science of Food and Agriculture*, **92**, 557–563. Doi:10.1002/jsfa.4606.
- [54] WILLS, R.B.H., MCGLASSON W.B., GRAHAM D., LEE T.H. and HALL E.G., (1989). Post-harvest. An introduction to the physiology and handling of fruit and vegetables. Oxford London Edinburg, Boston Melbourne, pp.16-37.

- [55] DAVIES, J. N., HOBSON G. E. and MCGLASSON W. B. (1981). The constituents of tomato fruit — the influence of environment, nutrition, and genotype. *C R C critical reviews in food science and nutrition*, 15(3), 205–280. Doi:10.1080/10408398109527317.
- [56] MERZLYAK, M. N., SOLOVCHENKO A. E. and GITELSON A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biology and Technology*, 27(2), 197–211. Doi :10.1016/s0925-5214(02)00066.
- [57] FÉRET, J.-B. (2009). Apport de la modélisation pour l'estimation de la teneur en pigments foliaires par télédétection. *Thèse de doctorat de l'Université Pierre et Marie Curie*. 214p.
- [58] LICHTENTHALER, H. K. (1987). Chlorophylls and carotenoids: pigments of photosynthetic biomembranes. *Methods in Enzymology*, 148: 350-382. Doi:10.1016/0076-6879(87)48036-1.
- [59] http://perso.numericable.fr/cce2015/recits2016/sapcb_lumiere_vie/texte.html. (Consulté le 11 septembre 2019).
- [60] KUSUMIYATI, Y. H., INE E. P. and AGUS A. M. (2021). Multi-product calibration model for soluble solids and water content quantification in Cucurbitaceae family, using visible/near-infrared spectroscopy. *Heliyon* 7(8), e076772. Doi: 10.1016/j.heliyon.2021.e07677.
- [61] GOMEZ, A. H., HE Y. and PEREIRA A. G. (2006). Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *Journal of Food Engineering*, 77(2), 313–319. Doi: 10.1016/j.jfoodeng.2005.06.036.
- [62] AQEEL, M., JAMIL M. and YUSOFF I. (2011). Introduction to Remote Sensing of Biomass. *Biomass and Remote Sensing of Biomass*. Doi:10.5772/16462.
- [63] SUN, X., ZHANG H. and LIU Y. (2009). Nondestructive assessment of quality of Nanfeng mandarin fruit by a portable near infrared spectroscopy. *International Journal*

Agriculture and Biology Engineering, 2(1), 65–71. Doi: 10.3965/j.issn.1934-6344.2009.01.065-071.

[64] CEN, H. and HE Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, 18(2), 72–83. Doi : 10.1016/j.tifs.2006.09.003.

[65] GROTE, M, CADOT Y., POUSSIER A., LOONIS D., PIÉTRI E., DUPRAT F. et BARBEAU G. (2001). Détermination du degré de maturité des baies de raisin par des mesures physiques : Aspects métrologiques. *Journal International des Sciences de la Vigne et du Vin*, 35, 2, 87-98.

[66] ASKOURA, M. L. (2016). Étude expérimentale et théorique de la propagation de la lumière dans les tissus de la pomme. Application à un procédé optique de spectroscopie laser résolue dans l'espace. Biophysique [physics.bio-ph]. *Université d'Angers, Français*. <NNT : 2016VENAM79356>. <Tel-01393074>.

[67] DOSSOU J., SOULÉ I. et MONTCHO M. (2007). Evaluation des caractéristiques physico-chimiques et sensorielles de la purée de tomate locale produite à petite échelle au Bénin. *Tropicultura*, 25, 2, 119-125.

[68] SARANWONG, S., SORNSRIVICHAI J. and KAWANO S. (2004). Prediction of ripe-stage eating quality of mango fruit from its harvest quality measured nondestructively by near infrared spectroscopy. *Postharvest Biology and Technology*, 31(2), 137–145. Doi: 10.1016/j.postharvbio.2003.08.007.

[69] TIWARI, G., SLAUGHTER D. C. and CANTWELL M. (2013). Nondestructive maturity determination in green tomatoes using a handheld visible and near infrared instrument. *Postharvest Biology and Technology*, 86, 221–229. Doi: 10.1016/j.postharvbio.2013.07.009.

[70] BEGHI, R., GIOVENZANA V., TUGNOLO A. and GUIDETTI R. (2017). Application of visible/near infrared spectroscopy to quality control of fresh fruits and vegetables in large-scale mass distribution channels: a preliminary test on carrots and

tomatoes. *Journal of the Science of Food and Agriculture*, 98(7), 2729–2734. Doi:10.1002/jsfa.8768.

[71] CUBERO, S., ALEIXOS N., MOLTO E., GOMEZ-SANCHIS J. and BLASCO J. (2010). Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables. *Food and Bioprocess Technology*, 4(4), 487–504. Doi:10.1007/s11947-010-0411-8.

[72] BARNES, R. J., DHANOA M. S. and LISTER S. J. (1989). Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy*, 43(5), 772–777. Doi :10.1366/0003702894202201.

[73] Wiki Stat, Analyse en Composantes Principales (ACP). (Septembre 2016).

[74] WOLD, H. (1975). Soft Modelling by Latent Variables: The Non-Linear Iterative Partial Least Squares (NIPALS) Approach. *Journal of Applied Probability*, 12(S1), 117–142. Doi :10.1017/s0021900200047604.

[75] TENENHAUS, M. (1998). La régression PLS : théorie et pratique. Editions Technip, 252p.

[76] WOLD, S., SJÖSTRÖM M. and ERIKSSON L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. Doi:10.1016/s0169-7439(01)00155-1.

[77] ANDERSSON, M. (2009). A comparison of nine PLS1 algorithms. *Journal of Chemometrics*, 23(10), 518–529. Doi :10.1002/cem.1248.

[78] RUMELHART, D. E., HINTON, G. E., and WILLIAMS, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536. Doi:10.1038/323533a0.

[79] KALIRAJ, G. and BASKAR S. (2010). An efficient approach for the removal of impulse noise from the corrupted image using neural network based impulse detector. *Image and Vision Computing*, 28(3), 458–466. Doi: 10.1016/j.imavis.2009.07.007.

- [80] AGRAWAL, R. K. and BAWANE N. G. (2014). Optimized neural network for classification of multispectral images. *ACEEE International Journal on Signal and Image Processing*, 5(1): 65-70.
- [81] COULIBALY, P., ANCTIL F. and BOBEE B. (1999). Préviation hydrologique par réseaux de neurones artificiels : état de l'art. *Canadian Journal of Civil Engineering*, 26(3), 293–304. Doi :10.1139/l98-069.
- [82] KOUASSI, E. K.A., SORO Y., VACA-GARCIA C., KOUASSI B. Y. (2017). Valorisation des déchets d'agroressources par bioproduction d'acide citrique. *Journal de la Société Ouest-Africaine de Chimie*, 044 : 36 – 42.
- [83] DADZIE B.K. et ORCHARD J.E. (1997). Evaluation post-récolte des hybrides de bananiers et bananiers plantain : critères et méthodes. *Guides Techniques Inibap*.
- [84] AFNOR AOAC (1990). Official Methods of Analysis, (15th Edn). *Association of Official Analytical Chemists*: Washington DC. 774p.
- [85] HACHANI, S., HAMIA C., BOUKHALKHAL S., SILVA A. M. S., DJERIDANE A. and YOUSFI M. (2018). Morphological, physico-chemical characteristics and effects of extraction solvents on UHPLC-DAD-ESI-MSn profiling of phenolic contents and antioxidant activities of five date cultivars (*Phoenix dactylifera* L.) growing in Algeria. *NFS Journal*. Doi: 10.1016/j.nfs.2018.10.001.
- [86] TIRKEY, B., PAL U. S., BAL L. M., SAHOO N. R., BAKHARA C. K. and PANDA M. K. (2014). Evaluation of physico-chemical changes of fresh-cut unripe papaya during storage. *Food Packaging and Shelf Life*, 1(2), 190–197. Doi: 10.1016/j.fpsl.2014.02.002.
- [87] PARK, S. K. and HUCK F. O. (1977). Estimation of spectral reflectance curves from multispectral image data. *Applied Optics*, 16(12), 3107-3114. Doi:10.1364/ao.16.003107.
- [88] NOORDAM, J. C., VAN DEN BROEK W. H. and BUYDENS L. M. (2005). Detection and classification of latent defects and diseases on raw French fries with multispectral imaging. *Journal of the Science of Food and Agriculture*, 85(13), 2249–2259. Doi:10.1002/jsfa.2226.

- [89] ZHANG, B., DAI D., HUANG J., ZHOU J., GUI Q. and DAI F. (2017). Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: A review. *Critical Reviews in Food Science and Nutrition*, 58(12), 2099–2118. Doi:10.1080/10408398.2017.1300789.
- [90] CHRISTEN D., BESSE S., GUELAT L.-M., COUTANT C., DUCRUET J., ROSSIER J. et CAMPS C. (2011). Outils d'aide à la décision pour la récolte des abricots Luizet à distiller. *Revue Suisse Viticulture, Arboriculture, Horticulture*, 44 (2), 114–119.
- [91] PURWANTO, Y. A., SARI H. P. and BUDIASTRA I W. (2015). Effects of preprocessing techniques in developing a calibration model for soluble solid and acidity in 'Gedong Gincu' mango using NIR spectroscopy. *International Journal of Engineering and Technology*, 7(5):1921-1927.
- [92] VESALI, F., GHARIBKHANI M. and KOMARIZADEH M. H. (2011). An approach to estimate moisture content of apple with image processing method. *AJCS*, 5(2):111-115.
- [93] HRUSCHKA, W. R. (1987). Data analysis: Wavelength selection methods. In: Williams P., Norris K. (Eds.), *Near-infrared technology in the agricultural and food industries*, St. Paul, MN: *American Association of Cereal Chemists*, pp. 35–55.
- [94] RIEPPO, L., SAARAKKALA S., NÄRHI T., HELMINEN H. J., JURVELIN J. S. and RIEPPO J. (2012). Application of second derivative spectroscopy for increasing molecular specificity of fourier transform infrared spectroscopic imaging of articular cartilage. *Osteoarthritis and Cartilage*, 20(5), 451–459. Doi: 10.1016/j.joca.2012.01.010.
- [95] ZHU, H., CHU B., FAN Y., TAO X., YIN W. and HE Y. (2017). Hyperspectral Imaging for Predicting the Internal Quality of Kiwifruits Based on Variable Selection Algorithms and Chemometric Models. *Scientific Reports*, 7(1). Doi:10.1038/s41598-017-08509-6.

- [96] CAO, D.-S., LIANG Y.-Z., XU Q.-S., LI H.-D. and CHEN X. (2009). A new strategy of outlier detection for QSAR/QSPR. *Journal of Computational Chemistry*, 31(3): 592-602. Doi :10.1002/jcc.21351.
- [97] YUN, Y.-H., WANG W.-T., TAN M.-L., LIANG Y.-Z., LI H.-D., CAO D.-S. and XU Q.-S. (2014). A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Analytica Chimica Acta*, 807, 36–43. Doi: 10.1016/j.aca.2013.11.032.
- [98] NACHAR N. (2008). The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology*, 4(1), 13-20.
- [99] KENNARD, R. W. and STONE L. A., (1969). Computer Aided Design of Experiments. *Technometrics*, 11(1), 137–148. Doi:10.1080/00401706.1969.10490666.
- [100] YAO, B., ASSIDJO E., ADO G., AZZARO-PANTEL C., DAVIN A. (2007). Modelling and Optimization of M-cresol Isopropylation for Obtaining N-thymol: Combining a Hybrid Artificial Neural Network with a Genetic Algorithm. *International Journal of Chemical Reactor. Engineering*, 5 (1), 87 - 97.
- [101] NGORAN, E. B.Z., ASSIDJO N. E., KOUAMÉ P., DEMBELE I., YAO B. (2009). Modelling of Osmotic Dehydration of Mango (*Mangifera Indica*) by Recurrent Artificial Neural Network and Experimental Design. *Research Journal of Agriculture and Biological Sciences*, 5(5), 754-761.
- [102] GROSS, L., THIRIA S., FROUIN R. and MITCHELL B. G. (2000). Artificial neural networks for modeling the transfer function between marine reflectance and phytoplankton pigment concentration. *Journal of Geophysical Research: Oceans*, 105(C2), 3483–3495. Doi:10.1029/1999jc900278.
- [103] SHOKOUFE, T., TAYEBE B. L. and REZA R. (2013). Applying Neural Network to Dynamic Modeling of Biosurfactant Production Using Soybean Oil Refinery Wastes Iranica. *J. Energy & Environ.*, 4 (2): 161-170.

- [104] MALLEY D. F., MARTIN P. D. and BEN-DOR E., (2004). Application in analysis of soils. *Near-infrared spectroscopy in agriculture, (nearinfraredspe)*,729-784.
- [105] ALENAZI, M. M., SHAFIQ M., ALSADON A. A., ALHELAL I. M., ALHAMDAN A. M., SOLIEMAN T. H. I. and AL-SELWEY W. A. (2020). Improved functional and nutritional properties of tomato fruit during cold storage. *Saudi Journal of Biological Sciences*. Doi: 10.1016/j.sjbs.2020.03.026.
- [106] DE OLIVEIRA, G. A., DE CASTILHOS F., RENARD C. M.-G. C. and BUREAU S. (2014). Comparison of NIR and MIR spectroscopic methods for determination of individual sugars, organic acids and carotenoids in passion fruit. *Food Research International*, 60, 154–162. Doi: 10.1016/j.foodres.2013.10.051.
- [107] KLEE, H. J. and GIOVANNONI J. J. (2011). Genetics and Control of Tomato Fruit Ripening and Quality Attributes. *Annual Review of Genetics*, 45(1), 41–59. Doi:10.1146/annurev-genet-110410-132507.
- [108] WLODARSKA, K., KHMELINSKII I. and SIKORSKA E., (2018). Evaluation of Quality Parameters of Apple Juices Using Near-Infrared Spectroscopy and Chemometrics. *Journal of Spectroscopy*, 1–8. Doi:10.1155/2018/5191283.
- [109] GOVERNICI, J. L., FARIA R. M., TININI R. C. D. R. and MEDEROS B. J.T. (2017). Tomatoes Maturation Analysis with Reflectance Spectral Images. *Journal of Agricultural Science and Technology B* 7, 426-431. Doi: 10.17265/2161-6264/2017.06.007.
- [110] WILLIAMS, P. and NORRIS, K. H., (2001). Variable affecting near infrared spectroscopic analysis. In P. Williams and K. H. Norris (Eds.), *Nearinfrared technology in the agriculture and food industries*, 2nd ed., pp. 171–185.
- [111] KOUAKOU A. K., BAGUI O. K., AGNEROH T. A., SORO A. P. and ZOUEU J. T. (2016). Cucumber mosaic virus detection by artificial neural network using multispectral and multimodal imagery. *Optik - International Journal for Light and Electron Optics*, 127(23), 11250–11257. Doi: 10.1016/j.ijleo.2016.09.035.

- [112] WÜLFERT, F., KOK W.T. and SMILDE A.K. (1998). Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Analytical Chemistry*, 70(9), 1761–1767. Doi :10.1021/ac9709920.
- [113] CAMPS, C. (2010). Mesure non destructive de la qualité des tomates par spectroscopie NIR. *Revue Suisse Viticulture, Arboriculture, Horticulture*, 42 (5), 298 -303.
- [114] WILLIAMS P. C., (2001). Implementation of near-infrared technology. *Near-infrared technology in the agricultural and food industries*, 2, 143-167.
- [115] PHUANGSOMBUT K., PHUANGSOMBUT A. and TERDWONGWORAKUL A. (2020). Empirical approach to improve the prediction of soluble solids content in mango using near-infrared spectroscopy. *International Food Research Journal*, 27(2): 217 – 223.
- [116] LU, R. (2001). Predicting firmness and sugar content of sweet cherries using near-infrared diffuse reflectance spectroscopy. *Transactions of the ASAE*, 44(5).
Doi :10.13031/2013.6421.
- [117] LEIVA-VALENZUELA, G. A., LU R. and AGUILERA J. M. (2013). Prediction of firmness and soluble solids content of blueberries using hyperspectral reflectance imaging. *Journal of Food Engineering*, 115(1), 91–98. Doi: 10.1016/j.jfoodeng.2012.10.001.
- [118] SLAUGHTER D. C. and CRISOSTO C. H. (1998). Nondestructive internal quality assessment of kiwi fruit using near-infrared spectroscopy. *Seminars in Food Analysis*, 3, 131-140.
- [119] ZHANG, X., LIU F., HE Y. and Gong X. (2013). Detecting macronutrients content and distribution in oilseed rape leaves based on hyperspectral imaging. *Biosystems Engineering*, 115(1), 56–65. Doi: 10.1016/j.biosystemseng.2013.02.007.
- [120] MIREEI, S. A., MOHTASEBI S. S., MASSUDI R., RAFIEE S., ARABANIAN A. S. and BERARDINELLI A. (2010). Non-destructive measurement of moisture and soluble solids content of Mazafati date fruit by NIR spectroscopy. *AJCS*, 4(3):175-179.

- [121] WANG, S., LI W., LI J. and LIU X. (2013). Prediction of soil texture using ft-nir spectroscopy and pxf spectrometry with data fusion. *Soil Science*, 178(11), 626–638. Doi:10.1097/SS.0000000000000026.
- [122] PU, H., LIU D., WANG L. and SUN D.-W. (2015). Soluble Solids Content and pH Prediction and Maturity Discrimination of Lychee Fruits Using Visible and Near Infrared Hyperspectral Imaging. *Food Analytical Methods*, 9(1), 235–244. Doi :10.1007/s12161-015-0186-7.
- [123] JHA, S. N., JAISWAL P., NARSAIAH K., GUPTA M., BHARDWAJ R. and SINGH A. K. (2012). Non-destructive prediction of sweetness of intact mango using near infrared spectroscopy. *Scientia Horticulturae*, 138, 171–175. Doi: 10.1016/j.scienta.2012.02.031.
- [124] BAIANO, A., TERRACONE C., PERI G. and ROMANIELLO R., (2012). Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. *Computers and Electronics in Agriculture*, 87, 142–151. Doi: 10.1016/j.compag.2012.06.002.
- [125] MOGHIMI A., AGHKHANI M. H., SAZGARNIA A. and SARMA M. (2010). Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosystems Engineering*, 106(3), 295–302. Doi: 10.1016/j.biosystemseng.2010.
- [126] XU, L., LI J. and ZHANG D. (2018). Near-infrared light penetration depth analysis inside melon with thick peel by a novel strategy of slicing combining with least square fitting method. *Journal of Food Process Engineering*, e12886. Doi:10.1111/jfpe.12886.
- [127] LI H. D., XU Q. S. and LIANG Y. Z. (2014). libPLS: An Integrated Library for Partial Least Squares Regression and Discriminant Analysis. *PeerJ PrePrints 2* : e190v1. Doi: 10.7287/peerj.preprints.190v1.

- [128] DAVIS, A. R., FISH W. W. and PERKINS-VEAZIE P. (2003). A rapid spectrophotometric method for analyzing lycopene content in tomato and tomato products. *Postharvest Biology and Technology*, 28(3), 425–430. Doi:10.1016/s0925-5214(02)00203-x.
- [129] LIU, C., LIU W., CHEN W., YANG J. and ZHENG L., (2015). Feasibility in multispectral imaging for predicting the content of bioactive compounds in intact tomato fruit. *Food Chemistry*, 173, 482–488. Doi: 10.1016/j.foodchem.2014.10.052.
- [130] FLORES, K., SANCHEZ M.-T., PEREZ-MARIN D., GUERRERO J.-E. and GARRIDO-VARO A. (2009). Feasibility in NIRS instruments for predicting internal quality in intact tomato. *Journal of Food Engineering*, 91(2), 311–318. Doi: 10.1016/j.jfoodeng.2008.09.013.
- [131] DEAK, K., SZIGEDI T., PEK Z., BARANOWSKI P. and HELYES L. (2015). Carotenoid determination in tomato juice using near infrared spectroscopy. *International Agrophysics*, 29(3), 275–282. Doi:10.1515/intag-2015-0032.
- [132] GOMEZ A. H., HE Y. and PEREIRA A. G., (2006). Non-destructive measurement of acidity, soluble solids and firmness of Satsuma mandarin using Vis/NIR-spectroscopy techniques. *Journal of Food Engineering*, 77(2), 313–319. Doi: 10.1016/j.jfoodeng.2005.06.036.
- [133] WONG C.K. and EASTON M.C. (1980). An Efficient Method for Weighted Sampling without Replacement. *SIAM Journal on Computing*, 9(1), 111–113. Doi: 10.1137/0209009.
- [134] LAMMERTYN J., NICOLAÏ B., OOMS K., DE SMEDT V. and DE BAERDEMAEKER J. (1998). Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Transactions of the American Society of Agricultural Engineers*, 41(4), 1089–1094. Doi :10.13031/2013.17238.
- [135] PONGPRASERT, N., SEKOZAWA Y., SUGAYA S. and GEMMA H., (2011). A novel postharvest UV-C treatment to reduce chilling injury (membrane damage,

browning and chlorophyll degradation) in banana peel. *Scientia Horticulturae*, 130(1), 73–77. Doi: 10.1016/j.scienta.2011.06.006.

[136] MERZLYAK, M. N., SOLOVCHENKO A. E. and GITELSON A. A. (2003). Reflectance spectral features and non-destructive estimation of chlorophyll, carotenoid and anthocyanin content in apple fruit. *Postharvest Biology and Technology*, 27(2), 197–211. Doi:10.1016/s0925-5214(02)00066-2.

[137] MERZLYAK M. N. and SOLOVCHENKO A. E. (2002). Photostability of pigments in ripening apple fruit: a possible photoprotective role of carotenoids during plant Senescence. *Plant Science*, 163(4), 881–888.

[138] UWADAIRA, Y., SEKIYAMA Y. and IKEHATA A., (2018). An examination of the principle of non-destructive flesh firmness measurement of peach fruit by using VIS-NIR spectroscopy. *Heliyon*, 4(2), e00531. Doi: 10.1016/j.heliyon. 2018.e00531.

[139] WALSH, K. B., BLASCO J., ZUDE-SASSE M. and SUN X. (2020). Visible-NIR “point” spectroscopy in postharvest fruit and vegetable assessment: The science behind three decades of commercial use. *Postharvest Biology and Technology*, 168, 111246. Doi: 10.1016/j.postharvbio.2020.111246.

[140] XIE, L., YING Y., SUN T. and XU H. (2010). Influence of temperature on visible and near-infrared spectra and the predictive ability of multivariate models. *Sensing for Agriculture and Food Quality and Safety II*. Doi:10.1117/12.849948.

[141] CLÉMENT, A., DORAIS M. and VERNON M., (2008). Nondestructive Measurement of Fresh Tomato Lycopene Content and Other Physicochemical Characteristics Using Visible–NIR Spectroscopy. *Journal of Agricultural and Food Chemistry*, 56(21), 9813–9818. Doi:10.1021/jf801299r.

[142] LAMMERTYN, J., PEIRS A., DE BAERDEMAEKER J. and NICOLAI B. (2000). Light penetration properties of NIR radiation in fruit with respect to non-destructive

quality assessment. *Postharvest Biology and Technology*, 18(2), 121–132. Doi:10.1016/s0925-5214(99)00071-x.

[143] ZHANG, D., XU Y., HUANG W., TIAN X., XIA Y., XU L. and FAN S. (2019). Nondestructive measurement of soluble solids content in apple using near infrared hyperspectral imaging coupled with wavelength selection algorithm. *Infrared Physics & Technology*, 98, 297–304. Doi: 10.1016/j.infrared.2019.03.026.

[144] JHA, S. N., NARSAIAH K., JAISWAL P., BHARDWAJ R., GUPTA M., KUMAR R. and SHARMA R., (2014). Nondestructive prediction of maturity of mango using near infrared spectroscopy. *Journal of Food Engineering*, 124, 152–157.

Doi: 10.1016/j.jfoodeng.2013.10.012.

[145] OLAREWAJU, O. O., MAGWAZA L. S., NIEUWOUDT H., POBLETE-ECHEVERRIA C., Fawole, O. A., TESFAY S. Z. and OPARA U. L., (2019). Model development for non-destructive determination of rind biochemical properties of “Marsh” grapefruit using visible to near-infrared spectroscopy and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 209, 62–69.

Doi: 10.1016/j.saa.2018.10.027.

[146] KUSUMIYATI, HADIWIJAYA Y., PUTRI I.E. and MUBAROK S. (2019). Water content prediction of ‘crystal’ guava using visible-near infrared spectroscopy and chemometrics approach. *IOP Conference Series: Earth and Environmental Science* 393, 012099. Doi: 10.1088/1755-1315/393/1/012099.

[147] KOUAKOU B. K., ASSOI E.K. N., GBOGBO A. Y. and ZOUEU J. (2021). Entomological Remote Dark Field Signal Extraction by Maximum Noise Fraction and Unsupervised Clustering for Species Identification. *Preprints*, 2021030352.

Doi: 10.20944/preprints202103.0352.v1.

[148] CEROVIC Z. G., SAMSON G., MORALES F., TREMBLAY N. and MOYA I. (1999). Ultraviolet-induced fluorescence for plant monitoring: present state and prospects. *Agronomie, EDP Sciences*, 19 (7), pp. 543-578. Doi: 10.1051/agro:19990701.

[149] KRAUSE, G. H. and WEIS, E. (1991). Chlorophyll Fluorescence and Photosynthesis: The Basics. *Annual Review of Plant Physiology and Plant Molecular Biology*, 42(1), 313–349. Doi:10.1146/annurev.pp.42.060191.001525.

[150] KOUAME N., ASSIDJO N.E., DICK A.E. (2017). Prédiction de la croissance du bananier plantain (musa sp., aab, cultivar corne 1) à partir de modèles mathématiques. *Agronomie Africaine*, 29 (2), 135 – 147.

PUBLICATIONS

Nouvel outil de mesure non destructive de la qualité des mangues par télédétection

**Eliane K. N. ASSOI^{1*}, Wilfried DIBI¹, Anicet K. KOUAKOU¹, Benoit Kouakou KOUAKOU¹,
Thomas KOFFI² et Jérémie T. ZOUEU¹**

¹ *Institut National Polytechnique-Houphouët Boigny, Laboratoire d'Instrumentation Image et Spectroscopie,
BP 1093 Yamoussoukro, Côte d'Ivoire*

² *Université Nangui Abrogoua, Laboratoire de Physique Fondamentale Appliquée,
02 BP 801 Abidjan 02, Côte d'Ivoire*

* Correspondance, courriel : akeneli@yahoo.com

Résumé

Nous avons développé dans ce travail une méthode d'évaluation de la qualité des fruits basée sur la spectroscopie par télédétection passive. Utilisant la lumière solaire, un télescope couplé à un spectromètre a permis d'acquérir les spectres de réflectance de mangues présentant deux niveaux de maturité. Ces spectres ont été corrélés à l'aide de la régression PLS à des paramètres chimiques de qualité mesurés sur ces mangues à savoir : La teneur en matière soluble, le pH et la teneur en eau. Les résultats obtenus montrent d'une part que le modèle de prédiction de la teneur en sucre est réussi avec un RPD de 3,8 et une valeur de coefficient de corrélation de 0,93 en validation croisée. D'autre part, les modèles de prédiction du pH et de la teneur en eau obtenus sont acceptables avec des RPD de pH de 2,14 et RPD de la teneur en eau de 1,9. Les valeurs de coefficient de corrélation du pH et de la teneur en eau confirme la performance de ces modèles. L'utilisation de cette technique permet donc d'avoir une appréciation rapide et précise de la qualité des mangues.

Mots-clés : *télédétection, qualité des fruits, spectroscopie, réflectance optique.*

Abstract

New tool for non-destructive measurement of mango quality by remote sensing

This work focuses on the development of an alternative method of assessing fruit quality based on remote sensing spectroscopy. Using solar light, a telescope coupled to a spectrometer allowed the acquisition of reflectance spectra of mangoes with two levels of maturity. These spectra were correlated using PLS regression to quality chemical parameters measured on these mangoes namely : Soluble matter content, pH and water content. The results obtained show, on the one hand, that the prediction model of the sugar content is successful with a RPD of 3.8 and a correlation coefficient value of 0.93 in cross-validation. On the other hand, the predicted pH and water content prediction models are acceptable with RPDs of pH 2.14 and water content RPD of 1.9. The correlation coefficient values of pH and water content confirms the performance of these models. The use of this technique allows to have a quick and accurate appreciation of the quality of mangoes.

Keywords : *remote sensing, fruit quality, spectroscopy, optical reflectance.*

1. Introduction

Les fruits constituent une importante ressource nutritionnelle et économique pour les pays en développement. Ces fruits deviennent, avec le développement des civilisations, l'objet de soins particuliers. En effet, ils sont de véritables sources de micronutriments pour l'équilibre alimentaire des populations. En complément à leur rôle nutritif, ils fournissent à l'organisme des vitamines, des sels minéraux, des fibres alimentaires et des acides organiques, éléments indispensables au bon fonctionnement de l'organisme [1, 2]. Les déficiences en micronutriments sont à l'origine de diverses maladies et dysfonctionnements de l'organisme. Le récent rapport FAO / OMS sur l'alimentation et les maladies chroniques indique clairement que la consommation de fruits peut diminuer les risques d'obésité, de maladies cardiovasculaires, de diabète de type 2 et de cancer [3]. Malgré les innovations technologiques, les techniques actuelles d'analyse de qualité des fruits restent encore manuelles voir destructives. Ces techniques sont basées sur des méthodes nécessitant l'extraction du jus des fruits et cela introduit parfois des artefacts [4, 5] au niveau des paramètres de qualité. De plus, la détermination de chaque paramètre de qualité requiert un protocole spécifique souvent difficile à réaliser. Les résultats de ces analyses sont parfois obtenus plusieurs jours après analyses. Des méthodes non destructives sont alors explorées. Elles présentent de nombreux atouts en comparaison aux méthodes destructives. Leur mise en œuvre est rapide et non destructive sur les échantillons de fruits et permettent une analyse globale de la qualité des fruits. Une des techniques repose sur un système portable Sinclair IQ doté d'un capteur piézoélectrique qui génère un signal électrique proportionnel à la dureté du fruit (dur ou mou) pour déterminer sa fermeté. La spectroscopie proche infrarouge semblerait être une alternative efficace.

De nombreuses études mettent en évidence le potentiel de la spectroscopie proche infrarouge appliquée aux fruits et légumes et en promouvant ainsi son utilisation comme technique d'analyse rapide et non destructive des caractéristiques internes et externes [6, 7]. [8] ont utilisé cette technique pour suivre l'évolution des abricots au verger et déterminer leurs paramètres physico-chimiques. [9] ont mis en place un appareil portatif, le SpectronTM utilisant la spectroscopie proche infrarouge pour déterminer le taux de sucre, l'acidité totale, la teneur en eau et la concentration en anthocyanes et prédire la maturité du raisin. Des éventuelles améliorations de cette technique ont été faites et plusieurs travaux le démontrent. Les travaux de [10] ont montré la possibilité de discriminer les fruits et de prédire certains descripteurs sensoriels à l'aide d'un appareil portatif. De plus, des études antérieures ont démontré le potentiel de cette technique pour la détermination non destructive de paramètres de qualité dans la mangue [11 - 13]. Ces différents outils et méthodes présentés permettent d'accéder aux paramètres de qualité. Cependant la plupart de ces techniques emploient des outils dont l'utilisation requiert assez de temps car l'analyse étant faite sur chaque fruit. La télédétection est une technique qui permet d'effectuer des mesures à distance et à grande échelle. Cette méthode est utilisée dans plusieurs domaines et en particulier dans le domaine de l'agriculture [14 - 16]. Il s'agit donc dans ce travail de développer une nouvelle instrumentation combinant télédétection et spectroscopie pour évaluer les paramètres de qualité au niveau des fruits tels que les mangues.

2. Matériel et méthodes

2-1. Matériel

2-1-1. Matériel végétal

Un total de 50 mangues recueillies à proximité du site de l'institut INP-HB ont constitué nos échantillons. Ceux-ci ont été récoltés à différents stades de maturité à savoir, fruits murs et non murs

2-1-2. Dispositif expérimental

Le dispositif est constitué d'un ordinateur portable (HP, corei3), d'un spectromètre ocean optics USB4000 et d'un télescope de Newton. Sur le télescope est placée une caméra qui capte l'image de l'échantillon pour être affichée à l'ordinateur pendant l'alignement. Le spectromètre est relié au télescope par une fibre optique. C'est un appareil de mesure qui permet la décomposition spectrale du flux lumineux provenant du télescope par la fibre optique puis d'associer une intensité à chaque longueur d'onde de la lumière. La visualisation et l'enregistrement des spectres des différents échantillons sont effectués à l'aide du logiciel Spectrasuite (Ocean Optics USB400). Nous pouvons résumer le système par la **Figure 1** ci-dessous :

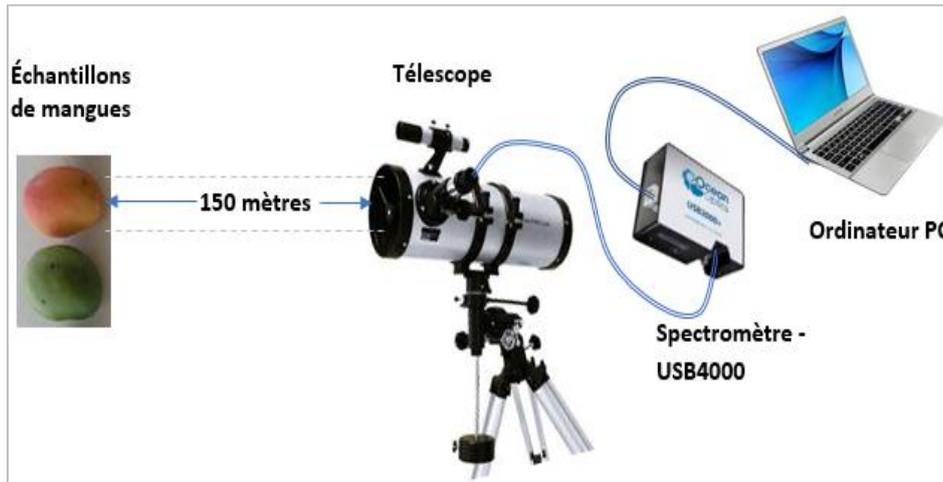


Figure 1 : Schéma du dispositif expérimental

2-2. Méthodes

2-2-1. Mesures classiques

Les mesures classiques sont des mesures effectuées dans le laboratoire de chimie pour déterminer les paramètres de qualité des fruits. Ces paramètres concernent le taux de sucre, la teneur en eau et le pH. La mesure de la teneur en eau des échantillons de mangues, fruits à pulpe consiste à déterminer la quantité d'eau contenue dans le fruit. Pour ce faire, une partie de l'échantillon a été prélevée et pesée. Elle a été ensuite séchée à l'étuve ventilée à 105°C pendant 24 heures. Après la sortie de l'étuve, elle est pesée à nouveau. La teneur en eau est calculée comme suit [17] :

$$T_c = \frac{m_D - m_E}{m_D} \times 100 \tag{1}$$

T_c étant la teneur en eau de l'échantillon en %, m_D la masse de l'échantillon avant séchage en g, m_E la masse de l'échantillon après séchage en g

Quant au pH et la teneur en sucre une masse de 10 g a été prélevée et broyée à l'aide d'une centrifugeuse pour en extraire le jus. Le jus obtenu a servi à déterminer le pH à l'aide d'un pH-mètre et la teneur en sucre à l'aide d'un réfractomètre.

2-2-2. Mesures spectroscopiques

2-2-2-1. Acquisition des spectres de réflectance

Les mesures spectroscopiques sont réalisées à l'aide du spectromètre USB 4000 *ocean optic* piloté par le logiciel Spectrasuite. Ce logiciel présente dans son interface des modes de mesures : L'absorption, la réflexion, diffusion et la transmission. Il permet le choix du mode de mesure et le contrôle du flux de photons qui arrive au détecteur du spectromètre. Dans le mode réflexion, cas de notre mesure, on active le mode « Réflexion ». L'obtention du spectre de réflectance nécessite trois types de mesures :

- Mesure de la référence R_b ;
- Mesure du bruit R_d ;
- Mesure brute de réflectance de l'échantillon R_s contenant du bruit dû aux réflexions de l'environnement et aux systèmes électroniques.

Ces trois signaux R_b , R_d et R_s sont acquis dans les mêmes conditions, c'est-à-dire l'échantillon et le réflecteur doivent être placés à la même position par rapport au dispositif expérimental et l'illumination provenant de l'environnement doit être la même. Le spectre corrigé R de l'échantillon est alors donné par la relation [18] :

$$R = \frac{R_s - R_d}{R_b - R_d} \quad (2)$$

La mesure de référence R_b est effectuée en mettant un diffuseur à la place de l'échantillon, car le spectre de réflectance se mesure par rapport à un réflecteur parfait. L'utilisation du diffuseur permet une distribution régulière d'intensité dans le champ entier de vision de la caméra, de sorte qu'une illumination légèrement irrégulière dans ce champ soit automatiquement corrigée [19]. Le signal R_d de bruit traduit les réflexions dues à l'environnement, au système optique et aux bruits électroniques. Le signal R_s bruité de l'échantillon est mesuré après avoir placé l'échantillon sur le site de mesure.

2-2-2-2. Prétraitement des spectres de réflectance

Les spectres obtenus des fruits ne peuvent pas être directement utilisés. Avant toute analyse, ils doivent subir certaines corrections ou transformations [20]. Ces corrections permettront de supprimer ou de réduire les variations non justifiées dans les spectres ou les interférences dans les spectres [10]. De plus, ces prétraitements permettent d'améliorer la précision des modèles de calibration tout en éliminant les informations non pertinentes des spectres dues à des sources inconnues telles que des irrégularités de surface, la variation de distance de l'échantillon et du détecteur [21]. Les méthodes de prétraitement utilisées dans cette étude sont le filtre Gaussien et dérivée seconde.

2-2-2-3. Analyse statistique des données

L'organigramme ci-dessus résume les différentes méthodes appliquées aux données pour aboutir aux résultats.

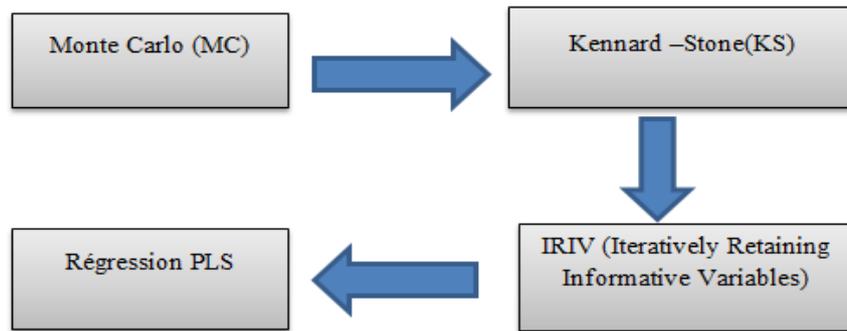


Figure 2 : Diagramme des différentes méthodes de traitement des données

2-2-2-3-1. Méthode Monte Carlo

La méthode MC est une approche qui consiste à repérer des échantillons douteux contenus dans un ensemble de données [22]. On utilise cette méthode afin d'augmenter la performance du modèle de prédiction. Le principe de la méthode consiste à étudier la distribution des erreurs de prédiction de chaque échantillon obtenu à partir de l'ensemble de données originale. Le nombre de composants principaux a été déterminé en utilisant la validation croisée dans les méthodes PLS et PCR (Principal Component Regression). Pour un nombre donné d'itération, 2/3 des échantillons sont choisis au hasard dans le jeu de données pour la calibration du modèle. Le reste des échantillons est utilisé pour le test du modèle établi. Ainsi, l'erreur de prédiction sur chaque échantillon est calculé et stocké. La distribution statistique de ces erreurs a été évaluée par la valeur moyenne $m(j)$ et l'écart type $s(j)$ pour le j ème échantillon. Les résultats pour tous les échantillons sont représentés dans un diagramme qui est capable de différencier les valeurs aberrantes et normales selon les coordonnées statistiques calculées. La valeur moyenne $m(j)$ et l'écart type $s(j)$ pour le j ème échantillon sont déterminés suivant les **Formules (3) et (4)** ci-dessous [22] :

$$m(j) = \frac{1}{k} \sum_{i=1}^k error(i) \tag{3}$$

$$s(j) = \left(\frac{1}{k-1} \sum_{i=1}^k (error(i) - m(j))^2 \right)^{1/2} \tag{4}$$

2-2-2-3-2. Algorithme Kennard-Stone

La méthode de Kennard et Stone est une approche utilisée pour partitionner une base de données en un ensemble de calibrations et un ensemble de tests [23]. Cette méthode sélectionne les échantillons en se basant sur l'évaluation de la distance euclidienne entre chaque paire (p, q) d'échantillons selon **l'Équation** suivante [23 - 25] :

$$ED_z(p, q) = \sqrt{\sum_{j=1}^N [z_p(j) - z_q(j)]^2} \quad , p, q \in [1, M] \tag{5}$$

où, N est le nombre de variables dans Z et M est le nombre d'échantillons. $z_p(j)$ et $z_q(j)$ sont la j ème variable pour les échantillons p et q , respectivement.

A chaque itération, l'algorithme conserve le point le plus éloigné des points précédemment choisis et cette opération est répétée jusqu'à l'obtention des n points souhaités.

2-2-2-3-3. IRIV

IRIV signifie Iterativement Retaining Informative Variable en anglais. Cette méthode considère une matrice de données x contenant N échantillons en lignes et p variables en colonnes, et une autre y , de taille $N \times 1$, qui désigne les paramètres mesurés [26]. L'échantillonnage en matrice binaire consiste à générer une matrice binaire M de dimension $K \times P$ qui comporte aléatoirement le nombre "1" ou "0" dans chaque colonne. Chaque ligne de la matrice binaire correspond à un échantillonnage individuel, où "1" indique les échantillons retenus pour la modélisation et "0" pour les non retenus. Par le processus d'inclusion et d'exclusion des valeurs binaires "1" ou "0", la première matrice est obtenue en changeant pour la première colonne tous les "0" à "1" et vice versa. Ensuite, un sous-ensemble (variable i ème colonne) est évalué par PLS et la distribution statistique du RMSECV à l'inclusion et à l'exclusion est obtenue. Ces distributions statistiques de RMSECV pour une variable sont testées avec l'hypothèse que la distribution des données est la même pour deux groupes afin d'identifier l'importance de cette variable. Selon la valeur de p obtenue, les variables sont rangées en 4 groupes: très informatif, faiblement informatif, interférant et non informatif. Après avoir supprimé les deux derniers types de variables, un nouveau sous-ensemble est généré et le même cycle de calcul est répété.

2-2-2-3-4. PLS

PLS signifie en anglais Partial Least Square. Elle est une méthode qui permet de construire les modèles de prédiction des différents constituants. Les modèles sont construits à partir de la combinaison linéaire des données spectroscopiques. La PLS construit les modèles sur la base d'un ensemble d'apprentissage sélectionnés par la méthode KS. Les validations des modèles construits sont effectuées avec le reste des données. Ainsi, le modèle général de la PLS multivariée est :

$$X = TP^1 + E \quad (6)$$

$$Y = UQ^T + F \quad (7)$$

où, X est une matrice de prédicteurs $n \times m$, Y est une matrice de réponses $n \times p$; T et U sont des matrices $n \times l$ qui sont respectivement des projections de X (le score X) et des projections de Y (les scores Y); P et Q sont respectivement les matrices de chargement orthogonales $m \times l$ et $p \times l$ et les matrices E et F sont les termes d'erreur, supposés être des variables normales aléatoires indépendantes et identiquement distribuées. Les décompositions de X et Y sont faites de manière à maximiser la covariance entre T et U

3. Résultats et discussion

3-1. Visualisation des spectres

La figure ci-dessous présentent les spectres de réflectance de trois échantillons de mangues non mures (**Figure 3A**) et trois échantillons de mangues mures (**Figure 3B**). L'observation de ces spectres indique une similarité entre ces deux groupes d'échantillons sur toute la plage de longueur d'onde. Les mangues non mures présentent un pic autour de 550 nm tandis que les mangues mures présentent un pic autour de 556 nm. Ce léger décalage du pic indique une faible variation de la coloration de ces fruits. La faible intensité de réflectance des mangues non mure par rapport aux mangues mures aux alentours de 675 nm traduit un taux de chlorophylle important au niveau des fruits non murs en comparaison aux fruits murs [27, 28]. Cette chute du taux de chlorophylle permet à certains pigments comme les caroténoïdes et xanthophylle de se manifester. On observe sur les différents spectres une absorption autour de 970 nm qui est liée à la teneur en eau des de ces fruits [29, 30].

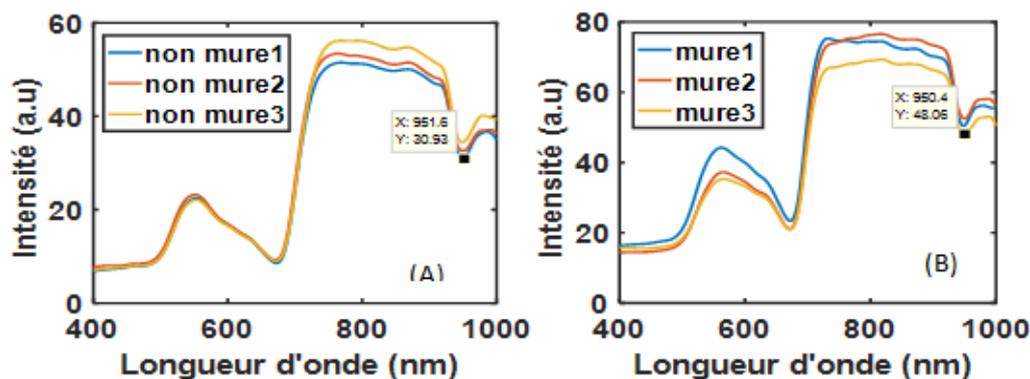


Figure 3 : Spectres de réflectances de trois échantillons de mangues non mures (A) et spectres de réflectances de trois échantillons de mangues mures (B)

3-2. Détection des données aberrantes par Monte Carlo

Les résultats du traitement par Monte Carlo ont permis d'observer parmi les valeurs de références celles qui semblent être douteuses et de les retirer par la suite. La **Figure** ci-dessous présente les différentes valeurs aberrantes au niveau des paramètres chimiques tels que l'eau, le sucre et le pH.

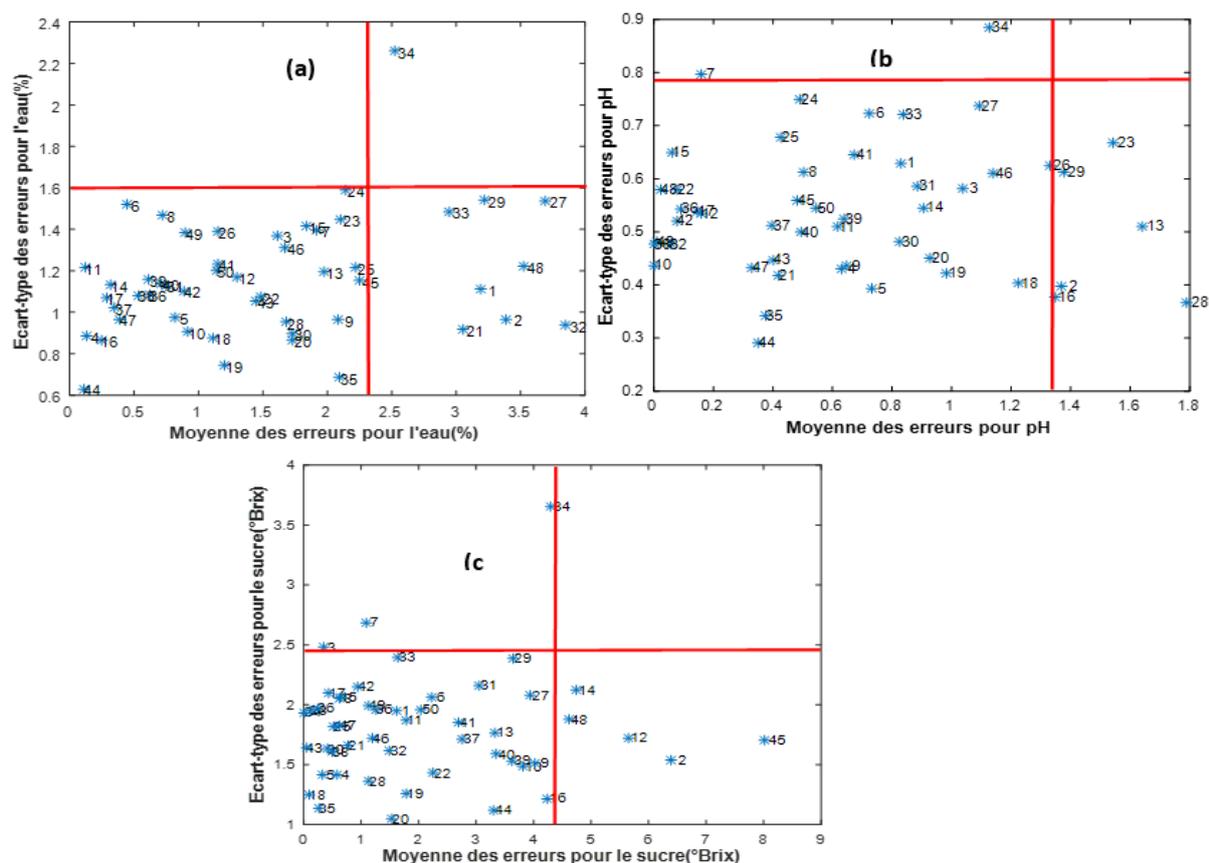


Figure 4 : Diagramme montrant la différence entre les échantillons aberrants et normaux de la teneur en eau (a) ; du pH (b) ; de la teneur en sucre (c)

Dans cette étude, l'utilisation de la méthode Monte Carlo, nous a permis de retirer respectivement 5 et 8 échantillons dans le jeu de données des teneurs en sucre, en pH et en eau.

3-3. Modélisation des paramètres de qualité par la régression PLS

La **Figure 5** permet de visualiser les valeurs des différents paramètres de qualité réellement mesurées par les méthodes de références par rapport aux valeurs obtenues par la modélisation PLS. Les modèles construits par la PLS peuvent être appréciés à partir de la distribution des nuages de points autour de la diagonale. Plus les points sont étirés et resserrés autour de cette diagonale plus la valeur prédite est proche de la valeur mesurée [31].

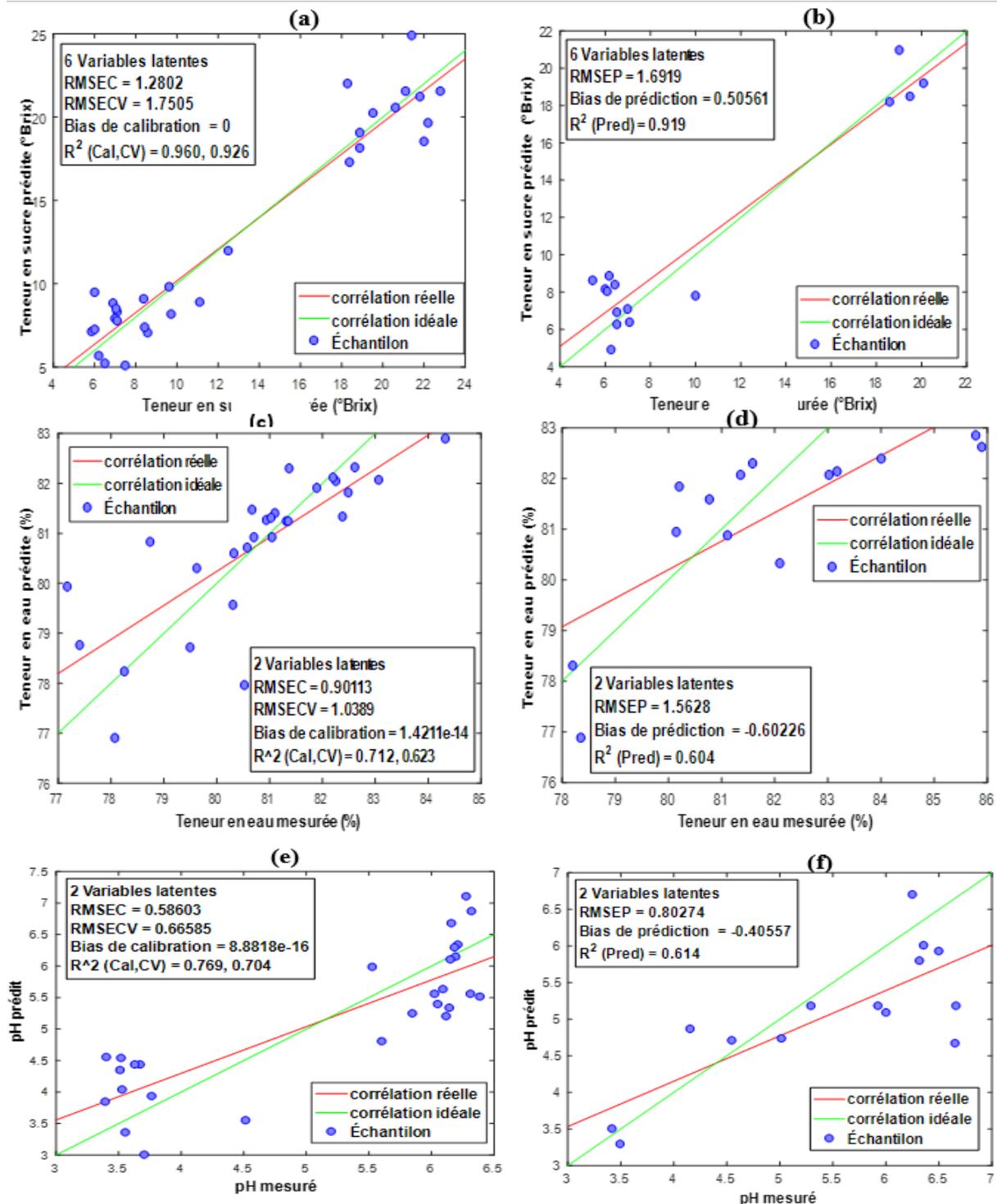


Figure 5 : (a) Calibration du sucre ; (b) Validation du sucre ; (c) Calibration de la teneur en eau ; (d) validation de la teneur en eau ; (e) Calibration du pH ; (f) Validation du pH

Les nuages de points de la teneur en sucre sont plus étirés et plus resserrés autour de la droite de régression. Cela signifie que la corrélation entre les valeurs de référence et les valeurs prédites par le modèle est beaucoup élevée. Par contre, ce resserrément est moins intense pour le pH et la teneur en eau. On observe une dispersion des points autour de la diagonale. Ainsi, cette dispersion se traduit par un coefficient de corrélation moyennement élevé. Les résultats de la modélisation sont rapportés dans le **Tableau 1**. Ce tableau résume les valeurs différentes critères statistiques d'évaluation la qualité des phases de calibration et du test.

Tableau 1 : Résultats des régressions PLS

Paramètres	Calibration				Test		
	R^2_C	R^2_{CV}	RMSEC	RPD	R^2_P	RMSEP	RPD
Teneur en sucre	0.96	0.93	1.28	5.06	0.919	1.69	3.58
Teneur en eau	0.71	0.62	0.9	1.9	0.60	1.56	1.52
pH	0.77	0.70	0.58	2.14	0.61	0.8	1.446

Les modèles construits sont en général tous acceptables du fait de leur coefficient de corrélation compris entre 0,77 et 0,96. Le modèle de calibrage de la teneur en sucre a été construit avec un coefficient de corrélation $R = 0.96$. Dans ce cas, l'erreur commise sur la prédiction est de 1.28, légèrement élevée. Le modèle ainsi construit est excellent et précis [32]. La robustesse de ce modèle se confirme avec les valeurs des RPD de la calibration et la validation qui sont respectivement de 5.06 et 3.58 > 3 . Le modèle est donc très performant en se référant aux travaux de [33]. La validation du modèle a donné un coefficient de corrélation de 0.91 et une erreur de 1.69. La prédiction de la teneur en sucre est donc correcte. Les résultats de la prédiction de la teneur en sucre obtenus dans ce travail sont meilleurs que ceux proposés par [10, 33 - 34]. Ces résultats montrent que la télédétection est une discipline qui présente plusieurs zones d'application en générale et apparaît comme un atout dans la détermination de la teneur en sucre des mangues. Cette technique peut donc être utiliser comme un moyen non destructif dans la détermination des paramètres de qualité des fruits. En ce qui concerne le pH et la teneur en eau, l'étalonnage des modèles a donné des RMSEC relativement faibles de 0.58 pour le pH et 0.9 pour la teneur en eau. Pour ces deux modèles, les coefficients de corrélation étaient moyennement élevés, 0.77 pour le pH et 0.71 pour l'eau lors du calibrage. Cela traduit une performance assez parfaite dans la prévision de ces paramètres. D'après le RPD obtenu pour le pH, le modèle est considéré comme utile 2.14 selon les travaux de [33]. La prédiction de ces deux paramètres est alors précise avec des RMSEP faibles. Ces deux modèles ont également pu être prédits avec de faibles variables latentes (2 variables latentes). Ces faibles nombres de variables latentes a permis d'améliorer les résultats des erreurs de calibration des modèles du pH et de la teneur en eau.

4. Conclusion

Dans cette étude, une technique spectroscopique par télédétection a été développée pour déterminer des paramètres de qualité au niveau de la mangue. L'étude a démontré qu'il est possible à distance d'accéder aux paramètres tels que le taux de sucre, le pH et la teneur en eau au niveau des fruits de mangue. La régression PLS a permis d'établir des modèles de prédiction de la teneur en sucre, en eau et le pH dans les échantillons de mangues. Cette technique a montré son intérêt vue la robustesse des modèles établit. Cela a été révélé par une grande précision dans la prédiction de la teneur en sucre ($R^2 = 0.96$) et une précision moyenne dans les prédictions du pH ainsi que la teneur en eau. Ces résultats obtenus montrent qu'il est possible d'évaluer les paramètres de qualité de la mangue de manière non destructive par télédétection. La spectroscopie par télédétection apparaît ainsi comme une nouvelle voie de détection des paramètres de qualité des fruits à grand échelle.

Remerciements

Cette recherche a été soutenue par le réseau AFSIN et l'Institut Polytechnique National Houphouët-Boigny (INP-HB). Nous tenons à exprimer notre gratitude à tous les laboratoires de chimie de l'INP-HB. Nous remercions également tous les collègues du laboratoire L2IS pour leur aide à l'accomplissement de ce travail.

Références

- [1] - HAGRETOU SAWADOGO-LINGANI et ALFRED S. TRAORE, " Composition chimique et valeur nutritive de la mangue Amelie", *J. Sci.*, Vol. 2, N° 1 (2001) 35 - 39
- [2] - JISSY K. JACOB, KRISHNARAJ TIWARI, JULIETA CORREA-BETANZO, AZIZAH MISRAN, RENU CHANDRASEKARAN, and GOPINADHAN PALIYATH, "Biochemical Basis for Functional Ingredient Design from Fruits ", *Annu. Rev. Food Sci. Technol.*, 3 (2012) 79 - 104
- [3] - FAO, "Fruits tropicaux — leurs valeurs nutritionnelles, leur biodiversité et leur contribution à la santé et à la nutrition", (Octobre 2003) *CCP: Ba/Tf03/15*
- [4] - J. DOSSOU, I. SOULE & MARCELLINE MONTCHO, "Évaluation des caractéristiques physico-chimiques et sensorielles de la purée de tomate locale produite à petite échelle au Bénin", *Tropicultura*, 25, 2 (2007) 119 - 125
- [5] - GEORGE ZONIOS, LEV T. PERELMAN, VADIM BACKMAN, RAMASAMY MANOHARAN, MARYANN FITZMAURICE, JACQUES VAN DAM, et MICHAEL S. FELD, "Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo", *applied optics*, Vol. 38, N°31 (1999)
- [6] - M. CASALE, L. BAGNASCO, M. ZOTTI, S. DI PIAZZA, N. SITTA, P. OLIVERI, A., "NIR Spectroscopy-based efficient approach to detect fraudulent additions within mixtures of dried porcini mushrooms", *Talanta*, 160 (2016) 729 - 734, <http://dx.doi.org/10.1016/j.talanta.2016.08.004>
- [7] - R. BEGHI, G. GIOVANELLI, C. MALEGORI, V. GIOVENZANA, R. GUIDETTI, "Testing of a VIS-NIR system for the monitoring of long-term apple storage", *Food Bioprocess Technol.*, 7 (2014) 2134 - 2143, <http://dx.doi.org/10.1007/s11947-014-1294-x>
- [8] - C. CAMPS et D. CHRISTEN, "Suivi des abricots avant récolte par spectroscopie proche infrarouge portable", *Revue suisse Vitic. Arboric. Hortic.*, Vol. 41 (3) (2009) 193 - 198
- [9] - V. GERAUDIE, J. M. ROGER et H. OJEDA, "Développement d'un appareil permettant de prédire la maturité du raisin par spectroscopie proche infrarouge", *Revue Française Œnologie*, 240 (2010) 2 - 8
- [10] - IRENE MAPPE, RENE SIR et FREDERIQUE JOURJON, Mélanie Blin, Céline Turbi Ilon, Emira Mehinagic, "Spectroscopie visible proche infrarouge pour décrire les fruits et prédire leurs paramètres sensoriels", *Revue suisse Viticulture*, Arboriculture, Horticulture, Vol. 248, 42 (4) (2010) 248 - 255
- [11] - C. WATANAWAN, T. WASUSRI, V. SRILAONG, C. WONGS-AREE and S. KANLAYANARAT, "Near infrared spectroscopic evaluation of fruit maturity and quality of export Thai mango (*Mangifera indica* L. var. Namdokmai)", *International Food Research Journal*, 21 (3) (2014) 1109 - 1114
- [12] - YOHANES ARIS PURWANTO, PUTRI WULANDARI ZAINAL, USMAN AHMAD, SUTRISNO MARDJAN, YOSHIO MAKINO, SEIICHI OSHITA, YOSHINORI KAWAGOE, SHINICHI KUROKI, "Non Destructive Prediction of pH in Mango Fruits cv. Gedong Gincu Using NIR Spectroscopy", *International Journal of Engineering & Technology IJET-IJENS*, Vol. 13, N°03 (2013)
- [13] - S. N. JHA, K. NARSAIAH, P. JAISWAL, R. BHARDWAJ, M. GUPTA, R. KUMAR & R. SHARMA, "Nondestructive prediction of maturity of mango using near infrared spectroscopy". *Journal of Food Engineering*, 124 (2014) 152 - 157
- [14] - CLEMENT ATZBERGER, "Advances in Remote Sensing of Agriculture : Context Description, Existing Operational Monitoring Systems and Major Information Needs *Remote Sens*", 5 (2013) 949 - 981

- [15] - AIYM ORYNBAIKYZY, URSULA GESSNER and CHRISTOPHER CONRAD, "Crop type classification using a combination of optical and radar remote sensing data : a review", *international journal of remote sensing*, (2019)
- [16] - DEEPA U PUJAR, UDACHAPPA U PUJAR, SHRUTHI CR, ANURADHA WADAGAVE and MADHU CHULAKI, "Remote sensing in fruit crops", *Journal of Pharmacognosy and Phytochemistry*, 6 (5) (2017) 2479 - 2484
- [17] - I. E. ELBATAWI, M. T. EBAID and B. E. HEMEDA, "Détermination of potato water content using NIR diffuse reflection method", *Misr J. Ag. Eng.*, 25 (4) (2008) 1279 - 1292
- [18] - M. A. AGNERO, J. T. ZOUEU and K. KONAN, "Characterization of a Multimodal and Multispectral Led Imager : Application to Organic Polymer's Microspheres with Diameter $\Phi = 10.2 \mu\text{m}$ ". *Optics and Photonics Journal*, 6 (2016) 171 - 183, <http://dx.doi.org/10.4236/opj.2016.67019>
- [19] - JACOB C. NOORDAM, WILLIE HAM VAN DEN BROCK et LUTGARDE MC BUYDENS, "Detection and classification of latent defects and diseases on raw French fries with multispectral imaging". *J Sci Food Agric*, 85 (2005) 2249 - 2259
- [20] - C. CAMPS, J.-P. SIEGRIST et D. CHRISTEN, "Qualité des fruits : mesure non destructive par spectroscopie proche infrarouge portable", *Revue suisse Vitic. Arboric. Hortic.*, Vol. 40 (3) (2008) 151 - 157
- [21] - H. CEN and Y. HE, "Theory and application of near infrared reflectance spectroscopy in determination of food quality". *Trends in Food Science & Technology*, 18 (2007) 72 - 83
- [22] - D.-S. CAO, Y.-Z. LIANG, Q.-S. XU, H.-D. LI, and X. CHEN, "A New Strategy of Outlier Detection for QSAR/QSPR". *Journal of Computational Chemistry*, 31 (2010) 592 - 602
- [23] - R. W. KENNARD and L. A. STONE, "Computer Aided Design of Experiments". *Technometrics*, 11 (1969) 137 - 148
- [24] - W. WU, B. WALCZAK, D. L. MASSART, S. HEUERDING, F. ERNI, I. R. LAST and K. A. PREBBLE, "Artificial Neural Networks in Classification of NIR Spectral Data : Design of the Training Set". *Chemometrics and Intelligent Laboratory Systems*, 33 (1996) 35 - 46
- [25] - R. K. H. GALVAO, M. C. U. ARAUJO, G. E. JOSE, M. J. C. PONTES, E. C. SILVA and T. C. B. SALDANHA, "A Method for Calibration and Validation Subset Partitioning". *Talanta*, 67 (2005) 736 - 740
- [26] - Y.-H. YUN, W.-T. WANG, M.-L. TAN, Y.-Z. LIANG, H.-D. LI, D.-S. CAO and Q.-S. XU, "A Strategy That Iteratively Retains Informative Variables for Selecting Optimal Variable Subset in Multivariate Calibration". *Analytica Chimica Acta*, 807 (2014) 36 - 43
- [27] - SUN XUDONG, ZHANG HAILIANG, LIU YANDE, "Nondestructive assessment of quality of Nanfeng mandarin fruit by a portable near infrared spectroscopy", *Int J Agric & Biol Eng*, Vol. 2, N° 1 (2009)
- [28] - D. C. SLAUGHTER, J. F. THOMPSON, E. S. TAN, "Nondestructive determination of total and soluble solids in fresh prune using near infrared spectroscopy Postharvest Biology and Technology", 28 (2003) 437 - 444
- [29] - SEYED AHMAD MIREEI, SEYED SAEID MOHTASEBI, REZA MASSUDI, SHAHIN RAFIEE, ATOOSA SADAT ARABANIAN, ANNACHIARA BERARDINELLI, "Non-destructive measurement of moisture and soluble solids content of Mazafati date fruit by NIR spectroscopy", *AJCS*, 4 (3) (2010) 175 - 179
- [30] - C. LIU, W. X. LIU, W. CHEN, J. YAN & L. ZHENG, "Potential of multispectral imaging for real-time determination of color change and moisture distribution of carrot slices during hot air dehydration". *Food Chemistry*, 195 (2016) 110 - 116
- [31] - CEDRIC CAMPS, "Mesure non destructive de la qualité des tomates par spectroscopie NIR", *Revue suisse Viticulture, Arboriculture, Horticulture*, Vol 42 (5) (2010) 298 - 303
- [32] - RANGANATHAN KAPILAN and VETHANAYAGAM CELESTY ANPALAGAN, "Determination of optimum maturity of north Sri Lankan Kilichondan Mango fruits (*Mangifera indica* L.) based on their biochemical properties", *Adv. Appl. Sci. Res.*, 6 (10) (2015) 105 - 113
- [33] - D. F. MALLEY, P. D. MARTIN and E. BEN-DOR, "Application in Analysis of Soils. Near-Infrared Spectroscopy in Agriculture", 44 (2004) 729 - 784
- [34] - CEDRIC CAMPS, "Mesure non destructive de la qualité des tomates par spectroscopie NIR", *Revue suisse Viticulture, Arboriculture, Horticulture*, Vol. 42, (5) (2010) 298 - 303
- [35] - YONG HE, YUN ZHANG, ANNIA G. PEREIRA, ANTIHUS H. GÓMEZ and JUN WANG, "Nondestructive Determination of Tomato Fruit Quality Characteristics Using Vis/NIR Spectroscopy Technique", *International Journal of Information Technology*, Vol. 11, N° 11 (2005)

Estimating maturity by measuring pH, sugar, dry matter, water and vitamin C content of cashew apple (*Anacardium occidentale*) from remote spectral reflectance data using neural network

Eliane K. Assoi^{1*}, Olivier K. Bagui¹, Benoit K. Kouakou¹, Adolphe Y. Gbogbo¹, Doudjo Soro², Jeremie T. Zoueu¹

¹Laboratoire d'Instrumentation Image et Spectroscopie, INP-HB, DFR-GEE, BP 1093 Yamoussoukro, Côte d'Ivoire

²Laboratoire des Procédés Industriels des Synthèses de l'Environnement et des Énergies Nouvelles, INP-HB, DFR-GCAA, BP 1093 Yamoussoukro, Côte d'Ivoire

*Corresponding author: akenel@yahoo.com

Abstract

In agricultural sector, maturity is the main decision criterion for starting the harvest. This criterion is usually revealed by a number of parameters such as pH, sugar, dry matter, water and vitamin C, which are informative but technically tedious to measure. The cashew apple is the hypertrophied peduncle which is attached to the cashew nut. It is a nutritious (very juicy fruit (85 to 90% water), sweet (7 to 13% carbohydrates), acidic and vitamin C content) fruit with high therapeutic and medicinal properties. The cashew apple is used as a raw material for many industrial applications (juice and alcohol). This research was conducted as a preliminary step towards the development of a real-time remote sensing technique for assessing the quality of tropical fruits. Spectral acquisitions were carried out from intact cashew apple using optical system composed reflector coupled with spectrometer *USB 4000 FL* from Ocean Optics (350-1100 nm). Immediately after spectral acquisition, the samples were analyzed by using chemical methods (sugar content, dry matter content, water content, vitamin C and pH). Preprocessing treatment method, bootstrap method was required to create statistical new samples and to increase the number of samples required. This method was used to improve the predictive performance of calibration model. Statistical models of prediction were developed using an artificial neural network (ANN) method. The results obtained from the models built by ANN showed strong relationships between predicted and experimental values: ($R_{\text{square}} = 0.9870$, $\text{RMSE} = 0.0262$) for pH, ($R_{\text{square}} = 0.9869$, $\text{RMSE} = 0.1392$) for Sugar, ($R_{\text{square}} = 0.9726$, $\text{RMSE} = 0.3333$) for water content, ($R_{\text{square}} = 0.9703$, $\text{RMSE} = 0.3464$) for vitamin C and ($R_{\text{square}} = 0.9922$, $\text{RMSE} = 5.0304$, $\text{RMSE} = 5.0304$) for dry matter. These results confirm the potential of visible spectroscopy to predict quality parameters of cashew apples remotely and make decisions about best harvest time.

Keywords: Cashew apple, Reflectance spectra, neural network, remote sensing.

Abbreviations: ANN_ artificial neural network; MC_moisture content; NIR_near-infrared; R_{square} _coefficient of determination; $\text{RMSE}_{\text{root}}$ mean square error; R_b _Reflectance spectrum of standard reflector; R_d _Reflectance spectrum of dark; R_s _Reflectance spectrum of sample;

Introduction

Measuring fruit quality makes it possible to control and to improve their sensory, nutritional and hygienic properties. This practice has given rise to the development of instruments and methodologies in the assessment of maturity and quality of fruits such as cashew apples over time.

Cashew tree (*Anacardium occidentale*) is an economic tropical plant originated from north-east Brazil. It produces nuts and cashew apple (Lautié et al. 2001). The cashew apple is the fleshy part of the cashew fruit attached to the cashew nut. It is a good source of carbohydrates, minerals, vitamins, carotenoids, organic acids, and antioxidants (Daramola, 2013, Olalusi et al., 2020). It contains anacardic acid which is an excellent source of antioxidants. The cashew apple is also known to have antiscorbutic, therapeutic and medicinal properties beneficial to health. The cashew apple is multi-criteria-based quality fruit.

In fact, the cashew apple is considered as a complex agricultural product when produced, processed or stored and in term of essential elements. It can be eaten fresh and utilized to produce syrup, alcoholic drink and candies (Hammed et al., 2008; Marc et al., 2011; Runjala and Kella, 2017). During harvesting or storage, cashew apple qualities are affected because of the moisture and the sugar content. It is one of the most perishable fruits and has very limited shelf life due to its high-water content. Conventional chemical analyses used to assess the quality of cashew apple such as total solids, soluble solids, titratable acidity, pH, dry matter and vitamin C content are limited (Gurunathan et al., 2010). These methods are laborious which require proper skill to produce consistent results and use toxic solvents. The application of these conventional analytical methods requires sample preparation and is time-consuming.

Alternatives to these "standard" analysis methods are now available (Uwadaira et al., 2018; Vasighi-Shojae et al., 2018; Eliane et al., 2019). The different non-destructive methods have been developed for fruit quality measurement. The near infrared spectroscopy is a rapid, low-cost, reliable and reproducible analytical method. It has become a more common method among other nondestructive methods for evaluation of fruits because of the advantages in quality evaluation, which are cost-effectiveness, robustness, reliability, and fruit safety (Seyed et al., 2010). In this study, we propose a remote sensing spectroscopy system as alternative technique to assess fruit quality. Remote sensing describes techniques to measure object characteristics without physical contact. It may be split into "active" remote sensing (using artificial light) and "passive" remote sensing (using sunlight). It is a multisectoral technical approach used in several fields, particularly in agricultural (Atzberger, 2013, Pujar et al., 2017; Orynbaikyzy et al., 2019). The objective of the present study was to develop remote and non-destructive rapid method based on visible spectroscopy to assess the maturity of cashew apple. The study demonstrates the potential to remotely and reach an informing decision about best harvest time using simple equipment.

This study proposes a remote sensing spectroscopy based on the measurement of backscatter signal and chemical parameters in conjunction with ANN to predict tropical fruit quality.

Results and Discussion

Visualization and analyzing of backscatter spectra

Figure 2 shows the reflectance spectra of the cashew apple. All the spectra have a similar profile. This justifies the fact that the fruit are the same species. The different intensities observed are due to the cross section of the cashew that backscatters the light. Spectra of cashew apple were characterized by absorption peaks between 400 nm and 575 nm, absorption peaks around 675 nm and absorption peaks around 970 nm. These spectra were also characterized by reflection peaks around 575 nm and 675 nm, and peak reflection around 720 nm.

The strong absorption between 400 nm and 575 nm is related to the presence of carotenoids. Carotenoids are the pigments responsible for the yellow-orange coloring of the fruit. There are several types of carotenoids divided into two main groups: carotenes (α -carotene and β -carotene) and Xanthophylls (Lichtenthaler, 1987). These pigments play a protective role for the plant by helping it to regulate the excess absorbed light. The incident light is also absorbed by the cashew apple around 675 nm. This is due to the chlorophyll content. The concentration of chlorophyll in each cashew apple is different because the intensity absorbed by each cashew apple is different. The general spectrum is dominated by reflectance maxima related to the presence of carotenoids responsible for fruit coloration. The absorption at 970 nm (Figure 2), are mainly due to the water contained in the fruits (Uwadaira et al. 2018). Noticeable reflections around 575 and 675 nm and 720 nm, could be due to the variations in carotenoids and chlorophylls for the cashew apple.

Prediction models

Five prediction models corresponding to the five quality parameters were established. Prediction model's accuracy is evaluated using root-mean-square error (RMSE) and the

coefficient of determination (R_{square}). Low values for RMSE ($\text{RMSE} < 1$) and high value of R_{square} indicate a good model (Lammertyn et al., 1998). Table 1 shows the statistical results of the ANN modelling. The table above shows the results of modelling cashew apple quality parameters of the artificial neural network. Observation of the table reveals that all predictive models are obtained with R_{square} greater than 0.97. This indicates that the models were constructed with very good predictive performance, according to the performance criteria of the Malley et al. (2004). It is also in agreement with low value of RMSE fluctuating between 0.02 and 0.3 for the fourth quality parameters. In the case of vitamin C, the R_{square} is 0.993 indicating that there is high accuracy between the experimental data and the predicted value generated by the ANN. But the RMSE is high ($\text{RMSE} = 4.72$). The inhomogeneous between the two values indicates an overfitting of the ANN. This appears when the number of layers is too high.

The second dataset was used to test the different models built from the first dataset. This second set of data was introduced into each equation in order to predict the different parameters. Comparing each predicted value to each experimental value provides information about the predictive power of the model. Figure 3 shows the correlations between predicted and experimental values for each quality parameter. The linear regressions are presented between experimental and predicted values in Figure 3. These figures show the actual values measured by the reference methods compared to the values predicted by the models. In general, prediction models are accurate when the set of points is stretched and tightened around the first diagonal of the regression line. In our case, the points are tightened around the regression lines. This means that the values predicted by the models are very close to the experimental values and that the correlations between the experimental values and the values predicted by the models are high. This reflects a good prediction of the different parameters by our models (ANN).

The second dataset was used to test the different models built from the first dataset. This second set of data was introduced into each equation in order to predict the different parameters. Table 2 summarizes the RMSE and R_{square} values for each model for the validation dataset. The R_{square} and RMSE of all the models indicate a very high accuracy of the models.

As stated earlier, the models for the quality parameters were developed using the NIR range 400–920 nm. This is similar to the range used in previous studies to develop models for predicting the soluble solids content of apples (Tran and Fukuzawa, 2020). The developed models showed good performances, with high predictive R_{square} values and low RMSEP values. These results were in agreement with previous values reported by Tian et al. (2020) for the prediction of soluble solids content. The good performances of models could possibly be because of the application of the Bootstrap method to increase the number of samples. In one study, Kuang and Mouazen (2012) suggested that the samples number of could influence the prediction error. The increasing of the number of samples will reduce the value of the error of prediction models. Furthermore, model of vitamin C, whereas it had high prediction errors ($\text{RMSE} = 4.72$). This could be due to the temperature impact on the vitamin C content of cashew apples (Islam and al., 1993). The fruit had higher content of vitamin C which was decreased with times of exposure to the sun. This may have affected the higher peaks in the spectral bands associated with vitamin C content.

Table 1. Statistical results of model's quality.

Quality Parameters	Statistical parameters	
	R _{square}	RMSE
pH	0.988	0.025
Sugar content	0.989	0.13
Water content	0.976	0.314
Dry matter	0.974	0.32
Vitamin C	0.993	4.72

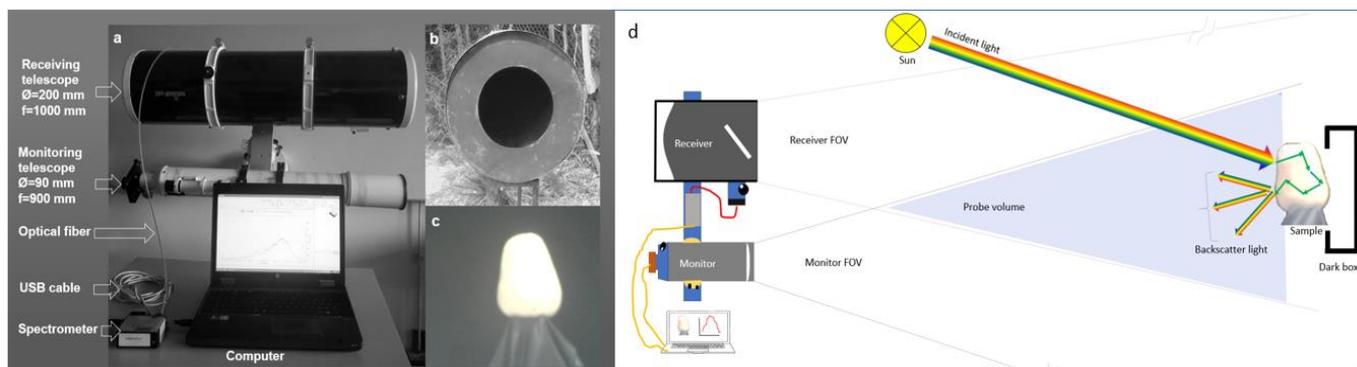


Fig 1. (a) Experimental setup. Receiving Telescope is a reflector coupled with a spectrometer mounted on one end of an aluminum barre collects the backscatter light from the sample and monitoring telescope is a refractor with a CCD camera fixed on another end monitoring the sample during acquisition. This can help to avoid saving unnecessary backscatter signal when the sample fall down. (b) Dark box used to reduce background signal and avoid saturation of the spectrometer. (c) Image of the cashew sample took with the monitoring telescope at 60 m from the setup. (d) Schematic view of the setup

Table 2. Statistical results of correlation between experimental and predictive values.

Model	Statistical parameters	
	R _{square}	RMSE
pH	0.987	0.026
Sugar content	0.987	0.139
Water content	0.972	0.333
Dry matter	0.970	0.346
Vitamin C	0.992	5.03

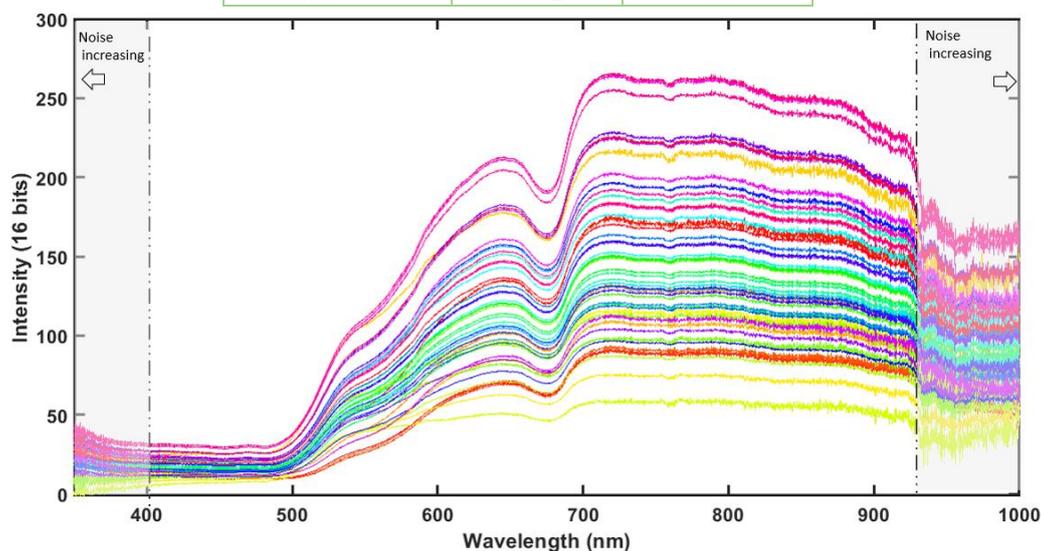


Fig 2. Spectra of intact cashew apple. Raw spectra denoising by singular value decomposition to avoid contribution of aerosol [see Benoit Kouakou et al, 2016; Benoit Kouakou, 2020 for more details]. Gray spectral band showing increasing noise at each end of the spectra are due to electrical noise.

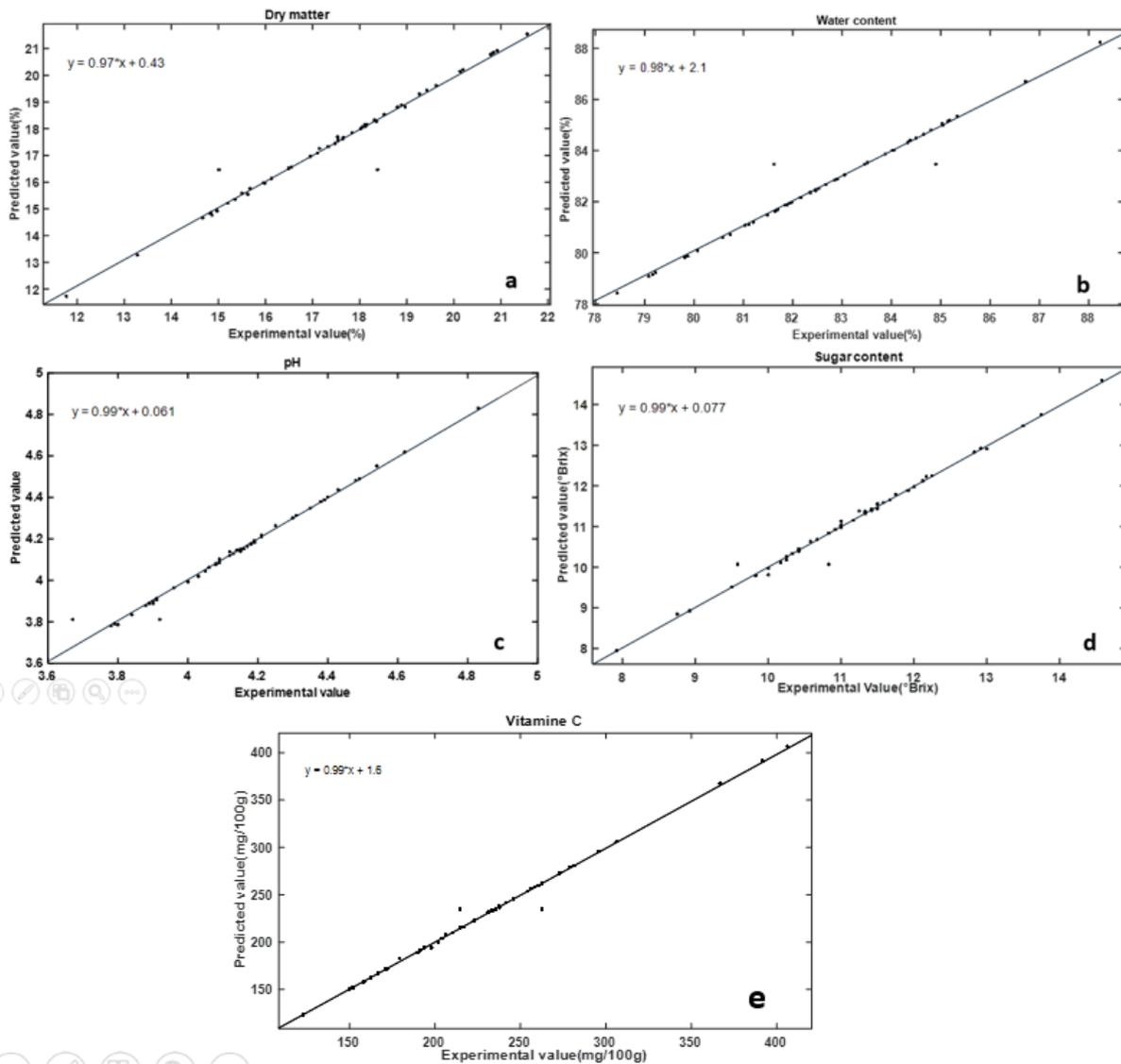


Fig 3. Correlation between experimental and predictive values for (a) dry matter, (b) water content, (c) pH, (d) sugar content and (e) vitamin C.

Materials and Methods

Sample preparation

During the measurement period, thirty (30) plants in a cashew nut plantation (6°52'14.6"N 5°12'14.2"W) were selected. Ten (10) samples were taken daily from the batch of selected plants, at a rate of two apples per plant for one week. They are then immediately transported to the laboratory after harvesting for spectral measurements and chemical analysis. In total, 70 cashew apple samples were used in this study. Before any measurement, the fruit was washed, wiped off and laid out on a table for later use. Each fruit is marked with an identification number. The measurement is made in the unmarked part of the fruit to avoid the presence of artefacts.

Spectral measurement

The data acquisition system has been discussed in (Benoit et al., 2016) as summarized in Fig 1.

The system is composed of two telescopes, the first one is a reflector and the second is a refractor. The two telescopes are aligned towards the samples which have the same field of view. We have connected the spectrometer *USB 4000 FL* from Ocean Optics to the reflector and a CCD camera to the refractor. Both

are controlled by a laptop. This system measured reflectance spectra in a wavelength ranging from 400 nm to 920 nm of seventy cashew apples. The reflectance spectra acquired from fruits most often contain noise (Eliane et al., 2019). They must then undergo certain corrections or transformations in order to reduce these noises. To do this, we applied a correction according to the equation below to the different reflectance spectra when acquiring the spectral data:

$$Reflectance = \frac{R_s - R_d}{R_b - R_d} \quad (1)$$

R_s : Reflectance spectrum of sample

R_b : Reflectance spectrum of standard reflector

R_d : Reflectance spectrum of dark

Chemical approach for parameters measurement (sugar content, water content, dry matter, vitamin C and pH)

The reference measurements are those carried out in Laboratory of Industrial Processes for Synthesizing the Environment and New Energies (LAPISEN) chemistry laboratory in order to evaluate the quality parameters of cashew apples. These parameters are related to sugar content, dry matter content, water content, vitamin C and pH. Sugar content of

each fruit was measured using Abbe refractometer (Novex 98.490). The unit of measurement for the sugar content is Brix. The pH of the samples was determined according to the method described by Adou et al. (2012). Following this method, the pH was directly measured on juice collected from each sample by means of a pH meter (HANA, HI98240, CHINA). Vitamin C is extracted in the presence of a solution of metaphosphoric acid and dosed with 2,6-dichlorophenol indophenol. For the determination, 1 ml of the sample is taken and assayed with 2,6-dichlorophenol indophenol solution until a persistent champagne pink coloration appears. The water content and dry matter content were determined according to the BIPEA (Analytiques, 1976) method. To determine the water content, a quantity of 5g of the sample was taken and weighed using a scale (SARTORUS ANALYTIC, Germany, precision 0.0001g) in a previously measured mass crucible (M_0). The set (crucible + sample) mass M_1 is put in the oven (MEMMERT 854 type B 40, West Germany) at 105 ° C for 24 hours. After cooling, the assembly is again weighed (M_2). The two parameters are then obtained from the following formula (Elbatawi et al., 2008):

$$\text{Dry mater content}(\%) = \frac{M_2 - M_0}{M_1} \times 100 \quad (2)$$

$$\text{Water content}(\%) = 100 - \text{Dry mater content} \quad (3)$$

Data analysis

Method for weighted resampling

The data analysis process consists of establishing mathematical model to predict chemical parameters, using reflectance spectra. But we do not have enough spectral data to do it. We proceeded to a resampling by the bootstrap method, which consists of creating statistical “new samples”, but only by drawing with discount, from the initial sample using the Wong and Easton algorithm (Wong and Easton, 1980).

Artificial neural networks

The data analysis process consists of establishing mathematical model to predict chemical parameters, using reflectance spectra. The visible reflectance spectra are used as inputs in the neural network application. ANN is computing systems inspired by biological neural networks. It is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another (Abiodun et al., 2018).

The artificial neuron receives signals and can process them appropriately. They automatically generate identifying characteristics from the learning material that they process. The original goal of the ANN approach is to solve problems in the same way that a human brain does.

Before processing the data (neural network application), they were divided into two sets: training set and validation set. Afterward, training set were split up into test set, validation set and training set, with a relative percentage of 15% (test set), 15% (validation set) and 70% (training test) (Lammertyn et al., 1998). Evaluation performance of a neural network model is highly dependent on partitioning of samples in dataset.

Training data is presented to the network during training, and the network is adjusted according to its error. The validation data is used to measure network generalization and to halt training when generalization stops improving.

The testing data have no effect on training and so provide an independent measure of network performance during and after training.

The developed prediction model uses two-layer feedforward network, with a tangent hyperbolic (tanh) transfer function in the hidden layer, a linear transfer function in the output layer and a Levenberg-Marquardt algorithm to reduce the time of computation. The default number of hidden neurons is 10.

The activation function used for the hidden neurons is hyperbolic tangent (tanh).

Thus, the relation between the predicted values, *Output*, and the input of the network, *Input*, which are spectral reflectance of the retain wavelength, is written as a combination function of weight of connection between input layer and hidden layer, *LW*, weight of connections between hidden layer and output layer, *IW*, skew of entry layer, B_1 , skew of output layer, B_2 in the form:

$$\text{Output} = (LW \times \tanh(IW \times \text{input} + B_1)) + B_2 \quad (4)$$

Selection of wavelength

In this study, each spectral measurement on a cashew apple corresponds to 3648 intensity values corresponding to 3648 variables. We selected only visible-near-infrared wavelengths (400 nm to 920 nm) because the spectrometer that we used in this study contained a lot of noise beyond the visible zone. Therefore, we limited ourselves to this zone. This selection will subsequently reduce noise and improve the predictive performance of our predictive models. Thus, we obtained 2000 variables per sample in the database. Each sample then contains 2000 spectral information.

Conclusion

In this study, a spectroscopic remote sensing technique was developed to evaluate the quality parameters of cashew apples. ANN technique was used to establish models for predicting quality parameters in cashew samples. The study has shown great accuracy in predicting different quality parameters. The RMSE were low for the global model from the selected wavelengths, giving an indication of the higher reliability of the calibration models. The results obtained also showed a strong correlation between the experimental values and the values predicted by the models. This means that the quality parameters of cashew apples can be predicted using remote sensing spectroscopy. The overfitting can be solved using a Bayesian Criterion Information to set a reasonable number of hidden layers in the ANN. This can be exploited for future studies.

Acknowledgements

SIDA (the Swedish International Development Cooperation Agency) through ISP (the International Science Program, Uppsala University) is acknowledged for financial support as well as the Third World Academy of Sciences (TWAS).

References

- Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA and Arshad H (2018) State-of-the-art in artificial neural network applications: A survey. *Heliyon*, 4(11), e00938.
- Adou M, Tetchi FA, Gbané M, Kouassi KN and Amani NGG (2012) Physico-chemical characterization of cashew apple juice (*Anacardium occidentale* L.) from yamoussoukro (Côte D’ Ivoire). *Innovative Romanian Food Biotechnology*. 11: 32–43.

- Analytiques BI d'Etudes (1976) Recueil des méthodes d'analyse des communautés européennes, BIPEA Gennevilliers, France 140p.
- Atzberger C (2013) Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. *Remote Sensing*. 5(2): 949-981. doi:10.3390/rs5020949
- Assoi EKN, Dibi W, Kouakou AK, Kouakou B, Koffi T, Zoueu JT (2019) Nouvel outil de mesure non destructive de la qualité des mangues par télédétection. *Afrique Science*. 15(2), 106-116.
- Daramola B (2013) Assessment of some aspects of phytonutrients of cashew apple juice of domestic origin in Nigeria. *Afr J Food Sci*. 7(6), 107-112
- Elbatawi IE, Ebaid MT, Hemeda BE (2008) Determination of potato water content using NIR diffuse reflection method. *Misr J Ag Eng*. 25: 1279-1292.
- Hammed LA, Anikwe JC, Adedeji AR (2008) Cashew nuts and production development in Nigeria. *Amer-Eurasian J Sci Res*. 3 (1): 54-61
- Islam MN, Colon T, Vargas T (1993) Effect of prolonged solar exposure on the vitamin C contents of tropical fruits. *Food Chemistry*. 48: 75-78.
- Kouakou BK, Bagui OKJ, Zoueu T (2016) Développement de techniques optiques d'identification d'insectes volants basées sur la spectroscopie par télédétection. Application à l'activité des insectes nuisibles aux cultures vivrières. *Afrique Science*. 12(5): 25-33.
- Lammertyn J, Nicolai B, Ooms K, De Smedt V, De Baerdemaeker J (1998) Non-destructive measurement of acidity, soluble solids, and firmness of Jonagold apples using NIR-spectroscopy. *Transactions of the American Society of Agricultural Engineers*. 41(4): 1089-1094.
- Lautié E, Dorniera M, Filhoc MDS, Reynesa M (2001), Les produits de l'anacardier: Caractéristiques, voies de valorisation et marchés. *Fruits*. 56(4): 235-248, doi: 10.1051/fruits:2001126.
- Lichtenthaler HK (1987) Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes. *Plant Cell Membranes*. 350-382, doi: 10.1016/0076-6879(87)48036-1
- Malley DF, Martin PD, Ben-Dor E (2004) Application in analysis of soils. *Near-infrared spectroscopy in agriculture, (nearinfrared spe)*. 729-784.
- Marc A, Achille TF, Mory G, Niaba Koffi PV and Georges AN (2011) Minerals composition of the cashew apple juice (*Anacardium occidentale* L.) of Yamoussoukro, Cote d'Ivoire. *Pakistan Journal of Nutrition*. 10: 1109-1114.
- Mireei SA, Mohtasebi SS, Massudi R, Rafiee S, Arabanian AS, Berardinelli A (2010) Non-destructive measurement of moisture and soluble solids content of Mazafati date fruit by NIR spectroscopy. *Aust J Crop Sci*. 4(3):175-179
- Olalusi AP and Ajayi OC (2020) Erinde, Selected Physio-Chemical Properties for the Processing of Cashew Apple (*Anacardium occidentale*). *Earth and Environmental Science*. 445: 012012
- Orynbaikyzy A, Gessner U, Conrad C (2019) Crop type classification using a combination of optical and radar remote sensing data: a review. *Int J Remote Sensing*. 1-43.
- Pujar DU, Pujar UU, Cr S, Wadagave A (2017) Remote sensing in fruit crops. *J Pharmacognosy Phytochemistry*. 6(5):2479-2484.
- Runjala S and Kella L (2017) Cashew apple (*Anacardium occidentale* L.) therapeutic benefits, processing and product development: An overview. *The Pharma Innovation*. 6(7):260-264.
- Sivagurunathan P, Sivasankari S, Muthukaruppan SM (2010) characterization of cashew apple (*Anacardium occidentale* L.) fruits collected from ariyalur district. *J Biosciences Res*. 1(2):101-107
- Tran NT and Fukuzawa M (2020) A portable spectrometric system for quantitative prediction of the soluble solids content of apples with a pre-calibrated multispectral sensor chipset. *Sensors*. 20 (20): 5883.
- Tian X, Li J, Yi S, Jin G, Qiu X, Li Y (2020) Nondestructive determining the soluble solids content of citrus using near infrared transmittance technology combined with the variable selection algorithm. *Artificial Intelligence in Agriculture*. 4: 48-57.
- Uwadaira Y, Sekiyama Y, Ikehata A (2018) An examination of the principle of non-destructive flesh firmness measurement of peach fruit by using VIS-NIR spectroscopy. *Heliyon*. 4(2): e00531.
- Wong CK and Easton MC (1980) An efficient method for weighted sampling without replacement. *SIAM J Computing*. 9(1): 111-113.

Entomological remote dark field signal extraction by maximum noise fraction and unsupervised clustering for species identification

Benoit Kouassi Kouakou^{*†}, Eliane Koko Nadine Assoi[†], Adolphe Yatana Gbogbo[†] and Jeremie Zoueu[†]

Laboratoire d'Instrumentation, Image et Spectroscopie; Institut National Polytechnique Houphouët Boigny, BP 1093 Yamoussoukro, Côte d'Ivoire

*Author for correspondence (benoitkouakou@gmail.com).

†Present address : Laboratoire d'Instrumentation Image et Spectroscopie, Institut National Polytechnique Houphouët Boigny, BP 1093, Yamoussoukro, Côte d'Ivoire

1. Abstract

Characterization of flying insects *in-situ* measurement using remote sensing spectroscopy is an emerging research field. Also, most analysis techniques in remote sensing spectroscopy are based on the use of an intensity threshold which introduces indeterminacies in the number of detected specimens. In this manuscript, we investigated the possibility of analyzing passive remote sensing spectroscopy measurement data using the maximum noise fraction method. The results obtained show that this analysis technique can help to overcome the measurement of background noise in spectroscopic measurements.

Keywords: Remote sensing; spectroscopy; blind source separation; unsupervised clustering; insects

2. Introduction

Remote sensing spectroscopy is an indirect measurement technique used to characterize aerial fauna and atmospheric particles. This measurement technique is done through instrument generally composed by reflector to collect backscattered light by object of interest, refractor to extend incident beam and sensors to measure backscattered light [1]. Several remote sensing spectroscopy instruments are been developed in recent decades to measure the spectral footprint [2,3], wing beat frequency [4], optical cross section of flies species [5,6]. These instruments are calibrated sometimes to measure fly direction [4] and detection range of insects [7,8]. These measurement instrument do not capture simultaneously image and identification parameters of the specimens studied, namely spectral footprint, wing beat frequency and optical cross section. Also, backscatter intensities due to insects are extracted using intensity threshold which sometimes omits intensities backscattered by small specimens [2]. Intensity threshold derived in conjunction of median intensities, the interquartile range of intensities and the choice of an arbitrary constant related to the signal to noise ratio of each data file has been proposed [9]. The arbitrary choice of this constant is subjective and leads to loss of information. Blind source separation methods seem to be an attractive alternative [10–12]. These information extraction techniques have been widely used in biomedical engineering, speech recognition systems and telecommunications. An extension of this technique has been proposed by some authors under the name of maximum noise fraction (MNF) [13]. In order to overcome the shortcomings of conventional indirect measurement systems, we are investigating a technique for simultaneous indirect measurement of the spectral footprint and image capture of flying specimens and analysis techniques based on the MNF for feature extraction. An unsupervised cluster analysis was done to identify insects' species.

3. Experimental set up and synchronous acquisition of footprint and image

The equipment is composed of two receivers a reflector (Skywatcher telescope ; D=200 mm ; F=1000 mm ; coated optic) and a refractor (Bresser Messier ; AR 90 Achromatic refractor telescope 90/900 f/10.0 ; coated optics) , mounted in parallel on an aluminum bar with a separation distance L (L=500 mm). The whole was fixed on an equatorial mount. The position of the counterweight is adjusted to determine the equilibrium position in the xy plane and in the yz plane. It is often necessary to move on the position of the telescope to adjust the balance in the yz plane. The determination of the equilibrium point allows flexibility in the handling of the instrument. This also allows the instrument to remain fixed at low atmospheric pressures, which minimizes the margins of error in measurements (Fig. 1).

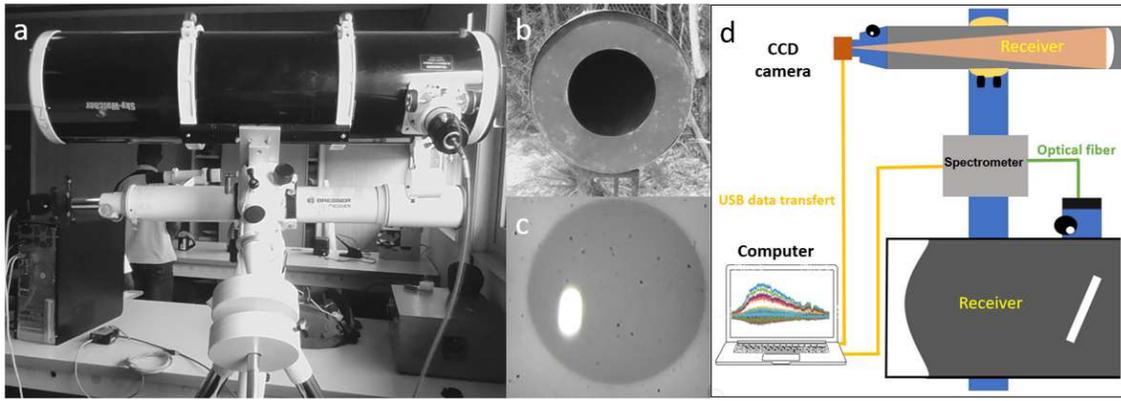


Figure 1: Panoramic view of the setup. (a) Passive remote sensing equipment consisting with two receivers, a reflector to measure the spectral footprint and the reflector for imaging. (b) Dark box to reduce the background backscatter and avoid saturation of spectrometer. (c) Image of a Lambertian sphere falling in front of the dark box to measure the reference intensity. (d) A schematic diagram of the setup.

Before any measurement, the reflector and the refractor are aligned on the black box located 1 meter above the ground, about 60 m from the experimental device (Fig. 1). This alignment phase consists of first, using an eyepiece, then a CCD camera connected to the computer to observe the orifice of the black box. The alignment phase of the telescope and the telescope allows the spectral and spatial information of the flying specimen to be acquired synchronously. Once the alignment phase is complete, the CCD camera is connected to the telescope then to the computer to record the image and the spectrometer on the telescope then to the computer to record the spectral imprint of the insect which is introduced into the volume defined by the intersection of the solid angle of the refractor and the reflector. The spectrometer and the camera are controlled by a computer program that we have developed using MATLAB for automated acquisition. The reference measurements are carried out by free-falling white Lambertian spheres in the focal plane of the black box. The release time is noted for future identification of the reference data stored on the computer. The integration time of the spectrometer has been set at 20 ms and the number of samples acquired is 1000.

4. Method for extracting and determining the spectral footprint

Raw data, $I_{raw,t}$, recorded is a 16 bits files, containing 3648 pixels x 1000 samples. Backscattered intensities due to dark box reflection and electronic noise intensities generated by electronic chain of spectrometer are add up to different measurements at each instant. These intensities represent the static intensities, $I_{static,t}$. Interquartile range intensity is insensitive to intensity flows backscattered by insects, atmospheric particles, plant leaves, aerosols and other undesirable factors such as smoke entering probe volume. This intensity has been calculated and represent $I_{static,t}$. Then, $I_{raw,t}$ was centered around $I_{static,t}$ to get the intensity I_t . Then, derived intensity I_t is supposed to contain backscattered intensity by an insect $I_{insect,t}$ and intensities backscattered in the volume probed by unknown objects considered as noise, $I_{noisy,t}$, at time t . $I_{insect,t}$ and $I_{noisy,t}$ are assumed to be independent and orthogonal and with zero means. We can write:

$$I_t = I_{insect,t} + I_{noisy,t} \quad (1)$$

The covariance of I_t , $I_t^T I_t$, is express as a function of the covariance of $I_{insect,t}$ and the covariance of $I_{noisy,t}$ [14]:

$$I_t^T I_t = I_{insect,t}^T I_{insect,t} + I_{noisy,t}^T I_{noisy,t} \quad (2)$$

The principle of the maximum noise fraction consists in finding a basic vector, ϑ , such that the maximum noise fraction, f_ϑ , define by equation (3) is optimal [13].

$$f_\vartheta = \frac{\|I_{noisy,t}\vartheta\|}{\|I_t\vartheta\|} \quad (3)$$

To determine the value of ϑ which makes optimal f_ϑ it is necessary to set to zero the derivative of f_ϑ with respect to ϑ . The calculus led to equation (4). This equation is the generalized eigenvalues problem.

$$I_{noisy,t}^T I_{noisy,t} = f_\vartheta I_t^T I_t \quad (4)$$

The solution of this equation to generalized eigenvalues leads to obtaining the values of the linear transformation matrix, B . This matrix separates and orders the non-correlated spectra by decreasing values of the maximum noise fraction, f_ϑ . The equation of this linear transformation links the uncorrelated spectral intensity, Y_t , with measured backscattered intensity, I_t :

$$Y_t = B^T I_t \quad (5)$$

Knowledge of background noise, $I_{noisy,t}$, in this methodology is required to derive B . Several existing techniques allow to estimate $I_{noisy,t}$. The method of differentiating neighboring pixels is generally used in image processing. $I_{noisy,t}$ was evaluated, by determining the difference of the intensities of the neighboring samples.

$$I_{noisy,t} = I_t - I_{t+1} \quad (6)$$

After extracting the spectra of interest, the spectral footprint, R_{insect} , characteristic of each species of specimen, was calculated based on the spectrum of the specimen, I_{insect} , and the spectrum of the Lambertian sphere, I_{ref} , according equation (7). In this equation, the subtraction of the intensity of the background noise does not appear, because the background noise was dissociated from the spectrum of interest during the extraction of the spectrum of interest (Fig. 2).

$$R_{insect} = \frac{I_{insect}}{I_{ref}} \quad (7)$$

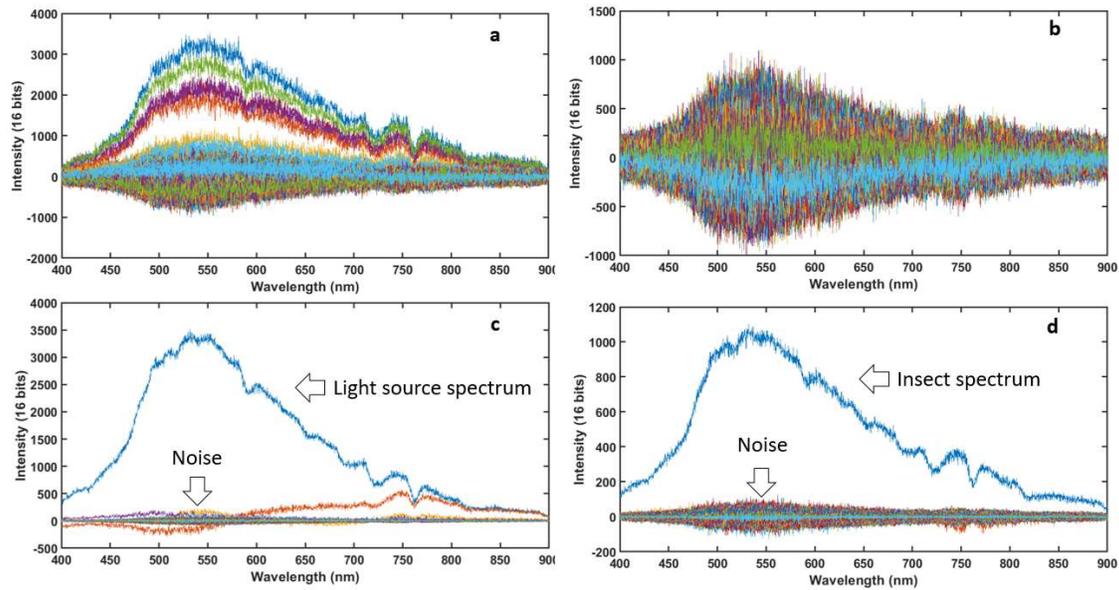


Figure 2: Examples of spectral of two events (a,b) Raw data centered around static intensity. (a) corresponds to reference spectra and (b) insect spectrum passing in the probe volume. (c,d) Spectra obtained after application of the maximum noise fraction method. The spectra were ordered according to the decreasing signal to noise ratio (SNR). The spectrum of interest is the high intensity spectrum (blue curves). In (c), highest intensity spectrum is sun spectrum which is used as illumination light; in (d), highest intensity spectrum is the spectrum of an insect detected by the acquisition system. Lower intensity spectra are considered to be background noise.

5. Identification of insects by their spectral footprints and their associated images by unsupervised classification

The spectral fingerprints determined were normalized and then grouped together. The Euclidean distance was then calculated. The decision criterion which allowed the determination of the number of classes (clusters) is a method based on the elbow detection called L-method [15–17]. The L-method consists of dividing the data relating to the associated distances into two batches and finding the best linear fit of the two batches (Fig. 3a) which minimizes the total mean square error, $RMSE_T$, (Fig. 3b) calculate according to the mean square error of the adjustment of each of the lots on the left (Fig. 3a, red curve), $rmse_L$, and right (Fig. 3a, green curve), $rmse_R$, the number of clusters, a , the maximum number of clusters, b , according equation (8).

$$RMSE_T = \frac{a-1}{b-1} rmse_L + \frac{b-a}{b-1} rmse_R \quad (8)$$

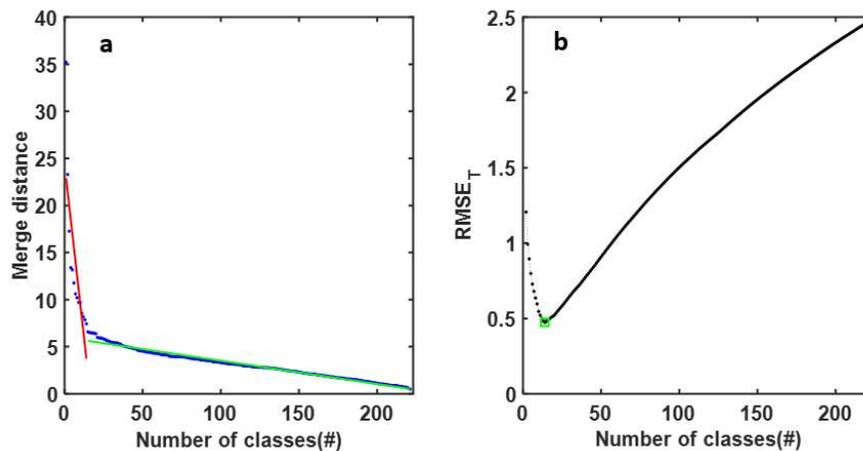


Figure 3: (a) Associated Euclidean distance vs number of classes. The intersection of the regression line in red and the regression line in green determines the decision boundary for the number of classes. (b) the minimum value of the total mean square error of regression of the lines in red and in green gives the value of the class number. The number of classes in this study was estimated at 14.

After determining the number of classes, the calculated Euclidean distances are used to construct the different classes according to their similarity. The determination of the corresponding image was carried out manually using the time of detection of the insect by the measurement system.

Figure Fig. 4 shows the central spectra of the 14 classes of insects identified and Fig. 5 shows the image associated with the central spectrum of each of the classes.

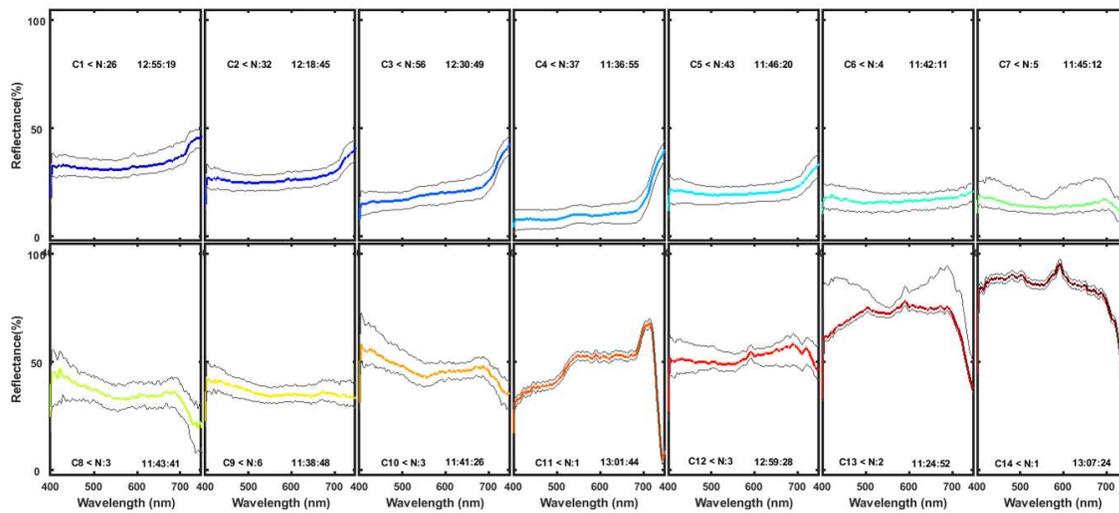


Figure 4: Central spectral fingerprints of the 14 species of flying specimens detected by the measurement system. In each sub-figure, the class of each species of flying specimens is indicated by C1, C2,..., The number of flying specimens in each class is indicated by N. On the far right at the top of the top sub-figures the time at which the flying specimen was detected is indicated and in the lower figures this time is written at the bottom right. The shaded curves indicate the interquartile range within a group.

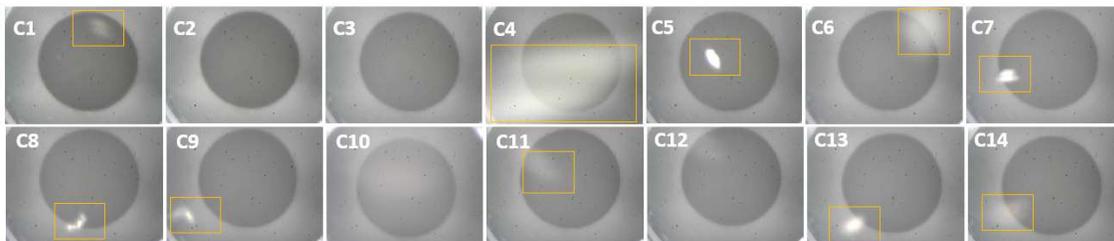


Figure 5: Raw images of flying specimens associated with different central spectral imprints. The captured specimens are identified on each sub-figure by a rectangle in orange. The images are saturated and do not identify the species detected. Certain fast flapping specimens have passed without being detected by the camera, for example in C2, C3, C10, etc. Others have passed far from the object focal plane. The image obtained in this case is blurred. This is the case of the images in C4, C6, C14, ..

6. Conclusion

We have implemented a device for synchronous acquisition of the image and the spectral imprint of flying specimens by passive remote sensing. We used the maximum noise fraction method to extract the spectra backscattered by the flying specimens in a set of 1000 samples containing the background noises. An unsupervised classification was used over a short measurement period to group the insect species detected by our system and thus identify each class of insect by an acquired image. As the camera used was not fast, we obtained saturated images. The simultaneous acquisition of the image and the spectral imprint being a major stake, this study can be deepened by using a fast camera.

Acknowledgments

The authors acknowledge Mikkel Brydegaard and Samuel Jansson for their technical support.

Funding Statement

General financial support from Swedish International Development Cooperation Agency (SIDA) through International Science Program (ISP), Uppsala University as well as the Third World Academy of Sciences (TWAS).

Data Accessibility

This article has no additional data.

Competing Interests

We have no competing interests.

References

1. Brydegaard M, Merdasa A, Gebru A, Jayaweera H, Svanberg S. 2016 Realistic Instrumentation Platform for Active and Passive Optical Remote Sensing. *Appl. Spectrosc.* (doi:10.1177/0003702815620564)
2. Runemark A, Wellenreuther M, Jayaweera HHE, Svanberg S, Brydegaard M. 2012 Rare events in remote dark-field spectroscopy: An ecological case study of insects. *IEEE J. Sel. Top. Quantum Electron.* (doi:10.1109/JSTQE.2012.2184528)
3. Brydegaard M, Jansson S. 2019 Advances in entomological laser radar. *J. Eng.* (doi:10.1049/joe.2019.0598)
4. Gebru A, Rohwer E, Neethling P, Brydegaard M. 2014 Investigation of atmospheric insect wing-beat frequencies and iridescence features using a multi-spectral kHz remote detection system. In *Remote Sensing and Modeling of Ecosystems for Sustainability XI*, (doi:10.1117/12.2060827)
5. Brydegaard M. 2015 Towards quantitative optical cross sections in entomological laser radar - Potential of temporal and spherical parameterizations for identifying atmospheric fauna. *PLoS One* (doi:10.1371/journal.pone.0135231)
6. Kirkeby C, Wellenreuther M, Brydegaard M. 2016 Observations of movement dynamics of flying insects using high resolution lidar. *Sci. Rep.* (doi:10.1038/srep29083)
7. Mei L, Brydegaard M. 2015 Atmospheric aerosol monitoring by an elastic Scheimpflug lidar system. *Opt. Express* (doi:10.1364/oe.23.0a1613)
8. Jansson S, Brydegaard M. 2018 Passive kHz lidar for the quantification of insect activity and dispersal. *Anim. Biotelemetry* (doi:10.1186/s40317-018-0151-5)
9. Malmqvist E, Jansson S, Torok S, Brydegaard M. 2016 Effective parameterization of laser radar observations of atmospheric fauna. *IEEE J. Sel. Top. Quantum Electron.* (doi:10.1109/JSTQE.2015.2506616)
10. Parra L, Sajda P. 2004 Blind source separation via generalized eigenvalue decomposition. In *Journal of Machine Learning Research*, (doi:10.1162/jmlr.2003.4.7-8.1261)
11. Pham DT, Cardoso JF. 2001 Blind separation of instantaneous mixtures of nonstationary sources. *IEEE Trans. Signal Process.* (doi:10.1109/78.942614)
12. Jutten C, Herault J. 1991 Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing* (doi:10.1016/0165-1684(91)90079-X)
13. Hundley DR, Kirby MJ, Anderle M. 2002 Blind source separation using the maximum signal fraction approach. *Signal Processing* (doi:10.1016/s0165-1684(02)00342-0)
14. Sun L, Rieger J, Hinrichs H. 2009 Maximum noise fraction (MNF) transformation to remove ballistocardiographic artifacts in EEG signals recorded during fMRI scanning. *Neuroimage* (doi:10.1016/j.neuroimage.2009.01.059)
15. Zagouras A, Inman RH, Coimbra CFM. 2014 On the determination of coherent solar microclimates for utility planning and operations. *Sol. Energy* (doi:10.1016/j.solener.2014.01.021)
16. Zagouras A, Kazantzidis A, Nikitidou E, Argiriou AA. 2013 Determination of measuring sites for solar irradiance, based on cluster analysis of satellite-derived cloud estimations. *Sol. Energy* (doi:10.1016/j.solener.2013.08.005)
17. Kouakou BK, Jansson S, Brydegaard M, Zoueu JT. 2020 Entomological Scheimpflug lidar for estimating unique insect classes in-situ field test from Ivory Coast. *OSA Contin.* (doi:10.1364/osac.387727)