University of Cape Coast https://ir.ucc.edu.gh/xmlui

UNIVERSITY OF CAPE COAST

ASSESSING THE IMPACT OF PREHARVEST AND POSTHARVEST FACTORS ON COCOA BEAN QUALITY USING STANDARD METHODS AND NOVEL HANDHELD NIR SPECTROMETER COUPLED WITH CHEMOMETRICS

ELLIOT KWAKU ANYIDOHO

2022

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

© Elliot Kwaku Anyidoho

University of Cape Coast

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

UNIVERSITY OF CAPE COAST

ASSESSING THE IMPACT OF PREHARVEST AND POSTHARVEST FACTORS ON COCOA BEAN QUALITY USING STANDARD METHODS AND NOVEL HANDHELD NIR SPECTROMETER COUPLED WITH CHEMOMETRICS

BY

ELLIOT KWAKU ANYIDOHO

Thesis submitted to the Department of Agricultural Engineering of the School of Agriculture, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Doctor of Philosophy degree in Food and Postharvest Technology

APRIL 2022

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature:..... Date:.....

Name: Elliot Kwaku Anyidoho

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature:..... Date:.....

Name: Rev. Dr. Ernest Teye

Co-Supervisor's Signature: ..... Date:.....

Name: Dr. Robert Agbemafle

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

ABSTRACT

Cocoa bean quality is influenced by preharvest and postharvest factors. Rapid evaluation is therefore required to aid in decision-making. In this study, four experiments were performed separately using a completely randomized design with age class, pollination type, production method, cocoa-producing region and geographical location as the main factors. The novel handheld NIR spectrometer combined with multivariate qualitative algorithms gave a 100 % classification rate for cocoa beans; from the seven cocoa-growing regions in Ghana, four geographical locations in Africa, fermented against unfermented, and organic against conventional. Quantitatively, the performance of the regression models for simultaneous prediction of fermentation index, pH, fat, polyphenols, flavonoids, and antioxidant capacity was in the range of: $0.87 \leq R^2_{cal} \leq 0.99$ and $0.88 \leq R^2_{pre} \leq 0.99$ in calibration and prediction sets respectively. Cocoa beans' physical, chemical and mineral properties were significantly impacted by age of the cocoa tree and pollination type. Calcium, magnesium, phosphorus and potassium were found in the range of 111.44 - 125.23, 238.79 - 249.05, 528.24 - 541.40 and 473.05 - 631.34 mg/100g, respectively whereas sodium, iron, copper and zinc were found in the range of 7.08 - 11.54, 5.80 - 8.83, 1.34 -3.33 and 2.36 - 5.14 mg/100 g, respectively for cocoa bean categories examined. Generally, the NIR spectroscopic technique developed correlated well with the wet chemistry method (R^2 = 0.93). The outcome of the study reveals that the handheld NIR spectroscopic technique can be used for rapid, non-destructive and on-site measurement of cocoa beans quality parameters qualitatively and quantitatively.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

LIST OF PUBLICATIONS

(1) **Elliot K. Anyidoho**, Ernest Teye and Robert Agbemafle, (2021a). Differentiation of organic cocoa beans and conventional ones by using handheld NIR spectroscopy and multivariate classification techniques, Hindawi *International Journal of Food Science.* doi:10.1155/2021/ 1844675.

(2) **Elliot K. Anyidoho**, Ernest Teye, Robert Agbemafle, Charles L. Y. Amuah and Vida Gyimah Boadu, (2021b). Application of portable near infrared spectroscopy for classifying and quantifying cocoa bean quality parameters, *Journal of Food Processing Preservation.* doi:10.1111/jfpp. 15445.

(3) **Elliot K. Anyidoho**, Ernest Teye and Robert Agbemafle, (2020). Nondestructive authentication of the regional and geographical origin of cocoa beans by using a handheld NIR spectrometer and multivariate algorithm. *Royal Society of Chemistry: Analytical Methods*. doi:10.1039/ d0ay00901f.

(4) Ernest Teye, **Elliot K. Anyidoho**, Robert Agbemafle, Livingstone K. Sam-Amoah, Chris Elliott., (2020). Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A review, *J. Infrared Physics and Technology*. doi:10.1016/j.infrared.2019. 103127.

(5) **Elliot K. Anyidoho**, Ernest Teye and Robert Agbemafle, (Under review), Effect of cocoa tree age and pollination type on some physical and biochemical properties of cocoa beans (Journal of Measurement: Food).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

KEY WORDS

Cocoa beans

Conventional cocoa beans

Handheld near-infrared spectroscopy

Organic cocoa beans

Qualitative classification

Quantitative estimation

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

ACKNOWLEDGMENTS

I am very grateful to the God Most High for the grace to make this study a success.

I would like to extend my sincere and deepest gratitude to my main supervisor and advisor, Rev. Dr. Ernest Teye for his constructive critiques, excellent guidance, suggestions and constant encouragement throughout my study at the University of Cape Coast. I have benefited from his rich and excellent knowledge, organisation and expertise. I also thank my co-supervisor Dr. Robert Agbemafle for his valuable advice and support with all the patience and care throughout my studies. I appreciate your meticulous supervision, which led to the publication of three papers (listed on the last page of this thesis) from this study.

Special appreciation also goes to Prof. Ernest Abano, Prof. Mrs. Sarah Darkwa, Rev. Prof. Joshua Owusu-Sekyere, Prof. Livingstone K. Sam-Amoah, Prof. Robert S. Amoah and Dr. Charles L.Y. Amuah for their encouragement and advice.

Many thanks to my fellow postgraduate students, research team and laboratory mate; Mrs. Roseline McArthur, Mr. Francis Padi Lamptey and Mrs. Vida Gyimah Boadu for their valuable assistance throughout the research work.

I would like to express my deepest appreciation to my lovely wife Yvonne Ibiele Anyidoho for her unwavering support, prayers and inspiration.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

DEDICATION

To my loving wife; Yvonne Ibiele Anyidoho and lovely children; Eyram, Klenam, Kekeli, and Esinam.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

TABLE OF CONTENTS

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

LIST OF TABLES

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

LIST OF FIGURES

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

LIST OF ABBREVIATIONS

AEAC	Ascorbic Acid Equivalent Antioxidant Content
ANN	Artificial Neural Networks
ANOVA	Analysis of Variance
CMC	Cocoa Marketing Company
CODAPEC	Cocoa Diseases and Pests Control Programme
DPPH	2,2-Diphenyl-1-Picrylhydrazyl
FD	First Derivative
FDA	Fishers Discriminant Analysis
FI	Fermentation Index
FmD	Fermentation Duration
g/kg	gram per kilogram
GAE	Gallic Acid Equivalent
GA-PLS	Genetic Algorithms Partial Least Square
HPLC	High performance liquid chromatography
i-PLS	Interval Partial Least Square
KNN	K-Nearest Neighbour
LBCs	Licensed Buying Companies
LDA	Linear Discriminant Analysis
LDA	Linear Discriminant Analysis
LID	Living Income Differential
MATLAB	MATrix LABoratory
MC	Mean Centering
mg/kg	milligram per kilogram

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

MSC	Multiplicative Scatter Correction
NIRS	Near-Infrared Spectroscopy
PCA	Principal Component Analysis
PCs	Principal Components
pH	Potential of Hydrogen
PLS	Partial Least Square
PLS-DA	Partial Least Square Discriminant Analysis
PLS-DA	Partial Least Square Discriminant Analysis
PLS-R	Partial Least Square Regression
PPRC	Producer Price Review Committee
QCC	Quality Control Company
QE	Quercetin Equivalent
R	Correlation Coefficient
RF	Random Forest
RMSECV	Root Mean Square Error of Cross-Validation
RMSEP	Root Mean Square Error of Prediction
SD	Second Derivative
Si-PLS	Synergy Interval Partial Least Square
SNV	Standard Normal Variant
SNV	Standard Normal Variant
SVM	Support Vector Machine
TAC	Total Antioxidant Capacity
TFC	Total Flavonoid Content
TPC	Total Polyphenol Content

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER ONE

INTRODUCTION

Background to the Study

Cocoa bean, simply referred to the seed obtained from the pod of the tree crop, is scientifically termed as *Theobroma cacao L*. It is currently the topmost commercially cultivated cash crop in more than 50 tropical and sub-tropical countries around the world (Dillinger et al., 2000). The combined value for exported cocoa beans, whether raw or roasted amounted to USD 8.6 billion in 2017 (Eghbal, 2018). It is projected that the world cocoa beans market will grow to reach USD 16.32 billion at a compound annual rate of 7.3% from 2019 to 2025 (Grand View Research, 2019). Africa accounts for 70% of the global cocoa beans production, with 68% coming from three Economic Community of West African States (ECOWAS) member countries namely Côte d'Ivoire, Ghana, and Nigeria. These three countries produced 3.5 million metric tons in the 2019/2020 crop season, out of a global total of 5.2 million metric tons (Santander et al., 2020). The economies of these countries benefited hugely from the cultivation and exportation of cocoa. For instance, in 2017 Côte d'Ivoire, Ghana, and Nigeria generated a total export value of USD 3.9 billion, USD 2.5 billion, and USD 0.8 billion, respectively (Voora, Bermúdez, & Larrea, 2019).

Ghana's cocoa beans enjoy a supreme reputation for the first-class quality in the global market and as such has become the standard by which cocoa beans from other countries are compared (Othman et al., 2007). The strict

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

measures put in place at the farmgate by the farmers, warehouses, and ports by the workers of the Quality Control Company of Ghana Cocoa Board might have yielded this result. In Ghana, there are seven cocoa-growing geographical regions (Eastern, Ashanti, Central, Volta, Brong Ahafo, Western North and Western South) where cocoa is cultivated on about 1.9 million hectares of land by some 800,000 farming families (Afoakwa, 2010). In these cocoa-growing regions Quality Control Company (QCC) of Ghana Cocoa Board (COCOBOD) has warehouses located at centralized points where cocoa bean quality is checked before onward transportation to the harbour for further quality check before shipment to the international market. Preharvest and postharvest treatments affect the quality of cocoa beans. Fresh cocoa beans are bitter, astringent with an unpleasant aroma and the elementary process to control these is through fermentation and drying to acquire the unique cocoa taste and flavour. The beans are scooped out of the cocoa pod after harvest and fermented for a period between four and six days before they are dried to a moisture content of 7%. The fermented dried cocoa beans are traded in the global market by a recognised institution.

Cocoa bean is a rich source of fats, proteins, fibre, carbohydrates, minerals, vitamins, and phytochemicals including polyphenols and flavonoids. Furthermore, studies have revealed that cocoa beans contain about three hundred and eighty (380) chemicals, ten (10) of which are psychoactive compounds that are very useful to the human body (Di Castelnuovo et al., 2012). These bioactive compounds are believed to possess physiological properties that have beneficial effects in human health such as treatment and prevention of

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

chronic ailments including cardiovascular diseases, inflammation, and cancer(Di Castelnuovo et al., 2012). Additionally, the availability of non-protein amino acids in cocoa beans is of great interest due to its biological activity in humans' and animals' health: acting as a vital inhibitory neurotransmitter of the central nervous system, improvement in the functioning of the kidney and liver, prevention of diuretic (water pills), diabetes, and lowering of blood pressure (Di Castelnuovo et al., 2012. Due to these nutritional and health benefits of cocoa, the consumption of cocoa bean and cocoa bean products has increased resulting in an increase in price. Similarly, cocoa bean quality control has moved to centre stage in recent years. In the global market stringent parameters are observed as a result of which cocoa beans from certain countries enjoyed premium price and others rejected due to poor quality, adulteration and mislabelling. Traditionally, methods used to evaluate cocoa beans for qualitative and quantitative assessments are sensory evaluation, cut test (Aculey et al., 2010), fermentation index (Tran et al., 2015), high-performance liquid chromatography (HPLC) (Ramli et al., 2001) and colourimetric test (Jonfia-Essien et al., 2008).

Conventional near-infrared (NIR) spectroscopy is an environmentally friendly and nondestructive instrumental technique that makes use of the naturally occurring electromagnetic spectrum for the acquisition of spectra of agricultural products (Ozaki et al., 2006). NIR spectroscopy covers a small portion of the electromagnetic spectrum (700-2500 nm). Teye et al. (2016) and Álvarez et al. (2012) used the traditional near-infrared spectroscopy to qualitatively and quantitatively analysis cocoa beans. The advancement in NIR

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

miniaturization has unlocked a new prospect for NIR applications suitable for industry, laboratories, and onsite analysis. This has led to the transformation of large, stationary laboratory-based NIR tools into handheld devices. Recent review articles describing the basics of portable NIR spectroscopy ascertained its potential in food control and inspection and the significance of developing and adopting it along the food supply chain has been emphasized and extensively reviewed for the inspection and control of meat (Kademi, Ulusoy & Hecer, 2019). Other authors evaluated a miniaturized spectrometer device for discriminating cultivars of barley, chickpea, and sorghum in Ethiopia (Kosmowski & Worku, 2018). Research works on the use of handheld NIR spectroscopy for cocoa bean analysis are insufficient and the few ones that are accessible considered very little about the cocoa beans from Africa where majority of the world's cocoa is produced. The examination of cocoa beans in the shortwave infrared spectral region of 900-1700 nm was seldom investigated and reported. If well developed, handheld NIR spectroscopy could be deployed for examining cocoa beans' quality.

Statement of the Problem

Cocoa beans are traded locally and internationally with high-quality standards of regulations (Adeyeye et al., 2010). This usually demands the need for examinations of the bean quality. Cocoa bean quality is linked to preharvest activities (production methods and age of the cocoa tree) and postharvest treatments (fermentation and drying). Inspection of cocoa bean quality is based on a destructive instrumental cut test, sensory evaluation (Aculey et al., 2010), fermentation index (Tran et al., 2015), colorimetric test (Jonfia-Essien et al.,

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

2008; Miller et al., 2006), and HPLC (Ramli et al., 2001) methods, which have several setbacks. These analytical methods are destructive, time-consuming, subjective, cumbersome, lack reproducibility and require the use of dangerous chemical solvents. Additionally, they are difficult to automate, require highly trained technical personnel. The next available technique for the evaluation of cocoa bean quality is the conventional NIR spectroscopic technology. The conventional NIR spectroscopy is expensive, stationary and laboratory-based. There is, therefore, the need to look for alternative techniques that can overcome all these problems.

Compared to the chemical methods and conventional NIR spectroscopy, handheld NIR spectrometers are less expensive and require minimal equipment and user involvement. Furthermore, handheld NIR spectrometers are easy to move to the field for on-site measurements and product-to-product examination and regulation.

However, upon literature search, no studies until now have been conducted to evaluate the impact of preharvest and postharvest treatments on cocoa bean quality using standard methods and handheld NIR spectroscopic technique in Africa and more specifically Ghana where a substantial proportion of the world's cocoa is produced. The result of this research will be useful to the producers of cocoa beans, processors, and consumers as well. It will transcend into production and supply of high-quality cocoa and could curb the rejection of Ghanaian cocoa beans in the international market. Subsequently, it will lead to the earning of more foreign exchange for the country and as such improve the socio-economic lives of the rural poor farmers.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Main objective

The main objective of this research was to assess the impact of some preharvest and postharvest factors on cocoa bean quality using standard methods and novel handheld NIR spectroscopic techniques.

Specific objectives

The specific objectives of the study were to:

1. determine and develop a novel rapid non-destructive authentication of regional and geographical cocoa beans.
2. determine and develop fermentation duration differentiation, and a real-time prediction of cocoa beans' quality parameters such as fermentation index, pH and moisture content.
3. determine and develop a novel qualitative and quantitative prediction of organic and conventional cocoa beans parameters; fats, polyphenols, flavonoids and antioxidants.
4. measure the effect of age class and pollination type on the physical and chemical properties of the cocoa beans.

Purpose of the Study

The purpose of the study is to assess the impact of some preharvest and postharvest factors on cocoa bean quality using traditional standard methods and novel handheld NIR spectrometer coupled with chemometrics. It examined the correlation between standard methods and novel handheld NIR spectrometer for examination of cocoa bean quality. Also, the study purposed to determine the effect of age of the cocoa tree and the type of pollination on some physical, proximate and phytochemical properties of cocoa bean.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Scope of the Study

The study aims at applying standard methods and a novel handheld NIR spectrometer to examine cocoa beans' quality. It identifies cocoa beans from all the producing-regions in Ghana, some African countries, organic cocoa beans, fermentation duration and simultaneously quantifies quality parameters. Also, the effect of cocoa tree age and type of pollination on some physical properties, proximate compositions and phytochemical properties of cocoa bean.

Significance of the Study

Quality is a crucial issue all over the world and this accounts for the reason why some cocoa beans sell at higher price than the others. And with the rising market of "healthy cocoa beans", consumers need to be informed of the quality, origin and veracity of cocoa beans. Also, in recent times, cocoa beans have been mislabelled for selfish financial gain. Processors and consumers have no fast way of detecting mislabelling, origin and quality of cocoa beans. The methods to detect these kinds of frauds and cocoa bean quality are rather: slow, chemical based, stationary laboratory-based and bench-top devices. In Ghana quality assessment before shipment is carried out at Tema and not at the farmyard because they are stationary laboratory based. All of these factors have underlined the need for fast and reliable techniques to assess cocoa bean quality.

Optimistically, NIR spectroscopy is a technology that has presented great potential to provide the needed solution. Some researchers have used NIR spectroscopy to: discriminate and quantify adulteration in cocoa beans (Teye et al., 2013), determine phytochemicals and other food quality parameters in bakery products (Bedini et al., 2013), determine fats, caffeine and theobromine

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

in unfermented criollo cocoa (Alvarez et al., 2012). However, upon a literature review, no work has examined the impact of preharvest and postharvest factors on cocoa beans' quality using standard methods and handheld near infrared spectroscopy. Especially in Africa and Ghana where about 70% of world's cocoa beans are produced. There is, therefore, a research gap that needs to be filled in this regard.

Delimitations

Although, the novel handheld NIR spectroscopy is fast, reliable and non-destructive, it depends on well-known reference methods closely tied to wet chemistry, it must be trained, require huge amount of previous measurements for the modelling, and reliance on sophisticated chemometric techniques for calibration or learning.

Limitations

This study is subject to a limitation that can be addressed in future studies. The cocoa bean samples were obtained from only four cocoa-producing countries in Africa and hence the findings can only be valid for only these countries.

General methodology

Study materials and area

The cocoa bean samples for the study were obtained from the seven cocoa-growing regions in Ghana namely Eastern, Ashanti, Central, Volta, Brong Ahafo, Western North and Western South Regions (appendix 1) and some four cocoa-growing countries in Africa including Uganda, Ivory Coast, Nigeria and Ghana (Appendix 2). The various categories of cocoa collected for

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

the study comprised unfermented, partially fermented, organic, conventional, naturally pollinated and artificially pollinated cocoa beans. Sample selection was randomized all through the study. The beans were transported to the Research Laboratory located at the Technology Village of the School of Agriculture, University of Cape Coast for investigations.

Instruments and experimental procedure

The handheld NIR spectrometer (Tellspec^{R}, UK) operated via a smartphone (Samsung A20) was deployed for scanning the cocoa bean samples in a clear zip-locked polythene bag as diagrammatically displayed in Figure 4.1. The handheld NIR spectrometer acquired the spectra of the experimental cocoa beans in a wavelength range of 900 to 1700 nm. The handheld NIR spectrometer took the spectrum of each cocoa bean sample in absorbance unit (i.e., log 1/R); where R = reflectance. Immediately after scanning (each sample was scanned three times), the cocoa beans were ground, and the ground samples were thoroughly pulverized and sieved through a 0.21 mm diameter sieve to obtain a fine cocoa beans powder. The powders were used for the wet chemistry analysis of proximate, biochemical and mineral compositions to obtain their reference values. The quality parameters were determined using internationally accepted standards, protocols and methods of the Association of Official Agricultural Chemists (AOAC, 2005) and International Office of Cocoa, Chocolate, and Confectionery (International Office of Cocoa & Confectionary, 1996).

Data processing and analysis

The NIR spectral data recordings with their matching reference values were downloaded unto the computer and imported into chemometric software

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

(Matlab version 9.6.0 1072779, License Number 968398, the MathWorks, Inc., USA) with windows 10 Pro software packages for processing. Data analysis was performed by using different multivariate methods to determine the most suitable procedure to achieve simultaneous classification and quantification analysis of cocoa bean categories.

Organisation of the Thesis

This thesis consists of seven chapters. Chapter 1 is the introduction, background of the study, main and specific objectives of the study, purpose of the study, scope of the study, significance of the study, delimitations, limitations and general methodology. Chapter 2 gives an overview of relevant literature and theoretical foundations on the research subject. Chapters 3, 4, 5 and 6 outline the completed research articles or papers that are purposive to achieve the research objectives: (1), (2), (3) and (4) respectively. The summary of the study, conclusions, novelty statements and relevant recommendations for future research are presented in chapter 7. Finally, references and a list of publications emanating from the entire research are stated.

Chapter Summary

The chapter gave an overview of the cocoa bean, its production, cocoa crop seasons, supply chain and near-infrared spectroscopy. The background to the study, significance of the study, delimitations and limitations of the study, general methodology and organisation of the thesis were included within the chapter.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter reviews the literature on cocoa bean, some prevarvest activites, postharvest treatments and near-infrared specctroscopy. Works that have been done around the filed of study and theoretical frameworks involved in chemometric multivarivate data analysis were also discussed.

Cocoa bean production

Cocoa is a historic tree crop of the lowland tropical rainforests. Cocoa production provides livelihoods for 40-50 million farmers, rural workforces and their families around the globe (Voora et al., 2019). The world's cocoa production stood at roughly 5.2 million metric tons in the 2019/2020 crop season (Santander et al., 2020). Côte d'Ivoire and Ghana, the globe's top two cocoa producers, together with Nigeria and Cameroon produced approximately 3.5 million metric tons (Figure 2.1). The backbone of the economies of these countries hugely depends on the production and export of cocoa beans for foreign exchange. In Ghana, the cocoa bean is the largest agricultural export commodity (Aculey et al., 2010).

There are three varieties of cocoa plants namely 'Criollo', 'Forastero', and 'Trinitario' (Wood & Lass, 2008). The 'Criolllo' group is more vulnerable to diseases, gives low yield, possesses a pleasant aroma, bland flavour, and is reported to have the maximum quality cocoa bean which is considered a delicacy. It has its origin in Mexico and is still mainly grown in Central America (Wood & Lass, 2008). Table 2.1 gives details of the varieties of cocoa plant.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

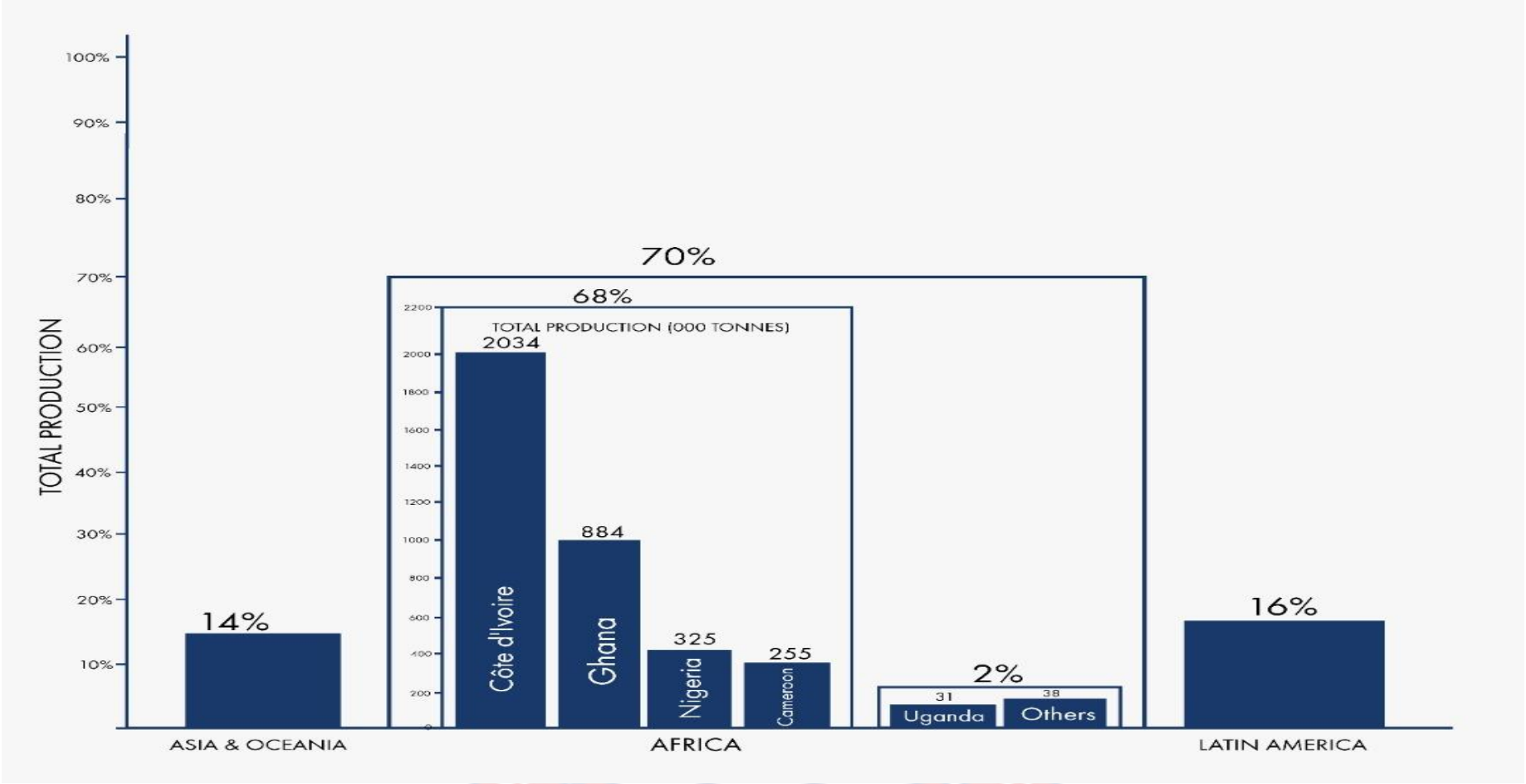


Figure 2.1 Global cocoa production in 2019/2020 (Shahbandeh, 2021)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 2.1 Varieties of cocoa plant and their characteristics, origin and observation

Variety	Characteristics	Origin	Observation
'Criollo' 	• Soft pod husk soft. • Warty and conspicuously furrowed pod surface. • Usually large and elongated fruit. • Red/green colour if unripe and yellow/orange colour if ripe. • Average number of beans per pod is 20-30. • Plumb seeds.	• Colombia • Guatemala • Mexico • Nicaragua • Indonesia • Venezuela	• Very susceptible to diseases. • Possesses pleasant aroma and bland flavour.
'Forastero' 	• Hard pod husk. • Pods are small and elongated. • Pod surface is not warty and furrows are inconspicuous. • Green fruit colour if unripe and yellow if ripe. • Average number of beans per pod is 30 or more. • Flat seeds.	• Brazil • Colombia • Ecuador • Peru	• Strong and high yielding. • Used for the selection of breeding varieties. • Leading planting material in cocoa cultivation. • Have harsh flavour with a bitter taste.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

'Trinitario' 	• Mostly hard pod husk. • Intermediate between 'Criollo' and 'Forastero'. • Red/purple fruit colour if Unripe and orange fruit colour if ripe. • Intermediate pod surface. • Average number of beans per pod is 30 or more. • Intermediate seeds.	• Trinidad • Venezuela	• Highly susceptible to Black pod. • Hybrid of 'Criollo' and 'Forastero'.

Source: Ghana Cocoa Board (2018)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

The 'Forastero' variety is very strong, high-yielding, widely used, and forms the bulk of world commercial cocoa bean mostly produced in Africa. The 'Amelonado' and Upper Amazon cocoa beans belong to this variety. 'Forastero' cocoa bean possesses a harsh flavour with a bitter taste. The 'Trinitario' variety is the hybrid of 'Criollo' and 'Forastero' groups. It is very heterogeneous, yields higher and of higher quality than the 'Forastero' variety. The 'Trinitario' variety was developed in Trinidad, hence its name, and today it is still the primary type of cocoa grown in the Caribbean. Fine flavour cocoa beans are obtained from 'Criollo' and 'Trinitario' varieties, whilst bulk cocoa beans are produced from 'Forastero' type (Ghana Cocoa Board, 2018).

Ghana's cocoa supply chain

Farmers or producers sell their cocoa beans through local buyers or cooperatives to exporters. Generally, a small quantity of cocoa beans is processed and sold to the domestic markets. The domestic marketing agents sell to exporters who are related to importers and processors in North America and Europe or who trade on the New York or London commodity exchanges. The processors grind cocoa beans into butter, liquor, and powder which they sell to the chocolate manufacturers and other food processors, who eventually sell their products to consumers.

Domestically, cocoa bean production and sale are under a well-structured organization known as COCOBOD. The entire cocoa supply chain of Ghana comprises farmers, input dealers, cooperatives, cocoa purchasing centres, Licensed Buying Companies (LBCs), transporters, local processors, worldwide manufacturers, local and international consumers, and COCOBOD.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

The monitoring and supervision of activities on the domestic divide are under the auspices of COCOBOD. The LBCs through their purchasing clerks purchase the cocoa beans at the minimum price determined by a Producer Price Review Committee (PPRC) from the farmers at the cocoa purchasing centres established in the cocoa-growing communities. The PPRC comprises the officials of COCOBOD, government representatives, farmers' representatives, and LBCs representatives. The LBCs after buying the cocoa beans invite the QCC to grade and seal the cocoa at a fee set by the PPRC. The graded and sealed cocoa beans are evacuated by the private hauliers to the designated take-over points specifically located at Takoradi, Tema, and Kaase (inland port at Kumasi). At the take-over points, Cocoa Marketing Company Limited (CMC) a subsidiary of COCOBOD takes charge of the graded and sealed cocoa beans. The CMC has the singular responsibility for the sale and export of Ghana's cocoa beans. Furthermore, the CMC is responsible for the procurement of graded and sealed cocoa from the LBCs at the take-over locations, cocoa beans stocking before shipment, collection of receipts, and management of sales (Bangmarigu & Qineti, 2018). The QCC, before the shipment of cocoa beans to the international market, inspects and fumigates all the cocoa consignments and shipping vessels. Roughly, 75% of the cocoa beans are exported whilst the remaining 25% is used in the production of confectionery products (Abbadi et al., 2019). As well, the export of domestically processed cocoa bean products to the international market is performed by CMC while those that are not exported are sold to local consumers. A simple diagram as shown in Figure 2.2 illustrates Ghana's cocoa supply chains.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

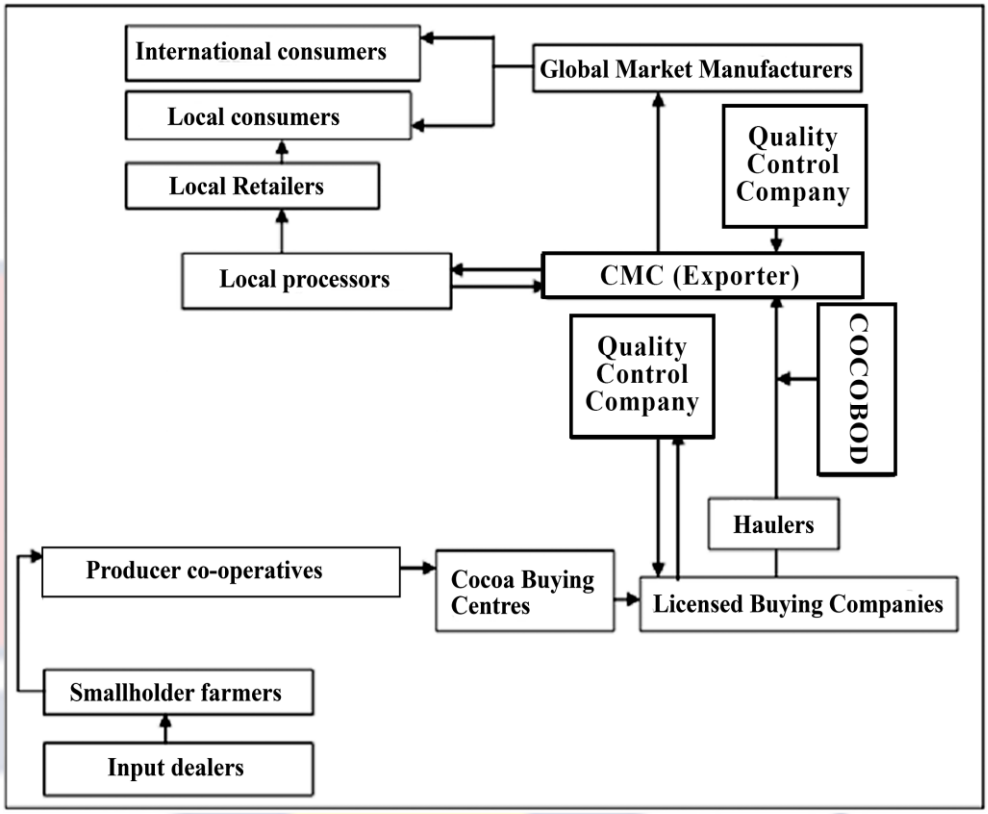


Figure 2.2 Ghana's cocoa supply chains

In 2020, irrespective of the devastating effects of COVID-19 on the world's economy and the seeming risks associated with the pandemic, COCOBOD secured a syndicated loan facility of USD 1.3 billion from a coalition of 28 institutions, made of 4 local and 24 international banks and financial institutions, including Standard Chartered, ABN Amro and Bank of China. The loan facility, which had an interest rate plus libor of 1.75%, was projected to finance the purchase of about 900,000 metric tons of cocoa beans and related operational activities for the 2020/2021 crop season, was repayable in seven (7) calendar months (Ghana Cocoa Board, 2020). COCOBOD instituted a lot of productivity enhancement measures to motivate cocoa farmers. Notable among these measures are the provision of free hybrid

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

seedlings, awarding of scholarships to farmers' children, rehabilitation of moribund and diseased farms, irrigation of cocoa farms, provision of extension and technical support, artificial hand pollination, and cocoa diseases and pests control programme (CODAPEC). These productivity enhancement measures coupled with monitoring and supervision of on-farm and postharvest activities by the various institutions of COCOBOD have consequently made Ghana's cocoa beans to be of superior quality and as such obtains a price premium ranging 4-6% per metric ton in the international market (Gilbert, 2009).

At the commencement of the 21st century, global media focused on the cocoa supply chain due to intensive family participation (i.e. 40-50 million people received incomes from cocoa cultivation) (Beg et al., 2017), product exportation of approximately $47 billion and cocoa industry's sustainability (International Cocoa Organization, 2018). Relative to this, some studies were carried out to identify unfair working conditions for cocoa workers largely in Africa. Also, a study conducted by the World Bank on the cocoa sector in Ghana revealed that several stages of the cocoa supply chain including production and market were at risk (World Bank, 2012). Based on these factors, the cocoa bean price fluctuation became a priority in the industry, along with the guarantee of a fair and decent livelihood for cocoa farmers in cocoa-growing communities. Côte d'Ivoire and Ghana being the most affected in 2019 introduced a "living income differential'' (LID) of USD 400 per ton of cocoa bean starting from 2020/2021 crop season, to increase the farmgate prices and better the living conditions of their smallholder farmers. The two countries fixed a minimum price of USD 2,600 per ton free-on-board (FOB) that chocolate companies must

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

pay if they want to access their cocoa beans. Funds generated from LID would be used to increase payments to farmers, with the goal being for them to get 70% of a USD 2,600 per ton FOB target price. Nevertheless, if market prices increase above USD 2,900 per ton FOB, proceeds raised from the LID would be deposited in a stabilization fund that would aim to ensure that the two governments pay farmers 70% of the USD 2,600 target price when market prices fell (Business & Human Rights Resource Centre, 2020).

For efficient management of the cocoa industry, the production of cocoa in Ghana has been zoned into seven cocoa-growing regions namely Ashanti, Volta, Eastern, Central, Brong Ahafo, Western south and Western north. Globally, Ghana continues to be the largest producer of premium quality cocoa beans probably because of the various institutions of COCOBOD working synergistically (Jonfia-Essien, et al., 2008).

Cocoa crop seasons

There are two (2) main seasons for cocoa cultivation in Ghana namely the 'main crop season' comprising 33 weeks (i.e., October-June) and 'light (minor) crop season' comprising 11 weeks (i.e., December-March). Cocoa beans are categorized by the QCC by size based on the number of beans weighing 100 g. Seven categories of cocoa beans such as super main (category H), main crop (category A), super light (category L), light (category F), small (category P), type 4 beans (category V) and remnants (category R) are declared for each crop harvesting season (Table 2.2). Usually, during the 'main crop season' the majority of the cocoa beans that are harvested fall within the super main (category H) and main crop (category A) categories. Conversely, during

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

the 'light crop season,' these two categories are produced in fewer quantities whereas the remaining five categories (smaller sized beans) are harvested in large quantities. The 'main crop season' cocoa beans are mostly exported to Europe and Asia, whilst the 'light crop season' beans are discounted to the local processing industries (though the quality of the beans is the same) and state-run cocoa processing company. Mostly, the 'light crop beans' are smaller in volume than the 'main crop beans.

Table 2.2 Categorization of Cocoa beans in Ghana

Categories	**Beans count per 100 g**
Super Main (Category H)	0 – 90
Main Crop (Category A)	91 – 100
Super Light (Category L)	101 – 110
Light (Category F)	111 – 120
Small (Category P)	121 – 130
Type 4 Beans (Category V)	131 – 150
Remnants (Category R)	151 – 180

(Source: Quality Control Company, COCOBOD, 2018/2019)

Postharvest operations for cocoa (Figure 2.3) begin after harvesting cocoa pod and comprise beans extraction through pod breaking, followed by fermentation, drying, cleaning and grading of beans. Processing of cocoa beans into various cocoa products begins with these on-farm and postharvest operations. These operations are very critical to the quality of the finished product as they initiate the development of chocolate flavour precursors and the brown colour of cocoa products (Afoakwa, 2010). These on-farm and postharvest operations are critically monitored by workers of COCOBOB before the cocoa beans are shipped to the international market.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

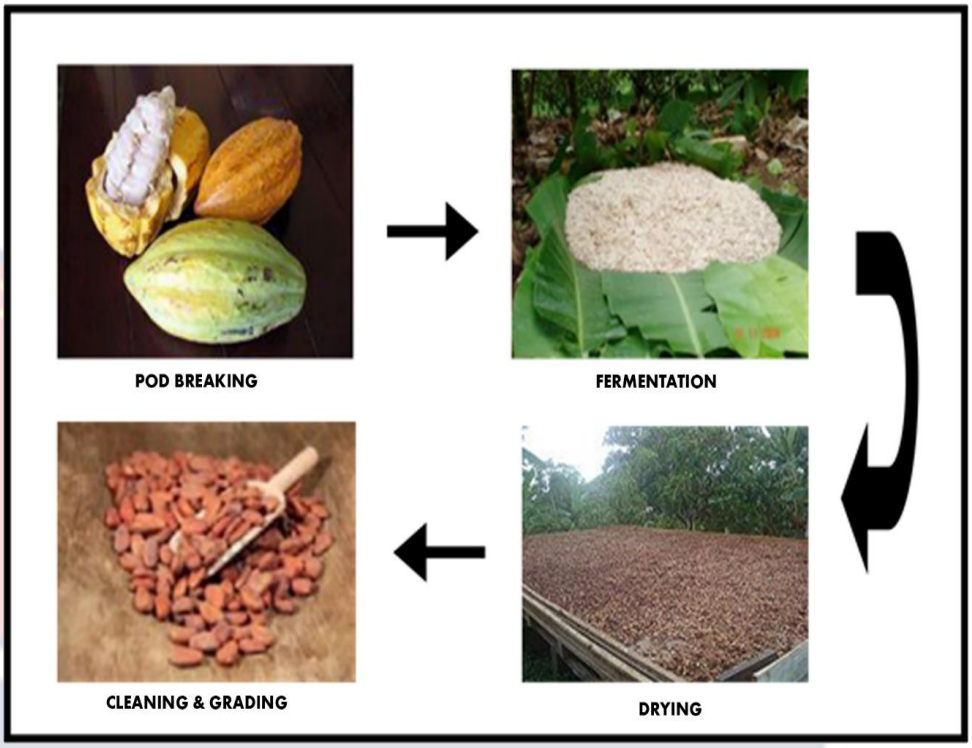


Figure 2.3 Postharvest operations for cocoa

Cocoa bean value addition and exports

Primarily, cocoa beans are processed into chocolate and cocoa derivatives such as cocoa butter, cocoa paste, cocoa liquor and cocoa powder. These products are hugely appreciated and consumed all over the world with North America and European Union showing a tremendous consumption growth rate. The demand for chocolate around the world increased by 2 - 5% per year and is presently near 3 million metric tons (Mota, El Makhloufi, & Scala, 2019). In 2012, a total export value of USD 8.4 billion was realized from 4.1 million metric tons of cocoa beans produced, representing a small fraction of the total value of the chocolate industry that was estimated at more than USD 83 billion (Markets & Markets, 2013). In 2017, the chocolate industry

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

consumed 43% of all cocoa and had a retail market value of USD 106.2 billion and is projected to grow by 2026 to USD 189.89 billion. However, Africa which produces more than 70% of the world's cocoa bean export the majority of its cocoa beans unprocessed. Cocoa-producing countries in Africa capture only 3% of the international chocolate industry. For instance, the world's leading producers such as Cote d'Ivoire locally process between 24-35%, Ghana 6-15%, Nigeria 6-14%, and Cameroon 10-27% of their cocoa beans, allowing them to capture an insignificant portion of the processed and finished cocoa products market (MarketWatch, 2019).

Health benefits of cocoa consumption

The consumption of cocoa bean and cocoa bean products has seen a tremendous increase over the past years as a result of their health benefits. The medicinal use of cocoa as primary therapy or as a means to convey other medicines had been traced from Aztec (Mexican) sources, and roughly 150 uses of cocoa medicine had been documented (Dillinger et al., 2000). Chocolate use was to stimulate the healthy function of the digestive systems and spleen. In the 17^{th}-18^{th} century, chocolate was frequently prescribed or mixed into medications for all kinds of sicknesses such as coughs, colds, to promote fertility, reinforce mental performance, stimulate the healthy function of the digestive systems, and as an anti-depressant (Di Castelnuovo et al., 2012).

The consumption of flavanol-rich cocoa improves blood flow and this could help to attain health benefits in hearts and other organs. Various studies have reported that prolonged intake of raw cocoa reduced the risk of cardiovascular ailments, and foods rich in cocoa reduced blood pressure.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Theobromine, the key methylxanthines in chocolate is a coronary dilator, diuretic, smooth muscle relaxant and myocardial stimulant. Furthermore, cocoa is rich in micronutrients such as copper which contributes significantly to human dietary intake (Di Castelnuovo et al., 2012).

Cocoa bean quality and quantity control

Cocoa bean quality and quantity control is a very important aspect in the cocoa trade and it is defined in the international cocoa market in four ways namely physical quality, biochemical quality, process quality, and origin quality. Physical quality focuses on bean size uniformity, moisture content, bean defectiveness, mouldiness, disease infestation and presence of foreign material (Jonfia-Essien et al., 2008). The biochemical quality relates to favour chemicals, fat content, toxic compounds, heavy metals and level of chemical residues left in the cocoa beans (Ramli et al., 2001). Process quality refers to the production procedure of cocoa including whether organic or conventional methods are used, whether the production procedure and subsequent rewards benefit the producing farmer and community and whether child labour is employed (Quarmine et al., 2012). Based on the quality reputation of the cocoa-producing country, cocoa bean is most often differentiated by the country of origin.

The raw unprocessed cocoa bean has a bitter, unpleasant taste, aroma, and flavour, and the basic postharvest operations that are carried out to improve the cocoa bean to the acceptable quality standard are fermentation and drying. Synergistically, fermentation and drying of cocoa beans promote cocoa flavour and aroma in the final cocoa bean products. For instance, poor cocoa bean fermentation leads to low chocolate flavour, bitterness and astringency of cocoa

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

bean whereas poor drying leads to the development of mould which eventually imparts unpleasant aroma and flavour on the final product (Minifie, 2012). These important basic operations provide decent cocoa bean quality if they are properly done in the country of origin. After harvesting the pod from the cocoa tree, the cocoa beans numbering 25-40 which are embedded in a mucilaginous pulp are scooped from the pod and fermented for about 4-7 days. The fermented beans are afterward dried to a safe moisture content from about 60% to not more than 8.5% (Abdullahi et al., 2018).

Well fermented and dried cocoa bean has a fermentation index above one and with optimum cocoa flavours and reduced acidity level (Whitefield, 2008). The quality and quantity of polyphenolic compounds content of cocoa beans are directly related to fermentation and drying operations. During the drying process, major phenolic oxidizing reactions are catalyzed by phenolic oxidases giving rise to new flavour components and membrane integrity loss, inducing the formation of browning colour. Development of chocolate flavour from cocoa bean precursors continues during drying with the formation of characteristic brown colour (Afoakwa, 2010). Fermentation and drying synergistically influence cocoa beans' quality parameters such as moisture content, pH, fermentation index, acidity, fat or butter content, total flavonoids content, polyphenols content and proximate compositions (Saltini et al., 2013). Therefore, measurements of these quality parameters have been used for the assessment of cocoa beans. For example, cocoa beans from far Eastern countries and Brazil were reported to be highly acidic, whereas those from

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

America showed low acidity and cocoa beans from Africa were observed to have intermediate acidity (Jinap & Dimick, 1990).

Practically, fermentation and drying differ from one country to the other and the variations may probably be due to climatic conditions, postharvest operations, variety grown, pod storage, pod breaking, batch size, and batch turning. It is also reported that different fermentation methods are employed contingent on the producing farmers, countries, location and these fermentation methods such as heaps, baskets, platforms and boxes differ from region to region. These aforementioned factors usually lead to the variation of the quality and quantity of cocoa beans in the global market. Cocoa beans from Ghana are very good examples of well fermented and dried, flavorful cocoa beans compared to other cocoa-producing countries, which are observed to be poorly or inadequately fermented and rapidly dried, and are of low chocolate flavour, high astringency, and bitterness (Miller et al., 2006).

Physical Quality Attributes

The physical attributes of agricultural materials such as cocoa beans are often needed for designing and developing postharvest systems and equipment for drying, cleaning, grading, transporting, processing, and storage (Bart-Plange & Baryeh, 2003).

Geometric dimensions

The geometric dimensions such as length, width and thickness of cocoa bean, like those of other agricultural grain determine the size and how much space it can occupy. Data on these geometric dimensions are required for the design of sorting sieves and conveying systems. Geometric dimensions of

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

agricultural materials like cocoa beans vary with location, variety, tree age, pollination type, growing season and moisture content.

Gravimetric attributes

Some gravimetric attributes of cocoa beans are bean mass, 1000-bean mass, bean density, bulk density and porosity. Bean mass is defined as the quantity of matter that an individual cocoa bean contains. It comprises the moisture of the bean and the dry matter of the bean. The moisture content is reliant on the void spaces and water holding capacity of the bean. The 1000-bean mass of the cocoa beans is an essential index to 'milling outturn' in determining the relative quantity of foreign material in a given lot of bean and the number of objectionable beans. Bean density is defined as the ratio of the mass of the cocoa bean to the volume occupied by the bean without void spaces. It is most often referred to as particle density or specific density or true density. The bulk density of cocoa beans is defined as the ratio of the mass of a group of individual cocoa beans to the volume of space occupied by the whole mass of the beans, together with the air space. Porosity is the proportion of total container volume being occupied by air spaces between the bean particles expressed normally in percentage. It relies on bean density and bulk density, hence the magnitude of variation of porosity relies on these two factors.

Repose angle

Repose angle is a mechanical attribute of the cocoa bean which defines the highest angle of a stable slope determined by cohesion, friction and shapes of the cocoa beans. The higher the repose angle, the higher the cohesion. Repose angle is influenced by the surface features, shape and moisture content of the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

grains. Repose angle is very essential for the computation of the number of cocoa beans that can be stored.

Moisture content

The entire constituent of cocoa bean is made up of moisture content and dry matter content, and the two summing up to 100% or unity. Cocoa bean storability is principally and directly connected to its moisture content. Moisture content examination is one of the physical quality assessments that is usually enforced both at the domestic and the international markets before the exportation of cocoa beans. Properly dried cocoa beans will usually have an optimum moisture content of less than 8.5% to avoid mould growth, development of microorganisms, astringency, bitterness, promote chocolate brown colour and for long shelf life. Among other Ghanaian standards, cocoa bean is classified as Grade I cocoa if its moisture is not higher than 7.5% (Quarmine et al., 2012).

pH

pH is the measure of negative logarithm of hydrogen ion concentration which usually spans between 1 and 10^{-14} gram-equivalent per litre into numbers between 0 and 14. An aqueous solution at 25°C with a pH less than 7 is considered acidic, while those with a pH higher than 7 are alkaline whereas a 7 is defined as "neutral". The pH is normally measured with a pH meter. The acidity of cocoa beans differs from country to country (Sadler & Murphy, 2010). Research works reported that chocolate samples obtained from cocoa beans of low pH (i.e. 4.8-5.2) and high pH (i.e. 5.5-5.8) had a low response in strong chocolate flavour, whilst chocolate samples obtained from Ghanaian and

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Nigerian cocoa beans that have medium pH values (i.e. 5.2-5.5) received a high response in strong chocolate flavour (Jinap, Dimick, & Hollender, 1995). Cocoa beans with low pH were reported to show more off flavour descriptors. Comparatively, the Ghanaian cocoa beans had a pH of 5.42 and flavour score obtained from these cocoa beans was set as a standard to measure flavour from other cocoa beans. The pH was used to group cocoa beans from different countries and highly acidic cocoa beans were characterised by high concentrations of lactic and acetic acids. The pH < 4.5 will lead to flavour precursors reduction and over-acidic product (Saltini et al., 2013).

Fat content

The fat content is a key quality parameter that chocolate producers take into consideration when selecting cocoa beans. In the chocolate formulations, it is the major constituent that constitutes one-third of the chocolate content. The cocoa bean has a high-fat content of about 45-57% of the dry weight of cocoa beans and it is accountable for the chocolate melting properties. Traditionally, fat content determination according to AOAC official methods involves the use of the soxhlet extraction apparatus (SOEP) (AOAC, 2005).

Fermentation index (FI)

FI is the measure of the degree of fermentation of cocoa beans. The method for the determination of FI was established by Gourieva and Tzerevitinov and it has been effectively used to measure the fermentation of cocoa beans ever since its invention (Gourieva & Tserevitinov, 1979). Studies observed that the colour of fermented cocoa bean correlates with total polyphenols and these polyphenols impart a purple to red colour with a

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

maximum absorbance at 500-550 nm, whereas absorbance values below 500 nm increase through fermentation. FI is assessed by the ratio of absorbance at 460 nm and 530 nm.

Antioxidant properties

Antioxidant properties of cocoa beans have seeming potential health effects on the human body: such as anti-depressant, anti-inflammatory, anti-carcinogenic, anti-radical properties, cerebral blood flow, insulin sensitivity and lipoprotein metabolism (Di Castelnuovo et al., 2012). The bioactive compounds found in cocoa beans that exhibit antioxidant properties are mainly polyphenols and flavonoids. The antioxidant activity of polyphenolic compounds found in cocoa beans depends on cultivated cocoa variety, soil, and postharvest handling such as fermentation and drying. Fermentation and drying of cocoa beans tend to decrease polyphenol content, flavonoid content and antioxidant capacity (Delgado-Ospina et al., 2020).

Many studies reported that flavonoids are a group of polyphenolic compounds that are found in vegetables, fruits, red wine, tea, cocoa beans and cocoa-based products. However, cocoa beans and cocoa-based products hold the highest quantity of flavonoids concentration among frequently consumed foods (Lee et al., 2003). Flavonoids in cocoa beans are capable of reducing blood clot and stroke, scavenging free radicals and lowering cardiovascular attacks. The cocoa bean has polyphenols content of about 12-18% of the dry weight of the whole cocoa bean and the total cocoa bean polyphenols content depends on cocoa variety, and postharvest activity such as fermentation, drying and roasting condition of cocoa bean (Hii et al., 2009). Practically, polyphenols

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

are astringent and bitter and the quantity of polyphenols in cocoa beans after bean fermentation and drying is a very important quality parameter.

The quantification of total polyphenols content, flavonoids content and antioxidant capacity of cocoa bean is done spectrophotometrically via the oxidation of phenolic compounds. Polyphenols content and flavonoids contents were determined using the Folin-Ciocalteu reagent. Antioxidant capacity was determined using 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical scavenging activity. These procedures involve the use of the colourimetric method with a standard acid equivalent (Gallic= GAE, Catechin= CAE, Ferrulic= FAE, Epicatechin= EAE, Ascorbic= AAE or vitamin C= VCAE) per gram. Among the various standard acid equivalents employed, the gallic acid equivalent is more preferred, because it is steady, pharmacologically active, and it is quantitatively equivalent to most polyphenolic compounds as well as gives consistent results (Abbe & Amin, 2008).

Ramli et al. (2001) reported that the total polyphenols content span between 34-60 mgGAE/g in cocoa beans, 20-62 mgGAE/g in cocoa powder and 45-52 mgGAE/g in cocoa liquor. Also, Hii et al. (2009) also detected 40-84 mgGAE/g of polyphenols among different cocoa varieties and countries of origin. Oliveira et al. (2011) also reported 233.3 mgGAE/g and 221.82 mgGAE/g in organically cultivated and conventionally cultivated cocoa beans, respectively. Jonfia-Essien et al. (2008) observed that the total antioxidant capacity of different cocoa varieties from Ghana ranged from 12.4-45.5 µmol TE/g, while Gu et al. (2006) reported values of 826 ± 103 µmol TE/g for natural cocoa powders. It was also reported by Gu et al. (2006) that the flavonoids

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

content (procyanidin) in natural cocoa powder and chocolate ranged from 45-517 mg/g. Furthermore, Lee et al. (2003) recorded total polyphenol content of 611 mg of gallic acid equivalents (GAE), total flavonoid content of 564 mg of epicatechin acid equivalents (EAE) and total antioxidant capacity of 836 mg of vitamin C equivalents (VCE) for cocoa.

Proximate composition

The dietary quality parameters of cocoa are largely determined by the biochemical composition of the cocoa and cocoa powder. The contribution of energy to daily nourishments is contingent on the quantum of proximate compositions viz. fats, proteins, carbohydrates, ash and fibre in the cocoa bean as well as cocoa-based product (Minifie, 2012). Considering the health effects; reduction of chronic bowel disorders, diabetes and obesity, associated with the consumption of proximate compositions (i.e. dietary fibre, protein, fats, carbohydrates) in cocoa, there has been a renewed interest in the determination of proximate compositions for the functional food industry. The protein content of cocoa helps to improve the functioning of the kidney and liver, aspartate aminotransferase, alanine aminotransferase of cocoa helps to monitor and detect cardiac ailments. The proximate compositions found in cocoa bean depends on the following factors: climatic condition, soil chemical composition, production process (i.e. organic or conventional), tree age, geographical origin, fermentation and drying (Afoakwa et al., 2013b). For instance, it was reported that cocoa bean protein constitutes 11–13% and may differ from one geographical origin to another between 11.8% and 15.7% (Kongor et al., 2016).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

The determination of proximate compositions (proteins, carbohydrates, fibre) involves the use of procedures of the Association of Official Analytical Chemists (AOAC, 2005). Total Carbohydrates were assessed by difference. Nitrogen was evaluated by the micro-Kjeldahl technique, and the percentage nitrogen was changed to crude protein by multiplying by 6.25. Adeyeye et al. (2016) reported the proximate compositions of natural cocoa powder, natural cocoa liquor and alkalized cocoa powder (g/100g) as protein, 10.9-24.3; total ash, 3.73-10.8; fibre, 1.03-17.3; and carbohydrate, 20.5-48.9. It was reported that protein content of cocoa beans varied from different geographical origins ranged from 11.7 ± 0.24- 13.35 ± 0.27 (Caprioli et al., 2016).

Reference methods

The applications of spectroscopic analytical techniques rely on reliable and trusted reference data or values (Chen, Zhao, & Lin, 2009). Thus, to build a quantitative regression model it is essential to have a reference analytical method to examine the attribute of analytes that will be employed in the computation. Such wet chemical analytical methods give what is mostly referred to as reference values of the analytes, and these values are used to calculate the model. For example, NIR spectroscopy reports data in terms of reference values. The reference data are employed for the development of a prediction model by the application of chemometric techniques.

Near-Infrared Spectroscopy

General overview

Near-infrared (NIR) is the first non-visible region of the electromagnetic spectrum. NIR spectroscopy is a type of vibrational spectroscopy that uses

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

photon energy in the energy range of 2.65 x 10^{-19} to 7.96 x 10^{-20} joules corresponding to the wavelengths in the range of 750 nm to 2,500 nm (wavenumber range: 13,300 cm^{-1} to 4000 cm^{-1}) (Pasquini, 2003). This wavelength range is positioned between the infrared and the visible regions (Figure 2.4).

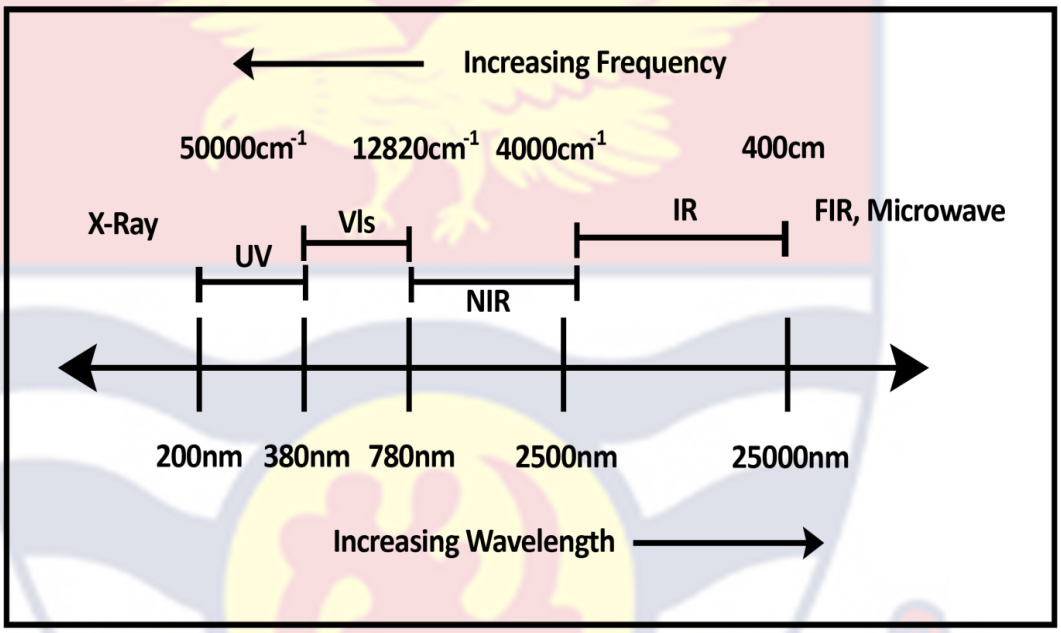


Figure 2.4 NIR in the Electromagnetic spectrum. Source: Pasquini (2003) with slight modification

The absorption of bio-molecules in this wavelength range originates from the overtones of C-H, S-H, O-H, and N-H. It involves the stretching-bending combination and stretching vibrations from involving bonds as shown in Figure 2.5. The NIR spectroscopy presents several important merits over the chemical methods (Veselá et al., 2007).

The main disadvantages are the wet chemistry dependence laborious and time-consuming calibration technique and the sophistication in the selection of data treatment by chemometric techniques.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

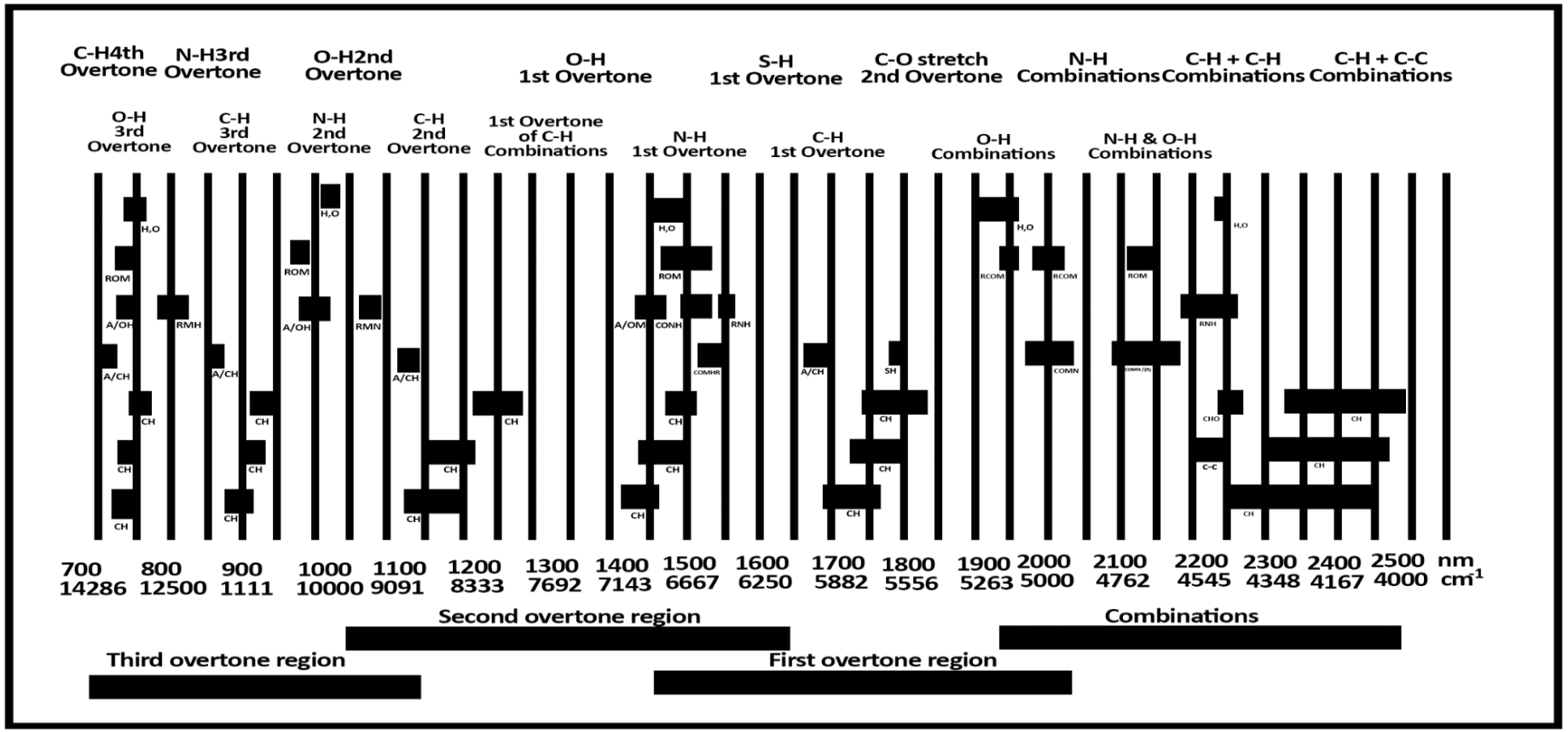


Figure 2.5 NIR spectra profiles with their major chemical functional groups. Source: Sandorfy et al. (2007) with slight modification.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Principles of near-infrared (NIR) Spectroscopy

Theoretical information of NIR spectroscopy is very significant for the apprehension of the NIR spectral information. The NIR spectra comprise a great deal of chemical and physical characteristics emanating from the reaction of electromagnetic radiation with the specimen or the biological material. The NIR radiation acts like a wave with simple harmonic properties. The properties was be defined by Sandorfy et al. (2007) as:

$$\text{Frequency of vibration} = \frac{angular\ velocity}{2} \tag{2.1}$$

$$\text{Wavelength} = \frac{light\ velocity}{frequency} \tag{2.2}$$

The NIR spectrum emerges from radiation energy which is transferred to mechanical energy that is related to the motion of atoms held together by chemical bonds in the molecule (Sandorfy et al., 2007). Spectroscopy is a word acquired from the Greek root skopia (i.e. to view) and the Latin root spectrum (i.e. image, appearance), and so by definition; spectroscopic measurement means to view a light appearance or image arising from a specimen.

Therefore, this phenomenon can be best explained by the law of conservation of energy, because it is explicit that the specimen does not produce light itself. Vibration spectroscopy has to do with energy transfer between matter and light energy. When light energy from the source is transferred to a molecule the fundamental vibration energy rises at a given wavelength (λ), for which the energy (E_p) can be expressed:

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

$$E_p = \mathrm{hv} = \frac{hc}{\lambda} \quad (2.3)$$

Where: h = Planck constant, v = frequency of light wave and c = velocity of light (2.998×10^8 m/sec in a vacuum).

The NIR absorption bands can be detected as a response to molecular vibrations of chemical bonds O-H, C-H, N-H, S-H and C-H. This vibration frequency of light is the function of the masses of two spherical atoms (m_1 and m_2) and the bond strength, k. The parabolic relationship that exists between interatomic distance and potential energy and Hook's law defines this energy (E) mathematically as:

$$\mathrm{E} = \frac{h}{2\pi}\sqrt{\frac{k}{u}} \quad (2.4)$$

Where u = the reduced mass, defined as

$$\mathrm{u} = \frac{m_1 m_2}{m_1 + m_2} \quad (2.5)$$

Instrumentation

NIR spectroscopic instrumentation has dramatically evolved in response to the need for pliability in adjusting to different states of specimens and speed in analyses. Spectrophotometers that are usually employed for the recording of NIR spectra are fundamentally equal to those used in other regions of the electromagnetic spectrum. Generally, NIR spectroscopic instrumentational configuration is composed of a source of radiation, grating monochromator, specimen presentation interface or specimen holder, wavelength selection (by the spectral encoder), and detector, allowing for reflectance or transmittance measurements (Figure 2.6).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

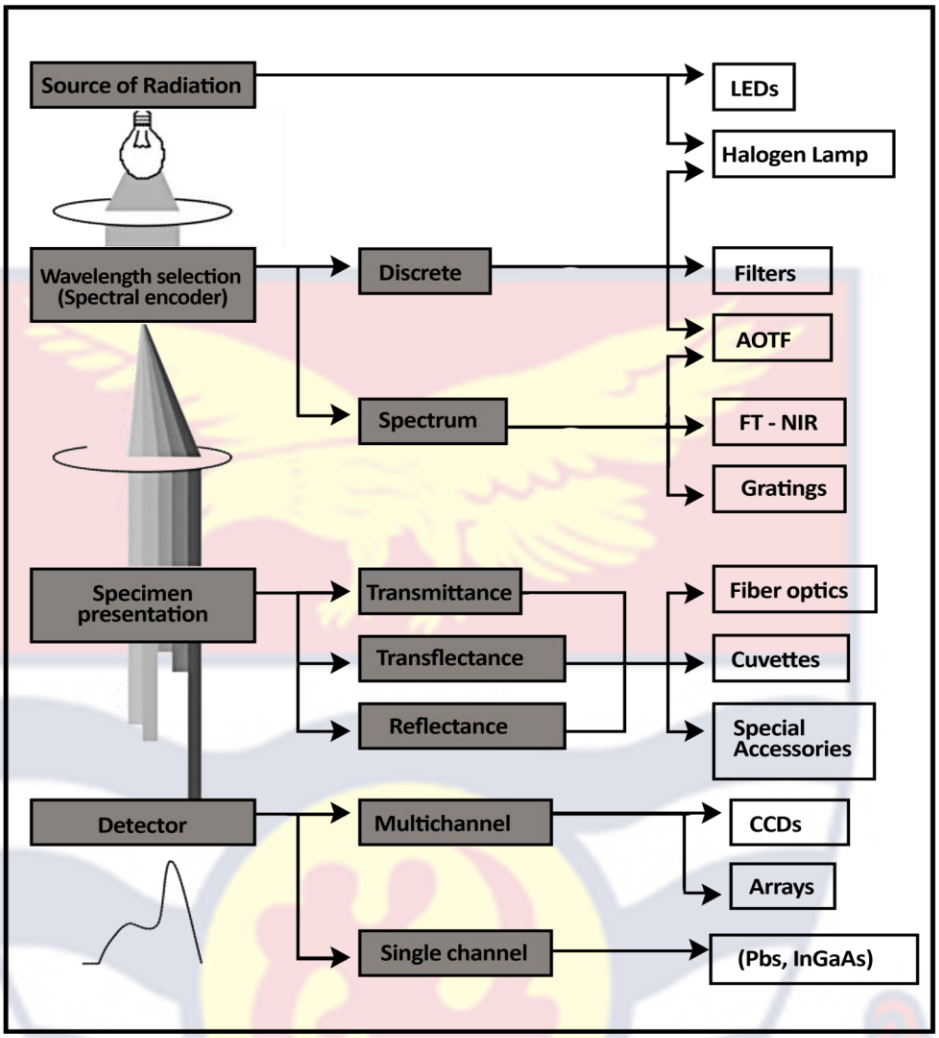


Figure 2.6 Basic and essential features of NIR spectroscopy kit. Source: Ozaki et al. (2006) with slight modification.

The specimen is illuminated with a light or radiation source in the NIR range, and after the interaction of electromagnetic radiation with the specimen, the resultant absorbed radiation or resultant reflected radiation is directed towards a sensitive detector (Ozaki et al., 2006). The light source which is normally small and rugged is a tungsten halogen lamp. The types of detectors include indium gallium arsenide (InGaAs), lead sulfide (PbS) and silicon. Silicon detectors are small, fast, of low noise, and highly sensitive from the visible region to 1100 nm, whereas PbS detectors are steady and slow, but very popular

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

and are sensitive from 1100 - 2500 nm and provide good signal-to-noise features. The InGaAs detector is the most expensive which combines the size and speed characteristics of the PbS detector with the wavelength range of the silicon detector (Kawata, 2009).

Source of radiation

The two sources of radiation in the NIR spectrometer are thermal and non-thermal sources. The thermal source includes a tungsten halogen lamp, quartz halogen lamp that offers high energy from the visible region to 2,526 nm. These thermal lamps are more stable due to the cleaning action of the halogen and they have a longer life span. The non-thermal sources comprise light-emitting diodes, discharge lamps and laser diodes. Relatively, the non-thermal sources are more effective and discharge much narrower bands of radiation than the thermal sources because most of the energy consumed appears as emitted radiation over a narrow range of wavelengths and they can be adjusted electronically which leads to efficient power consumption (Osborne, 2006).

Wavelength selection

Wavelength selection by a spectral encoder in NIR spectrophotometers is made up of two types viz. discrete wavelength and whole spectrum. In discrete wavelength spectrophotometers, samples are irradiated with only a handful of wavelengths. These few wavelengths can be chosen by using light-emitting diodes (LEDs) which radiate narrow bands or light source filters which permit the passing of changeably broad wavelength bands (Figure 2.6). The whole spectrum wavelength selection normally comprises diffraction grating. Recently, other wavelength selection instruments incorporated NIR

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

spectrophotometers including Acousto-Optic Tunable Filters, (AOTFs) (Blanco & Villarroya, 2002). These select wavelengths by using radio frequency signals to modify the refractive index of anisotropic crystal so that it completes a wavelength scan more quickly than with the aforementioned designs.

NIR spectroscopy employs InGaAs, PbS and silicon devices for the detection of the interaction of electromagnetic radiation with the specimen and these elements are arranged in focal planes array; charged coupled devices (CCDs) or rows array; diode arrays to measure several wavelengths at once and increase the pace at which spectral data can be collected.

Specimen presentation modes

The most significant recommendation for obtaining successful results in the development of calibrations for quality factors is to be sure that the NIR irradiating area and the sampling area match. This requirement is connected to the mode which largely determines the depth of penetration of the NIR energy into the sample. The different types of specimens used for food analysis namely solids, powder, liquids, and slurries need different modes to make the application of NIR spectroscopy possible for a broad range of materials. When the radiation of light travels through a substance the light is reflected or absorbed, transmitted and the total amount of the light interacting with the substance, according to the law of conservation of energy, must be equivalent to the total radiant energy illuminated on the specimen.

Depending on the sample type and the chemical being analysed, the most common measurement modes used in NIR spectroscopy are transflectance, transmittance, interactance, transmittance through a scattering

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

medium and diffuse reflectance (Pasquini, 2003). Figure 2.7 shows the different types of measurement modes used in NIR spectroscopy as described by Pasquini (Pasquini, 2003) with a few alterations. Each of these measurement techniques provides different advantages and weaknesses. Nevertheless, in contemporary NIR analysis, the two widespread techniques mostly employed are built on reflectance mode and transmittance technique (NIR transmittance). Interaction of radiation with a sample may result in absorption, reflection, or transmittance.

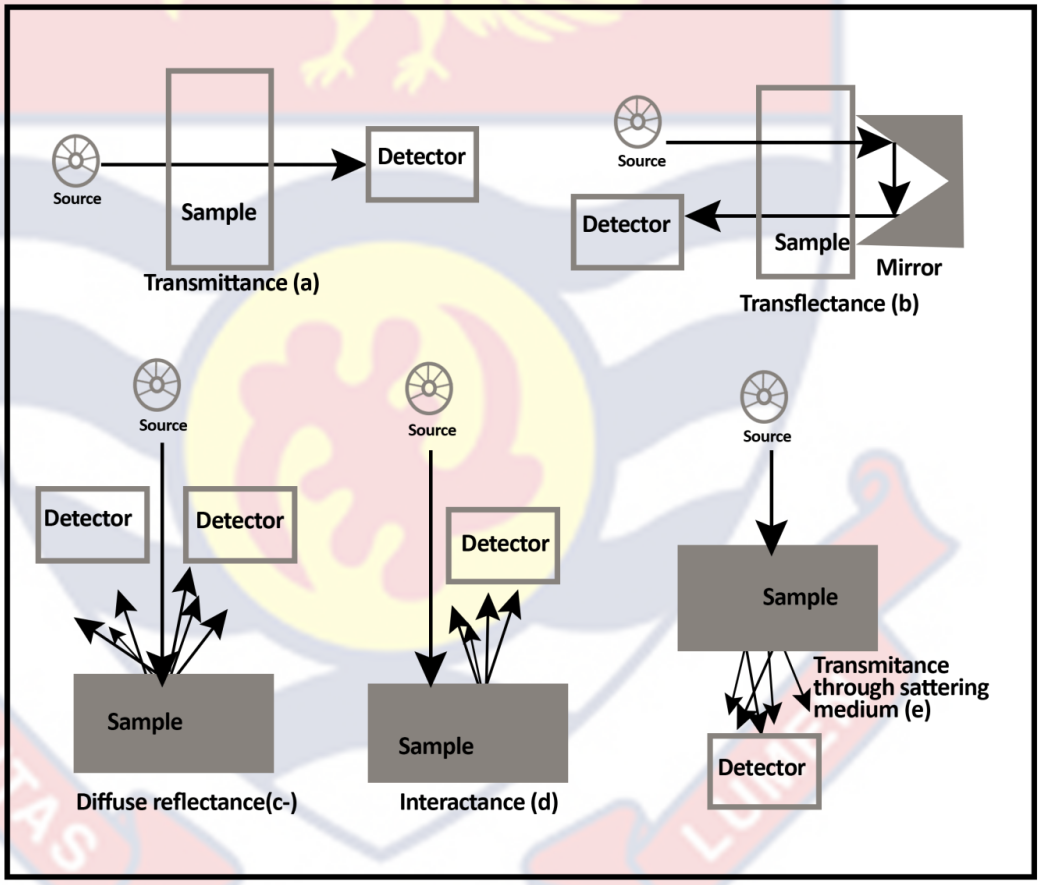


Figure 2.7 Measurement modes employed in NIR spectroscopy (Pasquini, 2003).

In a classical spectroscopic experiment, reflection is eradicated so that the proportion of radiation attenuated by the specimen may be assessed as transmittance. Measurements of diffuse transmittance are performed in the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

range of 800-1100 nm. The initial research in the NIR spectroscopy region was performed in solution by using transmission spectroscopy and the relationship between absorbance and concentration according to the law of Lambert-Beer was used to quantify the results. Lambert-Beer's law could just be deployed for explicit transparent liquid samples once there is no light scattering. This is because scattering alters the route length via which the radiation travels and the amount of scattering varies from specimen to specimen. For instance, if milk which is a fluid specimen is employed, the drops of fat in the milk would disperse the light and this invalidates the law of Lambert-Beer. Also, diffuse transmittance measurement mode has been normally employed to liquid specimens although some specimens such as meat grain and cheese have been applied (Davies & Grant, 1987). Raw diffuse transmittance is transformed to absorbance by using the mathematical expression:

$$A=\log\left(\frac{1}{T}\right) \tag{2.6}$$

Where: A= absorbance, T=transmittance

Reflectance spectroscopy is a type of measurement technique applied in NIR spectroscopy where radiation energy is reflected from the surface as diffuse (called regular) or specular and no absorption occurs. In a wavelength region of 1100-2500 nm.

This situation is termed diffuse reflectance and this is because almost all the incident radiation is reflected. Diffuse reflectance measurement mode is usually applied for granular or solid samples. The broadly accepted model for the treatment of the signal acquired in diffuse reflectance mode was formulated

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

by Kubelka and Munk (1931) and later on modified by Kortüm (2012) using the expression.

$$\frac{K}{S}=\left(\frac{1-R^2}{2R}\right) \tag{2.7}$$

Where: K= molar absorption coefficient, S= the coefficient of scattering, R= diffuse reflectance of an infinitely thick sample.

For this equation, the coefficient of scattering is correlated and affected by the size of the particle of the samples, thus it is important to take this into account in all applications. However, equation 2.7 is hardly used and it has been replaced by a more pragmatic equation (2.8), which has also been employed in this study.

$$\mathrm{A}=\log\left(\frac{1}{R}\right) \tag{2.8}$$

Where: A= absorbance, R= Reflectance

Detectors

There are two groups of detectors namely single-channel detectors and multichannel detectors (Blanco & Villarroya, 2002). The silicon (Si) detectors, PbS detectors, and InGaAs detectors are the single-channel detectors usually used in the NIR spectroscopy analysis. The Si detectors are of low noise fast, and extremely sensitive from the visible (400 to 1100 nm); the PbS detectors are comparatively slow, but sensitive and have an excessive signal to noise attributes (1100-2500 nm) and indium gallium arsenide (InGaAs) detectors, are both quick and extremely sensitive (800-1700 nm), but a lot more costly than PbS and Si detectors. Diode array NIR spectrometers use an array of infrared emitting diodes that operate as both the source of light and wavelength selection

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

technique. The diode array equipment typically covers 400 nm to 1700 nm and they are non-invasive and extremely fast (i.e., one spectrum per second).

Application of NIR Spectroscopy

The application of NIR spectroscopic technology is multi-purpose in analysis. NIR spectroscopic techniques have been deployed to predict several food properties. The application of NIR spectroscopy in the analysis of food can be grouped into three major headings viz. quantitative, qualitative, and simultaneous measurements. Tables 2.3, 2.4, and 2.5 show the summary of studies conducted by others on the application of the NIR spectroscopic technique in food analysis. Specifically, Table 2.6 shows a more comprehensive summary of the recent applications of NIR spectroscopic technology in the cocoa industry.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 2.3 Application of NIR spectroscopy for quantitative examination in food

Food	Type of study	Spectral mode	Reference
Apple	Prediction of total polyphenol and sugar content	Reflectance	(Pissard et al., 2013)
Bayberry	Prediction of titratable acidity, malic and citric acid	Transmission	(Xie et al., 2011)
Cotton seed	Estimation of oil content and fatty acid composition	Reflectance	(Quampah et al., 2012)
Grape	Screening of extractable polyphenols	Reflectance	(Nogales-Bueno et al., 2015)
Milk powder	Detection of melamine in milk	Reflectance	(Lu et al., 2009)
Orange	Vitamin C Content measurement	Reflectance	(Junfang et al., 2007)
Rice	Detection of grain moisture	Reflectance	(Lin et al., 2019)
Sapodilla fruit	Water content determination	Reflectance	(Hadiwijaya & Putri, 2018)
Soybean	Determination moisture, fat and protein content	Reflectance	(Zhu et al., 2018)
Tea	Determination of total polyphenols content	Reflectance	(Chen et al., 2008)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 2.4 Application of NIR spectroscopy for qualitative examination in food

Food	Type of study	Spectral mode	Reference
Bayberry	Varietal discrimination	Reflectance	(Li, He, & Fang, 2007)
Black beans	Discrimination of varieties	Reflectance	(Sun et al., 2016)
Coffee	Variety identification	Transmittance	(Zhang et al., 2018)
Green coffee	Geographical origin determination	Reflectance	(Giraudo et al., 2019)
Green tea	Discrimination by geographical origin	Reflectance	(Chen, Zhao, & Lin, 2009)
Melon	Distinguishing genotypes	Reflectance	(Seregely, Deak, & Bisztray, 2004)
Paddy seeds	Discrimination of storage age	Reflectance	(Li, He, & Wu, 2008)
Rice flour	Genotype identification	Reflectance	(Sampaio et al., 2020)
Rice	Authenticity of quality grades and country of origin	Reflectance	(Teye et al., 2019a)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 2.5 Application of NIR spectroscopy for simultaneous examination in food

Food	Type of study	Spectral mode	Reference
Bakery products	Simultaneous analysis of Xanthines and polyphenols as bitter taste markers	Reflectance	(Bedini et al., 2013)
Grapes berries	Simultaneous evaluation of qualitative and quantitative of various development stages	Reflectance	(Musingarabwi et al., 2016)
Milk	Simultaneous quantification of adulterants	Reflectance	(Borin et al., 2006)
Orange	Classification and sugar content prediction	Reflectance	(Shao et al., 2009)
Pineapple	Quality parameters authentication & quantification	reflectance	(Amuah et al., 2019)
Green tea	Determination of total polyphenols content	Reflectance	(Chen et al., 2008)
Vinegar	Simultaneous measurement of total acid and soluble salt-free solid content	Transmittance	(Chen et al., 2012a)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 2.6 Application of NIR spectroscopy technology in cocoa bean and cocoa bean products

Application	Task	Wavelength range	Chemometrics analysis		Performance/Rate %		Reference
			Preprocessing	algorithms	R cal	R pre	
Categorization of cocoa beans	Discriminating of geographical location of cocoa beans	400 - 1000 cm^{-1}	MC, MSC, Detrend, 2-derivative	KNN, LDA, BPANN, SVM	100	100	(Teye et al., 2013)
	Clustering fermentation degree of cocoa beans	800 - 2498 nm	SNV-MC	PCA-PLS	-	-	(Aculey et al., 2010)
	Identifying cocoa bean varieties	400 - 1000 cm^{-1}	SNV	LDA, SVM	100%	100%	(Teye, 2016)
	Classifying cocoa beans quality grades	900 - 2500 nm	PCA-SNV	LDA, KNN, SVM, ELM	94%	94%	(Kutsanedzie et al., 2017)
Chemical compositions of cocoa beans	Ammonia nitrogen content	400 - 2500 nm	PCA	m-PLS	0.975	0.938	(Hue et al., 2014a)
	Characterization of changes in flavan-3-ol derivatives (epicatechin)	400-2500 nm	SNV, detrend, SG, 2^{nd} derivatives, PCA	PLS	0.95-0.98	0.93-0.96	(Hue et al., 2014b)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Procyanidins content; monomers, epicatechin, catechin and oligomers of flavan-3-ol monomers	400-2500 nm	SNV, detrending, first derivative	m-PLS	0.983	0.98	(Whitacre et al., 2003)
Biochemical quality parameters; phenol, organic acid, epicatechin, lactic acid, theobromine	3600-12500 cm^{-1}	MSc, SNV, 1st der, min & max-normal	PCA-PLS	0.88-0.94	0.87-0.98	(Krähmer et al., 2015)
Fermentation index,pH, total polyphenols	3600-12500 cm^{-1}	Vector normal, MSC, 1st der, straight-line subtraction, min & max-normal	PLS	0.76-0.88	0.76-0.88	(Sunoj, Igathinathane, & Visvanathan, 2016)
Measuring pH, fermentation index and fermentation groups	4000-10000 cm^{-1}	SG, MSC, MC, SNV, 1st derivative, PCA	PLSDA, BPANN, & PLS, iPLS, SiPLS, BPANNR, SiBPANNR	0.98-1	0.98-0.997	(Teye et al., 2015)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

	Predicting total fat content	4000-10000 cm^{-1}	MC, MSC, 1st der, SG	Si-PLS, SVMR	0.984	0.971	(Teye & Huang, 2015)
	To determine fat, caffeine, theobromine and epicatechin in unfermented criollo	780-2500 nm	SNV, 2nd derivative, Savitzky-Golay	m-PLS	0.88-96	0.88-0.98	(Álvarez et al., 2012)
	Protein, moisture, fat, ash, carbohydrates, and colour (L*, a*, b*) of whole cocoa beans and ground cocoa beans	400-2498 nm	MSC, SNV, 1st & 2nd derivatives	PLS	0.9-0.98	0.80-0.99	(Barbin et al., 2018)
	Measuring protein, pH, acidity, fat, shell content, moisture, total phenolics, caffeine and theobromine	4020-10000 cm^{-1}	SNV, MSC, SG, 1st Dv	PLS	0.819-0.943	0.824-0.943	(Hashimoto et al., 2018)
Chemical composition of	Sucrose, lactose, fat, moisture, viscosity and yield of chocolate	910-2500 nm/ 11,000-4000 cm^{-1}	First and second derivatives	PLSR	0.98-0.89	0.90-0.57	(Tarkošová & Čopíková, 2000)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

cocoa bean products	Quantification of fat, nitrogen and moisture content of cocoa powder	1100-2500 nm	PC transform	PLSR	0.94-0.96	0.94-0.98	(Veselá et al., 2007)
	Predicting nutritional parameters; carbohydrates, fat, protein, energy value and cocoa content in chocolate samples	714-2857 nm (14000-3500 cm^{-1})	PCA, 1st & 2nd der.	ANNS, PLSR	-	-	(Moros et al., 2007)
	Analysis of fat, protein, sugar and water content in chocolate base	400-12500 cm^{-1}	EMSC, PCA	PLSR	-	0.989-0.998	(Stohner et al., 2012)
	Sucrose content in chocolate mass	1100-2500 nm	PCA, MC, Savitzky & Golay, SNV, MSC	PLSR, MLR, GA-MLR	0.997	0.998	(da Costa Filho, 2009)
Fraud and safety	Identifying and quantifying adulteration of fermented cocoa beans	400-1000 cm^{-1}	SNV, DOSC	SVM, PLS, Si-PLS	100%, & 0.99	100% & 0.98	(Teye et al., 2014a)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

	Fraud detection (qualitative & quantitative) of carob flour in cocoa powder	1100-2500 nm	2nd derivative, SG, OSC, PCA	PLS-DA, & PLSR	100%, & 0.980	100%, & 0.974	(Quelal-Vásconez et al., 2018)
	Detection of cocoa shell in cocoa powders	1100-2500 nm	EMSC, SNV, 2nd Dv, SG, OSC, PCA	PLS-DA, PLSR	100%, & 0.99	100%, & 0.96	(Quelal-Vásconez et al., 2019)
	Measurement of total fungal count in cocoa beans	1000-2500 nm	SNV, MSC	PLS, Si-PLS, Si-GAPLS, CARS-PLS	0.95	0.97	(Kutsanedzie et al., 2018)
Sensory quality	Sensory analysis of chocolate quality	1100-2500 nm	PCA, MSC	CVA	-	0.86	(Davies et al., 1991)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Chemometrics

Chemometrics is described as the science of correlating measurements taken on a chemical process or system to the state of the process or system through the application of statistical and mathematical approaches (Cen & He, 2007). The emergence of chemometrics came from the recognition that traditional univariate statistics were no longer sufficient to model and describe chemical experiments. It was thus developed to overcome the weaknesses of univariate statistics by using experimental design, multivariate calibration obtained from statistics, mathematics, and computer science (Geladi, 2003).

Chemometric techniques for the analysis of NIR spectroscopic data can be categorized into three main methods viz.:

1. preprocessing or pretreatment of spectroscopic data,
2. development or building of calibration models for qualitative and quantitative analysis,
3. model testing and transfer.

Preprocessing techniques

In NIR spectroscopy, preprocessing of spectral data is of utmost importance because the data obtained from NIR spectrometer oftentimes suffer from the challenge of unwanted spectral variations besides the characteristic information of the analyte. The sources of these unwanted spectral variations could be:

1. Interaction of molecules e.g., via intermolecular hydrogen bonds
2. Scattering of light from cloudy liquids and solid specimens

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

3. Poor reproducibility in the measurement process of NIR spectra, caused by variations of pathlength
4. Variations in the environment i.e., temperature, particle size, and density of the sample
5. Spectral distortions attributable to spectrometer hardware including:
    a. baseline drifts,
    b. wavelength drifts,
    c. effects from detector stray light,
    d. noises from the amplifier, detector, or analog-digital (AD) converter.

Such intrusions may contravene the presumptions upon which model equations are established (Siesler et al., 2008). For instance, the simple linear correlation between absorbance and concentration of the constituent as stated by Lambert-Beer's law (equation 2.8) does not hold entirely true, and the additive effects of individual spectral responses cannot be guaranteed.

$$A=\varepsilon lc \tag{2.9}$$

Substituting equation 2.6 into equation 2.9 will give rise equaion 2.10

$$A=-\log_{10}(T)=\varepsilon lc \tag{2.10}$$

Lambert-Beer's law (i.e., equation 2.10) is exclusively valid only for pure transmittance systems or processes with no scatter effect. For reflectance measurements, equation (2.10) must be revised in analogy to transmittance measurements as:

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

$$A=-\log_{10}(R)=\varepsilon lc \quad (2.11)$$

Where; A = absorbance molar absorptivity, ε = molar absorptivity, l = effective pathlength of light through the sample matrix, c = concentration of the constituent(s) of interest, T = light transmittance and R = reflectance detected.

The objective of preprocessing (before chemometric modelling) is to eliminate and transform physical phenomena in the spectral data to improve the subsequent multivariate classification, regression model or exploratory to adhere to Lambert-Beer's law (Cen & He, 2007). Thus, preprocessing methods are intended to compensate for these variations from linear relationships and hence boost the linear correlation between the analyte concentrations and spectral signals. Preprocessing methods that are most usually employed in NIR spectroscopy in both transmittance and reflectance study mode can be divided into two main categories viz. those that directly use obtainable reference values for the preprocessing process (reference-dependent preprocessing techniques), and those which do not (reference-independent preprocessing techniques) (Rinnan et al., 2009a). The reference-independent preprocessing techniques provide more general tools appropriate for operations such as exploratory analyses, for instance, in a situation where no reference value is available. The reference-independent preprocessing techniques can further be split into two subgroups namely scatter-correction methods and spectral derivatives. The scatter-corrective preprocessing methods include Standard Normal Variate (SNV), multiplicative signal correction (MSC), also known as Multiplicative Scatter Correction (MSC), inverse MSC, extended MSC, inverse extended

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

MSC, mean centering (MC), de-trending, normalisation and baseline correction whilst spectral derivatives or derivation methods include Savitzky–Golay, finite difference and Norris–Williams. The reference-dependent preprocessing techniques primarily comprise those methods that orthogonalize the spectral data concerning a reference of interest. Such techniques are not applicable as the response variables are actively used in chemometric modelling. The reference-dependent preprocessing techniques include orthogonal signal correction (OSC), and optimized scaling preprocessing. Most of these preprocessing techniques have been used in cocoa beans and cocoa beans products analysis and it was observed that the original spectra of cocoa beans and cocoa bean products changed when different preprocessing methods were applied (Teye & Huang, 2015). Different preprocessing methods displayed their advantage for different challenges that were investigated (Quelal-Vásconez et al., 2019). Thus various preprocessing methods could be tried for optimum results as cocoa beans are heterogeneous with distinctive differences. Preprocessing techniques are mostly applied on NIR spectra for baseline correction, resolution enhancement, normalisation and noise reduction.

Baseline correction techniques

The notable baseline correction techniques include MSC, SNV, OSC, normalisation, and derivative methods (i.e., first derivative- FD and second derivative- SD). These methods are designed to remove or decrease the physical variability between samples as a result of scattering, and baseline shifts between samples. The MSC is deployed to compensate for additivity (baseline shift) and

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

multiplicativity (tilt) effects in the spectral data, which are encouraged by physical effects such as particle size, refractive index, and non-uniform scattering on the radiation wavelength. The MSC eliminates nonlinearities, artifacts, or imperfections in the NIR spectra before modeling (Rinnan et al., 2009b). SNV is used for the elimination of slope variation and correction of scattering effect. The scatter correction concepts behind normalisation are the same as that of SNV. Derivative methods are used for the removal of additive and multiplicative effects and baseline variations in the NIR spectra. The FD technique which is calculated as the difference between two successive NIR spectral measurement points only eliminates the baseline. The SD is then computed by calculating the difference between two subsequent points of the first-order derivative NIR spectra to eliminate both linear trend and baseline correction (Rinnan et al., 2009b).

Noise reduction

The NIR spectra are characterised by various kinds of noise that arise from several interfering physical and chemical processes. The commonly used techniques for noise reduction are smoothing, wavelets, artificial neural networks (ANN), and eigenvector reconstruction. (Rinnan et al., 2009b). In the moving average technique, which is the simplest type of smoothing, the absorbance (A_i) at each variable i=1, 2 …k is substituted by a weighted average of itself and its nearest neighbours. From i - n to i + n:

$$A_i = \sum_{k=-n}^{n} w_k A_{i+k} \qquad (2.12)$$

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Where: A_i = Absorbance, w_k = convolution weights that define the smoothing.

In the Savitzky-Golay technique, w_k is calculated by fitting the NIR spectrum with low degree polynomials by using least squares regression. Savitzky-Golay determined w_k for the varied orders of polynomials and N(N = $2n$ + 1), and if for example, N is equivalent to 5, the values for smoothing can be found by substituting w_k = −3⁄35, 12⁄35, 17⁄35, 12⁄35, −3⁄35 (k = −2, −1, 0, 1, 2) into equation 2.12. The wavelet technique eliminates both low and high-frequency noise in addition to the removal of localized noise due to scattering effects, unlike smoothing which can only remove high-frequency information. ANN is built to mimic the structure of biological neural networks that form the human brain and nervous system and the manner they perform or are supposed to perform (Siesler et al., 2008).

Resolution enhancement

Resolution enhancement technique is used in disengaging overlapping bands and elucidating the presence of obscured bands. Mean centering (MC) is a very powerful resolution enhancement tool. It is an operation that adjusts a spectral dataset to relocate the centroid of the data to the origin of the coordinates system (Ozaki et al., 2006). MC can be defined as a procedure in which from every element of the jth row spectrum the column mean is deducted:

$$X_{jcent} = X_j - \left(\frac{1}{n} \sum_{j=1}^{n} X_{ij} \right) \qquad (2.13)$$

Where $\boldsymbol{X}j$ = an element of the j th spectrum, $\boldsymbol{X}i\ j$ = an element of data matrix $\boldsymbol{X}$.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Normalisation

Normalisation technique adjusts a spectral dataset that equalises the magnitude of each specimen. There are two normalisation methods used in spectral data preprocessing: vector normalisation and mean normalisation (Ozaki et al., 2006). The vectors are normalised to a constant Euclidean norm (square root of the total squared sum) i.e.

$$\mathrm{x}_{jnorm} = \mathrm{x}_j / ||\mathrm{x}|| \quad (2.14)$$

Whereas, in the mean normalisation technique all spectral points of the *j*th spectrum are divided by their mean value (sum of spectral values) i.e.

$$\mathrm{X}_{jnorm} = \mathrm{X}_j / \left(\frac{1}{m} \sum_{i=1}^{m} X_{ij} \right) \quad (2.15)$$

Where: ||**x**|| = Euclidean norm, X = spectral vector, *m* = total number of spectral points. Essentially, all the spectra will have the same area after the mean normalisation operation.

Model development for quantitative and qualitative analyses

Model development procedure involves the employment of multivariate calibration algorithms which interrelate matrix X and matrix Y. Where the matrix X is made up of known information (chemical information) and the matrix Y is made up of spectral data of the range of wavelength used (Ozaki et al., 2006).

Qualitative analysis (QA) by NIR spectroscopy

QA is a pattern recognition method that allows analytes with similar attributes to be clustered to develop classification techniques for unknown analytes. QA, basically, means the classification of analytes based on their NIR

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

spectral information. Mostly, classification methods are divided into two categories namely unsupervised algorithms and the supervised algorithms. The unsupervised algorithms are pattern recognition techniques that do not require any prior knowledge about the analytes to be classified, except the NIR spectra, but instead produce the clustering itself. Whereas the latter category classifies analytes based on prior knowledge, that is, the category membership of analytes is needed. Hence, the classification model is established on a calibration set of analytes with known categories and the model performance evaluation is achieved by comparing the classification predictions to the true categories of the validation analytes (Roggo et al., 2007).

Principal components analysis (PCA)

PCA is an unsupervised pattern recognition technique that is mostly deployed to visualize data trends in a scatter plot or a dimensional space. It is a feature (variable) reduction procedure that reduces the dimension of the data matrix and compresses the data into interpretable scores or variables known as principal components (PCs) which are linear combinations of the original data. The top three PCs (PC1, PC2, and PC3) will show the most relevant information and eliminate the non-useful ones. Thus, similar analytes are grouped closer to each other and vice versa. Graphical presentation of PCA results can give an initial output for determining the possible variations and similarities among the analyte set. PCA was used on NIR spectroscopy to discriminate transgenic tomatoes (Xie et al., 2007).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Linear discriminant analysis (LDA)

LDA is a linear technique with differentiating characteristics, that centre on finding optimum boundaries among classes. It is a variable reduction technique that chooses the directions that yield maximum separation among the given classes (Roggo et al., 2007). It assumes prior knowledge of the group membership of each analyte in a training set. LDA principle is based on the determination of linear discrimination functions that clearly show the ratio between-class variance and it reduces the ratio within-class variance (Chen et al., 2012a). LDA has been used on NIR spectra to differentiate fishmeal batches composed of different fish species (Cozzolino et al., 2011).

K-nearest neighbour (KNN)

KNN is a linear and non-parametric technique. In KNN, the matrix of the distance of the validation set analytes to all the others in the training set is calculated, and the neighbours of an unknown analyte are the analytes possessing the smallest Euclidean norms. The predicted group is the group presenting the majority of objects amidst the K neighbours (Roggo et al., 2007). Chen et al. (2009) used KNN on NIR spectra to discriminate roast green tea with a 96.3% correct identification rate.

Partial least square discriminant analysis (PLS-DA)

PLS-DA is a linear and parametric pattern recognition technique that focuses on discrimination among classes. It is a multivariate method for modeling a correlation between independent variables or scores (**X**) and dependent variables (***Y***). The principle of partial least square is to discover the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

components in the input matrix X that represent most possible the relevant differences in the input scores and simultaneously have a maximum correlation with the target in matrix value Y. PLS-DA aims at finding the variables and directions in the multivariate workspace that differentiate the proven groups in the calibration set. In PLS-DA, the Y matrix is built with zeros and ones and the ***X*** matrix comprises the preprocessed data. PLS-DA was employed for the classification of specimens out of dissolution specifications and to check the identity of the blistered tablets (De Maesschalck & Van den Kerkhof, 2005).

Support vector machine (SVM)

SVM is a supervised learning method established on the statistical learning theory for two-class classification problems (Cortes & Vapnik, 1995). It is applicable to scope with both qualitative and quantitative problems. For classification analysis, SVM is a technique for obtaining the optimum boundary of two groups in a vector workspace individually on the probabilistic distribution of calibration vectors in the dataset. If the groups are linearly parted, the goal of SVM is to discover the optimum hyperplane boundary that exactly parts both groups, classifying not only the calibration set but also the unknown specimens. The training vectors nearest to the boundary are known as support vectors. The minimum distance from the separating hyperplane to the nearest data points is known as margin. Thus, SVM machine learning looks for an optimum separating hyperplane where the margin is maximum. When the groups are parted by nonlinear boundary, the kernel technique is employed to find the boundary. The study by Zhao et al. (2006) disclosed that NIR

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

spectroscopy combined with the SVM could be effectively applied for rapid identification of the tea categories.

Quantitative analyses by NIR spectroscopy

Once the qualitative classification of analytes has been accomplished it could be beneficial to know more precisely in what amount analytes are different. So when one is interested in the quantity of a constituent present in a sample, it is referred to as quantitative analysis. Thus the development of a model for quantitative analysis appears very essential (Burns & Ciurczak, 2007). Historically, the first-ever application of NIR spectroscopy for the quantitative predictions was carried out on the moisture content of seeds by Hart, Norris, & Golumbic (1962).

Partial least square (PLS)

The PLS is the type of quantitative regression technique that is mostly used for linear data analysis. It uses information from all the wavelengths in the entire NIR spectrum to determine sample composition. PLS technique uses data reduction procedures to decrease a huge quantity of variables to a smaller quantity of new variables that contribute to most of the variation in the samples. These new variables known as factorial coordinates are beneficial concerning the chemical values which need to be predicted. The quantity of a component in samples can be determined by these new variables. The PLS technique performs computation on both spectral data and chemical (concentration) data simultaneously. The first coordinate is calculated to predict the spectral data and concentration variable as much as possible. The part of the spectral data and

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

concentration variables that were not determined, referred to as residuals, is then computed. The second coordinate best determines the residuals of the spectral data and concentration variables. PLS aims to create a linear relationship between spectral data (X) and reference values (Y). This method is modeling both X and Y to discover the variables in the X matrix which will describe the Y matrix (Roggo et al., 2007).

Interval Partial least square (i-PLS)

The i-PLS is introduced if the spectral data are highly related, windows of variables ought to be employed instead of doing variable selection on each variable separately. In i-PLS, the entire sample spectrum is split into smaller equidistant intervals, and then PLS regression models are applied on each of these intervals individually (Nørgaard et al., 2000). Afterward, an average error is computed for each subinterval and also for the entire spectrum. The wavelength region with the smallest error is selected. The main purpose is to find the interval(s) which give better prediction than the prediction obtained when using the entire spectrum. Thereafter, the performance of subintervals based on root mean square error of cross-validation (RMSECV) are compared to entire-spectrum RMSECV values.

Synergy Partial Least Square (Si-PLS)

The most vital spectral data for the quantitative regression are likely discontiguous. In this instance the selection of a single range is unsatisfactory, resulting in a bigger error in prediction. In this situation, the selection of spectral wavelengths regions ranges should be done judiciously to improve the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

predictive capability of the PLS regression model (Lee, Bawn, & Yoon, 2012). Therefore, variable selection by the Si-PLS regression technique could be employed to find out if the fusion of more than one interval could lead to models with better predictive ability. The Si-PLS regression technique performs by splitting the spectral dataset into a small number of intervals in order to compute all likely PLS regression model fusion of two or three or more intervals. Afterwards, the combined RMSECV of subinterval is compared to the entire spectrum RMSECV values.

Genetic algorithms partial least square (GA-PLS)

An alternative approach to copy a natural mechanism is the GA-PLS technique. According to the theory of Darwin, an evolutionary circle of reproduction and selection is the reason for the development and vanishing of species. The driving energy behind this situation is the struggle for life or survival of the fittest within an aggressive setting. The struggle for life ensures that only those species which are best adapted to their setting have an opportunity in the long run. If the likely solutions of an assumed problem are viewed as individual beings, the constituents that could impact the solution are their chromosomes (Siesler et al., 2008). The purpose is, to begin with, a handful of solutions with randomly selected constituents and permit and by so doing permit the individuals to evolve, i.e., to reproduce by fusing their chromosomes. An exceptional function, the fitness function, is employed to assess and rate the quality of the individuals. Depending on this rating a verdict is passed as to who "survives "and who "dies". In nature, individuals with "good genetic material

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

will live longer and will have better opportunities to procreate and to fuse their chromosomes, and thereby improve the genetic quality of new individuals. The theory of evolution circle of reproduction, mutation, and selection is a fundamental concept in the GA-PLS technique. It imitates selection in nature by evaluating models comprising certain fusions variables in many generations. It creates a vector called a chromosome which corresponds to the number of variables. In the spectroscopic multivariate calibration problems, a chromosome could possess the values of all the constituents one will manually differ to optimize the result. The fitness function will compute the root mean square error of prediction (RMSEP).

Calibration set development

In model development, the calibration procedure is constantly included for scientific purposes. Also, calibration offers all the three rudimentary functions of analytical wet chemistry viz. identification, separation, and quantification (Geladi, 2002). The function of the calibration procedure is to develop a model that matches up reference results to independent spectral data. For example, calibration is used for the development of the connection between the spectral data of NIR spectroscopy and the parameters of interest determined by a reliable analytical reference method. In the development of a robust model, the calibration set must contain a wide range of samples that includes different chemical compositions, variety, location, maturity stage, and preprocessing methods among others. The most common sample selection (i.e., partition) is that two-third of the analytes are employed in the calibration set whereas one-

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

third is employed as the internal validation set. Figure 2.8 shows a brief and schematic representation of the calibration model development process for NIR spectroscopy of the cocoa beans.

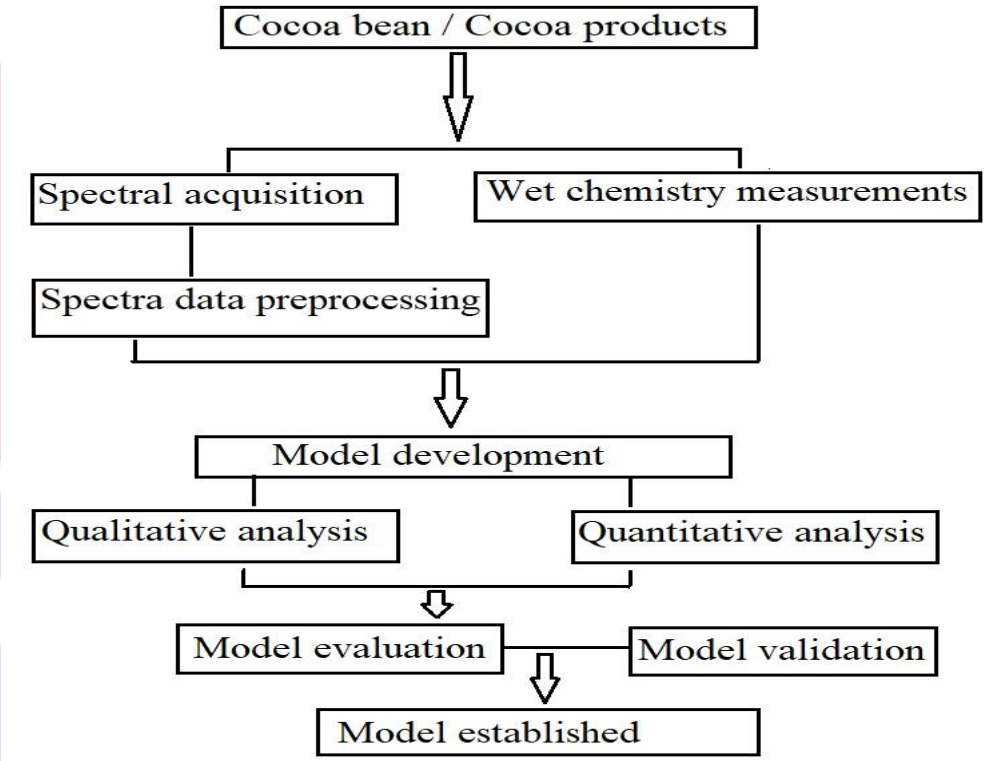


Figure 2.8 Schematic representation of the calibration model development process for NIR spectroscopy.

Prediction set development

After the development of the calibration model, the model is tested with the prediction set to establish the practicality, robustness and effectiveness of the model for future evaluations. Lin et al. (2011) used the leave one out cross-validation (LOO-CV) technique to test the model. The LOO-CV technique is performed as follows: one sample in the calibration set is taken out and the remaining data is used to build the model and then the left or deleted out sample

Digitized by Sam Jonah Library

is predicted with this model and this process is repeated with leaving out each of the samples of the calibration set.

Evaluation of models

The application of statistical calculations in scientific research has become an essential requirement. This is because it helps in drawing meaningful deductions from large data and for making decisions when challenged with variability. Usually, statistical evaluation is the last step in model development and it is very essential for results interpretation and evaluation of the accuracy and efficiency of the calibration model. The statistical tool generally employed for evaluating the performance of the method in multivariate algorithms are correlation coefficient (R), root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP) and bias (trueness) (Zhang et al., 2004) as detailed in Table 2.7.

Table 2.7 Statistical tools for performance evaluation

Equation	Parameter	Recommendation
R	Correlation coefficient	Close to one as possible
RMSECV	Precision	As small as possible
RMSEP	Tested precision	As small as possible or similar to RMSECV
Bias	Trueness	Close to zero as possible

Source: (Zhang et al., 2004)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Transfer of calibration models

Calibration building or development is rigorous, expensive, and time-consuming due to the reference analysis involved. Thus, it is inappropriate to keep on reproducing new calibration for each new spectrometer. The objective of calibration transfer is to apply a unique model on several spectrometers with equal prediction errors. The model is built on an initial spectrometer termed the master instrument and is used on other spectrometers known as slave instruments. Three different methods are available to transfer the calibration between instruments: The first approach is the development of a robust model that can give us a precise prediction on several devices. The calibrations are constructed with spectral information acquired on several spectrometers. The second method is connected with the predicted value correction. Here, a slope and/or bias between the predicted values of master and slave spectrometers is calculated. Bias correction modifies the differences between devices. Mathematical;

$$\text{Bias}=\sum_{j=1}^{n}(y_{master,j}-y_{slave,j})/n \qquad (2.16)$$

Where; y_{master} = values predicted with master device, y_{slave} = values predicted with slave device, n = number of samples in the calibration transfer data set

The third approach is spectral data correction. In this case, the master and slave spectra are compared. An arithmetic correction is then used on the slave spectra so that they are made similar to master ones and therefore to apply the master calibration on the slave devices (Roggo et al., 2007).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Conclusion

Cocoa bean, NIR spectroscopy and chemometric techniques have been reviewed. The consumption of cocoa beans is connected to numerous health benefits and this has necessitated the move for more research leading to an increase in its consumption. Qualitative and quantitative analysis of merchantable quality cocoa beans is very necessary after primary processing (fermentation and drying) of the beans, however, the various methods used are laborious, tedious, cumbersome, expensive, time-consuming, and require chemical usage. On the other hand, the application of portable handheld NIR spectroscopy combined with chemometric techniques to raw cocoa bean quality is very scanty or not available up to now. It could be concluded that there is vast potential for the application of handheld NIR spectroscopy combined with chemometric techniques for cocoa bean examination in the initial processing chain (fermentation and drying) before shipment to the international market.

Chapter Summary

This chapter reviewed the literature on cocoa bean, health benefits of cocoa consumption and standard methods for cocoa bean quality examination. Spectroscopy with emphasis on near-infrared were outlined and finally the theories involved in the chemometric data analysis discussed.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER THREE

NONDESTRUCTIVE AUTHENTICATION OF THE REGIONAL AND GEOGRAPHICAL ORIGIN OF COCOA BEANS BY USING A HANDHELD NIR SPECTROMETER AND MULTIVARIATE ALGORITHM

ELLIOT K. ANYIDOHO, ERNEST TEYE, ROBERT AGBEMAFLE

Department of Agricultural Engineering, School of agriculture, College of Agriculture and Natural Sciences, University of Cape Coast
Department of Laboratory Technology, School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast

Published: The Royal Society of Chemistry: Analytical Methods, 2020, 12, 4150-4158. doi:10.1039/d0ay00901f

Statement of Contributions

Elliot K. Anyidoho (Candidate)

Writing and compilation of original manuscript, investigation, data collection and formal analysis, established methodology, preparation of tables and figures

Ernest Teye (Principal Supervisor)

Data curation, supervised manuscript compilation, editing, and co-author of manuscript

Robert Agbemafle (Co-Supervisor)

Supervised manuscript compilation, editing and co-author of manuscript

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Introduction

The safety and quality of food are directly connected to people's health and socio-economic importance. Therefore, consumers and manufacturers are looking for quality seals and trust marks on food. Furthermore, the geographical origin of a specific food is very vital for traceability and quality control. These issues have underscored the need for developing a dependable technique to assess the geographical origin of food. It is therefore very important to develop a rapid, efficient and onsite technique for tracing the geographical origin of food.

NIR spectroscopy is progressively becoming a widespread technique for fast and non-invasive evaluation of food. This ubiquitous nature of NIR spectroscopy is reflected in its usage in a wide range of fields including agriculture, food, chemistry, polymers, medicine, geology, petrochemistry, pharmaceuticals, and environmental science (Chen et al., 2008). It is based on the absorption of electromagnetic radiation at wavelengths in the range of 780–2500 nm. Being a measurement technique, its merits include rapid detection (between 15 and 90 seconds) and a stress-free and nondestructive measurement procedure without the need for multiple chemical reagents. The major advantage of NIR spectroscopy is that usually no sample preparation is needed, and hence the analysis is very simple, and it can be carried out online. It undoubtedly becomes essential as a technique for traceability and quality control for food.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Cocoa (*Theobroma cacao L.*) beans are among the most celebrated food commodities universally, and their consumption has gradually become day-to-day activity owing to their useful medicinal properties. Ghanaian cocoa beans are of advanced and premium quality internationally (Othman et al., 2007). In Ghana, the cocoa bean is produced in seven cocoa-growing regions: Eastern, Volta, Ashanti, Central, Western North, Western South, and Brong Ahafo. The disparity in the cocoa bean quality owing to geographical origin is very much appreciated, and it is an essential factor for price determination (Gilbert, 2009). This financial motivation continues to push producers and retailers to mislabel and adulterate cocoa beans. NIR spectroscopy has been applied for qualitative and quantitative analysis of cocoa bean and cocoa bean products (Teye et al., 2019b). Other studies have also confirmed the potential of using NIR spectroscopy for qualitative and quantitative analysis of cocoa beans: identification of variety (Teye, Uhomoibhi, & Wang 2016), authentication of cocoa beans and differentiation of cocoa beans (Teye et al., 2013), and measurements of pH and fermentation index (Teye & Huang, 2015).

This study concentrated on the authentication of the origin of cocoa beans using a handheld spectrometer. Other researchers used a laboratory-based NIR spectrometer for the classification of Ghana's cocoa beans from different producing regions and found the MC-SVM model to be the best tool for optimum classification for different geographical origins (Teye et al., 2013). Aculey and co-workers (Aculey et al., 2010) also used FT-NIR spectroscopy to differentiate and characterize cocoa beans from different cocoa-growing regions

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

of Ghana. Also, FT-NIR spectroscopy was used for non-destructive identification of cocoa bean cultivars (Teye, Uhomoibhi, & Wang, 2016) and SNV plus SVM was found to be an appropriate model. Other authors also used NIR to discriminate three cocoa bean quality grades in terms of degrees of fermentation (Kutsanedzie et al., 2017). These researchers found 94% accuracy using support vector machines (SVMs) and extreme learning machines (ELMs). All these studies revealed that NIR spectroscopy coupled with the right chemometric tool can be used for qualitative measurements of cocoa beans. Although the aforementioned conventional NIR analyses are considerably quicker and simpler than wet analytical methods, it is a challenge to rely on expensive and stationary laboratory-based spectrometers especially in developing countries. On the other hand, it must be noted that sometimes handheld spectrometers suffer in terms of data acquisition performances concerning laboratory bench spectrometers. The advancement in NIR miniaturization has unlocked a new prospect for NIR applications suitable for industry, laboratories, and onsite analysis. This has led to the transformation of large, stationary laboratory-based NIR tools into handheld devices. Compared to laboratory-based spectral measurement spectrometers, handheld NIR spectrometers are less expensive and require minimal equipment and user involvement. Other additional advantages are: easy to move to the field for on-field measurements and product-to-product examination and regulation.

Recent review articles describing the basics of portable NIR spectroscopy ascertained its potential in food control and inspection (Dos et al.,

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

2013) and the significance of developing and adopting it along the food supply chain has been emphasized and extensively reviewed for the inspection and control of meat (Kademi, Ulusoy & Hecer, 2019). Other authors evaluated a miniaturized spectrometer device for discriminating cultivars of barley, chickpea, and sorghum in Ethiopia (Kosmowski & Worku, 2018).

Also, a recent review on cocoa bean and cocoa bean product quality evaluation by NIR spectroscopy and chemometrics demonstrated the success of NIR spectroscopy technology in the qualitative and quantitative examination of cocoa beans and cocoa bean products (Teye et al., 2019b). These researchers recommended the transition of this important technology from stationary laboratory applications to real on-field usage for optimum worldwide benefits.

However, no studies have evaluated the application of handheld NIR spectroscopy for regional and geographical classification of cocoa beans. This study, therefore, seeks to evaluate a handheld near-infrared spectrometer device for discriminating Ghanaian cocoa beans and other cocoa beans from African countries.

Materials and Methods

Cocoa sample preparation

Cocoa bean samples (total 210; i.e., 30 samples from each region) were collected from seven cocoa-producing regions (Eastern, Western North, Western South, Ashanti, Volta, Central and Brong Ahafo) in Ghana. While 120 samples (30 each from four countries- Uganda, Ivory Coast, Nigeria and Ghana) were collected in May and August 2019 respectively with the help of the Ghana

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Cocoa Board. The cocoa bean samples used for this study were prepared to meet the internationally accepted merchantable quality standard: well fermented, thoroughly dry, free from smoky beans and abnormal or foreign odours, free from evidence of adulteration, reasonably free from living insects, uniform in size, free from broken beans and pieces of shell, and virtually free from foreign matter. The well-prepared and labeled samples were transported to the Department of Agricultural Engineering Laboratory, the University of Cape Coast for examination.

Reference measurement of cocoa bean quality

Some chemical properties such as fat, moisture content, fermentation index, pH, and polyphenols of the cocoa bean samples from the Ghana group and African group were measured (replicated three times) by using standard recommended reference analytical methods used by other authors (Huang et al., 2014; Teye et al., 2014b; Teye et al., 2015). These chemical properties were selected because they serve as better indicators to differentiate cocoa beans. This is because the major postharvest activities (fermentation and drying) have a great influence on cocoa bean quality.

Spectra acquisition using handheld NIR spectroscopy

A handheld NIR spectrometer (TellSpec, UK) was used to scan the cocoa bean samples. The spectral data were acquired at 1 nm resolution over a wavelength range of 900 to 1700 nm. The hand-held spectrometer gives the measurements in relative absorbance units (log 1/R), which were associated with chemical constituents (Martens, Nielsen, & Engelsen, 2003). The device

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

was operated using a smartphone (Samsung Galaxy A20) application with spectral data stored remotely. All the cocoa bean samples were carefully scanned five times through a transparent zip-locked polythene bag at different positions. The five scans were averaged for a sample. The transparent bag showed no significant interference with the NIR signals.

Spectral data partition

The spectral data-set for Ghana cocoa beans (210 samples) and African cocoa beans (120 samples) was downloaded separately and each category was partitioned into two subsets: a training set and a prediction set. For the Ghana cocoa bean samples, 140 samples and 70 samples were selected as the training set and testing set respectively, while for the African cocoa beans 80 samples and 40 samples were also selected as the training set and testing set respectively. The training set was used to develop the model while the testing set was used to evaluate the actual predictability of the model. The individual samples in each set were selected randomly to come to a 2/1 division of the training set and testing set as was done by other authors. The training set was used for training the techniques while the test set was used to independently evaluate the developed model.

Spectra preprocessing and chemometric data analyses

All the computations, chemometric analysis, and graphics were done with Matlab (The MathWorks, Inc., USA, version 9.6.0.) using the Windows 10 Pro software package. The preprocessing method was done to improve the raw dataset. This act is very important in reducing redundant information in the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

spectra (Jha & Garg, 2010). The preprocessing methods employed initially were: first derivative (FD), second derivative (SD), mean centering (MC), multiplicative scatter correction (MSC), and standard normal variant (SNV). After a trial-and-error procedure, three methods were selected for reporting namely MC, MSC, and SNV.

MC

MC aims to subtract the average value from every variable in the first step of preprocessing. This ensured that the results were much more interpretable as done by others who deployed this preprocessing technique in their work (Teye et al., 2013).

MSC

MSC is a spectral pretreatment algorithm that is useful for removing un-useful variability due to scattering, that is, it decreases the effects of light emission and minimizes diversity in the spectrum. The working principle of MSC is to examine each spectrum and correct it based on the reference spectrum by utilizing the results of simple linear regression estimation. It was also used for the correction of scattered light on different particle sizes and this procedure is used to correct additive and multiplicative effects (Dhanoa et al., 1994).

SNV

SNV transformation is applied to each spectrum in isolation and without any reference to the sample set. This transformation first centers the spectral values, (that is) subtracting the mean of the individual spectrum from each

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

value. These centered values are then scaled by the standard variation calculated from the individual spectrum values. Standard Normal Variate transformation removed the multiplicative interference of scattering, particle size, and the change of light distance. It corrects both multiplicative and additive scatter effects. To remove slope variations on an individual spectrum basis, each object was transformed independently.

Principal component analysis (PCA)

After preprocessing, PCA as a visualization method was performed. PCA works by reducing correlated variables and then converting them into new variables that are not correlated with each other. PCA modifies the large numbers of variables into the principal components (PCs), which are few but still can explain diversity in the original data (Rinnan et al., 2009a). The best-performing PCs normally show the most important information. Hence, similar samples are grouped nearer to each other and vice versa. The graphical outline of PCA results normally gives preliminary output for the determination of possible variances and similarities in a dataset. PCA can be used to detect combinations of variables that have the highest contribution to variations in the dataset, as these variables are kept in the first two or three PCs and the PCA loading plot explains these contributions.

Multivariate classification algorithms

After careful selection of optimum spectral preprocessing methods, different multivariate classification algorithms were studied systematically and

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

their results were compared. These classification methods included linear discriminant analysis and support vector machines.

Linear discriminant analysis (LDA)

Linear discriminant analysis is a supervised classification method that focuses on finding optimal boundaries among classes. LDA was used to find the linear combination of features and the resulting combination was used as a linear classifier. It works by finding the linear combination of features which brings out clearly the ratio of between-class variance and reduces the ratio of within-class variance. Some authors used LDA in the differentiation of soy sauce with mid and NIR spectra (Teye et al., 2019a).

Support vector machine (SVM)

Classification using support vector machines is a strong non-linear supervised classification method (Yu & Kim, 2012). SVMs have shown good performance for classifying high-dimensional data when a limited number of training samples are available. They process sensor data by obtaining the optimal boundary of two groups in a vector space independent of the probabilistic arrangements of vectors in the training set. When the linear boundary in the low dimension input space is not enough to separate the two classes, the SVM algorithm constructs a hyperplane or a set of hyperplanes in a high-dimensional space for classification. However, if the classes are separated by a non-linear boundary, then the kernel function is used to find the boundary by mapping the non-separable data into a higher dimensional space and causes the classes to become linearly separated (Yu & Kim, 2012). The strength of

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

SVMs over the others is that they can achieve higher generalization by maximizing the margin and they can support efficient learning of non-linear functions using the kernel trick. Among the three kernel functions (sigmoid kernel, polynomial kernel, and Gaussian kernel), the Gaussian kernel function is mostly employed because of its simplicity and speed during its computation and this was selected in this study.

Results and Discussion

Spectroscopic data presentation

Fig. 3.1, Fig. 3.1a, and Fig. 3.1b present the spectral information of raw spectral data and different preprocessing techniques of the cocoa bean samples from the seven cocoa-producing regions of Ghana and four African countries respectively. From these figures, it could be seen that each spectrum shows a unique profile. The selection of these preprocessing treatments was based on a trial and error approach involving several selections (Rinnan et al., 2009b). NIR spectroscopy is based on the absorption of electromagnetic radiation at wavelengths in the range of 780-2500 nm. In this novel investigation, the absorbance data were obtained from the measurement of spectra with wavelengths of 900-1700 nm. This wavelength range could provide relevant features for the classification of different cocoa bean categories. When light with various wavelengths is radiated on organic matter, a portion of light at certain wavelengths is absorbed. The amount of absorbed light depends on the composition of the irradiated organic material, in this case, cocoa beans. Each species of cocoa bean has a different composition from other cocoa beans such

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

as polyphenols, proteins, fat, alkaloids, volatile and non-volatile acid, moisture, etc. This causes different absorption responses in each species. During the NIR scanning of the cocoa bean samples, spectra were generated that showed multiple bands and a few peaks as seen from the spectral profile presented. These bands are made up of overtones and combinations of fundamental vibrations which correspond to organic and biochemical properties that provide exclusive characteristics or a fingerprint of the cocoa bean samples used. To the human eye, they look very alike, though there exist many variances with useful and non-useful information. To extract useful and exclusive information, multivariate mathematics (chemometric) techniques were employed. Fig. 3.1 presents the spectral information of raw and mean spectra profiles of the cocoa bean samples from the seven cocoa-producing regions of Ghana and four African countries respectively. From these figures, it could be seen that each spectrum showed a similar profile pattern. Also, locations that were closer or overlapping are not far from each other as revealed in the mean spectral profile.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

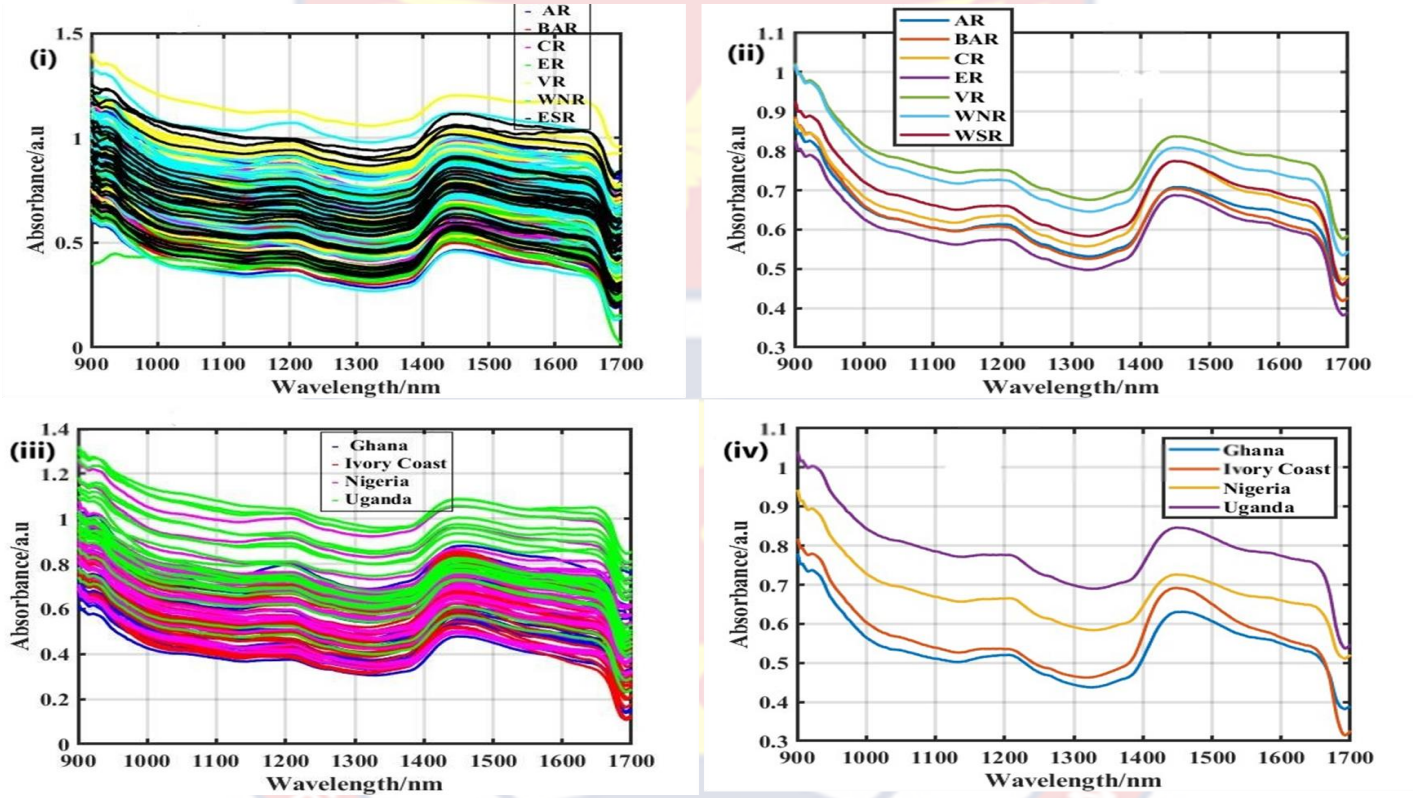


Fig. 3.1 Spectra of cocoa bean samples (i; raw & ii mean) of Ghana and samples (iii: raw & iv mean) of African countries (where: AR = Ashanti region, BAR = Brong Ahafo region, CR = Central region, ER = Eastern region, VR = Volta region, WNR = Western north region, and WSR = Western south region all in Ghana)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

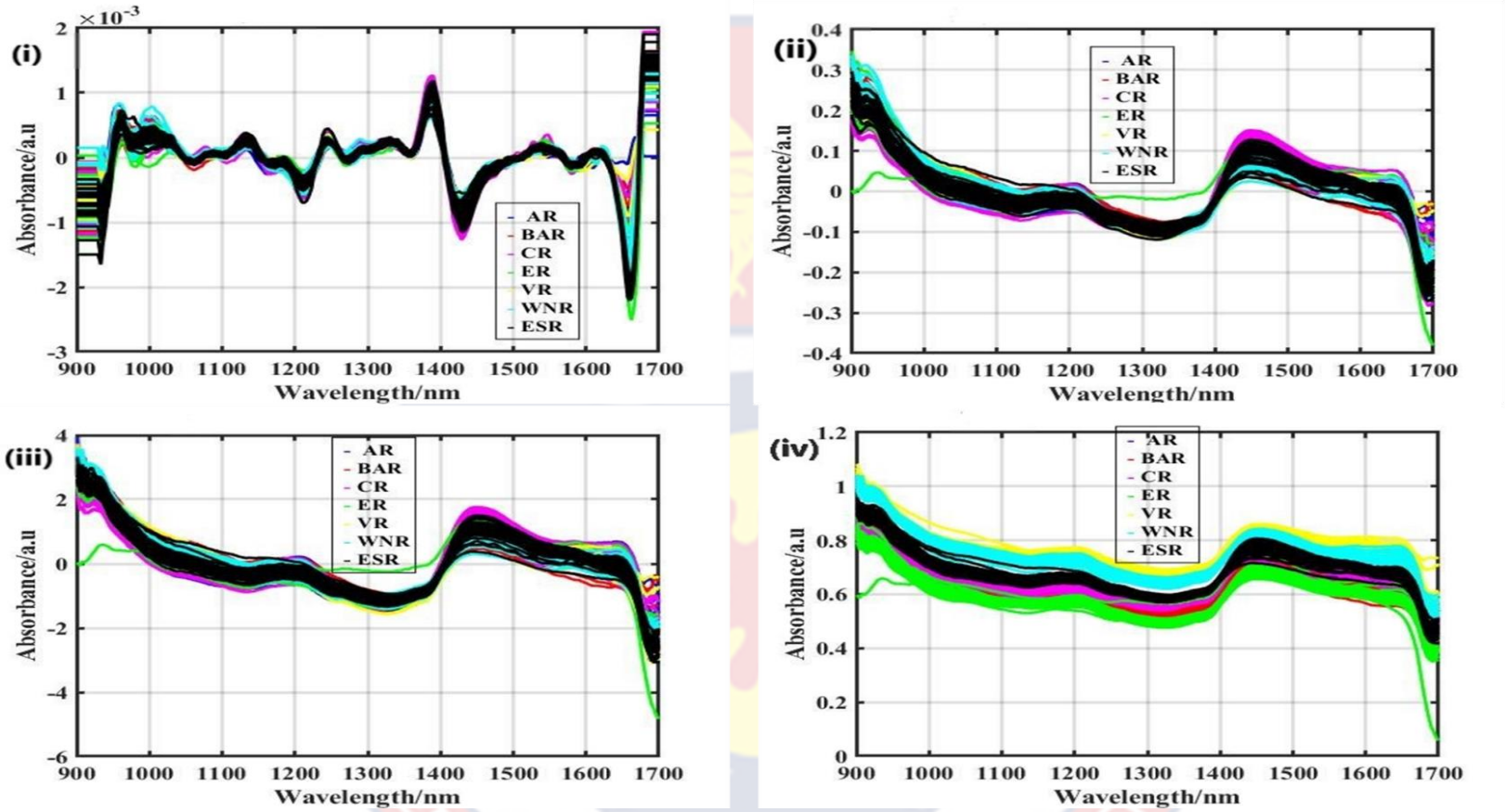


Fig. 3.1a Spectra profile of different preprocessing techniques (i) SD; Second derivative, (ii) MC; Mean centering, (iii) SNV; standard normal variant and (iv) MSC; multiplicative scatter correction for all the cocoa bean samples in Ghana

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

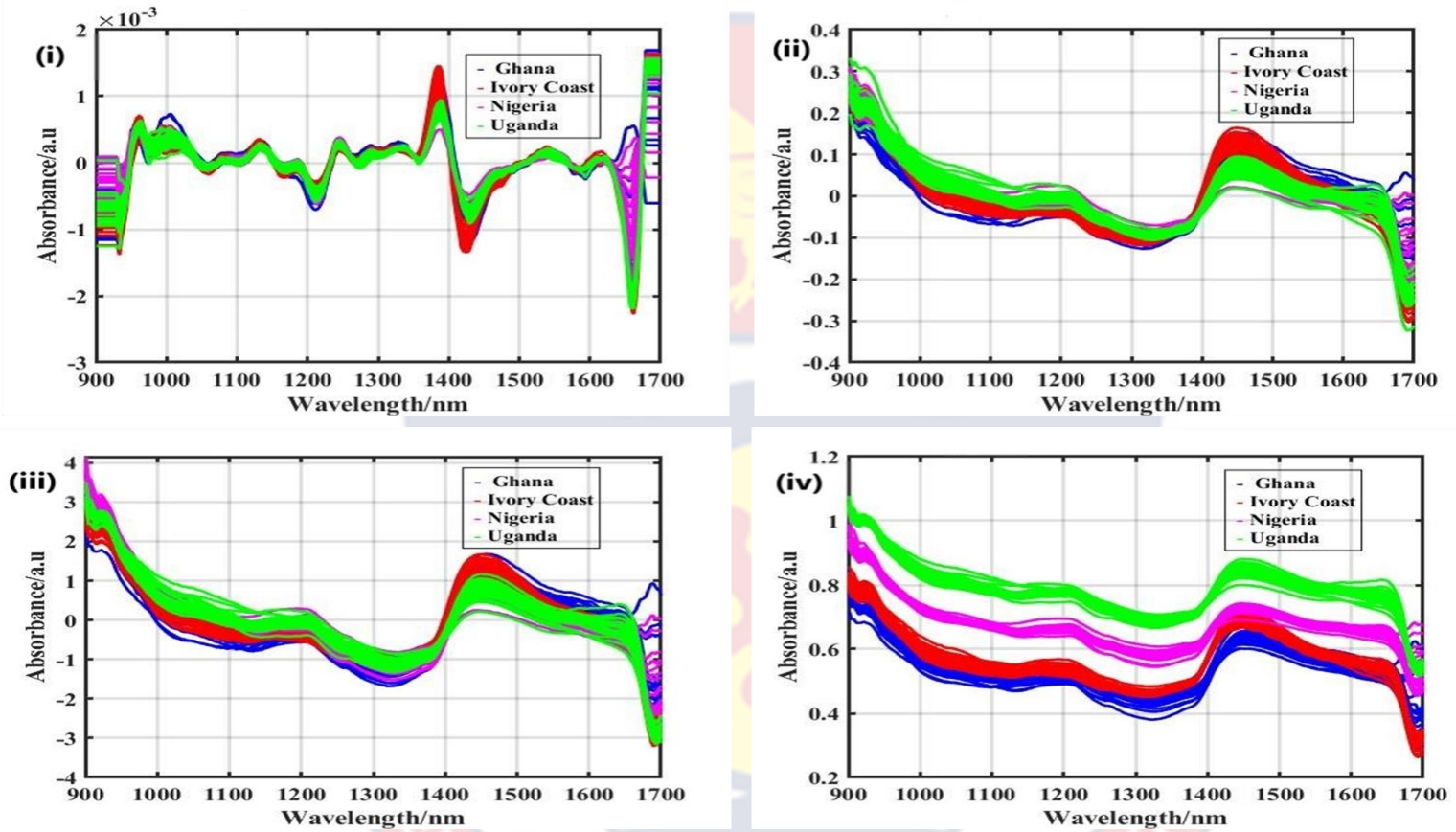


Fig. 3.1b Spectra profile of different preprocessing techniques: (i) SD, (ii) MC (iii) SNV and (iv) MSC for samples from African countries

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Specifically, in Fig. 3.1(ii) at wavelengths of 900–1100 nm, 1200 nm, and 1400–1500 nm Ashanti region and Eastern region samples were closer while Ashanti Region and Brong Ahafo samples were overlapping. These regions are also close neighbors concerning their location on the Ghana map. On the other hand, in Fig. 3.1(iv) at the same aforementioned wavelengths, the samples from Ghana and Ivory Coast showed similar patterns as these countries were close to each other compared to the other countries which are far from their location. This phenomenon revealed that geographical locations could be identified. It also means that similar pre and postharvest activities that influence cocoa bean quality attributes are not entirely different when the locations share the border. Further, wavelengths of 900–1000 nm corresponding to the third overtone of the C–H stretching vibration (910–950 nm) which represents carbohydrate, protein, and fat, while 980 and 1040 nm are associated with the second overtone of the O–H stretching and the second overtone of the N–H stretching of fat and protein respectively. In this study, various pretreatments were done to improve the classification model. The selection of these preprocessing treatments was based on a trial and error approach involving several selections (Rinnan et al., 2009b).

Principal Component Analysis

Ghana samples

The spectral data of cocoa beans from the seven cocoa growing regions of Ghana and four African countries were further processed with PCA. Although PCA is not a classification tool, its properties could provide spectral

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

information trends. The PCA results revealed that the first three principal components (PC1, PC2, and PC3) presented clear distinctions in all the attributes investigated with MSC providing the best pretreatment results in PCA compared with the others.

All the cocoa bean samples were grouped clearly along the first three PC planes for MSC as seen in the PCA score plot. PCA normally identifies the most essential directions of the variation in the data space and determined the principal phenomena in the dataset (Sun et al., 2017). The best performance of MSC could be attributed to the fact that it corrected the effect of scattering and additive and multiplicative effects (Dhanoa et al., 1994) better than the other pretreatments. From Fig. 3.2 (iv), the topmost three PCs contributed to 98.18% (PC1 = 91.18, PC2 = 4.52 and PC3 = 2.48) of the total variance in the dataset for cocoa beans in different locations in Ghana. This means that the top three PCs can explain 98.18% of the variance information from the spectral data which covers vital chemical fingerprints in the cocoa bean samples used (Teye et al., 2015). To explain this observation, the PCA loading plot as seen in Fig. 3.3 was examined as described by other authors (Cozzolino et al., 2011; Teye et al., 2015), and this loading plot was used to explain how much each wavelength (x variable) contributed to the meaningful variability in the data. It also facilitated the interpretation of variable relationships. Also, for the length and directions to be made meaningful the loading weights were normalized (Cozzolino et al., 2011).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

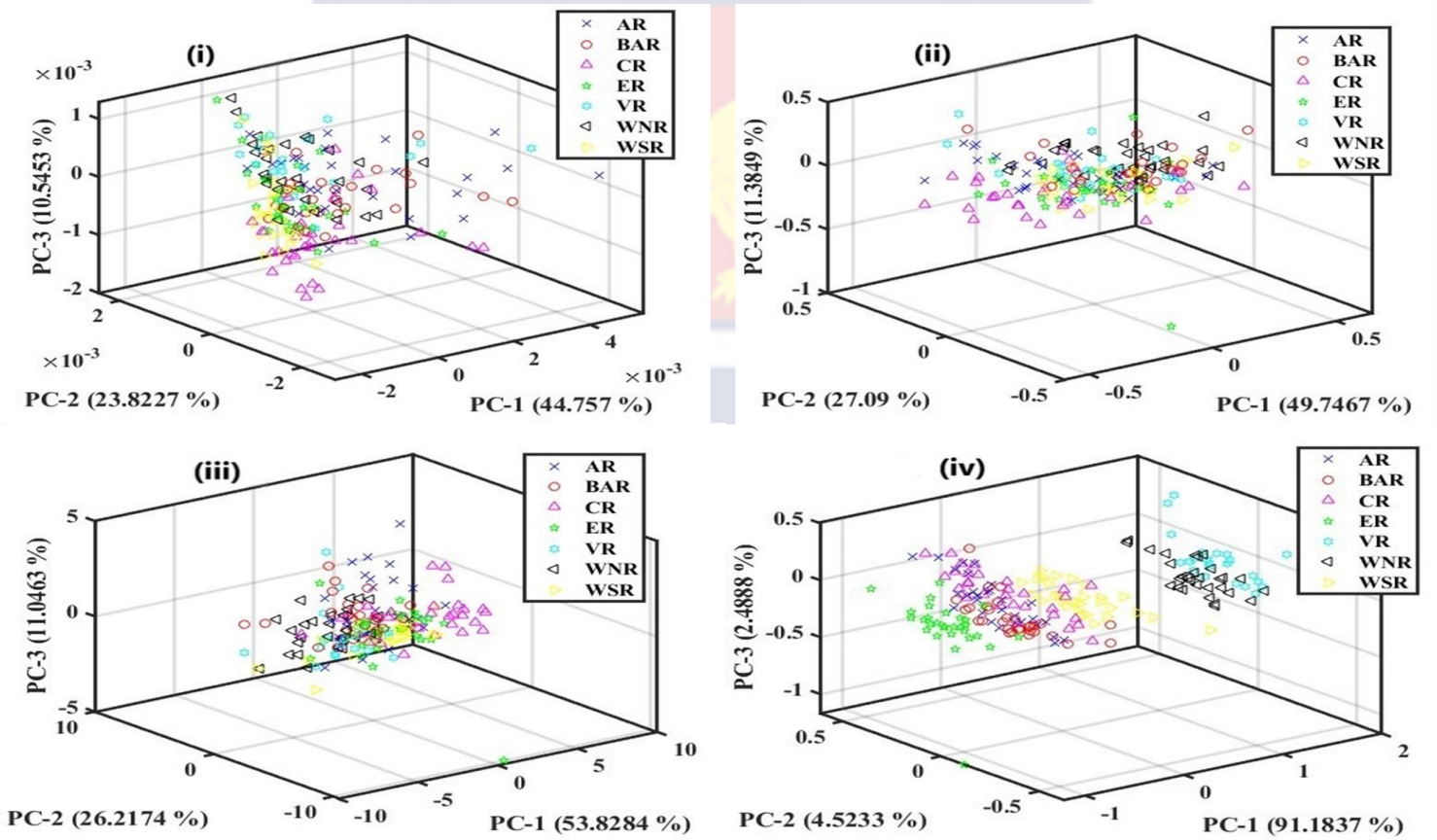


Fig. 3.2 PCA score plot of the first three PCs of cocoa beans from different locations in Ghana (i = raw, ii = MC, iii = SNV and iv = MSC).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

From Fig. 3.3 it could be observed that the significant absorption bands for the first component were around 1200 nm, 1450 nm, and 1650 nm and these wavebands represent the C–H 2nd overtone, N–H first overtone, and C–H first overtone respectively which could correspond to saturated and unsaturated triglycerides (Teye et al., 2015) and aromatic compounds (Veselá et al., 2007). The cocoa bean is made up of about 60% saturated fatty acids such as stearic and palmitic as well as 40% unsaturated fatty acids such as oleic, linoleic, and linolenic acids. Also, the wavelengths for the second component were around 1000 nm, 1150 nm, and 1420 nm which represent the N–H 2nd overtone, C–H 2nd overtone, and O–H first overtone which are associated with protein, organic acids (acetic, citric & lactic) and pH present in cocoa beans (Teye et al., 2015; Veselá et al., 2007).

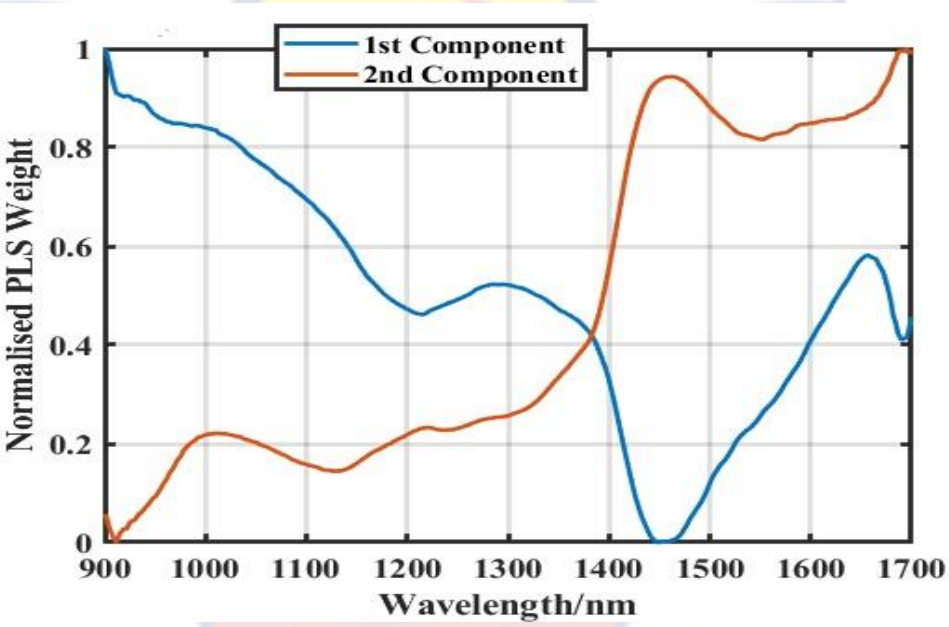


Fig. 3.3 PCA-loading weight of top two latent variables of seven regions of Ghana.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

African samples

From Fig. 3.4 it was observed that there was a well-ordered grouping of cocoa bean samples from the four African countries for the MSC PCA score plot compared to others. This may be attributed to the ability of MSC to remove the undesirable scatter effect of spectra from a data matrix before PCA (Coronel-Reyes et al., 2018).

The observed score distribution was similar to those done by other researchers for other commodities. Also, Giraudo and coworkers observed a similar trend for green coffee from American and Asian continents (Giraudo et al., 2019). Again, in Fig. 3.4 (iv) the observable cluster trend in the PCA score plot gave a total accumulation contribution of 99.22% variance for the 120 samples used. Cocoa bean samples from Nigeria and Uganda assumed positive score values of PC1; therefore, they were found on the right side of the scores plot. Conversely, all the cocoa bean samples from Ghana and Ivory Coast were found to assume negative score values of PC2; therefore, they were gathered on the left side of the score plot. Also, Ghana and Ivory Coast share the geographical border which influenced the closeness of their spectra as shown in Fig. 3.4. This could be basically attributed to the potential inherent variability.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

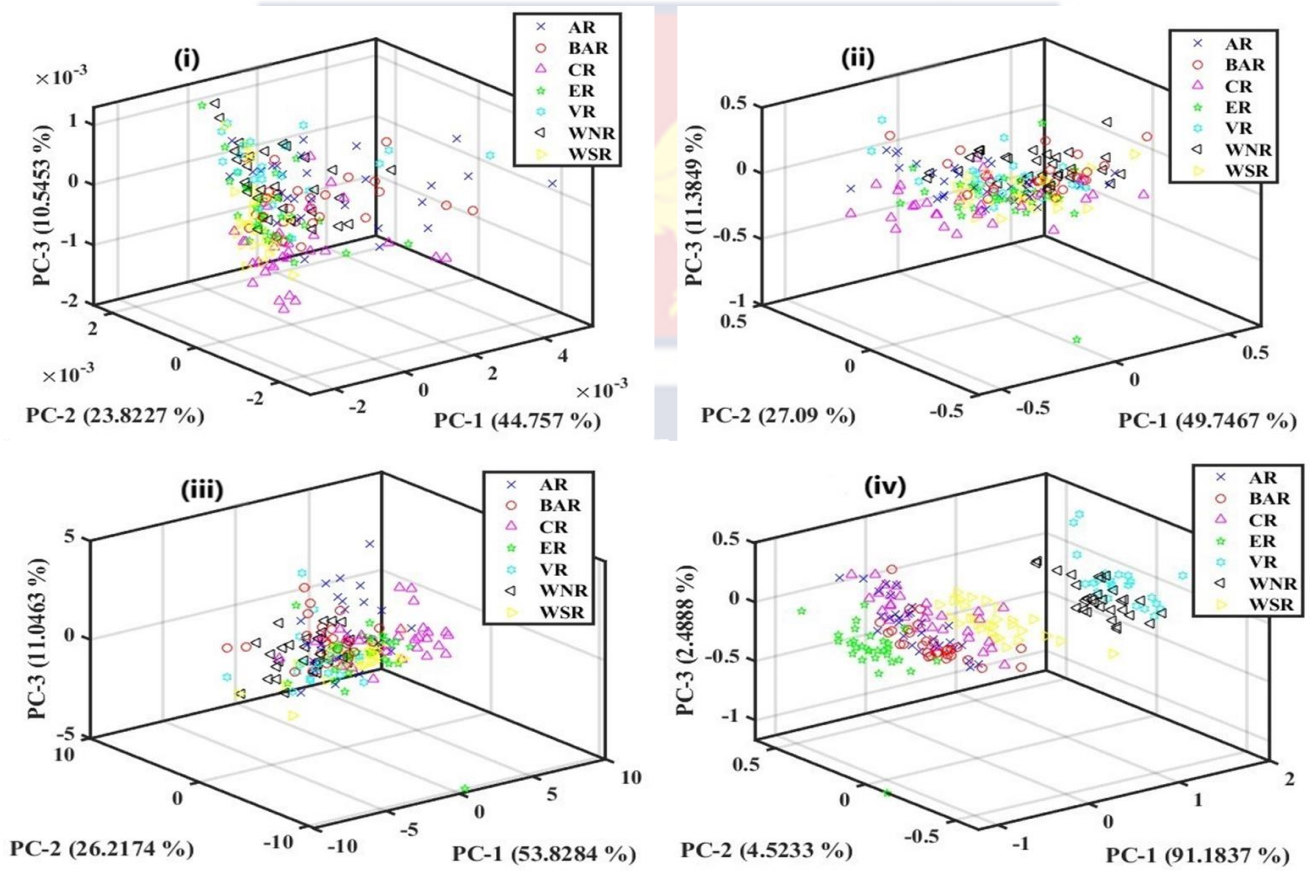


Fig. 3.4 PCA scores plot of the first three PCs of cocoa beans from four African countries preprocessed – raw (i), MC (ii), SNV (iii) and MSC (iv).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Specifically, PC1 explains 95.78% and PC2 explains 2.03% of the spectral variance of samples from the four African countries. To explain further, Fig. 3.5 shows significant peaks, for PC1 the wavelengths were 1050 nm, 1230 nm, 1450 nm, and 1650 nm which are associated with nonbounded water and aromatics, (Veselá et al., 2007) while PC2 showed major peaks around 950 nm, 1150 nm, 1450 nm, and 1690 nm which played a major role in the separation of cocoa bean groups. These bands represent H_2O, ROH, OH bonded aromatic groups, and NH_2 functional groups (Cen & He, 2007). Specifically, 1450 nm is associated with the O–H 2nd overtone that corresponds to ethanol while 1690 nm is related to the CH_3 stretching first overtone associated with O–H aromatic groups (Burns & Ciurczak, 2007). These observed wavelengths will be very useful for developing a classification model. These essential absorption bands closely correlated with the biochemical compositions and internal characteristics in cocoa beans which could provide a fingerprint for classification. For example, the wavelength at 920 is related to the chemical groups of C–H and O–H. The absorbance bands around 1208 nm, 1400 nm, and 1450 nm are caused by the C–H stretching 2nd overtone, C–H deformation, NH_2 functional groups, and O–H polymeric groups. These groups are characteristic of complex carbohydrates, water, aliphatic alcohols, fats, polyphenols, and acidic and aromatic compounds in the cocoa bean (Burns & Ciurczak, 2007; Workman & Weyer, 2012).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

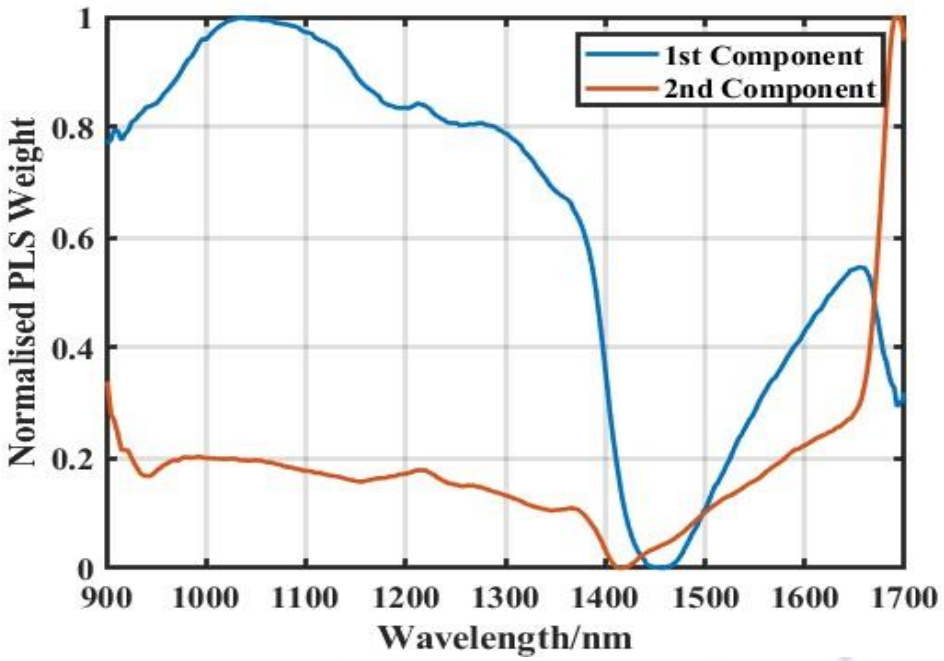


Fig. 3.5 PCA-loading weight of top two latent variables of four African countries.

Classification based on Ghanaian geographic regions

LDA and SVM were separately applied after PCA to perform the classification among the seven cocoa-producing regions of Ghana. The classification results are summarized in Table 3.1. From this table, it is observed that LDA and SVM gave an overall correct classification rate of 100% in the training set and testing set with 4 PCs. The biochemical compositions which contributed to the classification of cocoa beans according to the seven regions of Ghana are presented in Table 3.2. Although there were no significant differences among the parameters measured, the handheld NIR technology was able to differentiate the cocoa beans based on the obtained analyzed spectroscopic data. The values obtained for moisture, pH, fermentation index,

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

fat, and polyphenols were following data obtained by other authors (Afoakwa et al., 2013a; Aremu et al., 1995). Furthermore, the spectral differentiation could be due to the quality of physicochemical parameters measured in Table 3.2, not necessarily the amount as revealed in the distinct peaks observed in the spectra profile. Also, it showed that the handheld NIR spectrometer could be useful for classifying cocoa beans according to different locations in Ghana.

Table 3.1 Performance of the classification model based on seven regions of Ghana

Models	Preprocessing	PCs	Training set	Testing set
LDA	Raw	5	44.72	32.50
	MC	2	39.13	32.50
	MSC	5	100.00	100.00
	SNV	3	37.27	32.50
	SD	4	34.78	32.50
SVM	Raw	5	36.65	22.50
	MC	2	36.02	20.00
	MSC	3	100.00	100.00
	SNV	2	37.89	20.00
	SD	3	16.77	7.50

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 3.2 Biochemical parameter examination results for seven regions of Ghana

Parameter	Region	AR	BAR	CR	ER	VR	WNR	WSR
Moisture (%)	Average	7.244	8.329	8.266	7.754	7.208	7.724	8.140
	SDV	0.054	0.219	0.084	0.139	0.106	0.090	0.107
Fat (%)	Average	40.002	39.468	38.340	39.094	42.976	40.689	38.674
	SDV	0.060	0.049	0.036	0.115	0.078	0.078	0.062
Fermentation index (%)	Average	0.937	1.169	0.916	0.862	1.033	0.982	0.899
	SDV	0.024	0.053	0.041	0.013	0.026	0.057	0.035
pH (value)	Average	5.288	5.188	5.102	5.073	4.968	5.442	5.070
	SDV	0.113	0.022	0.038	0.040	0.052	0.131	0.063
Polyphenols (mg g^{-1})	Average	63.787	57.664	63.312	65.851	60.560	59.919	57.703
	SDV	2.205	0.820	0.981	1.216	2.786	2.975	0.806

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Classification based on four African countries

After the classification of cocoa beans from the Ghanaian geographic regions, the possibility to distinguish between cocoa beans from different African countries was studied. This is important mainly because some chocolate manufacturing companies and consumers prefer cocoa beans from other countries over others. Additionally, cheap importations are competing with high-quality beans. The variance in the data was mainly due to the country of origin (Ghana, Ivory Coast, Nigeria, and Uganda). LDA and SVM were separately applied after PCA to perform differentiation among the four African countries. The outcomes of the classification rate are summarized in Table 3.3. From this table, it could be seen that all the derived models (LDA and SVM) gave 100% overall correct classification rate both in the training set and testing set. The biochemical parameters which contributed to the classification of cocoa beans according to the four African countries are presented in Table 3.4. The concentration levels obtained for moisture, pH, fermentation index, fat, and polyphenols were in accordance with data obtained in the literature (Afoakwa et al., 2013c; Aremu et al., 1995). Even though there were no significant differences among the parameters measured the handheld NIR scanner was able to differentiate the cocoa beans based on the obtained spectroscopic data. This means that the NIR spectrometer can differentiate the inherent quality attribute of cocoa beans in addition to the quantity observed in Table 3.4.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 3.3. Performance of the classification model based on four African countries

Models	Preprocessing	PCs	Training set	Testing set
LDA	Raw	4	82.50	80.00
	MC	3	82.50	85.00
	MSC	4	100.00	100.00
	SNV	5	87.50	80.0
	SD	3	95.00	70.00
SVM	Raw	5	75.00	75.00
	MC	4	68.75	60.00
	MSC	3	100.00	100.00
	SNV	5	90.00	85.00
	SD	3	32.50	20.00

Table 3.4 Biochemical parameter examination results for four African countries

Parameter	Country	Ghana	Cote d'ivore	Nigeria	Uganda
Moisture (%)	Average	7.738	8.333	7.077	7.219
	SDV	0.080	0.119	0.121	0.113
Fat	Average	43.456	37.458	38.742	38.810
	SDV	0.494	0.087	0.091	0.113
Fermentation index (%)	Average	1.031	0.875	0.857	0.854
	SDV	0.030	0.014	0.008	0.009
pH	Average	5.055	5.374	5.489	4.971
	SDV	0.198	0.066	0.065	0.034
Polyphenols (mg g^{-1})	Average	60.691	66.637	64.684	73.599
	SDV	0.701	1.515	0.624	0.910

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Conclusion

This work revealed, for the first time, that handheld near-infrared spectroscopy coupled with the appropriate multivariate classification technique in chemometrics could be used for rapid and non-destructive classification of cocoa beans. The systematic selection of different preprocessing methods (MC, SNV, and MSC) with PCA and modeling with LDA and SVM multivariate calibration models showed that LDA or SVM showed superiority in identifying cocoa beans from seven regions of Ghana with a 100% classification rate for all categories of cocoa bean samples studied. On the other hand, LDA, and SVM gave a 100% excellent classification rate for cocoa beans from four African countries. Generally, it could be concluded that hand-held spectroscopy together with appropriate multivariate techniques could be exploited for fast and nondestructive detection of cocoa bean authenticity and traceability. Furthermore, there is a high potential for this model to be imported into smartphones for effective quality control measurements in the cocoa industry. Additional studies are needed for the estimation of quality attributes in cocoa beans.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER FOUR

APPLICATION OF PORTABLE NEAR-INFRARED SPECTROSCOPY FOR CLASSIFYING AND QUANTIFYING COCOA BEAN QUALITY PARAMETERS

ELLIOT K. ANYIDOHO, ERNEST TEYE, ROBERT AGBEMAFLE, CHARLES L. Y. AMUAH, VIDA GYIMAH BOADU

Department of Agricultural Engineering, School of agriculture, College of Agriculture and Natural Sciences, University of Cape Coast
Department of Laboratory Technology, School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast

Published: Journal of Food Processing and Preservation. 2021, doi:10.1111/jfp p. 15445.

Statement of Contributions

Elliot K. Anyidoho (Candidate)

Writing and compilation of original manuscript, established methodology, data collection and formal analysis, funding, preparation of tables and figures

Ernest Teye (Principal Supervisor)

Conceptualization, funding, supervision and co-author of manuscript

Robert Agbemafle (Co-Supervisor)

Supervised manuscript compilation, editing and co-author of manuscript

Charles Lloyd Yeboah Amuah (Collaborator)

Software, formal analysis and co-author of manuscript

Vida Gyimah Boadu (Collaborator)

Funding and co-author of manuscript

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Introduction

Cocoa bean is the seed obtained from the pod of a tree crop botanically referred to as *Theobroma cacao* L. It is a famous cash crop with beneficial nutritional and valuable medicinal characteristics. Cocoa has fascinated the world for so many years and is a big component of chocolate and other confectionary foodstuffs. Some of the valuable medicinal benefits of cocoa include enhancement of brain and heart health, reduction in inflammation, control of weight and blood sugar, and development of healthy skin and teeth. Studies show that cocoa bean contains polyphenols, flavonoids, and procyanidin and it is an intensive source of antioxidants properties that can reduce blood clots and the risk of stroke and cardiovascular attacks (Di Castelnuovo et al., 2012). Quality valuation is of countless relevance for producers, manufacturers, and buyers (Atlas, 2010).

The two most important postharvest activities after harvesting cocoa pods are fermentation and drying. These activities involve the scooping of cocoa beans which are implanted in a white, sweet mucilaginous pulp out of the broken cocoa pod and then fermented and later dried. These postharvest activities significantly influence cocoa beans' quality, flavor, and aroma. For instance, although the optimum fermentative quality and pH of cocoa beans are obtained at 6 days of fermentation, fermentation index (FI) and pH have a direct correlation; thus, pH rises with cocoa beans fermentation duration (Afoakwa et al., 2012). Studies showed that cocoa beans with FI of one (1) and above are fermented well (between 5 and 6 days) and such beans are of good quality

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

attributes and show a good characteristic aroma with optimum cocoa bean pH (Kongor et al., 2013). The fermentation process results in the death of cotyledon and improves the consumer acceptable quality (Afoakwa, 2010). Therefore, importers and processors constantly look for well fermented, dried, and unadulterated cocoa beans. According to Minifie, no manufacturing processing could modify blunders committed during these two all-important postharvest handlings (Minifie, 2012). Unfermented or partially fermented cocoa beans will result in bitterness and astringency with no chocolate flavor and aroma (Afoakwa, 2016). On the other hand, drying prevents mold growth, to prevents unfriendly flavor in the cocoa beans (Afoakwa, 2016). Therefore, cocoa beans fermentation and drying work in synergy for the attainment of high-quality beans.

The quality of cocoa beans is therefore vigorously examined by cut-test procedure or sensory estimation by a trained panel (Aculey et al., 2010). The cut-test involves cutting cocoa bean lengthways in halves and evaluating its interior color using a score grounded on brown and purple beans (Amoa-Awua, 2014). Visual examinations or sensory estimation analyses are subjective and frequently not consistent as a result of human errors arising from the tiredness or disposition of the evaluator. On the other hand, direct and indirect procedures are mostly used in the determination of cocoa bean moisture content; nonetheless, the complexity and time-consuming process make it difficult for onsite usage.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Furthermore, the analytical approaches for evaluating cocoa beans are cumbersome, boring, time-consuming, costly, and destructive and need thorough sample preparation (Ramli et al., 2001). To circumvent these disadvantages associated with the traditional wet chemistry analytical methods, recent studies have focused on developing precise, sensitive, rapid, and nondestructive analytical approaches founded on NIR spectroscopic technology. Owing to the capability of the NIR spectroscopic technique to deliver a typical fingerprint of a specific sample, it provides the basis for its application in the agri-food industry for an effective quality control tool for quality detection in cocoa beans. It has been applied in the estimation of major nutrients, such as fats, proteins, carbohydrates, and moisture, and other minor functional compounds, such as catechins, organic acids, and theobromines. Generally, the application of NIR spectroscopic techniques for the evaluation of cocoa beans and cocoa bean product evaluation has been well-reviewed (Teye et al., 2020). Teye and coworkers (Teye et al., 2013, 2016) used NIR spectroscopy to differentiate Ghana cocoa beans and identify different varieties of cocoa bean respectively. For quantitative examination, NIR spectroscopy has been used; Whitacre and co-workers (Whitacre et al., 2003) determined the predictive analysis of cocoa procyanidins; Hue and others (Hue et al., 2014a) estimated ammonia nitrogen content and fermentation levels in cocoa, Krähmer et al. (Krähmer et al., 2015) in the determination of biochemical quality parameters such as protein, free amino acids, carbohydrates, organic acids, methylxanthines, and phenolic substances, and Hashimoto et al. (Hashimoto et

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

al., 2018) for the prediction of total phenolic compounds, fat, moisture, acidity, shell content, theobromine, and caffeine. Also, other authors deployed NIR spectroscopy for simultaneous measurements in the compositional characterization of different cocoa beans varieties (Barbin et al., 2018), whereas it was also used for the detection of adulteration of cocoa powder with cocoa shell and carob flour (Quelal-Vásconez et al., 2018, 2019).

All these applications employed the use of laboratory-based NIR spectroscopy with little or no considerations for portable NIR application, especially in sub-Saharan Africa. Although the laboratory-based NIR spectroscopic technique provided fast, nondestructive, and reliable results, they are often stationary and not applicable for onsite use. Furthermore, though its application holds a brighter future for developing countries such as Ghana and other cocoa-producing countries owning to the factor that these countries are challenged by the availability of laboratory infrastructure, their use is not readily available. Therefore, miniaturized NIR spectroscopy into portable devices provides the needed solution to make this aforementioned novel tool beneficial to developing countries where the majority of the global cocoa beans are produced.

Portable spectroscopic systems provide an extra advantage as an ideal device for in situ assessments of agri-food quality. Other researchers have used portable NIR spectrometers to determine fruit quality components in pineapple (Amuah et al., 2019) and mangoes (Marques et al., 2016).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Also, hand-held spectrometers have been used for predicting numerous quality constituents in intact oranges (Cayuela & Weiland, 2010). However, no investigator has examined the possibility of using a portable NIR spectrometer coupled with chemometric analysis for simultaneous classification of cocoa bean fermentation durations as well as estimation of FI, pH, and moisture content in cocoa bean samples. Therefore, the aim of this research was to assess the possibility of using a portable NIR spectrometer coupled with multivariate calibration for quick and non-destructive authentication and measurement of cocoa beans fermentation durations, FI, moisture content, and pH. For this investigation, multivariate algorithms such as LDA and SVM were deployed for the qualitative analysis, whereas partial least squares regression (PLS-R) was employed for quantitative estimation of quality parameters. Several researchers have also tried these multivariate techniques for developing reliable models for the classification and estimation of quality parameters of several foodstuffs.

Materials and Methods

Cocoa pod harvest and collection

Ripe and semi-ripe cocoa pods were harvested from a representative population of trees from a local farmer in the Western South Region of Ghana (Latitude: N 5° 11' 41", Longitude: W 2° 42' 36", Altitude: 14m). A total of 750 pods were selected and pretreated by storing them at an ambient temperature of 25°C - 30°C and relative humidity: 70% - 80% for two days. The stored pods were then split open. There were five treatments, and each treatment comprised

Digitized by Sam Jonah Library

of beans scooped from 150 pods which weighed 6.6kg. The beans were fermented by an expert in Ghana COCOBOD according to the treatments: 0, 48, 96, 144, and 192 hours. The beans were then sun dried separately according to their treatment. Each sun dried cocoa bean samples were collected in well-marked jute bags and transported to the School of Agriculture Laboratory, University of Cape Coast for further investigation.

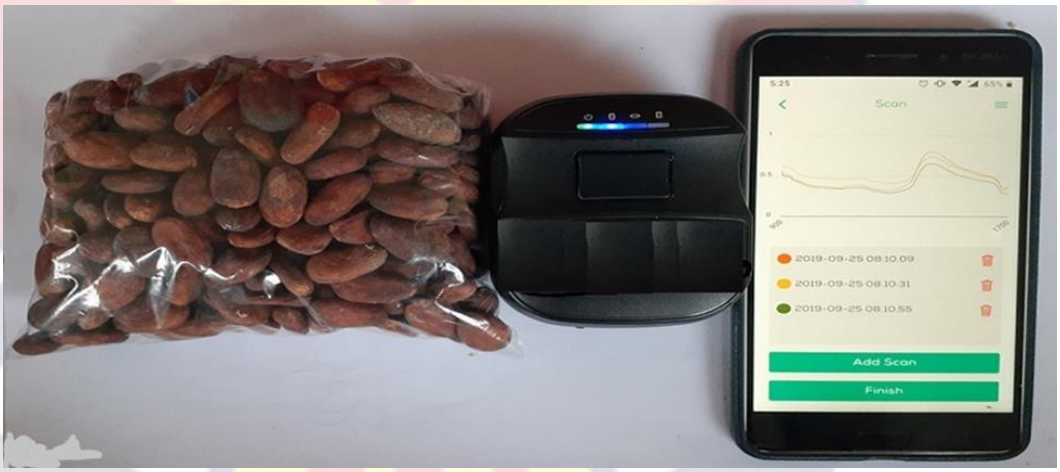


Figure 4.1. Set-up of scanning using handheld NIR spectroscopy and mobile phone

Instrument and spectra information

A handheld NIR spectrometer (Tellspec Ltd., UK) operated via a smartphone (Samsung A20) was deployed for scanning the whole cocoa bean samples (Figure 4.1; set-up of scanning) in a clear zip-locked polythene bag at three different spots. For each bag, three different portions were scanned independently after rotating it at 90°. The scanning was done at a temperature of 28 °C with a relative humidity of 65%. The NIR spectral information was acquired in a wavelength range of 900 to 1700 nm at 1 nm resolution (801

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

points). The portable spectrometer provided the measurements in proportionate absorbance parts (log 1/R), where R is the reflectance.

Reference examinations

The wet chemistry analysis for the examination of cocoa bean quality parameters adopted in this investigation involved destructive processes. Immediately after scanning, the cocoa bean samples were examined by using wet chemistry methods (International Office of Cocoa & Confectionary, 1996). Each parameter; fermentation index (FI), pH and moisture content was measured with three independent replications.

Moisture

Moisture content of the cocoa beans was determined according to the method used by others (Maalekuu & Teye, 2017). Two (2) grams of cocoa beans sample were weighed and placed into aluminum foil. The sample was dried in an oven at 110 ± 2°C for 24 hr, and it was then cooled in a desiccator and reweighed after 30 min. This process was repeated till it reached a constant weight, and the calculation of cocoa bean moisture content was achieved by using Equation (4.1).

$$MC_{Wb} = \frac{W_i - W_d}{W_i} 100\% \quad (4.1)$$

Where: MC_{wb} Moisture content on a wet basis (%), Wi is initial mass of bean sample (g), W_d is dried mass of bean sample (g).

Fermentation index (FI)

The FI examination was carried out by following the procedures laid down by Sunoj and co-workers (Sunoj et al., 2016). 0.5 g of milled sample of

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

cocoa beans was weighed into a 125 ml conical flask. 50 ml of methanol: hydrochloric acid in the ratio of 93:3 v/v was added and the mixture was kept in a refrigerator at 8°C ± 2°C for 18 hrs. Then the extract was sifted by the use of Whatman No. 1 filter paper. The absorption of the filtrate was collected at wavelengths; 460 nm and 530 nm using UV/Vis spectrophotometer (UVmini-1240; Shimadzu Corporation, Kyoto, Japan). The estimation of FI was obtained by using the ratio of absorbance at 460 nm and 530 nm. From equation 4.2 , FI above one (1) is considered well fermented. If FI is less than one (1) it is taken as partially-fermented.

$$FI = \frac{Absorbance\ at\ 460\ nm}{Absorbance\ at\ 500\ nm} \quad (4.2)$$

pH

The pH for the powdered cocoa bean was examined per the method proposed by the International Office of Cocoa & Confectionary, (1996). Five (5) grams of previously powdered samples were poured into conical flasks and 50 ml of boiling distilled water was added. The mixture was then allowed to stand for 30 minutes. Whatman filter paper No. 4 was used to filter the extract which was then allowed to cool to a temperature of 27°C. The pH value of the resultant filtrate was measured by using a digital pH meter (Jenway Model 3510, Wagtech Bibby Scientific Limited, UK).

Software tool

The recordings of spectra information stored in cloud-based data with their matching reference values were downloaded using a research license granted by Tellspec and imported to Matlab (version 9.8.0.1323502) with

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

windows 10 Pro software package for data processing involving all pretreatment and multivariate procedures. In this work, the calculation of the classification multivariate models (i.e., LDA & SVM) and regression models (i.e., PLS-R) was achieved by employing statistics and machine learning toolbox in Matlab, while the mean values, minimum values, maximum values, standard deviations, and correlation coefficients of the quality components estimated via wet chemistry analysis were analyzed using Microsoft Excel 2019 (Microsoft Corp., USA).

Data partitioning

The spectral data set for the different fermented cocoa beans (412 samples) was downloaded and partitioned into two subsets calibration set and prediction set. For the 412 samples, 290 samples and 122 samples were selected as calibration set and prediction set, respectively. The calibration set was used to develop the model, while the prediction set was used to evaluate the actual predictability of the developed model. To avoid bias in subset selection, the individual samples in each set were selected randomly into the two subgroups: calibration set and prediction as done by other authors (Teye et al., 2014b). Thus, for every five samples, three spectra with their corresponding reference data were randomly chosen as the calibration set, whereas the remaining sample was taken as the prediction set. The outcome was summarized in Table 4.1.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 4.1. Destructive measurements of FI, pH, and M% in the training and prediction sets

Quality parameter	Dataset group	No. of samples	Min.	Max.	mean	SDV
FI (%)	Training set	290	0.35	1.09	0.81	0.25
	Prediction set	122	0.36	1.07	0.80	0.25
pH	Training set	290	4.78	5.88	5.33	0.31
	Prediction set	122	4.88	5.77	5.34	0.30
Mc (%)	Training set	290	5.64	29.13	11.25	5.82
	Prediction set	122	5.78	28.34	10.53	4.86

Abbreviations: FI, fermentation index; Mc, moisture content; SDV, standard deviation.

Principal component analysis

Principal component analysis (PCA) was deployed to detect probable groupings in the numerous sensor data set. The PCA being an unsupervised pattern recognition technique extracted information from correlated matrix to visualize possible data leanings in a dimensional scatter plot. The unsupervised characterization terminology was used to show that the samples were categorized with no previous information, except the sensor indications. In the PCA analysis, the data sets together with spectra were transformed into a small number of unrelated but interpretable variables termed principal components (PCs). Similar samples were congregated nearer to each other and the other way round. The graphical outline of PCA outcomes provided a preliminary output for the determination of possible variances and similarities in a data set.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Multivariate systems

Following the careful selection of the best spectral pretreatment approaches, different multivariate characterization and quantification techniques were systematically studied, and selecting the suitable type is highly important in subsequent analyses. In this investigation, the multivariate systems deployed were support vector machine (SVM) and linear discriminant analysis (LDA) for the identification problem, whereas the regression tool, PLS-R, was used for the quantification problem.

Classification techniques

LDA is a linear and parametric supervised pattern recognition method that concentrates on noticing ideal borders among groups. LDA was applied to detect the linear combination of characteristics, and the resultant combination was used as a linear classifier. It performs its function by discovering the linear combination of characteristics that comes out with the ratio of between-class variance and minimizes the ratio of within-class variance. Some researchers used FT-NIR spectra data and LDA for the differentiation of Ghana cocoa beans (Teye et al., 2013). On the other hand, SVM is a nonlinear supervised technique that performs its function by finding the optimum border of two groupings in a vector space that is independent of the probabilistic provisions of vectors in the calibration set. SVM can generate a hyperplane that permits linear division in the advanced dimension feature space once the linear frontier in the low-dimension input space is inadequate to separate the two groups. This transformational technique changes data from a low-dimension input space to a

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

high-dimension feature space. For SVM as soon as the groupings are linearly parted, it optimizes a hyperplane boundary that splits both groups of the calibration set and unidentified sample. Nevertheless, if the groups are divided by nonlinear boundaries, the kernel function is employed to detect the boundary through mapping the inseparable data into a higher dimensional space and as a result of this, the classes become linearly separated. For more information, refer to Berrueta et al. (2007), which gives more details about SVM theory.

Quantification technique

In this experiment, simultaneous estimation of FI, pH, and moisture content were constructed by deploying the PLS-R technique. PLS-R is a classical estimation and linear multivariate regression technique that works on the entire spectrum. For this research, the PLS-R was performed on the whole spectrum of 900-to 1,700-nm wavelength range for all the samples used. This technique is the most extensively applied regression modeling technique in spectra data analyses as a result of its credibility and flexibility in handling redundant spectra data. PLS-R capitalizes on the variables comprising the most comprehensive data as well as recognizing noise via disintegrating and cleaning the spectra data; this will then be able to discover the linear combination of spectra data and the biochemical components. The flexible nature of PLS-R enables it to possibly establish a regression model even in situations where the quantity of samples is lesser than the number of variables (Xie et al., 2018).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Models' performance evaluation

The classification techniques (LDA & SVM) were evaluated based on the classification rate calculated by Equation (3). The closer the classification rate to 1% or 100% the better. The PLS regression models were optimized using the leave one out cross-validation principle, and the evaluation of model performance was done based on the correlation coefficient (*R*), root mean square error of prediction (RMSEP), root mean square error of calibration of cross-validation (RMSECV) and bias. The parameters used to evaluate (CR, R, RMSEP & RMSECV) were calculated using Equations (4.3)–(4.6) as done by other authors (Huang et al., 2014).

$$CR = \frac{N1}{N2} \text{ x } 100 \tag{4.3}$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{n}(y_{\backslash i} - y_i)^2}{n}} \tag{4.4}$$

$$RMSEP = \sqrt{\frac{\sum_{i}^{n}(y_i - y_i)^2}{n}} \tag{4.5}$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{4.6}$$

Where: CR = Classification rate (%), N1 = number of correctly identified samples, N2 = number of samples in the calibration or prediction set, n = the number of samples, y*i* = the reference measurement results for sample i, ŷ*i* = the estimated result for sample i when the model is constructed with sample i removed, ŷ*i*= the estimated results of the model for the sample i, ȳ = the mean of the reference measurement results for all samples.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Results and Discussion

Spectroscopic data presentation

For extraction of vital and precise sample information from the raw spectra which contained background data noise from the beneficial characteristics of the scanned samples, the NIR spectra data set was pretreated. Also, spectra pretreatment was deployed to minimize various influences such as shell thickness and skin disorder of the cocoa bean, which enabled and improved spectra classification (Jha & Garg, 2010).

In this work, the raw spectroscopic data collected using the portable NIR spectrometer were preprocessed independently by using first and second derivatives (FD and SD), mean centering (MC), standard normal variate (SNV), multiplicative scatter correction (MSC). Figure 4.2 depicts the raw spectra profile and pretreated spectra techniques for all the cocoa bean samples. These pretreatment techniques were carried out to ensure that the outcomes from the models were based on biochemical thumbprints from the spectra data acquired. The derivative techniques were deployed for baseline correction and resolution enhancement. The FD chemometric technique was deployed to eliminate standardization problems from the original NIR spectra data. The SD transformation technique is applied to separate superimposed peaks, enhance resolution, remove additive, and multiplicative baseline variations in the raw NIR spectra. The MSC is an excellent and powerful technique that is applied to correct additive baseline variation (i.e., vertical variations of the baseline) and multiplicative baseline variation (i.e., the inclination of the baseline). Generally,

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

MSC essentially improves the linearity and corrects offset in NIR spectroscopic data. The standard normal variant (SNV) is similar to MSC; it is a transformational spectral treatment technique usually employed to eliminate the multiplicative interferences of scatter, particle size and the change of light distance. In SNV, each individual spectrum is normalized to 0 mean and unit variance (Burns & Ciurczak, 2007). Figure 4.2 depicts the raw spectra profile and pretreated spectra techniques for all the cocoa bean samples. It revealed characteristic unique features.

From Figure 4.2, the abscissa denotes wave bands from 900 to 1,700 nm, while ordinate signifies the absorption values of the spectral. From the figure, the spectra share similar absorbance patterns in the range used. However, there exist numerous different information from the spectral profile. The wavelength range used corresponds first overtone and second overtone region. In these regions, there exist first overtone S–H, first overtone C–H, first overtone OH as well as second overtone C–H and second overtone N–H (Ozaki et al., 2006). Furthermore, these overtone regions reveal some useful functional groups such as CH_3, CH_2, CH, RNH_2, HOH, and ROH which are key constituents of water, pH (acidity), phenols, and cellulose of cocoa. FI, pH, and moisture content of cocoa beans are organic molecules which comprise C–O, C–H, O–H, and C–C bonds, and NIR spectroscopy can be employed to assess these molecules (Cen & He, 2007). Hence, with the help of chemometric techniques, prediction models were developed from the fingerprint of the cocoa beans spectra data set.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

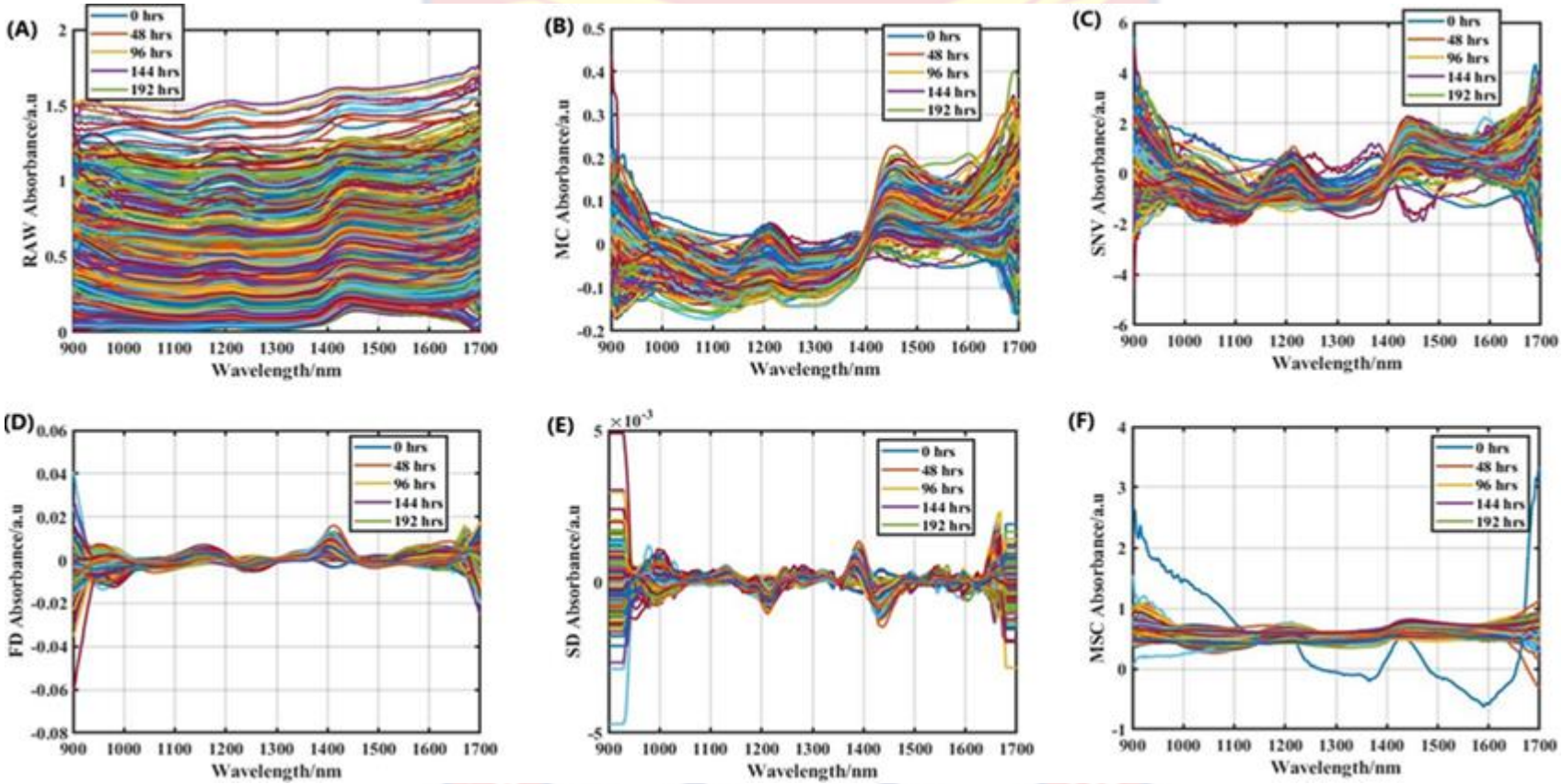


Figure 4.2. Raw spectra (A) and MC (B), SNV (C), FD (D), SD (E) and MSC (F) pretreated spectra for different fermentation days

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

PCA analysis for fermentation duration

In this study, PCA was deployed after pretreating with several preprocessing techniques independently to identify any cluster tendency trends in the cocoa bean samples used based on fermentation duration. The PCA results from the fermentation duration showed that the topmost first three PCs provided some useful groupings in all the cocoa bean fermentation durations investigated. Among the pretreatment methods, MSC was the best and superior to the others. This was because it gave clearer clustering trends along the first three PCs plane as seen in Figure 4.3 (F). It could be observed that well-ordered clustering of cocoa beans from different fermentation duration (0, 48, 96, 144, and 192 hrs) was detected by PCA + MSC, whereas a weak grouping was observed by the others. This occurrence could be attributed to the ability of the MSC technique to correct the spectroscopic scattering effect and eliminate undesirable scattering results from the spectral data matrix before modeling. MSC technique works by circling each spectrum and then modifying the spectrum to find similarities to the standard spectrum (Coronel-Reyes et al., 2018). Also, in PCs analysis, PC1 describes the greatest variability, while PC2 and PC3 explain the remaining variability. Again, in Figure 4.3 (F), It was observed that there was three neat groupings: group one (192 fermentation-duration), group two (unfermented cocoa, 48 fermentation duration, 96 fermentation duration), and group three (144 fermentation duration). This revealed that the recommended fermentation duration of 144 hrs (6 days) separated from the others. This phenomenon could be useful for the initial clustering of the recommended

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

fermented cocoa beans for quality assurance purposes. PCA is an unsupervised data description and dimension reduction technique. There thus was the need to apply various classification algorithms comparatively.

Classification model

Machine learning techniques such as LDA and SVM were experimentalized towards the development of classification models for characterizing cocoa bean samples according to their fermentation duration. The cocoa samples used in this study were made up of 0, 48, 96, 144, and 192 hrs of fermentation groupings. From the classification results (Table 4.2) it could be observed that the resultant models MSC + PCA + LDA and MSC + PCA + SVM gave a 100% overall classification rate both in the training set and prediction set. This means that the classification of the cocoa bean according to their fermentation duration is practically feasible by using portable NIR spectroscopic technology. These outcomes have established that the technique could be automated and incorporated in mobile phone to aid quality control.

More importantly, SVM performed slightly better than LDA as it used only three PCs (Table 4.2). The advantage of using SVM technique in solving nonlinear classification problems cannot be ignored as exemplifies structural risk reduction value where the upper bound is diminished on the anticipated risk (Chen et al., 2009). Also, it is well-known that cocoa beans of different fermentation duration are multifaceted mixtures that are interrelated, and a nonlinear classification model such as SVM could be superior and very efficient.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

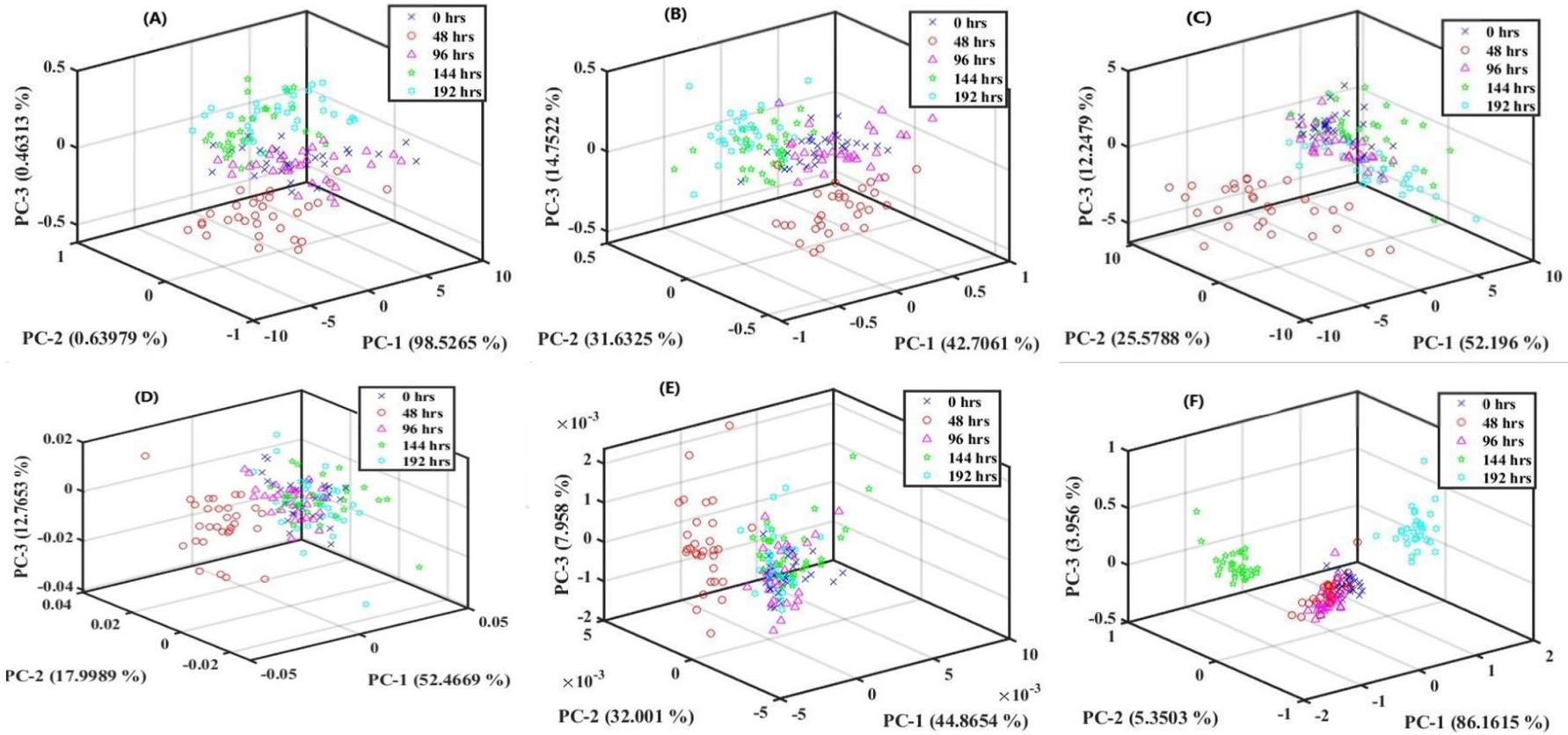


Figure 4.3. PCA cluster plot of unfermented, and fermented samples at different days. Raw (A), MC (B), SNV (C), FD (D), SD (E) and MSC (F).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

The performance of LDA was very good, but at a high number of PCs, the model's effectiveness is influenced. From Figure 4.4, the mean spectra profile revealed a clear separation at certain wavelength ranges. These spectroscopic ranges contained very beneficial information that contributed to providing accurate classification results. From the figure, the major peaks observed to have contributed to the classification of the cocoa bean samples are observed at 1,150–1,250 and 1,400–1,500 nm. These essential absorption bands are closely correlated to the biochemical compositions and internal characteristics of cocoa beans which provided the fingerprint for accurate classification. For example, these significant absorbance bands are associated with the chemical group of O–H, C–H, C–H_2, CH_3, and NH_2 functional groups, These groups are characteristics of complex carbohydrates, water, aliphatic alcohols, fats, polyphenols, acidic, and aromatic compounds in cocoa bean (Workman& Weyer, 2012). Furthermore, the good results by the model can be the effect of nutritional compositions and complex flavonoids in unfermented, fully fermented, over fermented, and partially fermented cocoa beans. For example, fully fermented cocoa bean is brown with high chocolate aroma and taste, low in starch and have low epicatechin content, whereas unfermented cocoa bean is gray, coupled with high astringency and bitterness and high polyphenols with a pH value above 5.5 (Aculey et al., 2010; Aikpokpodion & Dongo, 2010). Also, the recommended duration of fermentation is 6 days (144 hr) and as it possesses useful biochemical and health properties (Kongor et al., 2013; Sunoj et al., 2016).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 4.2. Classification models based on fermentation duration

Models	Pretreatment	No. of PCs	Training set (290)	Prediction set (122)
LDA	Raw	4	78.33	80.00
	MC	5	88.33	80.00
	MSC	5	100.00	100.00
	SNV	5	82.50	83.33
	FD	2	56.67	40.00
	SD	5	73.33	80.00
SVM	Raw	4	65.83	56.67
	MC	5	68.33	60.00
	MSC	2	100.00	100.00
	SNV	4	84.17	80.00
	FD	3	22.50	10.00
	SD	3	22.50	10.00

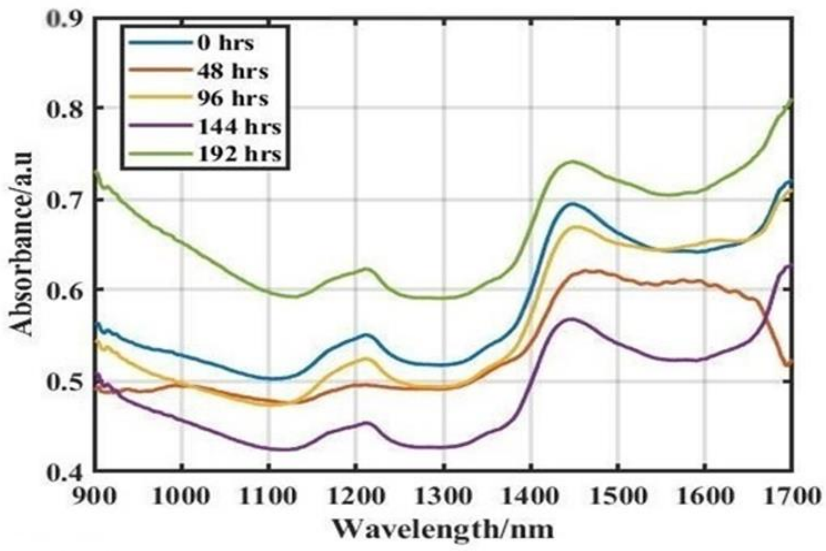


Figure 4.4. Mean spectra profile of cocoa bean at different fermentation duration.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Cocoa quality by reference measurement

The conventional reference methods applied in this investigation provided the results for the quality parameters such as FI, pH, and moisture content in cocoa bean samples. These parameters are recognized by the cocoa industry as crucial cocoa quality components. There was a wide range of FI, pH, and moisture content in the samples studied and is in agreement with other research findings (Apriyanto & Harmayani, 2016; Teye et al., 2015) and also presented a factual picture of cocoa bean samples in the market place. Similarly, the samples' reference estimation outcome for FI, pH, and moisture content in the calibration set encompass the range in the prediction set.

FI is a quality component usually evaluated in a research laboratory. The fermentation degree at the industrial level is correlated with variances in the biochemical component. For instance, fermentation brings huge variations in the cocoa polyphenol profile. On the other hand, the estimation of fermentation degree is normally done by evaluating the color changes inside the cocoa bean by applying the cut-test procedure. In this study, FI ranged from 0.354 to 1.087, and a mean FI of 0.806 with a standard deviation of 0.247. These outcomes are beneficial for real applications, as it covers a wider representative range. According to earlier authors, FI above one (1) is considered fully fermented and below one (1) is regathered unfermented (Gourieva & Tserevitinov, 1979). The unfermented sun dried cocoa beans have a very high pH value and polyphenol content (Afoakwa, 2016). For full fermentation, the fermentation duration of

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

144 hrs is required, and it is reported that increasing fermentation duration had s significant effect on the FI (Nazaruddin et al., 2006).

The evaluated pH of cocoa bean samples obtained had a minimum pH value of 4.78 for cocoa bean samples fermented for 144 hr, while the maximum pH value of 5.88 was recorded for unfermented cocoa bean samples. Afoakwa and coworkers explained that the reduction and the increasing tendency of pH during fermentation periods are a result of the conversion of pulp into alcohol and acid (Afoakwa et al., 2013b). The moisture content of the cocoa beans used ranges from 5.64 to 29.13. This range is wide enough to encompass the practical range of cocoa beans in the entire value chain. This further demonstrates the diversity of the amount of data being examined. Furthermore, this is very important as moisture content affects cocoa beans' storage life, quality, transportation, and selling price.

Quantification models

Quantification model for FI

FI provides a measure for cocoa bean quality. PLS model was used to find a correction between the spectral data set and the referenced measured value. From Figure 4.5 (A), the measured values correlated linearly with the NIR predicted values with some outliers which subsequently affected the PLS model. From Table 4.3, it could be observed that among the pretreatment methods used, SD spectral preprocessing was the best. Comparatively SD-PLSR gave optimum results of R = 0.885 and RMSEC = 0.183 in the calibration set, while R = 0.870 and RMSEP = 0.121 in the prediction set. This supports the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

accession that the influence of a unique preprocessing technique is vital for improving the PLS model. The results also prove that portable NIR spectroscopy could be employed for cocoa bean quality control evaluation. From Figure 4.5 (B), the performance of the model for FI estimation could be explained that the major peaks at 1,190, 1,212, 1,420, and 1,460 nm were found accountable for the prediction of FI in the cocoa beans. For instance, the peaks around 1,200–1,460 are related to CH_3, CH_2, and CH (Cen & He, 2007), which correlates to phenolics, acidity, and aromatic elements in cocoa beans, and these chemicals are enhanced during fermentation.

Table 4.3. PLS models for predicting FI, pH and MC of cocoa samples

Parameter	Pretreatment	Training			Prediction		
		R	RMSEC	Bias	R	RMSEC	Bias
FI (%)	Raw	0.585	0.199	0.018	0.618	0.194	0.035
	MC	0.601	0.197	0.018	0.694	0.177	0.032
	MSC	0.486	0.216	0.019	0.704	0.175	0.032
	SNV	0.506	0.213	0.019	0.715	0.172	0.031
	FD	0.639	0.189	0.017	0.716	0.172	0.031
	SD	0.885	0.183	0.017	0.870	0.121	0.022
pH	Raw	0.519	0.256	0.023	0.592	0.238	0.043
	MC	0.515	0.257	0.023	0.632	0.223	0.042
	MSC	0.362	0.279	0.025	0.648	0.225	0.041
	SNV	0.380	0.277	0.025	0.651	0.225	0.041
	FD	0.552	0.250	0.023	0.671	0.219	0.040
	SD	0.836	0.227	0.021	0.815	0.171	0.031
Moisture (%)	Raw	0.878	2.763	0.223	0.874	2.870	0.478
	MC	0.979	2.752	0.222	0.899	2.931	0.488
	MSC	0.385	5.317	0.429	0.575	4.840	0.806
	SNV	0.755	3.781	0.306	0.725	4.070	0.678
	FD	0.908	2.407	0.194	0.889	2.711	0.452
	SD	0.849	3.043	2.246	0.864	2.981	0.496

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

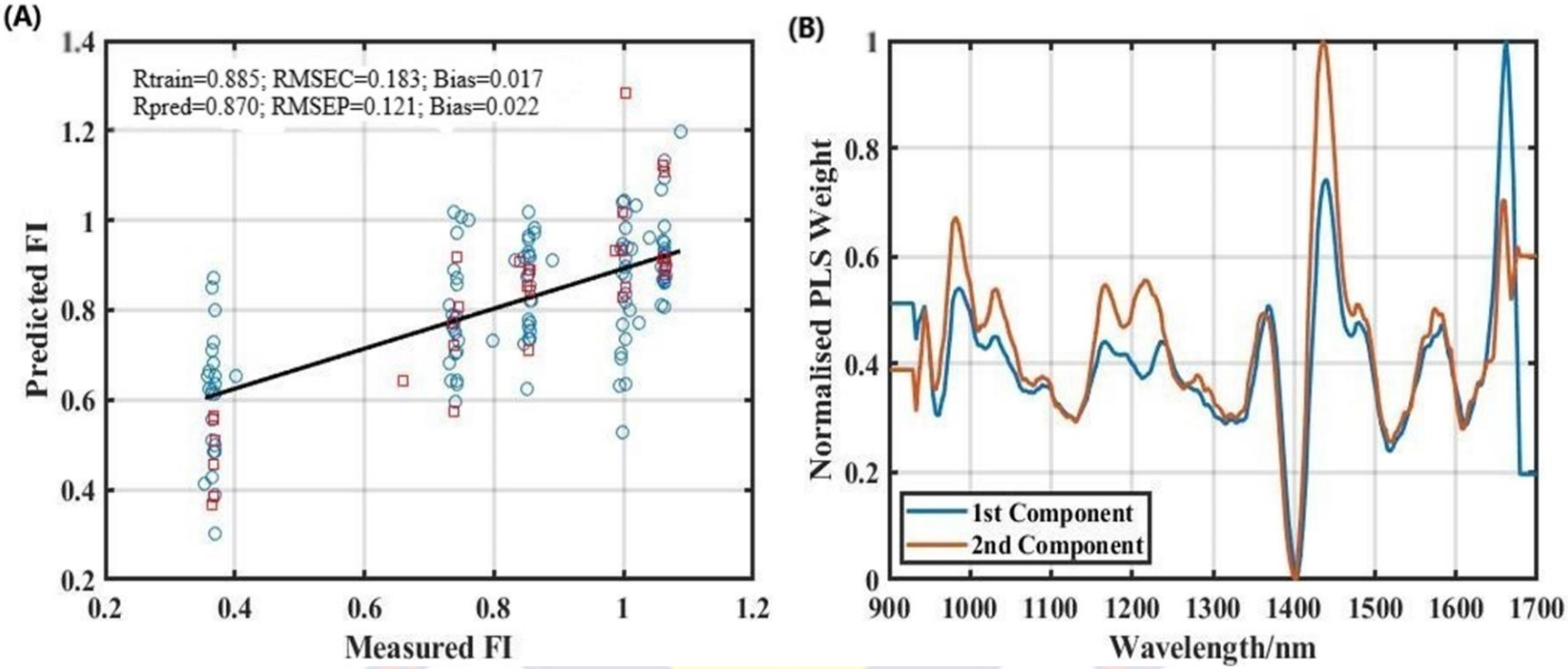


Figure 4.5. NIR-PLSR predicted versus measured FI (A) and (B) PLS loadings plot

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Quantification model for pH

The best performance obtained for predicting pH in cocoa beans was derived by SD-PLSR compared to the others as seen in Table 4.3. The results of R = 0.836 and RMSEC = 0.227 in the calibration set while R = 0.815 and RMSEP = 0.171 were obtained in the prediction set. These results were similar to those observed by other researchers (Krähmer et al., 2015). On the other hand, even though other authors recorded slightly better results for pH in cocoa beans, these researchers conducted their experiments with powdered cocoa bean samples or used a bigger laboratory-based NIR spectroscopy (Hue et al., 2014b; Krähmer et al., 2015; Teye et al., 2015). From Figure 4.6 (B), the measured values correlated linearly with spectral measurements. This proved that PLSR was useful as it is a well-known linear model. From Figure 4.6 (B), the absorption bands at 1,120, 1,180, 1,390, and 1,400 nm provided useful information for modeling pH (Cen & He, 2007) in cocoa beans. These peaks could represent organic acids. More specifically, the absorption bands within the range of 1,120 and 1,180 nm are associated with second overtone stretching of C–H groups, and rocking of O–H softly bonded water, whereas the range between 1,390 and 1,400 nm is associated with first overtone stretching of C–H groups (CH_3 and CH_2 stretching modes), and their combination corresponding to acidity which is related to pH (Cen & He, 2007; Ozaki et al., 2006; Sadler & Murphy, 2010).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

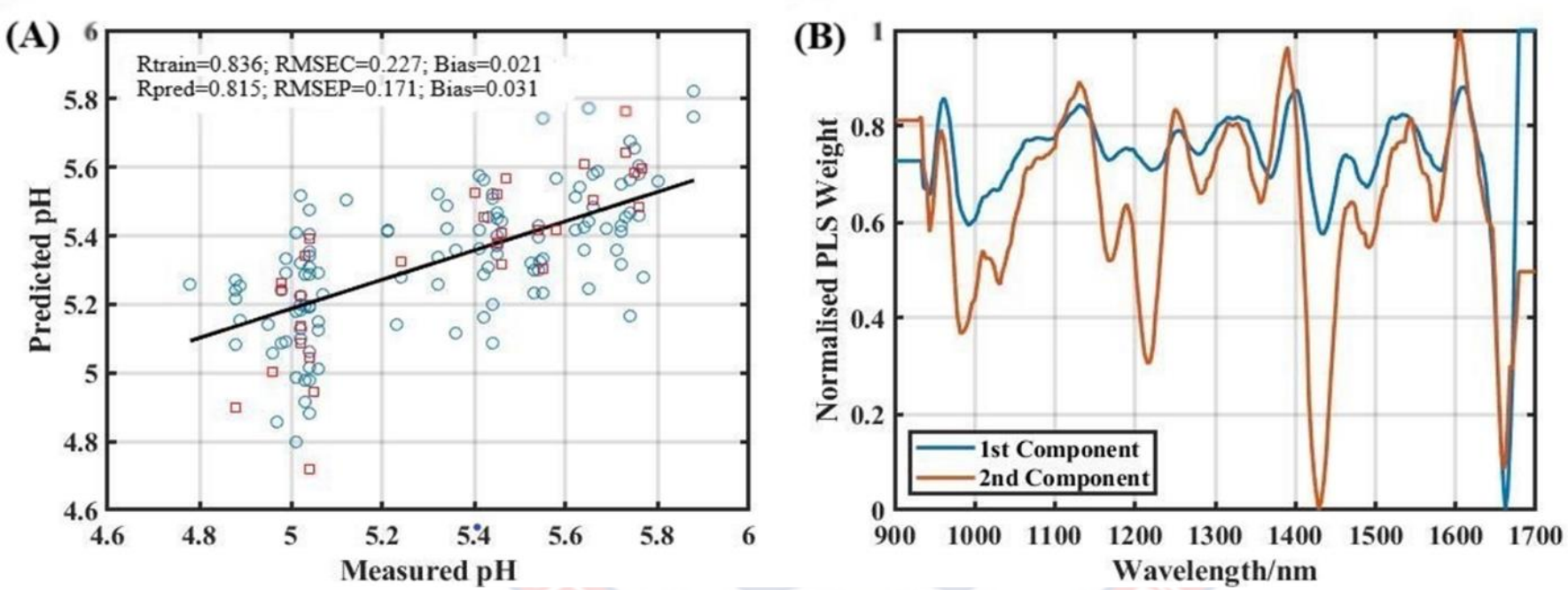


Figure 4.6. NIR-PLSR predicted versus measured pH (A) and (B) PLS loadings plot

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Moisture quantification model

Moisture content is vital for cocoa bean storage and conveyance. From Table 4.3, it could be observed that NIR spectral results correlated well with reference measured results. The results showed that the MC pretreatment technique plus PLSR was the most suitable for building a model for cocoa bean moisture content determination. From Figure 4.7 (A), a linear correlation was used to obtain optimum performance such that, the correlation coefficient (*R*) values of 0.979 and 0.899 were obtained in the training and prediction set respectively. Similarly, RMSEC value of 2.752 and RMSEP value of 2.931 were obtained, indicating that the values determined by the multivariate analysis model portrayed good agreement with the reference analysis. The corresponding bias values obtained for the training and prediction models were 0.222 and 0.488 respectively. To explain this phenomenon, it could be seen from Figure 4.7 (B) that the principal peaks between 1,420 and 1,492 nm are related to ROH and OH in the NIR absorption bands could be associated with moisture content in cocoa beans (Cen & He, 2007).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

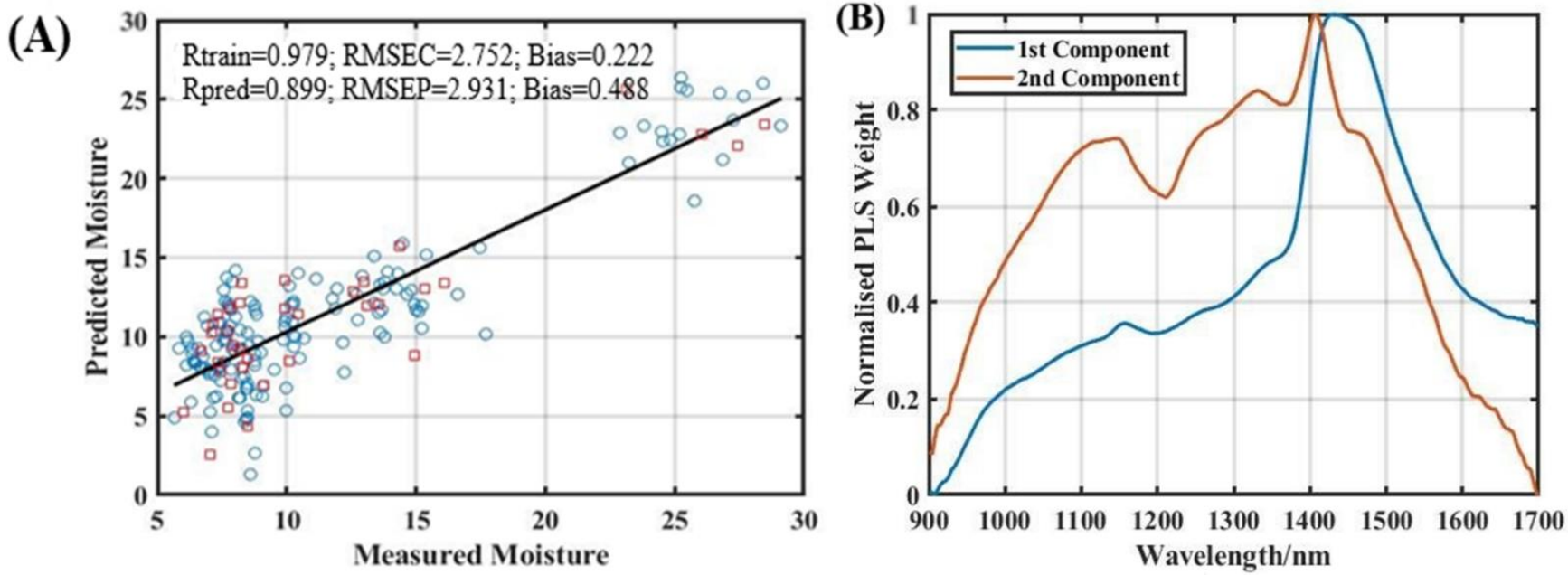


Figure 4.7. NIR-PLSR predicted versus measured moisture content (A) and (B) PLS loadings plot

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Conclusion

The study has shown for the first time that, cocoa beans of different fermentation duration and pH, FI, and moisture content can be determined non-destructively and rapidly by using portable NIR spectroscopy and multivariate data analysis with results between 100% for classification performance and 0.81–0.89 for prediction coefficient. The overall results have proven that portable NIR spectroscopy could be incorporated in the cocoa beans value chain especially in developing countries where the majority of production is done and are challenged with laboratory infrastructure. This phenomenon is the advantage of portable NIR over laboratory-based NIR spectroscopy. However, there is a slight trade-off in terms of prediction accuracy for quantitative parameters (between 81% and 89%) due to its narrow range. This, therefore, calls for further studies to improve the accuracy of portable NIR spectroscopy.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER FIVE

DIFFERENTIATION OF ORGANIC COCOA BEANS AND CONVENTIONAL ONES USING NIR HANDHELD SPECTROSCOPY AND MULTIVARIATE CLASSIFICATION TECHNIQUES

ELLIOT K. ANYIDOHO, ERNEST TEYE, ROBERT AGBEMAFLE

Department of Agricultural Engineering, School of agriculture, College of Agriculture and Natural Sciences, University of Cape Coast
Department of Laboratory Technology, School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast

Published: International Journal of Food Science. Volume 2021. Article ID 1844675. doi:10.1155/2021/1844675.

Statement of Contributions

Elliot K. Anyidoho (Candidate)

Investigation, writing and compilation of original manuscript, established methodology, data collection and formal analysis, preparation of tables and figures

Ernest Teye (Principal Supervisor)

Data curation, editing, supervision of manuscript compilation and co-author of manuscript

Robert Agbemafle (Co-Supervisor)

Editing, supervised manuscript compilation and co-author of manuscript

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Introduction

Several modern-day environmental challenges are inherent in agrifood schemes. These schemes are partly held accountable for the decrease in ecosystem destruction, water pollution, global warming, and biodiversity. Hence, the greening of agrifood production, processing and marketing could be an important contribution to quality, safety, and sustainability. The advent of post-Fordism has put environmental issues and quality matters at the heart of agri-food provisioning (Oosterveer et al., 2011).

The enhancement of sustainability performance in the cocoa industry is developing as a strategy within universal product value chains. In making the global cocoa chain and network sustainable, both private and public players have introduced many initiatives at different levels. The main driver of this trend is the emerging consumer demand for socially fair and eco-friendly products. For instance, sales of organic chocolate reached the USA$304 million in 2005, representing an increase of 75% in comparison to 2002 sales (Berlan & Bergés, 2013). Much attention has to be shifted to West Africa because it produces more than 70% of all cocoa and is the location of many organic initiatives. Ghana the second largest exporter of cocoa started the exportation of organic cocoa in 2005 to the global market. More than 20,000 smallholder farmers are currently involved in the organic cocoa network, as well as other stakeholders at national levels, such as non-governmental organizations, farmers' organizations, several public institutions, licensed buying companies, and importers. Inferentially, the most important bean category which influences and drives consumers'

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

preference, nutritional composition, quality, and safety is the organic cocoa beans category (Glin et al., 2015).

Organic cocoa beans unlike conventional ones are cocoa beans produced following the farming practices and principles that do not allow the use of growth-stimulating elements, herbicides, synthetic pesticides, and fertilizers (Glin et al., 2015). Concerns about growth stimulating elements, herbicides, synthetic pesticides, and fertilizers have given additional motivation to organic cocoa beans demand, as consumers progressively query the quality and safety of conventional cocoa beans. Relative to the aforementioned factors the demand for organic cocoa beans by chocolate producers and consumers has increased, and the production of organic cocoa beans is more lucrative due to the higher price it receives (Pay, 2009). The higher price for the organic label as compared to the conventional cocoa beans has led to mislabeling which is regarded as fraud to gain undeserved economic advantage. The international market, thus, call for trust tags for organically produced cocoa beans. Therefore, screening of organic cocoa beans before export, marketing, and processing to prevent mislabeling has become very necessary.

Currently, the techniques for ensuring the integrity and quality of organic cocoa beans are time-consuming, expensive, and require highly skilled personnel and are often not applicable in low-resource countries. They include chromatographic methods such as gas chromatography (GC) or high-performance liquid chromatography (HPLC) (Aculey et al., 2010), and various conventional approaches including wet chemistry (Afoakwa et al., 2013a). The use of handheld NIR spectrometer for ensuring the integrity and authenticity of

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

organic cocoa beans from conventional ones could provide a help. This would offer a rapid, and less expensive technique for the assessment of organic and conventional cocoa beans for quality control and assurance purposes.

NIR spectroscopy technique provides a non-destructive analytical tool, more especially for the assessment of chemical composition and physical quality characteristics of cocoa bean and cocoa products (Teye et al., 2019b). This is due to its sensitivity to OH, CH, and NH absorptions associated with cocoa bean components. The NIR spectroscopy has been used for the quantification of moisture content and fat of cocoa powder (Veselá et al., 2007), prediction of procyanidins in cocoa (Whitacre et al., 2003), differentiation of Ghana cocoa beans and cocoa bean varieties (Teye et al., 2013), verification of cocoa powder authenticity, classification and determination of chemical quality parameters (Barbin et al., 2018; Kutsanedzie et al., 2017) and estimation of cocoa bean parameters (Teye et al., 2015). A critical study of recent applications of the use of NIR spectroscopic technique in the cocoa bean industry showed that it has also been applied in the rapid detection of cocoa beans adulterations and fraud (Quelal-Vásconez et al., 2019; Quelal-Vásconez et al., 2018) and quality control of commercial cocoa beans (Hashimoto et al., 2018). Therefore, NIR spectroscopy offers a reliable alternative for the assessment of organic cocoa bean integrity and quality.

Additionally, advancement in NIR instrumentation has led to the miniaturisation of stationary NIR spectrometers into lightweight handheld spectroscopic devices that are simple, relatively less expensive, and provide extra speed. Their portability makes them ideal instruments for *in situ*

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

assessments of agricultural products. However, no studies have investigated the application of handheld NIR spectrometer for screening and ensuring the integrity of organic cocoa beans non-destructively. Also, no information is available on the application of different multivariate classification algorithms for effective and accurate discrimination of organic cocoa beans.

Therefore, the objective of this work was to use a handheld NIR spectrometer and chemometric multivariate techniques to non-destructively identify organic cocoa beans from conventional ones and simultaneously predict quality characteristics such as crude fat, total polyphenol content, total flavonoid content, and total antioxidant capacity of organic cocoa beans and conventional cocoa beans. Specifically, the study aimed to determine the ideal multivariate classification algorithm and effective variable selection algorithm for accurate differentiation and prediction of quality characteristics of organic and conventional cocoa beans.

Materials and Methods

Cocoa bean samples

A total of 120 organic cocoa bean samples ready for exportation were obtained from the Cocoa Research Institute of Ghana (CRIG) and Yayra Glover Limited, a licensed organic cocoa producing and marketing company in Ghana. Whilst 140 conventional cocoa bean samples were collected from the seven cocoa-producing regions of Ghana under the guide of the Quality Control Company and Cocoa Marketing Company of COCOBOD. The two categories of cocoa beans (organic and conventional) according to the producers were fermented for 6 days using heap protocols similar to those described by other

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

authors (Afoakwa et al., 2013c). The cocoa bean samples were well labelled and transported in marked jute bags to the Department of Agricultural Engineering Research Laboratory, the University of Cape Coast for further examination. Spectral measurements were taken on the whole cocoa beans whilst chemical examinations were conducted on the ground samples.

Sample spectral measurement

The handheld NIR spectrometer (Tellspec^{R}) was used to take the spectrum of each cocoa bean sample in an absorbance unit (log 1/R); R = reflectance. The NIR spectroscopic data set was developed in a wavelength range of 900-1700 nm. The instrument was operated using a smartphone application and spectroscopic data stored in the cloud remotely was downloaded onto the laptop. All the cocoa beans samples were scanned three times in a zip lock bag at different sides, and the spectrum for each sample was the mean of the three scans. Scanning of the samples was carried out at an ambient temperature of 25 ± 1°C with a humidity of 60%.

Reference measurement

The wet chemistry techniques adopted for the prediction of bean quality parameters were done using recommended methods after milling the samples into powder (500 mm) using a laboratory mill (Model Brook Crompton, Series 2000, Glen Creston Ltd, UK). Each chemical analysis was performed in three independent replications and the mean values were reported.

Fat measurement

Quality parameters such as crude fat content and crude fibre content were measured by using recommended techniques (AOAC, 2005). The crude

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

fat content was quantified based on the AOAC 963:15 method, which consisted of a Soxhlet extraction method.

Total polyphenol content (TPC)

The TPC concentration was assessed using the Folin-Ciocalteu reagent technique as previously defined by others (Arslan et al., 2017) with minor modification. 1 ml of bean extract was added to 5 ml of Folin-Ciocalteu reagent (ten-fold diluted) and 4 ml (75 g/l) of Na_2CO_3. The resultant mixture was allowed to incubate in absence of light for 2 hours. The absorbance of the mixture was obtained at 765 nm with a Shimadzu UV-VIS spectrophotometer (UNmini-1240, Shimadzu Corporation, Tokyo-Japan). Gallic acid solution (20-100µg/ml) was applied as a standard for the creation of a calibration curve. Results obtained were stated as milligrams gallic acid equivalents (GAE) per gram (mg GAE g^{-1}) of the cocoa bean.

Total flavonoid content (TFC)

TFC was measured by following methods used by other authors (Ordonez et al., 2006). The cocoa bean was extracted in 80% ethanol and kept for 24 hours. Briefly, 0.5 ml of 10% aluminum chloride ($AlCl_3$) and 2.5 ml of distilled water were added to 0.5 ml of cocoa bean extract and the reaction mixture was left to cool for 1 hour. The absorbance was recorded at 510 nm with a Shimadzu UV-VIS spectrophotometer (UNmini-1240, Shimadzu Corporation, Tokyo-Japan). The quercetin standard solution was applied to create the calibration curve and the required outcomes were stated as milligrams of quercetin equivalents (QE) per gram (mg QE g^{-1}) of the cocoa bean.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Total antioxidant capacity (TAC)

TAC of the cocoa bean sample was measured using the process by others (Sun et al., 2005). 1 ml of cocoa bean extract was mixed with 0.02 mg/mL 2,2-diphenyl-1-picrylhydrazyl radical (DPPH) solution. The glass tube containing the reaction mixture was cooled for 30 minutes to room temperature. Absorbance was determined at 590 nm with a UV-VIS spectrophotometer (UNmini-1240, Shimadzu Corporation, Tokyo-Japan). The ascorbic acid standard solution was applied to build the calibration curve and TAC was determined and stated as milligrams of ascorbic acid equivalent antioxidant content (AEAC) per gram (mg AEAC g^{-1}) of the cocoa bean.

Software tools

All preprocessing and analysis of the spectra data were performed using multivariate analysis software in Matlab version 9.6.0 (The MathWorks, Inc., USA) with windows 10 Pro software package for data treatment.

Dataset partitioning

The spectroscopic dataset obtained from 260 samples of organic and conventional cocoa beans was preprocessed with appropriate techniques. The spectral data obtained from the samples were randomly divided into two different datasets called: calibration set (spectroscopic data from 182 samples) and the prediction set (spectroscopic data from 78 samples). The calibration set which represented 70% of the data were used to construct the models whereas the remaining 30% of the data were used for the prediction set which was used to evaluate the predictive capability of the built models.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Theory of chemometric techniques

Spectral preprocessing approaches

The raw NIR spectra as shown in Fig. 5.1(a) contain beneficial, and non-useful information of the cocoa bean samples. This could be as a result of interferences from a scattering of light from the samples, temperature variations, and or background noises (Jha & Garg, 2010). Therefore, preprocessing of the dataset to acquire only the useful properties of samples whilst keeping the similarities and variations among the primary observations were adopted. To accomplish this, four spectral preprocessing approaches such as MC (mean centering), MSC (multiplicative scatter correction), FD (first derivative) and SD (second derivative) were comparatively employed in the Matlab version 9.6.0. MC is a spectral pre-processing approach is carried out by computing the mean spectrum of the dataset and deducting the mean from each spectrum. For correcting the scattered light on different particle size and multiplicative and additive effects MSC is mostly deployed (Rinnan et al., 2009a). FD pre-processing approach which is assessed as the difference between two consequent spectra measurement points eliminates baseline effects. SD transformation technique is applied to separate overlapped peaks, enhance resolution, remove additive and multiplicative baseline variations in the raw NIR spectra. Before the application of the SD preprocessing technique, the NIR spectra were smoothed using the Savitzky-Golay algorithm (Siesler et al., 2008). Generally, the Savitzky-Golay smoothing SD algorithm best improved the linearity and corrected offset in NIR data.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Principal component analysis (PCA)

Furthermore, the principal component analysis (PCA) was deployed on all the preprocessed NIR datasets to identify any cluster trend (to detect probable groupings). The PCA has been an unsupervised pattern recognition algorithm that extracted information from correlated matrices to see probable data leanings in a dimensional scatter plot. In the PCA analysis, the data sets coupled with spectra were converted into a small number of uncorrelated but explainable variables referred to as principal components (PCs). Similar samples congregated closer to each other and vice versa. The graphic profile of PCA results yielded initial output for the determination of possible variations and resemblances in a dataset. Usually, PC1, PC2, PC3, PC4, PC5, etc. explain and give relevant information in descending order.

Multivariate classification algorithms

RF

RF (Random Forest) is an ensemble procedure that is based on tree classifiers. It grows many classification trees to produce accurate discrimination. In RF each tree grows on an independent bootstrap sample obtained from the calibration sample/data (Svetnik et al., 2003). Classification of the new feature vector is achieved by classifying the input vector with each of the trees in the forest. A classification is given by each tree, often considered as that tree's vote for that class. The forest selects the classification with the maximum votes over all the trees in the forest (Kulkarni & Lowe, 2016). RF computations comprise two measures of variable importance (based on rough-and-ready measure and permutations) and measures of the resemblance of data

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

points that could be applied for graphical representation, multi-dimensional scaling, imputing missing values and clustering (Cutler et al., 2007).

KNN

KNN (K-nearest neighbours) is a non-parametric and linear learning algorithm where the distance between each of the samples of the calibration set and unknown sample is assessed and for more information refer (Thanh & Kappas, 2018). For the KNN approach, the parameter K has a huge influence on the classification rate of the K-NN model. The selection of K was optimised by computing the calibration ability with a preferably an old number of small K values. In this study, PCs were applied as input data in the KNN model. KNN model efficiency was examined by the number of parameters K and PCs (Berrueta, Alonso-Salces, & Héberger, 2007).

LDA

LDA (Linear discriminant analysis) is a linear and parametric supervised pattern recognition approach mostly applied to discover a linear combination of features and the resultant combination may be employed as a linear classifier. LDA concept is founded on the determination of linear discrimination functions that produce the ratio between-class variance and decrease the ratio of within-class variance. In the LDA approach, the classes are linearly separated and keep to a normal distribution (Chen et al., 2012a). Also, the LDA is viewed as PCA in which the number of PC (principal component) is key to the performance of the LDA classification model.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

PLS-DA

PLS-DA (partial least square discriminant analysis) is a linear differentiation technique that combines properties of partial least square regression with the discrimination presentation of a differentiation technique (Lee, Liong, & Jemain, 2018). The PLS-DA with k-fold cross-validation was deployed to screen out and differentiate organic cocoa beans from conventional ones, and to prevent overfitting of the calibration models. This combines the variables in the dataset to calculate factors that maximize the correlation value with the different classes. PLS-DA concurrently decomposes spectral and class matrices and extracts the spectral data most associated with the classes that can lead to the development of a reliable and accurate identification model (De Maesschalck & Van den Kerkhof, 2005).

Theory of different PLS algorithms

There are several quantitative multivariate data analysis techniques for assigning specific chemical components to spectroscopic features and the selection of spectroscopic variables of concern. For simultaneous assignment of specific biochemical components to spectroscopic features, a full spectral-based algorithm like partial least square (PLS), and spectral-variable selection procedures such as interval partial least square (i-PLS), synergy interval partial least square (Si-PLS), and genetic algorithm partial least square (GA-PLS) quantitative algorithms were attempted to measure crude fat, crude fibre, total polyphenol, flavonoid content and antioxidant capacity in cocoa beans.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Partial least square (PLS)

The PLS which is a classical multivariate technique and a full spectra-based quantitative algorithm was deployed to analyse the full spectrum. The PLS techniques use the spectroscopic information enclosed in the attained spectra and their matching reference data to establish the prediction model. Due to the complex relationship among samples and the spectroscopic data, the selection of full-spectrum could lead to the overfitting of the prediction model. The prediction accuracy and reliability of the PLS model are frequently influenced because full-spectrum comprises immaterial information other than the targeted variables enclosed in the determined spectra of the sample. Therefore, the selection of a suitable spectral range with various variables that relate to the targeted quality components is an important step towards the building of a precise prediction model. (Chen et al., 2008; Nørgaard et al., 2000) proposed interval partial least square (i-PLS) and synergy interval partial least square (Si-PLS) whilst (Kutsanedzie et al., 2017) stated that the genetic algorithm partial least square (GA-PLS) was most reliable for fungi count prediction in cocoa beans. Other investigators such as (Chen et al., 2012b) have demonstrated that Si-PLS is superior to i-PLS and PLS. Nevertheless, all of these quantitative multivariate tools have their distinctive strengths and shortcomings.

Interval partial least square (i-PLS)

The windows of variables are applied rather than performing variable selection on each variable discretely if the spectral data are highly interrelated. Hence the i-PLS technique was instituted and it has become one of the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

commonly deployed variable selection techniques. The i-PLS technique split the spectra into equidistantly smaller subintervals of the full spectrum region and then build a model for each subinterval by the PLS regression principle. The purpose of the i-PLS methods is to discover the interval that produces the best prediction concerning the situation when the full spectrum is employed. The evaluation between interval performances is generally founded on the root mean square (RMSECV). Although this approach is simple and attractive to wavelength selection inappropriate interval size selection in i-PLS may corrupt the predictive performance in the resultant model (Javidnia et al., 2013).

Synergy interval partial least square (Si-PLS)

The synergy interval- PLS is the modified i-PLS as a result of the use of different interval combinations. The Si-PLS principle of variable selection algorithm split the dataset into several intervals and then combine with 2, 3, or 4 intervals of the full spectra that directly correlated to the interested components (in this case crude fat, crude fibre, total polyphenol, flavonoid content and antioxidant capacity). This PLS technique usually chooses the lowest RMSECV value (Andersen & Bro, 2010).

Genetic algorithm partial least square (GA-PLS)

GA-PLS is an evolutionary selection algorithm whose principle is to determine exact or approximate answers to search and optimization hitches and by mimicking biological evolution concepts such as mutation, inheritance, crossover, and natural selection. In this algorithm, each discrete of the populace (signified by the chromosome of binary values) denoted a subgroup of descriptors. The number of genes at each chromosome was equivalent to the

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

number of descriptors. A gene was coded one if its conforming descriptor was included in the subgroup; if not, it was coded zero. The arbitrarily identified zeros and ones show the variables which ought to be incorporated. (Alma & Bulut, 2012; Nørgaard et al., 2000).

Performance assessment of multivariate data analysis algorithms

Qualitatively the performance of the PLS-DA classification model was assessed according to identification rate or accuracy, sensitivity, specificity and efficiency. Accuracy is the proportion of samples, either organic cocoa beans or conventional cocoa beans correctly identified in a population, either in the calibration set or prediction set. It computes the degree of closeness or veracity of the measured result to the true value or analytical sample. Sensitivity evaluates the capability of the model to correctly identify and classify samples belonging to the targeted class (i.e., organic cocoa bean class). It measures how good the model is at detecting and classifying an organic cocoa bean from a conventional cocoa bean. Specificity evaluates the capability of the model to correctly detect and reject samples belonging to both classes (i.e., organic cocoa bean class and conventional cocoa bean class). It assesses how likely conventional cocoa beans could be ruled out correctly from organic cocoa beans. Efficiency is defined as the geometric mean of sensitivity and specificity in both calibration and prediction sets. Sensitivity and specificity depend on the values of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) (Wang, Huang, & Wang, 2019). The assessment of the model's performance was done according to the methods described by (Chen et al., 2008). Quantitatively, the performance of each final model of PLS, i-PLS, Si-

Digitized by Sam Jonah Library

PLS, and GA-PLS was expressed in terms of root mean square error of cross-validation (RMSECV), root mean square error of prediction (RMSEP), coefficient of correlation in the calibration set (R^2_{cal}), and prediction set (R^2_{pre}) and bias. The assessment of the model's performance was done in harmony with the procedures explained by (Chen et al., 2008). Computation of these parameters was done using equations (5.1) - (5.10):

$$\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP} \tag{5.1}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5.2}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{5.3}$$

$$\text{Efficiency} = \sqrt{Sensitivity * Specificity} \tag{5.4}$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n_c}\left(\hat{y}_{ci} - y_{ci}\right)^2}{n_c}} \tag{5.5}$$

$$R^2_{cal} = \sqrt{1 - \frac{\sum_{i=1}^{n_c}\left(\hat{y}_{ci} - y_{ci}\right)^2}{\sum_{i=1}^{n_c}\left(\hat{y}_{ci} - \bar{y}_c\right)^2}} \tag{5.6}$$

$$\text{Bias} = \sqrt{\frac{\sum_{i=1}\left(y_{ci} - \hat{y}_{ci}\right)^2}{n_c}} \tag{5.7}$$

Where: n_c = number of samples in calibration set, y_{ci} = reference measurement value of the *i*th sample, $\hat{y}_{ci}$ = estimated value for the *i*th sample by the built

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

model while the *i*th sample is removed, $\overline{y}_c$ = average of all reference values of samples in calibration set.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_p}\left(y_{pi} - \hat{y}_{pi}\right)^2}{n_p}} \quad (5.8)$$

$$R^2_{pre} = \sqrt{1 - \frac{\sum_{i=1}^{n_p}\left(\hat{y}_{pi} - y_{pi}\right)^2}{\sum_{i=1}^{n_p}\left(\hat{y}_{pi} - \overline{y}_p\right)^2}} \quad (5.9)$$

$$Bias = \sqrt{\frac{\sum_{i=1}\left(y_{pi} - \hat{y}_{pi}\right)^2}{n_p}} \quad (5.10)$$

Where: n_p = number of samples in prediction set, y_{pi} = reference measurement value of the *i*th sample in the prediction.

Results

Cocoa compositional quality characteristics

The determined compositional quality characteristics such as crude fat (CFA), total polyphenol content (TPC), total flavonoid content (TFC), and total antioxidant capacity (TAC) varied according to the cocoa category as presented in Table 5.1. For both categories of cocoa beans, crude fat was the major constituent. These discoveries are in agreement with results obtained by (Barbin et al., 2018). There were statistical differences at $p > 0.05$ between the organic cocoa beans and conventional cocoa beans for crude fat, total polyphenols, total flavonoid, and antioxidant contents. Nevertheless, there were no significant statistical differences at $p > 0.05$ between the two cocoa bean categories for crude fibre, although their obtained values numerically differed.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

The obtained CFA content varied from 38.64 to 42.81 and 39.37 to 44.09 percent for organic cocoa bean samples and conventional cocoa bean samples, respectively. These fat contents were comparatively lesser than the reported values of 53 to 59 percent for Forastero cocoas (Wood & Lass, 2008). Obviously, stress due to drought or heat, ambient temperature, genetic, differences in extraction methods, and geographic origins affected biosynthesis (Afoakwa, 2010).

The crude fibre content for organic and conventional cocoa beans range from 5.25 to 8.91 percent, being highest for organic cocoa bean samples as depicted in Table 5.1 and this was expressively higher than the literature values of 3.1 to 5.9 percent recorded for raw and fermented cocoa beans from Nigeria (Aremu, Agiang, & Ayatse, 1995). From the results, obtained mean value of 7.83 ± 0.08 for organic cocoa beans was statistically higher than the mean value of 7.56 ± 0.07 for conventional cocoa beans (Table 5.1). Although there were numerical differences between organic cocoa beans and conventional cocoa beans, no significant statistical variations existed between the two cocoa bean categories (p-value < 0.05).

The obtained values for TPC varied from 56.09 to 63.29 $mgGAEg^{-1}$, being the highest in organic cocoa beans (Table 5.1). TPC reference values obtained in this study are significantly high than literature values of 34.93±4.89 mg/g reported for Malaysian cocoa beans (Ramli et al., 2001). However, the obtained TPC values fall within the values of 59.55 to 153.58 mg/g reported for Ghana cocoa beans (Afoakwa et al., 2013c). The observed differences and similarities are to be possible because research has reported that TPC is

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

influenced by factors such as climatic condition, degree of fermentation, variety or genotype, farming management practices, and location (Aikpokpodion & Dongo, 2010).

The determined TFC varied from 79.44 to 82.61 $mgQEg^{-1}$ and 71.97 to 73.99 $mgQEg^{-1}$ for organic cocoa beans and conventional cocoa beans, respectively, being statistically highest in organic cocoa bean samples (Table 4.1). The TFC values obtained in the study fall within the values of 35.03 mg/g to 126.21 mg/g reported for cocoa bean samples from Indonesia, China, and Papua New Guinea (Gu et al., 2013). ANOVA showed that the comparison of TFC values between organic cocoa beans and conventional cocoa beans showed significant statistical differences at $p > 0.05$.

The measured TAC ranged from 107.02 to 109.90 $mgAEACg^{-1}$ and 127.63 to 131.94 $mgAEACg^{-1}$ for organic cocoa beans and conventional cocoa beans respectively, being statistically highest in conventional cocoa beans samples as presented in Table 5.1. The values obtained for TAC were notably higher than the TAC values of 52.29 to 72.92 mg/g reported by previous workers (Awarikabey, Amponsah, & Woode, 2014) probably due to differences in cocoa beans geographic location, variety, age, and differences in extraction methodologies. ANOVA showed that the variations in the TAC values between organic cocoa beans and conventional cocoa beans were statistically significant at $p > 0.05$.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 5.1: Proximate compositions (g/100g) and phytonutrient contents for organic and conventional cocoa samples (mean ± SD)

Samples	Crude fat (%)		Total polyphenol content (mgGAE/g)		Total flavonoid content (mgQE/g)		Total antioxidant capacity (mgAEAC/g)	
	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD	Range	Mean ± SD
Organic	38.64-42.81	41.75 ± 0.20^{b}	58.66-63.29	61.66 ± 0.16^{a}	79.44-82.61	81.44 ± 0.56^{a}	107.02-109.90	108.78 ± 0.85^{b}
Conventional	39.37-44.09	43.66 ± 0.15^{a}	56.09-60.99	59.73 ± 0.22^{b}	71.97-73.99	72.66 ± 0.69^{b}	127.63-131.94	130.03 ± 0.96^{a}

NB: Same letters show that there is no statistical difference ($p > 0.05$) among samples. GAE= gallic acid equivalent. QE= quercetin equivalent. AEAC= ascorbic acid equivalent antioxidant content. SD= standard deviation.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

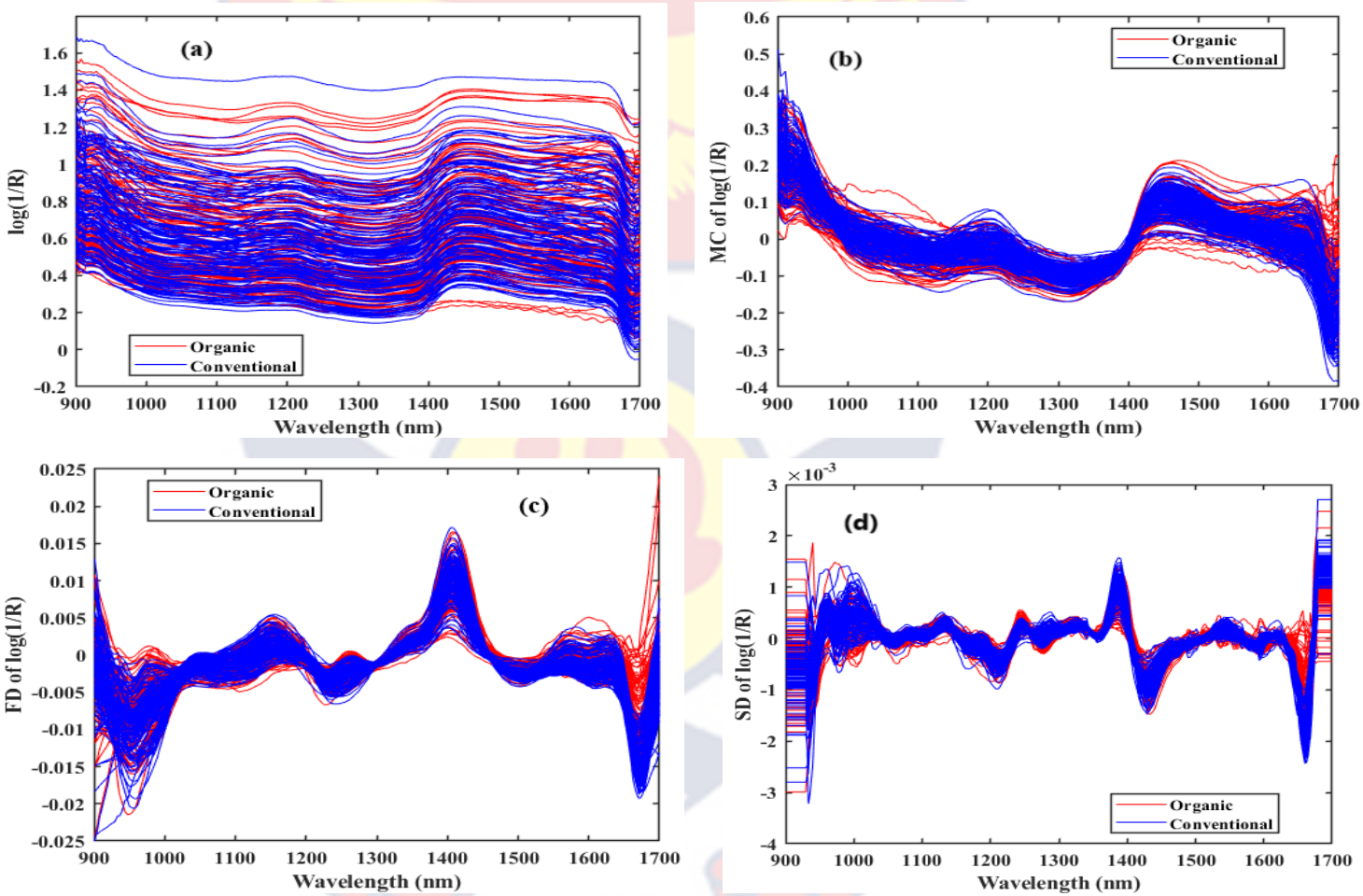


Fig. 5.1. Spectra (a) Raw spectra, (b) mean centering, (c) first derivative, and (d) second derivative of cocoa bean samples.

Digitized by Sam Jonah Library

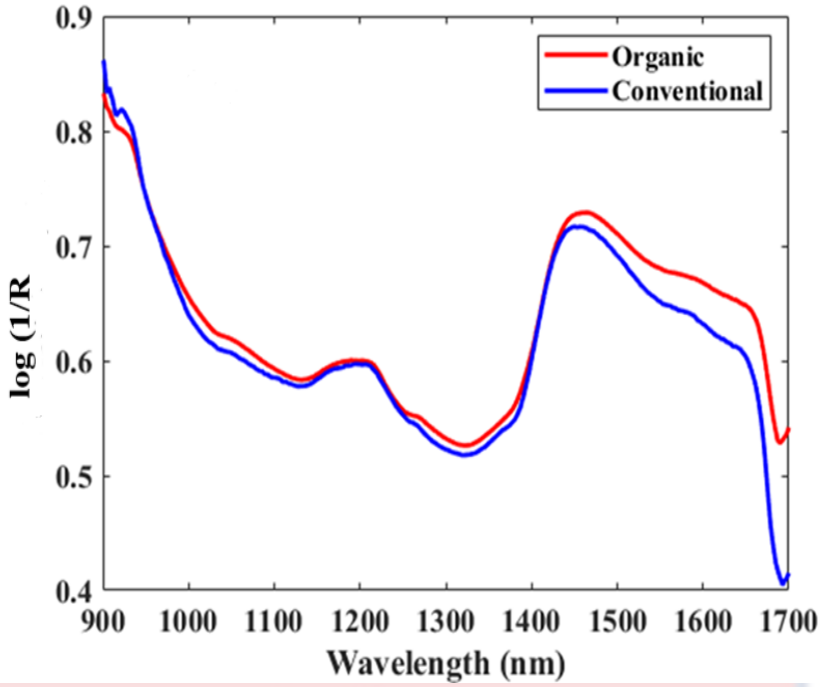


Fig. 5.2. Mean spectra profile of organic and conventional cocoa bean samples.

NIR spectra examination

The raw spectra of the 260 cocoa bean samples obtained in the wavelength range of 900 to 1700 nm were presented in Fig. 5.1 (a). This spectral wavelength range can offer useful features for the differentiation of organic and conventional cocoa bean samples, though the raw spectra profile seemed to be similar. This made it difficult to determine exact bands in the original spectra due to the high degree of overlapping of bands. Hence, chemometric preprocessing analysis of the dataset was applied to acquire only the useful properties of samples, build a reliable model while keeping the similarities and variations among the primary observations. Among the chemometric analysis applied the SD preprocessing approach best smoothen the original spectra and eventually led to satisfactory classification, and its spectra were presented in Fig. 5.1 (d). This contributed to clear and noticeable groupings as shown in the mean spectra profile in Fig. 5.2. It depicts specific absorption bands observed from main valleys and peaks that are related to vibrations of chemical bonds

Digitized by Sam Jonah Library

such as N-H, S-H, C=O, -CH_3 and CH_2 (Arslan et al., 2018). These chemical bond vibrations are associated with major biochemical constituents such as polyphenols, flavonoids, alkaloids, antioxidants, volatile and non-volatile acid, fats, proteins, carbonyl group, C-H deformation and C-H stretch (Table 5.2), and other composites present in the cocoa beans no matter the production method or origin. Specifically, the absorption band attributable to C-H bond of cocoa that is mainly connected to proteins and fats was found around 910 nm (Veselá et al., 2007). Absorption bands of 1000 - 1100 nm are attributable to C-H stretch 1st overtone, carbonyl groups (-CH_2, CH_3-, and -CH=CH-) (Siesler et al., 2008). An observable absorption band around 1440 nm might be associated with 1st overtone of starch, moisture and sugars (Veselá et al., 2007). These spectral wavelength bands might have significantly contributed to the classification of organic and conventional cocoa beans.

Spectral presentation and principal component analysis (PCA)

To observe a visible trend of the samples and evaluate the relation among samples, PCA was performed using the raw spectral data and the outcomes presented in Fig 5.3 (a) – (d). The PCA after second derivative preprocessing yielded a good cluster trend. The topmost three PCs extracted from the 260 samples were PC1 (68.03%), PC2 (16.71%), and PC3 (8.18%). It shows that the topmost three PCs can explain 92.92% of the variance information from the spectra dataset that covers the relevant biochemical information in the samples. PCA technique brings out useful relevant information and removes irrelevant ones so that bean samples with the same characteristics are clustered nearer to each other. Thus, the graphic output could

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

be used to discover the variances between the categories of cocoa bean samples used. Fig. 5.3 (d) depicts that two main groups of cocoa bean samples were used in the study. The groups cover a broader array of cocoa beans. The graphic plot offers relevant information that could be used for the determination of differences between organic and conventional cocoa bean samples. PCA is not a classification tool but it showed the data trend in visualizing dimension space (Sun et al., 2017). PCA loadings (Fig. 5.4) was performed to give an explanation as to how much each wavelength contributed to the significant variation in the data. It was observed that wavelengths corresponding to the biggest eigenvector loading values for PC1 (68.03 %) were situated around the range of 986 nm associated with pH 1st overtone absorption peak O-H stretching (Wang et al., 2015) and O-H stretch 2nd overtone of carbohydrate, 1280 nm, and 1417 nm are 2nd overtone bands C–H bond stretching and C-H combination (aromatic) respectively. The peak around 1200 nm could be attributed to the 1st overtone of C-H stretch (Bao et al., 2019). These absorption bands are characteristics of proteins, fats, and aromatic compounds found in cocoa beans. Observable absorption band around 1608 nm might be ascribed to 1st overtone of C-H stretching (Zhang, Liu & He, 2018). PC2 explains 16.71% of the variance, and the biggest vibration placed around 958 nm, 973, and 1395 nm correlated with 2nd overtone of OH stretch of carbohydrate, 2nd overtone of N-H stretching of fat, and 2 x C-H stretch + C-H deformation of protein respectively (Zhang, Liu & He, 2018).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

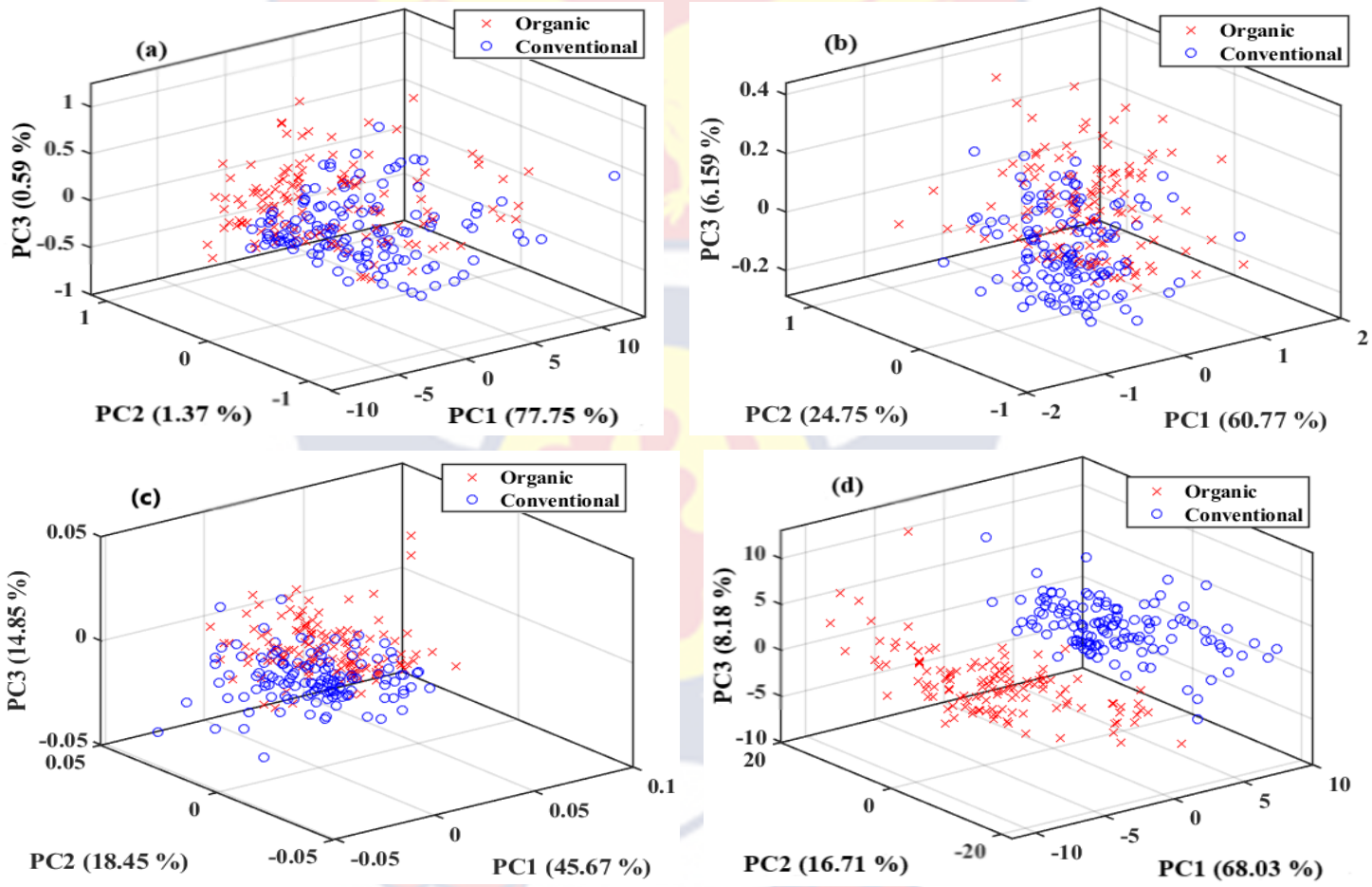


Fig. 5.3: PCA score plot of the first three PCs of organic and conventional cocoa beans (a) raw, (b) MC, (c) FD, and (d) SD

Digitized by Sam Jonah Library

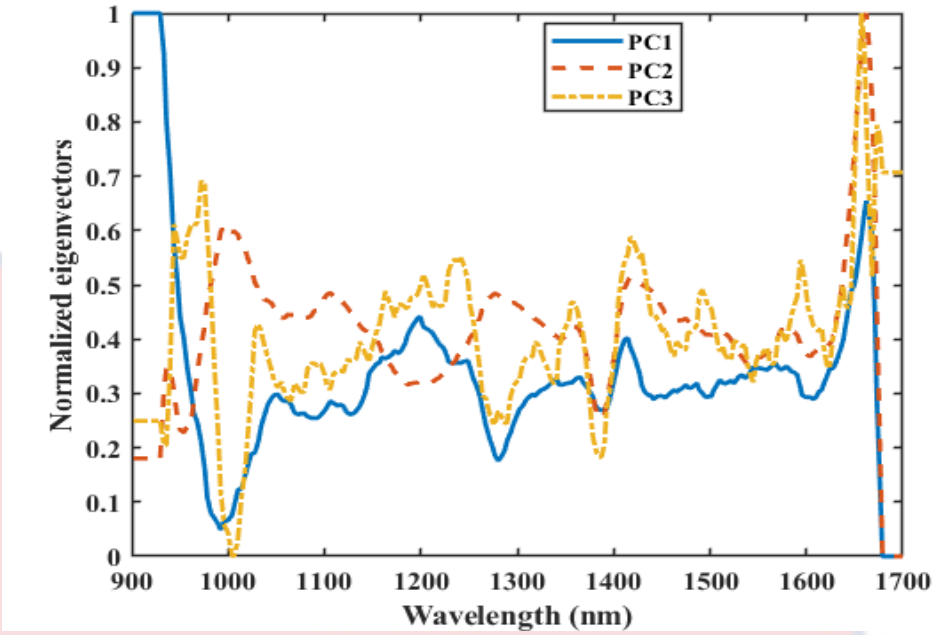


Fig. 5.4: PCA loadings of top three latent variables of cocoa bean samples.

PC 3 explains 8.18% of the variation, and it appeared to be the mirror image of the cocoa bean spectra, and this accounted for the slight differences in particle size. The differences correlated to compositional variations among the cocoa bean categories. It implied a particular chemical constituent alone or in combination with others contributed the largest influence that explained the basis for the detected variations between the cocoa bean samples.

Performance of classification models

The results from different classification models for the discrimination of organic and conventional cocoa beans were reported in Table 5.2. Every multivariate classification algorithm has its potentials and limitations. As shown in Table 5.2 the smooth-SD processing (17-point window, 2^{nd} -order polynomial) highly enhanced the performance of all the multivariate classification algorithms in both the calibration set and prediction set than MC, MSC and FD.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 5.2: NIR wavelength band assignments

Wavelength (nm)	Functional group	Band assignments
958	OH	O-H 2nd overtone stretch of carbohydrate (Phuangsombut et al., 2017)
973, 1005	NH_3	N-H 2nd overtone stretch associated with fat (Zhang, Liu, & He 2018)
986	O-H	O-H 1st and 2nd overtones stretch of absorption peak of starch (Phuangsombut et al., 2017)
1200	C-H	C-H 1st overtones stretch related to proteins and starch (Bao et al., 2019)
1280	C-H	C–H 1st overtone bond stretching corresponding to fats, and aromatic compounds (Rout, Acharya & Maji, 2017)
1483	O-H	O-H 2nd overtone stretch corresponding moisture (Zhang, Liu & He, 2018)
1417	O-H	H_2O band groups corresponding to weakly bounded water, aromatic compounds (Veselá et al., 2007; Arslan et al., 2018)
1440, 1460	O-H	O-H stretch 1st overtone of starch, water band and sugars (Li-Chan, Ismail, Sedman & Van de Voort, 2006; Veselá et al., 2007)
1608	C=O	C-O from COOH typical of amines, acidity (Kays et al., 2005; Wang et al, 2015)

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 5.3: Classification models with preprocessing algorithms for cocoa bean samples

Models	Preprocessing	PCs	Evaluation	Performance (%)	
				Calibration set	Prediction set
RF	Raw	3	Accuracy	67.79	67.31
			Sensitivity	69.23	69.23
			Specificity	66.35	65.38
			Efficiency	67.77	67.28
	MC	7	Accuracy	85.58	88.46
			Sensitivity	82.69	84.62
			Specificity	88.46	92.31
			Efficiency	85.53	88.38
	FD	7	Accuracy	89.90	88.46
			Sensitivity	90.65	91.30
			Specificity	89.11	86.21
			Efficiency	89.88	88.72
	SD	9	Accuracy	96.15	98.08
			Sensitivity	94.95	96.77

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

			Specificity	97.25	100.00
			Efficiency	96.09	98.37
KNN	Raw	3	Accuracy	65.38%	69.23
			Sensitivity	71.29%	62.07
			Specificity	59.81%	78.26
			Efficiency	65.30%	69.70
	MC	5	Accuracy	79.81	82.69
			Sensitivity	82.86	76.00
			Specificity	76.70	88.89
			Efficiency	79.72	82.19
	FD	5	Accuracy	80.77	80.77
			Sensitivity	75.76	83.87
			Specificity	85.32	76.19
			Efficiency	80.40	79.94
	SD	5	Accuracy	91.35	92.31
			Sensitivity	87.27	95.00
			Specificity	95.92	90.63

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

			Efficiency	91.49	92.79
LDA	Raw	3	Accuracy	69.71	69.23
			Sensitivity	72.38	64.00
			Specificity	66.99	74.07
			Efficiency	69.63	68.85
	MC	5	Accuracy	84.13	78.85
			Sensitivity	86.14	86.21
			Specificity	82.24	69.57
			Efficiency	84.17	77.44
	FD	5	Accuracy	78.85	80.77
			Sensitivity	83.65	88.46
			Specificity	74.04	73.08
			Efficiency	78.70	80.40
	SD	5	Accuracy	90.38	98.08
			Sensitivity	89.42	100.00
			Specificity	91.35	96.15
			Efficiency	90.38	98.06

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

PLS-DA	Raw	3	Accuracy	77.88	78.85
			Sensitivity	80.81	83.87
			Specificity	75.23	71.43
			Efficiency	77.97	77.40
	MC	5	Accuracy	92.79	94.23
			Sensitivity	96.97	90.32
			Specificity	88.99	100.00
			Efficiency	92.89	95.04
	FD	5	Accuracy	98.56	100.00
			Sensitivity	99.04	100.00
			Specificity	98.08	100.00
			Efficiency	98.56	100.00
	SD	5	Accuracy	100.00	100.00
			Sensitivity	100.00	100.00
			Specificity	100.00	100.00
			Efficiency	100.00	100.00

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

RF

The k-fold cross-validation results showed that the RF algorithm with 9 PCs on normalized data provided correct identification rates with 96.09% and 98.37% efficiency in the calibration set and prediction set respectively (Table 5.3). The optimum number of PCs was based on the best classification rate performed by k-fold cross-validation. In Table 5.3, the best classification rate by RF model for calibration set was 96.15% and 98.08% for the prediction set at an optimum number of 9 PCs.

KNN

In Table 5.3, the k-fold cross validation outcomes disclosed that the KNN algorithm with PCs equal to 5 on normalised data provided a correct classification rate with 91.49% efficiency for the calibration set and 92.79% for the prediction set. Table 5.3 demonstrates that the best classification rate for the calibration set was 91.35% and 92.31% for the prediction set.

LDA

In Table 5.3, the k-fold cross-validation results showed that the LDA algorithm with PCs equal to 5 on normalised data provided a correct classification rate with 90.38% calibration set efficiency and 98.06% prediction set efficiency. Table 5.3 demonstrates that the best classification rate for the calibration set was 90.38% and 98.08% for the prediction set.

PLS-DA

Table 5.3 shows the performance of the PLS-DA classification algorithm used in identifying organic and conventional cocoa beans. K-fold cross validation outcomes demonstrated that the PLS-DA technique with 5

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

principal components (PCs) on normalized data provided correct identification rates with 100% efficiency in the prediction set. Fig. 5.5 displays the performance of the PLS-DA model for solving the discrimination problems after k-fold cross-validation. The optimum number of PCs was based on the best classification accuracy achieved by k-fold cross-validation. In Table 5.3, the best classification rate for both calibration set and prediction set was 100.00% at an optimum number of 5 PCs.

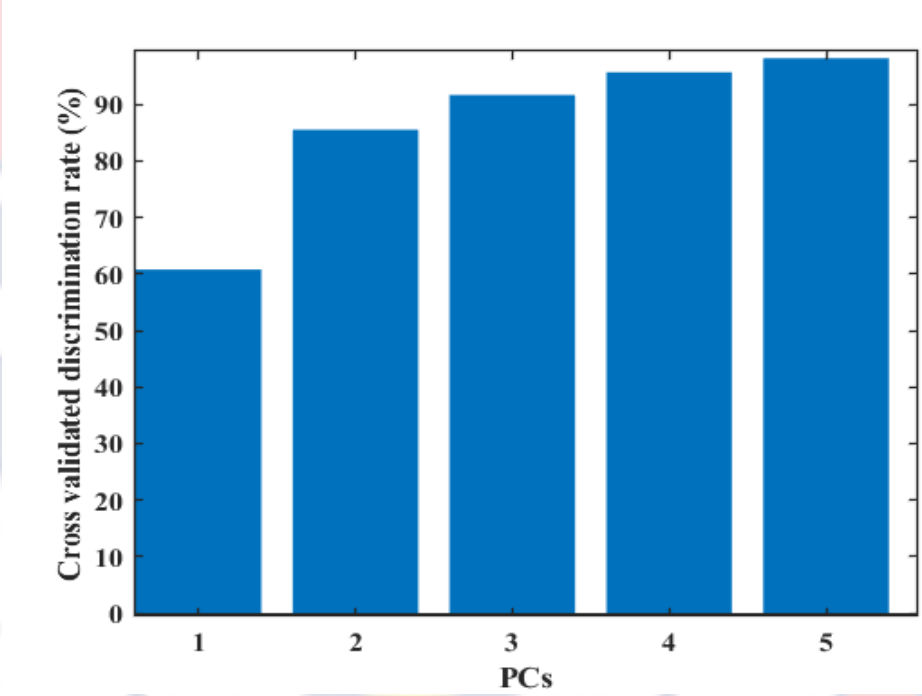


Fig. 5.5. The k-fold cross validation discrimination rates of the second derivative preprocessed PLS-DA models at 5 PCs.

Overall performance of classification algorithms

The identification rates of multivariate classification algorithms are presented in Table 5.4. In this table, we compare the classification accuracy of the RF, KNN, LDA and PLS-DA models. Comparatively, the results show that the performance of the PLS-DA established model was superior to others viz. RF, KNN and LDA (Table 5.4). The result is in agreement with that of

Digitized by Sam Jonah Library

Kosmowski & Worku, 2018 where the PLS-DA technique performed better in the identification of sorghum cultivars. The discrimination stability for all the cocoa bean samples investigated increased in the order of KNN<LDA<RF<PLS-DA denoted by identification rate

Table 5.4 Overall performance of classification algorithms.

Models	Total cocoa bean samples		PCs	Identification rate (%)	
	Calibration set	Prediction set		Calibration set	Prediction set
RF	182	78	9	96.15	98.08
KNN	182	78	5	91.35	92.31
LDA	182	78	5	90.38	98.08
PLS-DA	182	78	5	100.00	100.00

Regression models for chemical composition prediction

NIR wavelength selection and band assignment

Wavelength selection aims to restrict the spectral region used by the model to only areas essential for modelling the analyte of concern. The best combination of different spectral intervals was attained with the lowest root mean square error of cross-validation (RMSECV), the lowest bias, and the highest coefficient of correlation. In this analysis, the full spectrum (900-1700) was subdivided into 20 equidistant intervals with a combination of 2, 3, or 4 intervals for the optimum results. Table 5.5 displays the performance of different NIR wavelength regions selected for the prediction of crude fat (CFA), crude fibre (CFI), total polyphenol content (TPC), total flavonoid content (TFC), and total antioxidant capacity (TAC) by smooth-SD Si-PLS algorithm. The efficient combination of the selected spectral intervals corresponded to spectral wavelengths 1033-1079 nm, 1170-1210 nm, 1338-1375 nm for CFA; 1255-1293 nm, 1296-1334 nm, 1378-1415 nm, 1605-1635 nm for CFI; 1296-

Digitized by Sam Jonah Library

1334 nm, 1338-1375 nm, 1378-1417 nm, 1417-1453 nm for TPC; 1127-1167 nm, 1296-1334 nm, 1378-1415 nm for TFC and 1038-1079 nm, 1127-1167 nm, 1296-1334 nm, 1378-1415 nm for TAC in the full spectrum (Table 5.3). These spectral wavelengths selected by smooth-SD Si-PLS algorithm are associated with aromatic C-H stretch 1st overtone, carbonyl groups (-CH_2, CH_3-, and -CH=CH-), C=N and C=C combination, and 2nd overtone of C-H in CH_2 (Ozaki, McClure, & Christy, 2006; Siesler et al., 2008; Veselá et al., 2007) as explained in Table 5.3. These wavelength band vibrations are triggered by constituents such as fats, water, polyphenols, fibre, organic acids, alkaloids, polysaccharides, amines, and aromatic compounds found in cocoa beans (Arslan et al., 2018; Oliveira & Franca, 2011). Therefore, these spectral wavelength range selected might have significantly contributed to the classification and compositional estimation of cocoa bean categories used.

Comparison of different PLS regression models

The results from different regression models with full-spectrum and spectral-interval selection for the prediction of crude fat content, crude fibre content, total polyphenol content, total flavonoid content, and total antioxidant capacity in cocoa beans were reported in Table 5.6. Comparatively, the results show that the performance of the Si-PLS established model was superior to others viz. PLS, i-PLS, and GA-PLS (Table 5.6). The prediction stability for all the chemical parameters investigated increased in the order of PLS<i-PLS<GA-PLS<Si-PLS denoted by R^2_{pre}, RMSEP, and bias. It was observed that all the spectral-interval selection models performed better than the traditional PLS model.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 5.5 NIR wavelength regions selected by smooth-SD Si-PLS model for the prediction of CFA, CFI, TPC, TFC, and TAC and corresponding band assignments

Quality parameter	Selected wavelengths (nm)	Band assignments
CFA	1033-1079, 1170-1210, 1338-1375	-CH_2, -CH_3; 2nd overtone of asymmetric stretching associated with oils and fats (Ozaki et al., 2006)
CFI	1255-1293, 1296-1334, 1378-1415, 1605-1635	-CH_2, H_2O; 1st overtone of stretching vibration typical of amines, fibre, and weakly bounded water (Kays et al., 2005)
TPC	1296-1334, 1338-1375, 1378-1417, 1417-1453	-CH=CH-, -C=N; 1st overtone bands from polyphenols, tannins, sugars and aromatic rings (Veselá et al., 2007)
TFC	1127-1167, 1296-1334, 1378-1415	-CH_3, -CH=CH-; 2nd overtone of symmetric stretching corresponding to organic acids, flavonoids, alkaloids, and other aromatics found in cocoa beans (Siesler et al., 2008)
TAC	1038-1079, 1127-1167, 1296-1334, 1378-1415	-CH_3, -C=C; 1st overtone of carbonyl band groups attributed to antioxidants, aromatic compounds (Arslan et al., 2018)

CFA= crude fat, CFI= crude fibre, TPC= total polyphenol content, TFC= total flavonoid content, TAC= total antioxidant capacity

Digitized by Sam Jonah Library

This could be because the traditional PLS technique was applied to the full spectra range of 900-1700 nm with 256 variables to construct a regression model, and among this full spectrum existed irrelevant and redundant variables that were unrelated to the cocoa bean parameters of interest. These irrelevant and redundant spectra variables significantly weakened the performance of the PLS model. Among the variable selection algorithms, i-PLS was the least performing algorithm. The i-PLS gave overview spectra data to select the interesting spectra variable and eliminate some irrelevant and redundant variables but selected only one variable interval to build the PLS model. Some relevant spectra variables were eliminated and thereby reduced the general performance of the model. Si-PLS algorithm just like GA-PLS variable selection algorithm has a consistent concept that divided the full spectrum into several equidistant sub-intervals to build the PLS model. However, the model built using GA-PLS chose several variables that may not contain variables about chemical parameters (viz. fat, fibre, polyphenol, flavonoid, and antioxidant capacity) of the cocoa bean samples. The general performance of the GA-PLS model was thus weakened by these several unwanted variables. The Si-PLS algorithm selected fewer variables from 20 equidistant intervals and combined them with 2, 3, or 4 efficient sub-intervals which were highly linked to the characteristics of the chemical parameters of the cocoa bean samples to build a PLS model. Thus, the Si-PLS established model was superior to the others.

Si-PLS model for crude fat

The best Si-PLS model for crude fat was attained when 8 PCs were used to build the model with the combination of 3 sub-intervals from 16 sub-intervals

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

(Table 5.6). The best combination of selected sub-intervals are [4 7 11] as shown in Fig. 5.6 (a) corresponding to three wavelength ranges: 1033-1079, 1170-1210, and 1338-1375 nm in the full spectrum region, and the wavelength features contained in these ranges and corresponding band assignments are stated in Table 5.5. The optimum performance of Si-PLS model for crude fat was R^2_{cal} = 0.972, RMSECV = 0.232 and bias = 0.003 in the calibration set. It was realized that when the performance of the smooth-SD Si-PLS model was evaluated by the independent samples in the prediction set, R^2_{pre} = 0.981, RMSEP = 0.191 and bias = -0.008 (Table 5.6). It was observed that the relationship between reference measured and NIR values for crude fat in the calibration and prediction sets were excellent. The outcome shows that the wavelength ranges chosen and modeled by the Si-PLS algorithm have an excellent correlation with crude fat values.

Si-PLS model for total polyphenol content

From Table 5.6, the ideal Si-PLS model for total polyphenol content was achieved when 6 PCs were used for building the model with the combination of 4 sub-intervals from 12 sub-intervals. The ideal combination of selected sub-intervals are [10 11 12 13] as shown in Fig. 5.6 (b) corresponding to three wavelength ranges: 1296-1334, 1338-1375, 1378-1417, and 1417-1453 nm in the full spectrum region, and the wavelength features contained in these ranges and corresponding band assignments are stated in Table 5.5. Here, the best performance of smooth-SD Si-PLS model for total polyphenol was R^2_{cal} = 0.962, RMSECV = 0.268 and bias = 001 in the calibration set.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

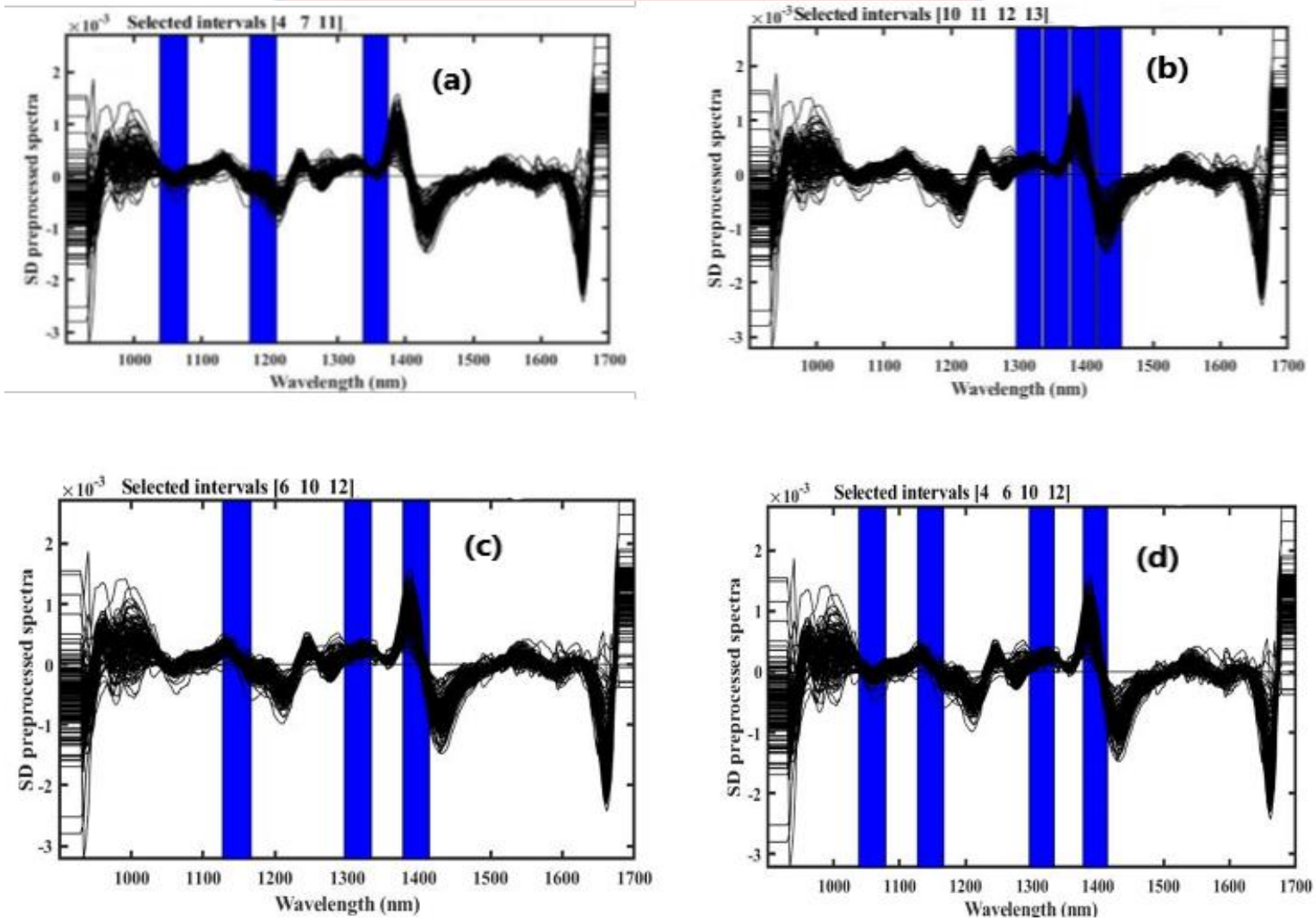


Fig. 5.6: The optimum spectral intervals selected by the Si-PLS algorithm for the prediction of crude fat (a), total polyphenol content (b), total flavonoid content (c) and total antioxidants capacity (d).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 5.6 Comparison of results based on different regression models for crude fat, crude fibre, total polyphenol content, total flavonoid content, and total antioxidant content. Bold values represent the best results.

QP*	Algorithms	VN*	PCs	Calibration set			Prediction set		
				R^2_{cal}	RMSECV	Bias	R^2_{pre}	RMSEP	Bias
Crude fat									
	PLS	256	8	0.9151	0.3945	0.0292	0.8749	0.4687	0.0531
	i-PLS	104	10	0.9493	0.3079	0.0037	0.9645	0.2619	0.0447
	Si-PLS	**149**	**8**	**0.9715**	**0.2322**	**0.0030**	**0.9812**	**0.1909**	**-0.0080**
	GA-PLS	256	9	0.9510	0.3663	-0.0035	0.9712	0.2605	0.0439
Total polyphenol content									
	PLS	256	8	0.8865	0.3442	0.0023	0.9124	0.2876	0.0654
	i-PLS	138	10	0.9067	0.3223	0.0032	0.9271	0.2987	0.0877
	Si-PLS	**135**	**6**	**0.9621**	**0.2680**	**0.0008**	**0.9797**	**0.2097**	**-0.0529**
	GA-PLS	256	9	0.9567	0.2776	0.0021	0.9432	0.3002	0.0599
Total flavonoid content									
	PLS	256	7	0.9432	0.9875	-0.0655	1.0981	0.7698	0.0765
	i-PLS	124	9	0.9488	0.9889	0.0577	1.127	0.8756	0.0587
	Si-PLS	**148**	**5**	**0.9831**	**0.8115**	**-0.0485**	**0.9896**	**0.6364**	**0.0489**
	GA-PLS	256	8	0.9688	0.8754	0.0498	0.9899	0.7664	0.0563

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 5.6 continued

Total antioxidant content								
PLS	256	9	0.9654	1.5437	0.0876	0.9667	1.2658	0.1433
i-PLS	142	11	0.9705	1.6755	0.0655	0.9840	1.1228	0.2322
Si-PLS	**134**	**7**	**0.9931**	**1.2441**	**0.0583**	**0.9947**	**1.0999**	**0.0101**
GA-PLS	256	8	0.9894	1.4338	0.0765	0.9905	1.1887	0.1656

Note: QP; quality parameter, VN; variable number, PCs; principal components

Digitized by Sam Jonah Library

However, when the performance of the Si-PLS model was evaluated by the independent samples in the prediction set, R^2_{pre} = 0.980, RMSEP = 0.210 and bias = 0.053 (Table 5.6). It was observed that the correlation between reference measured and NIR values for total polyphenol in the calibration and prediction sets were very good. The outcome shows that the spectral regions chosen and modeled by the Si-PLS algorithm have an excellent relationship with total polyphenol values.

Si-PLS model for total flavonoid content

From Table 5.6, the ideal Si-PLS model for total flavonoid content was achieved when 5 PCs were used for building the model with the combination of 3 sub-intervals from 16 sub-intervals. The ideal combination of selected sub-intervals are [6 10 12] as shown in Fig. 5.6 (c) corresponding to three wavelength ranges: 1127-1167, 1296-1334, and 1378-1415 nm in the full spectrum region, and the wavelength features contained in these ranges and corresponding band assignments are stated in Table 5.5. Here, the best performance of Si-PLS model for total flavonoid content was R^2_{cal} = 0.983, RMSECV = 0.812 and bias = -0.049 in the calibration set. Nevertheless, when the performance of the smooth-SD Si-PLS model was evaluated by the independent samples in the prediction set, R^2_{pre} = 0.990, RMSEP = 0.636 and bias = 0.049 (Table 5.6). It was observed that the relation between reference measured and NIR values for total flavonoid content in the calibration and prediction sets were good. The outcome shows that the spectral regions chosen and modeled by the Si-PLS algorithm have a very good correlation with total flavonoid characteristic values. The result shows that the spectral regions

Digitized by Sam Jonah Library

selected and modeled by the Si-PLS algorithm have an excellent relationship with total flavonoid values.

Si-PLS model for total antioxidant capacity

From Table 5.6, the optimal Si-PLS model for total antioxidant capacity was achieved when 5 PCs were used for building the model with the combination of 4 sub-intervals from 17 sub-intervals. The ideal combination of selected sub-intervals are [4 6 10 12] as shown in Fig. 5.6 (d) corresponding to three wavelength ranges: 1038-1079, 1127-1167, 1296-1334, and 1378-1415 nm in the full spectrum region, and the wavelength features contained in these ranges and corresponding band assignments are stated in Table 5.5. Here, the best performance of smooth-SD Si-PLS model for carbohydrate was R^2_{cal} = 0.993, RMSECV = 1.244 and bias = 0.058 in the calibration set. When the performance of the Si-PLS model was evaluated by the independent samples in the prediction set, R^2_{pre} = 0.995, RMSEP = 1.100 and bias = 0.010 (Table 6.6). It was observed that the relationship between reference measured and NIR values for total antioxidant capacity in the calibration and prediction sets were good. The outcome shows that the spectral regions chosen and modeled by the Si-PLS algorithm have an excellent correlation with total antioxidant characteristic values. The result shows that the spectral regions selected and modeled by the Si-PLS algorithm have a brilliant relationship with total antioxidant capacity (Arslan et al., 2018).

Discussion

There were observed differences in chemical compositions of organic cocoa beans and conventional ones (as seen in Table 5.1). These could largely

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

be attributed to the influence of production methods and partly to the reaction of inherent compositions of the organic and conventional cocoa beans to the fermentation process (which was carried out using the same protocols for both categories of cocoa beans). Chocolate flavour compounds do not only originate by character precursor formation during fermentation but could also be generated during production management systems (Afoakwa, 2010). Thus, the composition of organic and conventional cocoa beans interacted with the fermentation process in the formation of cocoa flavour quality constituents. The use of synthetic fertilizers and chemicals in conventional methods contributed to variations in the cocoa bean biochemical composition that could lead to a distinct cluster trend.

The spectra obtained from scanning of the organic and conventional cocoa bean samples with the handheld NIR spectrometer produced a spectral profile that displayed multiple wavelength bands and peaks as shown in Fig. 5.1 (d). The bands consisted of overtones and combinations of fundamental vibrations that matched the chemical compositions which provided exclusive fingerprint of the cocoa bean categories employed in this study. The pre-processing of the spectra profile into mean was performed and two groupings were representing the two distinct cocoa bean samples used as seen in Fig. 5.2. This is due to the unique biochemical and physical properties of each bean group to give a well-defined separation trend.

The comparative analysis of the PCA cluster using different pre-processing techniques revealed that the second derivative treatment performed better by showing a clear cluster trend as shown in Fig. 5.3. The clustering can

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

be explained by the biochemical compositions in each of the cocoa bean samples as a result of differences in the categories of the cocoa bean either been organic or conventionally produced cocoa bean. The contributions of the topmost three PCs were 92.92% for the total variance in the original data. Nevertheless, PCA does not give definite identification because it is not a classification tool, however, it preserves much variance in a high dimensional space by reducing dimensionality. PCA loading plot in Fig. 5.4 shows that the most important wavelength bands which contributed to the cluster trend of the cocoa bean samples and were located at around 986, 1200, 1280, 1417 and 1068 nm for PC1; 958, 973, 1395 and 1460 nm for PC2; 1005, 1440 and 1483 nm for PC3. The wavelengths at 958, 973, 986, 1440 and 1460 nm are due to 1st overtone and 2nd overtone of O-H/O-H stretch; 1005 and 1483 nm are attributable to 2nd overtones of N-H stretch; 1200 and 1608 nm are related to C-O from COOH typical and 1st overtone of C-H stretch; 1280nm band might be characterised by 2nd overtone bands C–H bond stretching; 1375 and 1417 nm could be associated with C-H vibration modes, 1395 nm and 1417 nm absorption band might correspond to 2 x C-H stretch + C-H deformation and combination. These observable wavelengths are principally characterised by the asymmetric stretching, overtones and combinations of vibrations of C-H, N-H, O-H, C=O which are triggered by constituents such as fats, water, polyphenols, fibre, organic acids, alkaloids, polysaccharides, amines, and aromatic compounds found in cocoa beans (Arslan et al., 2018; Oliveira & Franca, 2011; Zhang et al., 2018). Table 5.2 gives additional information on the observable absorption bands and their associated chemical constituents. These spectra

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

observations echoed the outcome of the chemical compositions of the two categories of cocoa beans studied. These spectral wavelength bands might have significantly contributed to the classification of organic and conventional cocoa beans as seen in Table 5.3.

Four other pattern recognition algorithms which are known to have potentials in solving identification problems were applied. The pattern recognition algorithms- RF, KNN, LDA and PLS-DA were applied to build a classification model and to ensure their stability cross-validation was done. PLS-DA model produced classification accuracy of 100 % in both the calibration set and prediction set whereas the classification accuracy for the calibration set and prediction set were 96.15% and 98.08% for RF, 91.35% and 92.31% for KNN and 90.38% and 98.08% for LDA (Table 5.4). The experimental outcomes showed that the PLS-DA algorithm was superior to RF, KNN and LDA algorithms. This can be due to the fact that the PLS-DA algorithm possesses stronger and added potential of self-adjusting and self-learning properties. For cocoa beans categories used in this work, the biochemical compositions and complex organoleptic properties can explain why RF, KNN and LDA could not deliver the optimum solution. The PLS-DA delivered its best performance at 5 PCs. The high number of PCs as seen in the RF model may result in low generalization in the performance lowering the efficiency of its model.

Generally, the optimum classification accuracy (100%) received could largely be attributed to the influence of production methods and partly to the reaction of inherent compositions of the organic and conventional cocoa beans

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

to the fermentation. Cocoa bean flavour compounds do not only originate by character precursor formation during fermentation but could also be generated during production management systems (Afoakwa, 2010). Thus, the composition of organic and conventional cocoa beans interacted with the fermentation process in the formation of cocoa flavour quality constituents. The use of synthetic fertilizers and chemicals in the conventional method contributed to variations in the cocoa bean biochemical composition leading to the distinct cluster trends and differentiation of the cocoa samples used in this experiment. Also, according to other authors, organically produced foods show high polyphenols and ascorbic acid contents as a response to stress stimuli (Hurtado-Barroso et al., 2019). More so, organic crops often grow more slowly compared to synthetic fertilized crops with readily available minerals nutrients and this might reduce their water content leading to a higher concentration of some plant compounds. It is therefore expected that organically produced cocoa will have higher concentrations of some compounds (polyphenols, protein, carbohydrate, fibre and total flavonoids) as recorded in this study. This phenomenon might have contributed to the accurate classification of the different categories of cocoa beans used in this study by the handheld NIR spectroscopy.

Quantitatively, the performance of a handheld NIR spectrometer and four different regression algorithms was investigated. Table 5.6 displays the comparison of results of PLS, i-PLS, Si-PLS, and GA-PLS regression algorithms. The Si-PLS algorithm was superior to all the other regression algorithms, and the classical PLS algorithm was the least. All the variable

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

selection algorithms (i-PLS, Si-PLS, and GA-PLS) performed better than the classical PLS algorithm. Such phenomena could be attributed to the fact that the classical PLS was performed on the full spectrum (900-1700 nm) involving 256 variables which may contain redundant information. Thus, the 256 variables contain irrelevant and redundant variables that were uncorrelated to the parameters studied which eventually weakened the performance of the model. Although, i-PLS removed some irrelevant information it, also chose only one spectral interval to build the PLS model while leaving other useful ones. The chosen variable may not be the only variable that correlated to the parameters studied. Other useful spectral variables that could be associated with these quality parameters were abandoned. So, the resultant i-PLS model was certainly reduced. Si-PLS improved the strength of i-PLS (i.e. selection of multiple spectral intervals which correlated to quality parameters studied) and removed the weaknesses of the classical PLS algorithm (Lei et al., 2014). Si-PLS removed immaterial and redundant variables that could have affected the established model. Though GA-PLS performed better than PLS and i-PLS it was incomparable to Si-PLS. This may be due to fewer variables used by Si-PLS with 2, 3, or 4 intervals combinations whereas GP-PLS deployed all the 256 variables magnified the computational challenge and influence unpredictability and undependability (Chen et al., 2012b). It means that the irrelevant and redundant variables in the full spectrum declined the model's performance as they may be uncorrelated to crude fat, total polyphenol content, total flavonoid, and total antioxidant capacity.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Conclusion

This work represents the first study to successfully evaluate the application of a low-cost handheld NIR spectrometer and chemometric classification techniques for rapid non-destructive screening and authentication of organic and conventional cocoa beans produced in Ghana. The PCA score plot exhibited the feasibility of identifying cocoa bean categories. Four different chemometric classification algorithms viz. RF, KNN, LDA and PLS-DA were comparatively performed for the construction of classification models. PLS-DA being a dimensionality reduction and variant squares regression algorithm exhibited superior performance over the others (RF, KNN and LDA) after second derivative (SD) preprocessing for the differentiation of organic cocoa beans from conventional ones. PLS-DA model yielded classification accuracy of 100% in both calibration set and prediction set. Variable selection model by using Si-PLS revealed its superiority for simultaneous prediction of crude fat, total polyphenol content, total flavonoid content, and total antioxidant capacity in cocoa beans with the range of: $0.871 \leq R^2_{cal} \leq 0.993$ and $0.882 \leq R^2_{pre} \leq 0.995$ in calibration and prediction set respectively. The application of a handheld NIR spectrometer and effective variable selection algorithms could be employed as a simple, cost-effective, reliable, rapid, ecofriendly, and non-destructive technique for the identification of organic cocoa beans from conventional ones to prevent fraud and ensure the integrity of organic cocoa beans, as well as for simultaneous prediction of quality parameters.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER SIX

INFLUENCE OF AGE OF COCOA TREE AND POLLINATION TYPE ON SOME PHYSICAL AND BIOCHEMICAL PROPERTIES OF THE COCOA BEANS

ELLIOT K. ANYIDOHO, ERNEST TEYE, ROBERT AGBEMAFLE

Department of Agricultural Engineering, School of agriculture, College of Agriculture and Natural Sciences, University of Cape Coast
Department of Laboratory Technology, School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast

Submitted (under review): Measurement: Food

Statement of Contributions

Elliot K. Anyidoho (Candidate)

Writing and compiling of original manuscript, investigation, data collection and formal analysis, established methodology, preparation of tables and figures

Ernest Teye (Principal Supervisor)

Supervised manuscript compilation, editing, and co-author of manuscript

Robert Agbemafle (Co-Supervisor)

Supervised manuscript compilation, editing and co-author of manuscript

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Introduction

Cocoa (*Theobroma cacao L*) belongs to the botanical family Malvaceae. Its seed, cocoa bean is the single most important tropical agricultural merchandised commodity in the world (Afoakwa et al., 2013a). Many cocoa-growing countries especially West African countries derive huge economic benefits from cocoa production. Specifically, foreign earnings from the cocoa bean trade have been the backbone of Ghana's economy for a century, contributing 6 - 7 % of gross domestic product (GDP). The cocoa crop contributes more than 67% of household income for Ghana's producing households of over 2 million people (Ghana Cocoa Board, 2018).

Commercial cocoa bean is an essential raw material in confectionery industries and businesses which deal with the production of several kinds of food items such as chocolates, biscuits, sweets, ice-creams and cakes. It is the main ingredient for cocoa powder, cocoa butter, and cocoa liquor makers and this has led to the continuously growing demand for cocoa beans. Cocoa beans are traded locally and internationally with quality standards of regulations on the physical and biochemical properties of the beans (Adeyeye et al., 2010). This usually demands the need for aggressive inspections of the bean quality. Hence, commercial cocoa beans are excepted to be of high physical and chemical quality.

Physical properties of cocoa beans just like other agricultural seeds and grains are required during the design and development of structures and instruments for planting, harvesting, conveying, separation, cleaning, grading, handling and storage (Bart-Plange & Baryeh, 2003). Also, the popularity of and

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

demand for cocoa beans' products are related to their exceptional biochemical properties that offer numerous nutritional benefits and health effects. The biochemical properties are key parameters that determine the cocoa bean quality. This is contingent on the quantum of proximate compositions, phytochemicals and minerals found in cocoa beans (Adeyeye et al., 2010). Therefore, to ensure quality, the cocoa bean industry desires to measure the physical and biochemical properties of cocoa beans.

However, there are several factors that affect the physical and biochemical quality properties of cocoa beans that must be considered. These factors include variety, production method, geographic location, climate, fermentation, type of pollination and age of the tree (Santander et al., 2020; Kongor et al., 2016). Among these factors, the pollination type and tree age are the most important. Nevertheless, studies conducted on pollination and cocoa tree age focused on the impact on yield with no study on the physical and chemical properties of the beans. This research was therefore designed to study the physical, biochemical and mineral properties of cocoa beans as influenced by age of the tree and the type of pollination.

Materials and Methods

Sites and cocoa bean samples for determination of the influence of tree age and pollination type

Artificially pollinated cocoa bean samples and naturally pollinated ones for the study were obtained from the fields of the Seed Production Division (of Ghana Cocoa Board) at Saamang. Artificially pollinated samples of 40 kg beans for 10 years, 40 kg for 15 years and 40 kg for 20 years cocoa trees making a

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

total of 120 kg for all three age groups were collected. This category of cocoa beans was obtained from artificial hand pollination. Samples from artificially hand pollinated 10 years trees, 15 years trees and 20 years trees were named AP10, AP15 and AP20, respectively. Likewise, 40 kg naturally pollinated cocoa bean samples for 10 years, 15 years and 20 years cocoa trees making a total of 120 kg for all three age groups were collected. This category of cocoa beans was obtained from natural pollination. Samples from naturally pollinated 10 years, 15 years and 20 years trees were named NP10, NP15 and NP20, respectively. The samples were thoroughly mixed for each category for examination of physical, biochemical and mineral properties. This study was performed using a completely randomized design with age class and pollination type as the principal factors.

Biochemical Properties Measurement

Protein, fat, fibre and carbohydrate content measurements

The cocoa beans were analysed for the content of moisture, crude fat, crude fibre, crude protein, ash and carbohydrate based on the methodology described in the AOAC, (2005). The moisture content was determined by oven drying procedure at 130°C±2 by following AOAC 931:04 method till constant weight was attained. Crude fat content was quantified based on the AOAC 963:15- a Soxhlet extraction method. The crude fibre content was assessed by AOAC 991 method, crude protein content was determined by measuring the total nitrogen content with Kjeldahl method (AOAC 970:22). The ash content was assessed by AOAC 972:15 and carbohydrate content was obtained by

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

difference of constituents (Abdullahi et al., 2018). The outcomes were stated in percentage (g/100 g) of the sample.

pH

The pH for the powdered cocoa bean was examined per the method proposed by the International Office of Cocoa & Confectionary, (1996). Five (5) grams of previously powdered samples were poured into conical flasks and 50 ml of boiling distilled water was added. The mixture was then allowed to stand for 30 minutes. Whatman filter paper No. 4 was used to filter the extract which was then allowed to cool to a temperature of 27°C. The pH value of the resultant filtrate was measured by using a digital pH meter (Jenway Model 3510, Wagtech Bibby Scientific Limited, UK).

Total polyphenol content (TPC)

The TPC was estimated using the Folin-Ciocalteu reagent method as described in Arslan et al. (2017) with slight modification. A 1 ml of cocoa bean extract was added to 5 ml of Folin-Ciocalteu reagent (ten-fold diluted) and 4 ml (75 g/l) of Na_2CO_3. The mixture was allowed to incubate in the dark for 2 hours. The absorbance of the reaction mixture was measured at 765 nm using the UV-VIS spectrophotometer (UNmini-1240, Shimadzu Corporation, Tokyo-Japan). Gallic acid solution (20-100µg/ml) was used as a standard for the creation of a calibration curve. The attained results were expressed as milligrams gallic acid equivalents per gram (mg GAE g^{-1}) of the cocoa bean.

Total flavonoid content (TFC)

The TFC was measured by following the method described by Ordonez et al. (2006). The cocoa bean was extracted in 80% ethanol and kept for 24

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

hours. Briefly, 0.5 ml of 10% aluminum chloride ($AlCl_3$) and 2.5 ml of distilled water were added to 0.5 ml of cocoa bean extract and the reaction mixture was allowed to stand for 1 hour. The absorbance was recorded at 510 nm using UV-VIS spectrophotometer (UNmini-1240, Shimadzu Corporation, Tokyo-Japan). The quercetin standard solution was used to create the calibration curve and the required results were expressed as milligrams of quercetin equivalents per gram (mg QE g^{-1}) of the cocoa bean.

Total antioxidant capacity (TAC)

The TAC of the cocoa bean sample was determined using the method of Sun et al. (2005). 1 ml of cocoa bean extract was mixed with 0.02 mg/mL 2,2-diphenyl-1-picrylhydrazyl radical (DPPH) solution. The glass tube containing the reaction mixture was placed in the dark for 30 minutes at room temperature. Absorbance was measured at 590 nm using UV-VIS spectrophotometer (UNmini-1240, Shimadzu Corporation, Tokyo-Japan). The ascorbic acid standard solution was used to draw the calibration curve and total antioxidant capacity was determined and expressed as milligrams of ascorbic acid equivalent antioxidant content per gram (mg AEAC g^{-1}) of the cocoa bean.

Mineral Properties Measurements

To analyse the mineral properties, the cocoa beans were milled and stored in 50 ml glass. 0.3 g of the ground cocoa beans were used for acid digestion by perchloric nitric solubilization. Mineral properties were analysed from the obtained solutions by dry-ashing the milled samples at 550 °C to get white ashes. The white ashes were dissolved into a flask using hydrochloric acid solution 2 cm^3 of 70% HCIO and distilled water. Ca, Mg, K, Na, Fe, Cu and Zn

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

were analysed employing atomic absorption spectrophotometer whilst P was calorimetrically analysed using a Spectronic 20 UV/ VIS Spectrophotometer (AOAC, 2005). Chemicals used for the analysis were of standard grade. Before the analysis Spectrophotometer was calibrated using a metal analytical solution.

Cocoa Beans Physical Properties Measurements

In the determination of some physical properties of cocoa beans, the following equipment, tools and arithmetic algorithm were deployed:

Geometric properties

i. To measure the geometric dimensions, 100 individual cocoa beans were randomly picked from the (remainder of the) 12 kg of each age group sample and the three major geometric dimensions namely length (L), width (W) and thickness (T) were measured by digital micrometer screw gauge reading to 0.01 mm.

ii. Sphericity (Ø) of the cocoa beans were determined by using the classical equation proposed by Mohsenin (1980):

$$Ø = \frac{(LWT)^{\frac{1}{3}}}{L} \quad (6.1)$$

Gravimetric properties

i. Cocoa bean mass and the 1000-bean mass were measured using a digital electronic balance with a sensitivity of 0.01g. Twenty samples each of 100-beans from the three age categories were randomly picked and weighed. The measured mass of each sample was divided by 100 to obtain the cocoa bean mass. The mass of 1000-cocoa beans was attained by counting twenty 1000-bean samples for the desired age category and weighed. To obtain the 1000-bean mass of each age category of samples, the weights were averaged. Similar procedures have been applied by other researchers including Bart-Plange et al. (2011) and Sandoval et al. (2019).

Digitized by Sam Jonah Library

ii. In the determination of bean density, 100 cocoa beans were randomly selected from each age category sample and the mass weighed. The beans were used to displace methanol in a measuring cylinder. The bean density was found as the average of the ratio of beans masses to the volume of methanol displaced by the cocoa beans. For the bulk density, an empty predefined container of 1000 ml was weighed and was subsequently filled with the desired cocoa bean age category from a height of about 1.5 m at a constant flow rate and the top leveled with a ruler without manual compaction. The bulk density for desired cocoa bean age category was determined as the ratio of the measured mass of the beans only to the volume occupied by the beans. For each age category, three replications were carried out and the average was reported.

iii. Porosity (P), a void space in the beans which is not occupied by the bean was computed from the values of bean density (ρ_p) and bulk density (ρ_b) by following the arithmetic algorithm in equation (6.2) proposed by (Mohsenin, (1968) and employed by other researchers such as Abano & Amoah, (2011), Bart-Plange & Baryeh, (2003).

$$P=\left(1-\frac{\rho_b}{\rho_p}\right)100 \qquad (6.2)$$

Repose angle

i. In determining the repose angle (Θ), a hollow cardboard cylindrical mould of 200 mm diameter and plywood plate were used. The cylinder was filled with the desired cocoa bean category and removed slowly in the upwards direction permitting the beans to fall gradually till it was empty to form a natural conical heap of cocoa beans. The height (H) and diameter (D) of the conical heap were recorded. The repose angle was computed by using the classical trigonometry rule as shown in equation (6.3). This procedure was employed previously by Bart-Plange et al. (2003) for maize.

$$\theta=\tan^{-1}\left(\frac{Opposite}{Adjacent}\right)=\tan^{-1}\left(\frac{2H}{D}\right) \qquad (6.3)$$

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Software tool

All the data were analysed using the GenStat Release 12.1 statistical software. Analysis of variance (ANOVA) and Duncan's multiple-comparison test ($p < 0.05$) were used to compare statistical differences between the means of the results. Matlab version 9.6.0 1072779, License Number 968398, the MathWorks, Inc., USA) with windows 10 Pro software packages for processing was used to plot graph for the physical properties.

Results and Discussion

Physical properties of cocoa beans

The mean moisture content, in all measured values, were 7.24, 7.21 and 7.25 % wet basis for cocoa beans obtained from N10, N15 and N20 and 7.23, 7.22 and 7.20 % (wet basis) for cocoa beans obtained from A10, A15 and A20, respectively (Table 6.1). The mean moisture content of all the cocoa bean categories examined was within the internationally acceptable critical limits of 6 - 8 percent wet basis for safe storage of cocoa beans (Wood & Lass, 2008). The obtained moisture content depicted that the cocoa beans would not get mouldy (below 7.50 %) nor get cracked easily because they were above 6.00 %

Geometric properties (length, thickness, width, sphericity), gravimetric properties (bean mass, 1000-bean mass, bean density, bulk density, porosity) and repose angle of each cocoa bean category were investigated to develop suitable equipment for separating, sizing, transportation, packaging, sorting, grinding and quality checks which will help decrease post-harvest losses and hence boost productivity. Cocoa bean shape can be determined in relation to its sphericity which influences the flow rate attributes. Bean mass, 1000-bean

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

mass, bean density and bulk density of the cocoa beans could be used in determining the storage bin size and also influence the structural loads. In designing transporting and storage structures the repose angle is an important factor. Porosity affects the airflow resistance through the bulk cocoa bean bed and hence data on it are essential in designing the drying process.

Influence of tree age and pollination type on cocoa bean physical properties

Geometric properties

Length, width and thickness

Data on geometric dimensions such as length (L), width (W) and thickness (T) of cocoa beans are needed to develop suitable equipment for efficient packaging, sorting, separation, transportation, sizing, grinding and quality checks which will help decrease post-harvest losses and hence boost productivity. Cocoa bean physical properties as influenced by age of tree and type of pollination are presented in Tables 6.2. Cocoa bean L, W, and T were significantly ($p < 0.05$) influenced by tree age and pollination type. The largest mean geometric dimensions- L (24.64±1.21 mm), W (13.63±1.13 mm) and T (9.32±1.21 mm) were found from the A20, whereas least mean geometric dimensions- L (21.03±1.36 mm), W (10.26±1.42 mm) and T (6.56±1.04 mm) were obtained from the N10. Variation in L, W & T with tree age and pollination type are shown in Fig. 6.1 (A)-1(C). It could be observed that these geometric dimensions increased with the age of the cocoa tree and the type of pollination. The measured geometric dimensions for A20 are statistically larger than Category B cocoa beans examined by (Bart-Plange & Baryeh, 2003) which were 22.50, 12.86 and 7.70 mm for L, W and T, respectively. However, these

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

values were not statistically different from 23.80 mm of L, 13.50 mm of W and 9.60 mm of T reported by (Sandoval et al., 2019) for Venezuelan Trinitario cocoa beans at a moisture content of 6.51 percent wet basis. ANOVA analysis showed that cocoa bean geometric dimensions were significantly ($p < 0.05$) influenced by age of the cocoa tree and pollination type.

Sphericity

Data presented in Table 6.1 showed that A20 gave the highest mean sphericity value of 0.59±0.03 though not significantly different from the mean value of 0.58±0.04 obtained from the N20 and mean value of 0.58±0.03 obtained from the A15. N10 gave the lowest mean sphericity value of 0.53±0.03. Variation in cocoa bean sphericity with cocoa tree age and type of pollination are shown in Fig. 6.1 (D). It could be observed that cocoa bean sphericity numerically increased with the age of the tree and type of pollination. The data from this study disclosed that cocoa bean shape is relatively different from a spherical shape and that age of cocoa tree and type of pollination did not influence cocoa bean sphericity. Cocoa bean sphericity indicates the ability of the bean to roll rather than slide in a delivery tube or bean hopper or conveying machine due to its ellipsoid shape.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 6.1: Measurement of cocoa bean physical properties as influenced by age of cocoa tree and type of pollination

Physical properties	Statistics	Natural Pollination			Artificial Pollination		
		N10	N15	N20	A10	A15	A20
Length (mm)	Range	17.53-24.55	18.79-25.51	17.83-25.82	18.81-25.02	20.00-25.54	21.25-26.68
	Mean	21.03±1.36^{e}	22.36±1.76^{d}	23.54±1.94^{b}	22.21±1.56^{d}	22.92±1.31^{c}	24.64±1.21^{a}
Width (mm)	Range	4.78-12.75	8.73-13.49	9.34-14.87	9.98-14.74	10.32-14.32	11.11-15.25
	Mean	10.26±1.42^{e}	11.43±1.18^{d}	12.52±1.32^{b}	12.10±1.15^{c}	12.48±0.95^{b}	13.63±1.13^{a}
Thickness (mm)	Range	4.97-9.65	5.02-11.35	5.97-11.48	5.64-9.86	6.24-9.89	7.27-11.11
	Mean	6.56±1.04^{e}	7.28±1.33^{d}	8.79±1.45^{b}	7.72±1.18^{c}	7.90±1.04^{c}	9.32±1.21^{a}
Sphericity (decimal)	Range	0.51-0.58	0.53-0.61	0.54-0.63	0.52-0.61	0.57-0.68	0.58-0.72
	Mean	0.53±0.03^{c}	0.54±0.02^{b}	0.58±0.04^{a}	0.54±0.03^{b}	0.58±0.03^{a}	0.59±0.03^{a}
Bean mass (g)	Range	1.23-1.26	1.26-1.32	1.31-1.35	1.26-1.31	1.30-1.33	1.32-1.37
	Mean	1.25±0.01^{d}	1.28±0.02^{c}	1.33±0.01^{b}	1.29±0.01^{c}	1.32±0.01^{b}	1.35±0.02^{a}
1000- bean mass (g)	Range	1236.32-1241.66	1238.89-1243.81	1240.98-1244.58	1238.31-1243.29	1241.58-1245.22	1244.44-1247.65
	Mean	1239.31±1.70^{d}	1241.89±1.64^{c}	1242.97±1.13^{b}	1241.05±1.53^{c}	1243.05±1.22^{b}	1245.81±1.24^{a}
Bean density (kg/m^{3})	Range	873.14-897.00	913.29-938.71	920.64-952.14	908.86-924.86	928.64-952.93	939.86-962.43
	Mean	883.53±6.90^{e}	929.25±8.38^{c}	939.33±10.81^{b}	916.33±5.38^{d}	939.69±7.61^{b}	953.60±7.36^{a}

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Bulk density (kg/m^3)	Range	671.07-673.25	693.30-697.70	707.96-709.73	696.40-710.38	706.33-709.97	718.19-723.40
	Mean	671.79±0.65^d	696.13±1.36^c	708.85±0.57^b	699.54±1.57^c	708.20±1.23^b	721.56±1.59^a
Porosity (%)	Range	19.70-21.76	21.45-23.59	23.17-27.77	22.52-22.64	22.95-25.25	24.98-31.11
	Mean	20.56±0.61^e	22.81±0.67^d	26.73±1.09^b	22.56±0.04^d	23.98±0.71^c	28.14±2.20^a
Repose angle (°)	Range	29.15-30.06	27.30-29.15	24.50-26.37	26.37-29.15	23.55-27.30	21.66-25.44
	Mean	29.51±0.46^a	27.76±0.56^b	25.11±0.63^c	27.67±1.10^b	25.53±1.25^c	23.22±1.12^d
Moisture (% wb)	Range	7.23-7.44	7.15-7.31	7.14-7.35	7.11-7.28	7.10-7.19	7.02-7.26
	Mean	7.35±0.05^a	7.21±0.04^a	7.26±0.05^a	7.19±0.06^a	7.15±0.05^a	7.17±0.06^a

Note: Means and standard deviations in each row not followed by the same superscripted alphabets are significantly different at $p < 0.05$.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

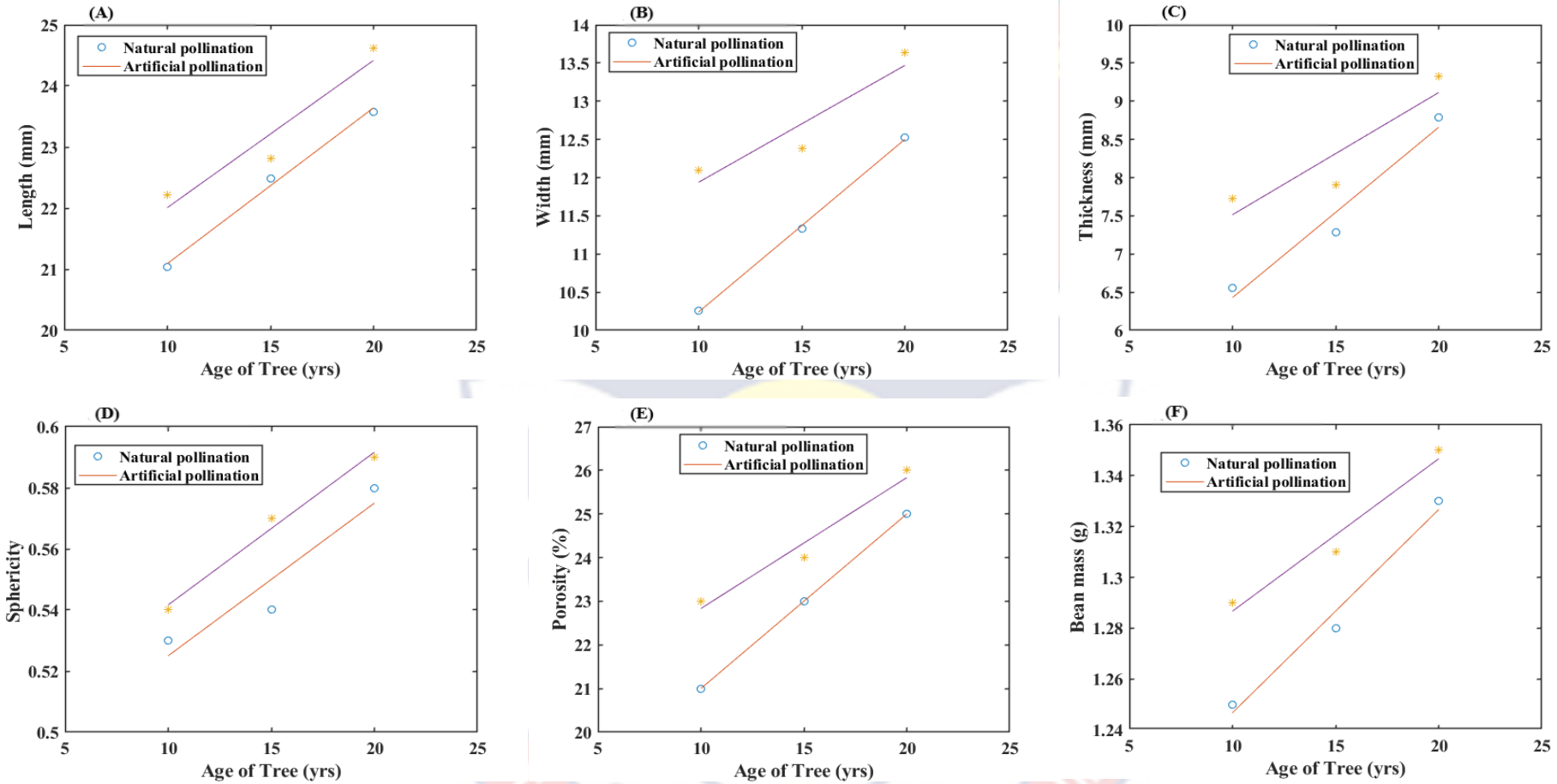


Fig. 6.1 Variation of (A) length, (B) width, (C) thickness, (D) sphericity, (E) porosity (F) bean mass (G) 1000-bean mass (H) bulk & particle densities and (I) repose angle with the age of tree and type of pollination

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

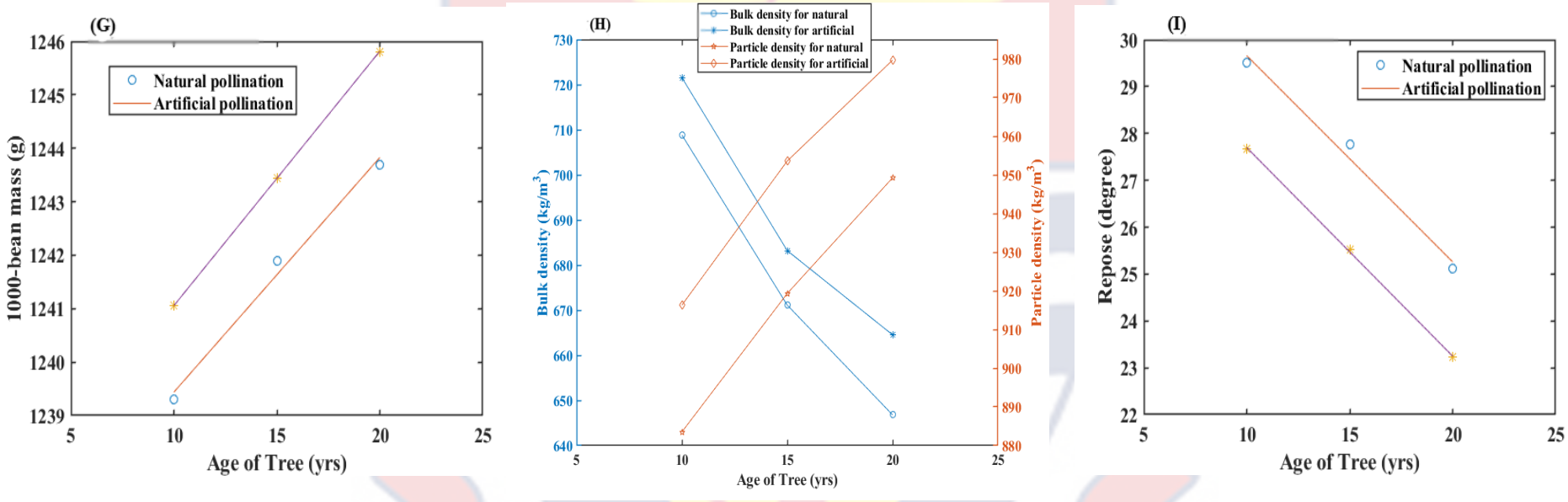


Fig. 6.1 Continued.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Gravimetric properties

Cocoa bean mass and 1000-cocoa bean mass

Cocoa bean mass and 1000-cocoa bean mass are attributes that help in the design and simulation of bean processes and they give information about bean quality and its acceptability by different groups of customers. Results displayed in Table 6.1 showed that cocoa bean mass and 1000-bean mass increased with an increase in tree age and pollination type. It was also observed that interaction of cocoa trees age and artificial pollination produced significant ($p < 0.05$) higher values of the cocoa bean mass and 1000-bean mass than the interaction of tree age and natural pollination. A20 recorded the highest mean bean mass and 1000-bean mass of 1.35±0.02 g and 1245.81±1.24 g whilst N10 recorded the least mean bean mass and 1000-bean mass of 1.25±0.01 g and 1239.31±1.70 g respectively. Variation in the cocoa bean mass and 1000-bean mass with the age of the tree and type of pollination are shown in Fig. 6.1 (F) and (G), respectively. Graphically, it could be seen that cocoa bean mass and 1000-bean mass increased with tree age and type of pollination. Different works have reported that average bean mass for Category B cocoa beans from Ghana ranged between 1.11g and 1.31g (Bart-Plange & Baryeh, 2003), average bean mass for Mexican Criollo Cacao ranged between 1.38 and 1.48 g (García-Alamilla et al., 2012), average 1000-bean mass for Category D cocoa beans varied between 948.75 and 1112g whereas average 1000-bean mass for Category B cocoa beans from Ghana varied between 1125 and 1247g (Bart-Plange et al., 2011). Cocoa bean mass and 1000-cocoa bean are essential because they mostly influence yield, revenue and consumer preference. It is

Digitized by Sam Jonah Library

expected that cocoa bean weight which is one of the critical quality indicators should not be less than one gram (Wood & Lass, 2008) and this standard has been met by all the categories of cocoa beans studied in this research.

Bean density and bulk density

Understanding the bean density is important for designing storage bags or systems for determination of purity of seeds, evaluation of maturity and measurement of dielectric attributes of grains (Karimi et al., 2009). Bulk density is an essential property for cleaning, prediction of loads and pressures on storage systems or structures. Table 6.1 showed that A20 recorded the highest mean bean density and bulk density of 953.60±7.36 kg/m^3 and 721.56±1.59 kg/m^3 whilst N10 recorded the least mean bean density and bulk density of 883.53±6.90 kg/m^3 and 671.79±0.65 kg/m^3, respectively. Fig. 6.1 (H) depicted the variation in bean particle and bulk density with cocoa tree age and pollination type. It was detected that age of tree and type of pollination had significant ($p < 0.05$) influence on cocoa bean density and bulk density.

Porosity

The resistance of bulk beans to the flow of air is the function of bean size and porosity. It ensures better exchange of heat, aeration during heating, cooling and drying operations. In Table 6.1, the highest porosity was found to be 28.14±2.20 % for the A20, whereas the least porosity value of 20.56±0.61 % was obtained for the N10. Fig. 6.1 (E) graphically displayed variation in porosity with cocoa tree age and type of pollination. It could be seen that porosity increased significantly ($p < 0.05$) with cocoa tree age and type of pollination. The mean porosity values of all the cocoa bean categories studied

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

are statistically lower than the 43.20 % obtained for Venezuelan Trinitario cocoa beans (Sandoval et al., 2019) and 45.60% for the thick bed of cocoa beans under stationary and transient inlet conditions but not significantly different from 31.59 percent obtained for Category B cocoa beans (Bart-Plange & Baryeh, 2003).

Repose angle

The repose angle tells the maximum angle at which a heap of the cocoa beans will stand without sliding. It determines the maximum angle of the pile cocoa bean on the horizontal surface. The knowledge of repose angle is useful for designing machines for mass flow, storage systems, and determining contour piles. Data presented in Table 6.1 showed that N10 recorded the highest mean value of repose angle (29.51±0.46°) whereas the A20 recorded the minimum mean value of repose angle (23.22±1.12°). Variation in repose angle with the age of the tree and type of pollination are shown in Fig. 6.1 (I). The interaction of tree age and pollination type had a decreasing influence on cocoa bean repose angle. The mean repose angle data obtained from all the cocoa bean categories studied were statistically higher than the mean value of 21.8° reported for Venezuelan Trinitario cocoa beans. (Sandoval et al., 2019). However, the results were within the average values of 23.74° and 33.81° recorded for Category B at a moisture content of 5.67% and 22.00% wet basis, respectively (Bart-Plange & Baryeh, 2003).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Influence of tree age and pollination type on proximate properties of cocoa bean

The mean moisture content, in all measured values, were 7.24, 7.21 and 7.25 % wet basis for N10, N15 and N20 and 7.23, 7.22 and 7.20 % wet basis for A10, A15 and A20 respectively (Table 6.2). The obtained mean moisture content of all the cocoa bean categories is within the internationally acceptable critical limits of 6 to 8 percent wet basis for safe storage of cocoa beans (Wood & Lass, 2008). The obtained moisture content depicted that the cocoa beans would not get mouldy (below 7.50 %) nor get cracked because they were above 6.00 %. The moisture content of the categories of cocoa beans examined was not significantly different.

Protein content varied from 12.56 – 15.29% which is relatively lower than the literature values of 16 - 22% reported by Aremu et al. (1995) and Quao, et al. (2013) but significantly higher than the literature data of 10.29 - 13.34 reported by Assa et al. (2017). It is seen in Table 6.2 that there were general increases in cocoa bean protein content with pollination type. Similarly, general increments in protein content were seen as cocoa trees get older. ANOVA check showed that tree age and pollination type significantly ($p < 0.05$) influenced cocoa bean protein content. N20 recorded the highest protein content of 15.29% while the N10 yielded the lowest protein content of 12.56%.

Fat content of merchantable cocoa is an essential quality indicator parameter for cocoa processors during buying of cocoa beans. Generally, fat content of cocoa cocoas as reported in literature fall between 53 and 59% (Wood & Lass, 2008). It was observed that data of cocoa bean fat contents (42.22 –

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

48.77 %) from samples examined in this study (Table 6.2) were comparatively lesser. However, fat concentrations recorded in this study were higher than the values (42.65 to 43.53 %) reported for cocoa beans from Ghana (Maalekuu & Teye, 2017). General increases were observed in cocoa bean protein content with pollination type and similar increments were seen as trees get older. ANOVA check indicated that protein content was statistically ($p < 0.05$) impacted by cocoa tree age and pollination type. Fat content in cocoa beans of artificial pollination increased from 43.35 to 48.77 % with the interaction of A20 recording the highest fat content of 48.77 %. Although numerically cocoa beans obtained from the N15 had a higher fat content of 44.92 % compared to the 45.52 % fat content obtained from the A15, there was no statistical difference between the two samples. Stress due to drought or heat, ambient temperature, pollination type, tree age, genetic and geographic origins affected biosynthesis (Afoakwa, 2010).

Fibre content varied from 7.88 to 9.49 %, being highest for the N10 and lowest for the A15 (Table 6.2). These values were significantly higher than the values of 3.1 to 5.9 % reported by Aremu et al. (1995). ANOVA analysis showed that fibre content was statistically ($p < 0.05$) influenced by cocoa tree age and pollination type. Fibre content in cocoa beans of natural pollination decreased from 9.49 to 8.09 % and artificial pollination decreased from 9.04 to 7.99 %. The fibre content in cocoa beans was lower for older trees (15 and 20 yrs) compared to younger cocoa trees (15 and 20 yrs). General reductions in cocoa bean fibre content with pollination type and similar reductions were seen as trees get older.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

The ash content ranged from 2.60 to 3.76 %, being lowest for N10 and highest for A20 (Table 6.2) and this was comparable to literature values of 2.50 to 3.70 % for Vietnamese cocoa beans (Lam Thi et al., 2016) but lower than literature values of 3.90 to 4.4 % for raw and fermented cocoa beans from Nigeria (Aremu, Agiang, & Ayatse, 1995). From Table 6.2, the ash content of naturally pollinated cocoa beans and artificially hand-pollinated cocoa beans showed a pattern of increasing ash content with an increase in tree age. ANOVA showed that the increase in the ash contents because of tree age and artificial hand pollination were significant ($p < 0.05$).

Carbohydrate content in the cocoa bean samples studied ranged from 18.79 to 25.87 % (Table 6.2) which is comparable to values of 15.47 to 24.93 % in literature (Afoakwa et al., 2013c). Although cocoa beans obtained from the A15 had higher carbohydrate content of 22.03 % compared to the 21.88 % carbohydrate content obtained from the N15, there was no significant difference between the two samples. Cocoa bean carbohydrate content reduced with pollination type, and similar reductions were noticed as the trees grow older. Carbohydrate content in cocoa beans was higher for younger trees (10 yrs) of natural pollination compared to older cocoa trees (15 and 20 yrs). There was a significant ($p < 0.05$) reduction in cocoa bean carbohydrate content due to the influence of age of tree and pollination type.

Influence of tree age and pollination type on phytochemical properties of cocoa bean

pH value of 4.87 to 4.98 was recorded for 10-year-old trees, 5.47 to 5.51 for 15-year-old trees and 6.05 to 6.40 for 20-year-old trees cocoa bean samples.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

ANOVA check showed that pH content was statistically ($p < 0.05$) influenced by cocoa tree age and pollination type (Table 6.3). pH content in cocoa beans of natural pollination increased from 4.87 to 6.05 while that of artificial pollination increased from 4.98 to 6.40 %. The pH examinations showed an increase in pH values as trees grow older.

TPC in the cocoa bean samples studied varied from 63.60 to 75.02 mgGAE/g (Table 6.3). These obtained values are significantly high than literature values of 34.93±4.89 mg/g reported for Malaysian cocoa beans (Ramli et al., 2001). However, the obtained TPC values fall within the values of 59.55 to 153.58 mg/g reported for Ghana cocoa beans (Afoakwa et al., 2013c). The observed differences and similarities are to be possible because research has reported that TPC is influenced by factors such as climatic condition, degree of fermentation, variety, farming practices, and location (Aikpokpodion & Dongo, 2010). Total polyphenol content in cocoa beans was higher for younger trees (10 yrs) of artificial pollination compared to older cocoa trees (15 and 20 yrs). There was a significant ($p < 0.05$) decline in cocoa bean TPC due to the influence of age of tree and pollination type.

The average TFC of the cocoa beans examined varied from 78.31 to 83.57 mg/g (Table 6.3). The TFC was highest in the A20 samples, followed by the N20, A15, N15, A10 and N10 samples. The TFC values obtained in the study fall within the values of 35.03 mg/g to 126.21 mg/g reported for cocoa bean samples from China, Indonesia and Papua New Guinea (Gu et al., 2013). Significant differences ($p< 0.05$) were found in TFC for all the cocoa bean categories studied, except for N10 and A10, N15 and A15, and N20 and A20

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

which showed an absence of significant changes ($p < 0.05$). TFC has antioxidant property that improves cardiovascular health (Awarikabey et al., 2014). Thus, the existence of these metabolites in the cocoa bean samples indicates a positive signal of improving human health.

The average TAC of cocoa beans examined in the study ranged from 96.30 to 128.06 mg/g (Table 6.3). The TAC was highest in A20. The values obtained for TAC were higher than the TAC values of 52.29 to 72.92 mg/g reported by Awarikabey et al. (2014) probably due to differences in cocoa beans geographic location, variety, age and differences in extraction methodologies. Naturally pollinated cocoa beans and artificially hand-pollinated cocoa beans showed a pattern of increasing the TAC with increasing cocoa tree age. Antioxidants are micronutrients that can inhibit or delay lipid oxidation and can also involve in scavenging reactive oxygen species and reactive nitrogen species which are responsible for the pathogenesis of several human disorders or diseases (Di Castelnuovo et al., 2012). Therefore, the outcomes of this study on the bioavailability of cocoa bean antioxidant capacity give a positive indication of promoting human health.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 6.2: Influence of tree age and pollination type on proximate compositions of cocoa beans

Age (Yrs)	Pollination type	Proximate compositions					
		Moisture (%)	Protein (%)	Fat (%)	Fibre (%)	Ash (%)	Carbohydrate (%)
10	Natural	7.24±0.04^{a}	12.56±0.04^{f}	42.22±0.14^{e}	9.49±0.02^{a}	2.60±0.04^{c}	25.87±0.26^{a}
	Artificial	7.23 ±0.08^{a}	12.90±0.01^{e}	43.35±0.05^{d}	9.04±0.03^{b}	2.82±0.08^{e}	24.62±0.28^{b}
15	Natural	7.21±0.04^{a}	14.34±0.04^{b}	44.92±0.21^{c}	8.13±0.02^{c}	3.48±0.04^{d}	21.88±0.27^{c}
	Artificial	7.22±0.08^{a}	13.91±0.02^{d}	45.52±0.32^{c}	7.88±0.05^{f}	3.69±0.09^{b}	22.03±0.44^{c}
20	Natural	7.25±0.05^{a}	15.29±0.10^{a}	46.55±0.08^{b}	8.09±0.03^{d}	3.67±0.10^{b}	19.81±0.52^{d}
	Artificial	7.20±0.06^{a}	14.17±0.12^{c}	48.77±0.14^{a}	7.99±0.02^{e}	3.76±0.08^{a}	18.79±0.27^{e}
Lsd		0.09	0.13	0.71	0.02	0.17	0.23

Means in each column not followed by the same superscripted alphabets are significantly different at $p < 0.05$.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 6.3: Influence of tree age and pollination type on phytochemical properties of cocoa beans

Age (years)	Pollination type	Phytochemical properties			
		pH (value)	TPC (mgGAE/g)	TFC (mgQE/g)	TAC (mgAEAC/g)
10	Natural	4.87±0.02^{d}	71.53±1.03^{b}	78.31±1.32^{b}	96.30±1.31^{d}
	Artificial	4.98±0.03^{d}	75.02±1.05^{a}	78.61±2.04^{b}	101.04±1.22^{c}
15	Natural	5.47±0.28^{c}	66.08±1.27^{d}	81.04±1.85^{a}	110.69.±2.03^{b}
	Artificial	5.51±0.26^{c}	70.18±1.04^{c}	81.24±2.23^{a}	111.81±2.25^{b}
20	Natural	6.05±0.28^{b}	63.60±1.18^{f}	83.43±1.68^{a}	127.23±2.11^{a}
	Artificial	6.40±0.22^{a}	65.09±1.05^{e}	83.57±1.34^{a}	128.06±3.02^{a}
Lsd		0.14	0.73	0.14	0.73

Means in each column not followed by the same superscripted alphabets are significantly different at $p < 0.05$.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Influence of tree age and pollination type on cocoa bean mineral properties

Ca content reduced from 125.23 to 111.44 mg/100 g in the cocoa bean samples studied (Table 6.4). Table 6.3 showed that Cu content in cocoa beans was higher for younger cocoa trees (10 yrs) compared to older cocoa trees (15 and 20 yrs). It could be seen that there were general decreases in Cu content of the cocoa beans with tree age. A similar pattern of declining Cu content of cocoa beans was seen as the trees get older. Although there was a numerical difference between N10 and A10, there was no significant difference between the two samples (Table 6.4). ANOVA analysis showed that the Cu content of the cocoa bean samples was significantly ($p < 0.05$) influenced by cocoa tree age.

Mg content of 238.79 to 241.49 mg/100 g was recorded for 10-year-old trees, 245.38 to 247.82 mg/100 g for 15-year-old trees and 248.38 to 249.05 mg/100 g for 20-year-old trees cocoa bean samples (Table 6.4). Although it there appeared that the Mg content of the cocoa bean samples increased numerically with cocoa tree age and pollination type, there was no significant statistical difference ($p < 0.05$) among the samples.

P content in the cocoa bean samples varied from 528.24 to 541.40 mg/100 g, being highest for the A15 and lowest for the A20 (Table 6.4). P content in the cocoa bean samples studied did not follow any specific trend. ANOVA examinations showed that cocoa tree age and pollination type did not have a significant ($p < 0.05$) influence on the P content of cocoa beans samples.

K content of 629.18 to 631.34 mg/100 g was found to be the highest in cocoa bean samples obtained from 10 yrs trees (Table 6.4). Generally, K content in the samples decreased with cocoa tree age and pollination type. ANOVA

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

analysis showed that the P content of the cocoa bean samples examined was significantly ($p < 0.05$) influenced by cocoa tree age pollination type.

Na content of the cocoa bean samples studied varied from 7.08 to 11.54 mg/100g, being highest for N10 and lowest for A20 (Table 6.4). A general pattern of reduction in Na content in cocoa bean samples with tree age and pollination type was noticed (Table 6.4). ANOVA revealed that cocoa tree age significantly ($p < 0.05$) influenced Na content in the cocoa bean samples. Similarly, pollination type also influenced the Na content in the cocoa beans.

Fe content of the cocoa bean sample generally did not follow any specific pattern. It ranges from 5.80 to 8.83 mg/100 g, being highest for the N15 and lowest for A20 (Table 6.4). ANOVA analysis demonstrated that cocoa tree age, pollination type and their interactions had a significant ($p < 0.05$) influence on the Fe content of the cocoa bean samples studied.

Cu content in the cocoa bean samples increased with tree age: 1.34 to 1.42 mg/100 g, 1.66 to 1.83 mg/100 g and 3.24 to 3.33 mg/100 g for cocoa trees at ages 10, 15 and 20 respectively (Table 6.4). ANOVA check revealed that pollination type had no significant effect on the Cu content of the cocoa beans.

Zn content in the cocoa bean sample generally increased significantly ($p < 0.05$) as the age of the trees increased. In a contrary manner, ANOVA analysis revealed that pollination type did not influence Zn content in the cocoa bean samples investigated. Zn content in cocoa beans was lower for younger trees (10 yrs) compared to older cocoa trees (15 and 20 yrs). Cocoa beans obtained from 20 yrs trees recorded the highest Zn content of 5.09 to 5.14 mg/100 g (Table 6.4).

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Table 6.4 Influence of tree age and pollination type on cocoa bean mineral properties

Age (Yrs)	Pollination type	Mineral properties in mg/100 g							
		Ca	Mg	P	K	Na	Fe	Cu	Zn
10	Natural	125.23±6.26^{a}	238.79±17.13^{a}	539.31±19.94^{a}	631.34±4.32^{a}	11.54±1.46^{a}	7.87±0.92^{b}	1.34±0.02^{d}	2.44±0.02^{c}
	Artificial	124.88±5.34^{a}	241.49±36.26^{a}	540.90±21.32^{a}	629.18±4.95^{a}	10.65±1.06^{b}	6.65±0.88^{c}	1.42±0.04^{cd}	2.36±0.08^{c}
15	Natural	116.83±7.24^{b}	245.38±18.83^{a}	537.82±13.44^{a}	586.75±5.10^{b}	9.21±1.21^{c}	8.83±0.44^{a}	1.83±0.08^{b}	3.13±0.03^{b}
	Artificial	115.84±6.44^{b}	247.82±31.82^{a}	541.40±26.66^{a}	584.80±3.32^{b}	8.64±1.34^{c}	8.66±0.31^{a}	1.66±0.01^{bc}	3.25±0.07^{b}
20	Natural	112.84±4.76^{b}	248.38±31.17^{a}	533.82±32.45^{a}	484.56±5.98^{c}	7.69±0.89^{d}	6.07±1.10^{d}	3.33±0.05^{a}	5.14±0.05^{a}
	Artificial	111.44±7.77^{b}	249.05±21.43^{a}	528.24±20.54^{a}	473.05±4.76^{d}	7.08±0.62^{d}	5.80±0.41^{d}	3.24±0.02^{a}	5.09±0.06^{a}
Lsd		6.857	18.850	13.180	4.077	0.749	0.497	0.288	0.220

Means in each column not followed by the same superscripted alphabets are significantly different at $p < 0.05$.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Conclusions

1. The cocoa bean length, width, thickness and were influenced by age of the cocoa tree and the type of pollination. The highest length, width, thickness and sphericity obtained were 24.64±1.21 mm, 13.63±1.13 mm, 9.32±1.21 mm and 0.59±0.03), respectively for A20.

2. Cocoa tree age and pollination type influenced cocoa bean mass and 1000-bean mass with A20 recording the heaviest cocoa bean mass and 1000-bean mass values of 1.35±0.02 g and 1245.81±1.24 g, respectively.

3. A20 produced the largest mean bean density and bulk density values of 953.60±7.36 kg/m^3 and 721.56±1.59 kg/m^3, respectively.

4. The maximum mean protein content measured was 15.29±0.10 % for N20.

5. The N10 recorded the largest fibre content of 9.49±0.02 % and carbohydrate content of 25.87±0.26 % for cocoa bean samples studied.

6. The highest mean fat content and pH value of cocoa beans was 48.77±0.14 % and 6.40±0.22 for A20, respectively.

7. The maximum TPC of cocoa beans was 75.02±1.05 mg/100 g for the A10 whereas the minimum value of 65.09±1.05 mg/100 g occurred for the A20.

8. Calcium, magnesium, phosphorus and potassium were found in large quantities in the range of 111.44±7.77 - 125.23±6.26, 238.79±17.13 - 249.05±21.43, 528.24±20.54 - 541.40±26.66 and 473.05±4.76 - 631.34±4.32 mg/100g, respectively whereas sodium, iron, copper and zinc were found in small quantities in the range of 7.08±0.62 - 11.54±1.46, 5.80±0.41 - 8.83±0.44, 1.34±0.02 - 3.33±0.05 and 2.36± 0.08 - 5.14±0.05 mg/100g, respectively for cocoa bean categories examined.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

CHAPTER SEVEN

SUMMARY, CONCLUSIONS, NOVELTY STATEMENTS AND RECOMMENDATIONS

Summary

Cocoa bean quality is the result of preharvest activities including production management methods and age of the cocoa tree as well as postharvest treatments such as fermentation and drying. Traditionally, examination of cocoa bean quality is based on a destructive instrumental cut test and wet chemistry methods. These methods are time-consuming, cumbersome, destructive, subjective, require the use of solvents and require highly trained technical personnel.

In this research, an assessment of the impact of some preharvest and postharvest treatments on cocoa bean quality using standard methods and handheld NIR spectroscopy were conducted. A rapid, nondestructive, reliable, and chemical-free measurement technique was developed in this work. It employs a handheld NIR spectroscopy and multivariate statistical algorithms for the differentiation and simultaneous quantification of cocoa beans. It allows reduction of inconsistency usually encountered during the human inspection and therefore is superior to the traditional analytical methods.

More specifically, handheld near-infrared spectroscopy has been developed for the measurement of moisture content, fermentation index, pH, total fat content, crude protein, total carbohydrate, total polyphenol content, total flavonoid content and total antioxidant capacity. Chemometric techniques such as the first derivative, smoothing second derivate, mean centering, standard

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

normal variate (SNV), multiplicative signal correction (MSC), linear discriminant analysis (LDA), K-nearest neighbour (KNN), support vector machine (SVM), and partial least square discriminant analysis (PLS-DA), principal component analysis (PCA) plus partial least square (PLS), interval partial least square (i-PLS), synergy partial least square (Si-PLS) and genetic algorithm partial least square (GA-PLS) were discovered as used in Matlab chemometric software.

Conclusions

1. The first specific objective of the research was to determine and develop novel rapid non-destructive authentication of regional and geographical cocoa beans. From the study it was observed that:

   a. Multivariate classification models (LDA and SVM) revealed a 100 % classification rate for cocoa beans from different growing regions in Ghana and for the African countries considered in this study.

   b. Among the pre-processing techniques employed, multiplicative scatter correction (MSC) performed better.

   c. The results give strong indications that handheld spectroscopy coupled with multivariate classification models could be employed to provide the quick, accurate, and non-destructive classification of cocoa beans according to different locations. This technique could improve the work of quality control inspectors both from industry and regulatory perspectives for effective and quick detection of cocoa bean fraud.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

2. The second specific objective of the research was to determine and develop fermentation duration differentiation and real-time prediction of cocoa beans' quality parameters such as fermentation index, pH, moisture content. From the study it was observed that:

a. Multivariate classification algorithms (MSC plus SVM and MSC plus LDA) gave an accurate classification of 100% for cocoa beans at different fermentation durations.
b. The partial least square regression model gave prediction coefficients of 0.87, 0.82, and 0.89 for predicting fermentation index (FI), pH, and moisture content (Mc), respectively.
c. The study has shown for the first time that, cocoa beans of different fermentation duration and pH, FI, and Mc can be measured non-destructively and rapidly by using novel handheld NIR spectroscopy and chemometrics especially in developing countries where the majority of production is done and are challenged with laboratory infrastructure. This phenomenon is the advantage of portable NIR over laboratory-based NIR spectroscopy.

3. The third specific objective of the research was to determine and develop a novel qualitative and quantitative prediction of organic and conventional cocoa beans parameters; fats, polyphenols, flavonoids and antioxidants. From the study it was observed that:

a. There were observed differences in chemical compositions of organic cocoa beans and conventional cocoa beans and this might

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

have contributed to the 100 % accurate differentiation of the different cocoa bean categories by Partial least square discriminant analysis (PLS-DA) algorithm in both the calibration set and prediction set.

b. Synergy interval partial least square (Si-PLS) revealed its superiority for simultaneous prediction of crude fat, total polyphenol content, total flavonoid content, and total antioxidant capacity in cocoa beans with the range of: $0.871 \leq R^2_{cal} \leq 0.993$ and $0.882 \leq R^2_{pre} \leq 0.995$ in calibration and prediction set respectively.

c. Generally, the results showed that a handheld NIR spectrometer coupled with an appropriate multivariate algorithm could be used in-situ for the simultaneous prediction and differentiation of organic cocoa beans from conventional ones to ensure food integrity along the cocoa bean value chain.

4. The fourth specific objective of the research was to measure the effect of age class and pollination type on the physical and biochemical properties of cocoa beans. From the study, it was observed that:

a. A20 had the highest mean length, width, thickness, bean mass, 1000-bean mass, bean density and bulk density of 24.64±1.21 mm, 13.63±1.13 mm, 9.32±1.21 mm, 1.35±0.02 g, 1245.81±1.24 g, 953.60±7.36 kg/m^3 and 721.56±1.59 kg/m^3, respectively.

b. The highest mean fat content and pH value of cocoa beans was 48.77±0.14 % and 6.40±0.22 for A20, respectively.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

c. The maximum total polyphenol content of cocoa beans was 75.02±1.05 mg/100 g for the A10 whereas the minimum value of 65.09±1.05 mg/100 g occurred for A20.
d. The N10 recorded the largest fibre content of 9.49±0.02 % and carbohydrate content of 25.87±0.26 % for cocoa bean samples.
e. N10 produced the highest mean of 125.23±6.26 mg/100g, 631.34±4.32 mg/100g, 11.54±1.46 mg/100g, for Ca, K, and Na contents, respectively.

It was observed from the study that NIR spectroscopy correlated well with the standard methods. It is therefore concluded that the novel handheld NIR spectrometer is a promising, reliable and non-destructive instrument for objective measurement of cocoa bean quality. The outcome of the study reveals that the handheld NIR spectroscopy coupled with chemometrics has a huge potential for rapid differentiation and quantification of cocoa beans. Also, the study shows that the age of the cocoa tree and pollination type influenced the physical and biochemical properties of cocoa beans.

Novelty Statements

This research has developed and established the link between handheld near-infrared spectroscopy and different cocoa bean categories and cocoa beans of different geographic locations for qualitative and quantitative evaluation. The specific novelty includes the development of spectral preprocessing tools such as mean centering, standard normal variant, support vector machine, smoothing

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

plus the second derivative, and synergy partial least square (Si-PLS) with smoothing second derivative.

The study has shown for the first time in real-time usage that, cocoa beans of different regional and geographical origin, classes (fermented, partially fermented and unfermented) and categories (organic and conventional) could be examined non-destructively and rapidly by using a handheld NIR spectrometer coupled with chemometrics. The overall results have proven that the novel handheld NIR spectroscopic technique could be incorporated in the cocoa beans value chain especially in Ghana and other developing countries where the majority of production is carried out and are challenged with laboratory infrastructure.

Unlike stationary laboratory-based wet chemistry technique or tabletop NIR spectroscopy, this study revealed that the relatively inexpensive handheld NIR spectroscopic technique could provide very fast (within 30 s) results in the routine onsite quantitative evaluation of cocoa beans parameters; moisture content, fats, fermentation index, pH, polyphenols, flavonoids and antioxidants on farmers field in Sub-Saharan Africa. Also, the study outcome highlights the potential application of handheld NIR spectroscopy based on machine learning for efficient classification of fermentation duration of cocoa beans, organic and conventional cocoa beans, regional and geographical cocoa beans in real-time usage. This phenomenon is the advantage of the handheld NIR spectroscopic technique over laboratory-based NIR spectroscopy and wet chemistry analytical methods.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Furthermore, cocoa beans are traded with high-quality standards of regulations on beans. They are excepted to be of high physical and chemical quality. For the first time as demonstrated in this study age of the cocoa tree and type of pollination has influenced the physical and biochemical properties of cocoa beans.

Recommendations

The potential and versatility of handheld near-infrared spectroscopy coupled with chemometric techniques have been demonstrated for cocoa beans analysis in this research. The major difficulty in handheld NIR spectroscopy is; developing a stable and reliable calibration model, requiring a huge number of previous measurements for the modelling steps. These steps are very important and a needful activity for NIR spectroscopic technique development and hence must be done very cautiously. From this study three recommendations are made.

a. Further study should be directed towards the feasibility of fusion of handheld NIR spectroscopy with electronic tongue technique for discrimination and estimation of quality parameters, harmful chemical residues and mycotoxins in cocoa beans.
b. A research into the age of tree and mould development during storage of cocoa beans by handheld near infrared spectroscopy is recommended.
c. Finally, COCOBOD should invest money into this novel handheld technology to make it available to all the 70 cocoa districts for rapid and non-destructive examination of cocoa beans quality.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

REFERENCES

Abano, E. E., & Amoah, K. K. (2011). Effect of moisture content on the physical properties of tiger nut (*Cyperus esculentus*). *Asian Journal of Agricultural Research, 5: 56-66.*

Abbadi, S., Senadza, B., Lieuw-Kie-Song, M., & Abebe, H. (2019). Assessing the employment effects of processing Cocoa in Ghana. *International Labour organization, Geneva, Switzerland*, pp. 1999-2947.

Abbe, M., & Amin, I. (2008). Antioxidant properties of cocoa powder. *Journal of Food Biochemistry*, *34*,111-128.

Abdullahi, G., Muhamad, R., Dzolkhifli, O., & Sinniah, U. R. (2018). Analysis of quality retentions in cocoa beans exposed to solar heat treatment in cardboard solar heater box. *Cogent Food & Agriculture*, *4*(1), 1483061.

Aculey, P. C., Snitkjaer, P., Owusu, M., Bassompiere, M., Takrama, J., Nørgaard, L., & Nielsen, D.S. (2010). Ghanaian cocoa bean fermentation characterized by spectroscopic and chromatographic methods and chemometrics. *Journal of Food Science*, *75*(6), S300-S307.

Adeyeye, E. I. (2016). Proximate, mineral and antinutrient compositions of natural cocoa cake, cocoa liquor and alkalized cocoa powders sourced in Nigeria. *Journal of Advanced Pharmaceutical Science and Technology*, 1(3), 12.

Afoakwa, E. O. (2010). Industrial chocolate manufacture-processes and factors influencing quality. *Chocolate Science and Technology*, *1*, 35-57.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Adeyeye, E., Akinyeye, R., Ogunlade, I., Olaofe, O., & Boluwade, J. (2010). Effect of farm and industrial processing on the amino acid profile of cocoa beans. *Food chemistry*, *118*(2), 357-363.

Afoakwa, E. O. (2016). *Chocolate science and technology*: John Wiley & Sons.

Afoakwa, E. O., Kongor, J. E., Takrama, J. F., Budu, A. S., & Mensah-Brown, H. (2013a). Effects of pulp preconditioning on total polyphenols, o-diphenols and anthocyanin concentrations during fermentation and drying of cocoa (Theobroma cacao) beans. *Journal of Food Science and Engineering*, *3*(5), 235.

Afoakwa, E. O., Kongor, J. E., Takrama, J. E., & Budu, A. S. (2013b). Changes in acidification, sugars and mineral composition of cocoa pulp during fermentation of pulp pre-conditioned cocoa (Theobroma cacao) beans *International Food Research Journal* 20(3): 1215-1222 (2013).

Afoakwa, E. O., Quao, J., Takrama, J. E., Budu, A. S., & Saalia, F. K. (2013c). Chemical composition and physical quality characteristics of Ghanaian cocoa beans as affected by pulp pre-conditioning and fermentation. *Journal of food science and technology*, *50*(6), 1097-1105.

Afoakwa, E. O., Kongor, J. Q., Budu, A. S., Jemmy, S., & Firibu, K.S (2012). Influence of pulp-preconditioning and fermentation on fermentative quality and appearance of Ghanaian cocoa (*Theobroma cacao*) beans. *International Food Research Journal*, *19*(1), 127.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Aikpokpodion, P., & Dongo, L. (2010). Effects of fermentation intensity on polyphenols and antioxidant capacity of cocoa beans. *Int. J. Sustain. Crop Prod*, *5*(4), 66-70.

AOAC, I. (2005). *Official methods of analysis of AOAC International*: AOAC International.

Alimuddin, S. (2018). Changes in phenol level and antioxidant activity of cocoa beans during fermentation and roasting. *Journal of Food Research*, *7*(4), 23-29

Alma, O. G., & Bulut, E. (2012). Genetic algorithm based variable selection for partial least squares regression using ICOMP criterion. *Asian Journal of Mathematics & Statistics*, *5*(3), 82.

Álvarez, C., Pérez, E., Cros, E., Lares, M., Assemat, S., Boulanger, R., & Davrieux, F. (2012). The use of near infrared spectroscopy to determine the fat, caffeine, theobromine and (−)-epicatechin contents in unfermented and sun-dried beans of Criollo cocoa. *Journal of near infrared spectroscopy*, *20*(2), 307-315.

Amoa-Awua, W. K. (2014). Methods of cocoa fermentation and drying. *Cocoa and Coffee Fermentation. CRC, Taylor & Francis, Boca Raton, USA*.

Amuah, C. L., Teye, E., Lamptey, F. P., Nyandey, K., Opoku-Ansah, J., & Adueming, P. O.-W. (2019). Feasibility study of the use of handheld NIR spectrometer for simultaneous authentication and quantification of quality parameters in intact pineapple fruits. *Journal of Spectroscopy*.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Andersen, C. M., & Bro, R. (2010). Variable selection in regression - a tutorial. *Journal of chemometrics, 24*(11-12), 728-737.

Apriyanto, M., & Harmayani, E. (2016). Study on effect of fermentation to the quality parameter of cocoa bean in Indonesia. *Asian Journal of Dairy and Food Research, 35*(2), 160-163.

Apshara, E., Bhat, V., Ananda, K., Nair, R., & Suma, D. (2009). Evaluation and identification of high yielding trees in Nigerian cocoa germplasm. *Journal Plantation Crops, 37*, 111-116.

Aremu, C., Agiang, M., & Ayatse, J. (1995). Nutrient and antinutrient profiles of raw and fermented cocoa beans. *Plant foods for human nutrition, 48*(3), 217-223.

Arslan, M., Rakha, A., Khan, M. R., & Zou, X. (2017). Complementing the dietary fiber and antioxidant potential of gluten free bread with guava pulp powder. *Journal of Food Measurement and Characterization, 11*(4), 1959-1968.

Arslan, M., Xiaobo, Z., Xuetao, H., Tahir, H. E., Shi, J., Khan, M. R., & Zareef, M. (2018). Near infrared spectroscopy coupled with chemometric algorithms for predicting chemical components in black goji berries (Lycium ruthenicum Murr.). *Journal of near infrared spectroscopy, 26*(5), 275-286.

Assa, A., Noor, A., Misnawi, M., Djide, M. N., & Muliadi, M. (2017). Protein and Mineral Contents of Cocoa Beans Fermentation from Luwu Timur

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Regency, South Sulawesi. *Pelita Perkebunan (a Coffee and Cocoa Research Journal), 33*(3), 196-202

Atlas, C. (2010). In C. Rohsius, S. Elwers, & R. Lieberei, *Cocoa atlas* (2010th ed.). Hamburg, Germany: German cocoa and chocolate foundation.

Awarikabey, E., Amponsah, I. K., & Woode, M. Y. (2014). The value of the cocoa bean shell (hull) and the effect of various processing methods on the phyto-constituents and antioxidant activity of the nib and shell.

Bangmarigu, E., & Qineti, A. (2018). *Cocoa Production and Export in Ghana*.

Bao, Y., Mi, C., Wu, N., Liu, F., & He, Y. (2019). Rapid classification of wheat grain varieties using hyperspectral imaging and chemometrics. *Applied Sciences, 9*(19), 4119.

Barbin, D. F., Maciel, L. F., Bazoni, C. H. V., da Silva Ribeiro, M., Carvalho, R. D. S., da Silva Bispo, E., & Hirooka, E. Y. (2018). Classification and compositional characterization of different varieties of cocoa beans by near-infrared spectroscopy and multivariate statistical analyses. *Journal of Food Science and Technology, 55*(7), 2457-2466.

Bart-Plange, A., & Baryeh, E. A. (2003). The physical properties of Category B cocoa beans. *Journal of Food Engineering, 60*(3), 219-227.

Bart-Plange, A., Dzisi, A., Addo, A., Teye, E., & Kumi, F. (2011). A comparative study of some physical properties of large and medium size cocoa beans from Ghana. *Small (Category G), 121*, 130.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Bedini, A., Zanolli, V., Zanardi, S., Bersellini, U., Dalcanale, E., & Suman, M. (2013). Rapid and simultaneous analysis of xanthines and polyphenols as bitter taste markers in bakery products by FT-NIR spectroscopy. *Food analytical methods, 6*(1), 17-27.

Beg, M. S., Ahmad, S., Jan, K., & Bashir, K. (2017). Status, supply chain and processing of cocoa-A review. *Trends in food science & technology, 66*, 108-116.

Berlan, A., & Bergés, A. (2013). Cocoa production in the Dominican Republic: sustainability, challenges and opportunities. *Commissioned by Green & Black's*.

Berrueta, L. A., Alonso-Salces, R. M., & Héberger, K. (2007). Supervised pattern recognition in food analysis. *Journal of chromatography A, 1158*(1-2), 196-214.

Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry, 21*(4), 240-250.

Borin, A., Ferrao, M. F., Mello, C., Maretto, D. A., & Poppi, R. J. (2006). Least-squares support vector machines and near-infrared spectroscopy for quantification of common adulterants in powdered milk. *Analytica chimica acta, 579*(1), 25-32.

Burns, D. A., & Ciurczak, E. W. (2007). *Handbook of near-infrared analysis*: CRC press.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Business & Human Rights Resource Centre, https://www.business-humanrights.org/en/latest-news/c%C3%B4te-divoire-ghana-set-a-fixed-living-income-differential-of-400-per-tonne-of-cocoa-to-relieve-farmer-poverty, date accessed: 18/12/2020 (2020).

Caprioli, G., Fiorini, D., Maggi, F., Nicoletti, M., Ricciutelli, M., Toniolo, C., Sagratini, G. (2016). Nutritional composition, bioactive compounds and volatile profile of cocoa beans from different regions of Cameroon. *International Journal of Food Sciences and Nutrition, 67*(4), 422-430.

Cayuela, J. A., & Weiland, C. (2010). Intact orange quality prediction with two portable NIR spectrometers. *Postharvest Biology and Technology, 58*(2), 113-120.

Cen, H., & 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.

Chen, Q., Ding, J., Cai, J., Sun, Z., & Zhao, J. (2012a). Simultaneous measurement of total acid content and soluble salt-free solids content in Chinese vinegar using near-infrared spectroscopy. *Journal of Food Science, 77*(2), C222-C227.

Chen, Q., Guo, Z., Zhao, J., & Ouyang, Q. (2012b). Comparisons of different regressions tools in measurement of antioxidant activity in green tea using near infrared spectroscopy. *Journal of Pharmaceutical and Biomedical Analysis, 60*, 92-97.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Chen, Q., Zhao, J., & Lin, H. (2009). Study on discrimination of Roast green tea (Camellia sinensis L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *72*(4), 845-850.

Chen, Q., Zhao, J., Liu, M., Cai, J., & Liu, J. (2008). Determination of total polyphenols content in green tea using FT-NIR spectroscopy and different PLS algorithms. *Journal of Pharmaceutical and Biomedical Analysis*, *46*(3), 568-573.

Coronel-Reyes, J., Ramirez-Morales, I., Fernandez-Blanco, E., Rivero, D., & Pazos, A. (2018). Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques. *Computers and Electronics in Agriculture*, *145*, 1-10.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Cozzolino, D., Cynkar, W., Shah, N., & Smith, P. (2011). Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food research international*, *44*(7), 1888-1896.

Cutler, D. R., Edwards Jr, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., & Lawler, J. J. (2007). Random forests for classification in ecology. *Ecology*, *88*(11), 2783-2792.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

da Costa Filho, P. A. (2009). Rapid determination of sucrose in chocolate mass using near-infrared spectroscopy. *Analytica chimica acta, 631*(2), 206-211.

Davies, A., & Grant, A. (1987). Near infra-red analysis of food. *International journal of food science & technology, 22*(3), 191-207.

Davies, A., Franklin, J., Grant, A., Griffiths, N., Shepherd, R., & Fenwick, G. (1991). Prediction of chocolate quality from near-infrared spectroscopic measurements of the raw cocoa beans. *Vibrational spectroscopy, 2*(2-3), 161-172.

De Maesschalck, R., & Van den Kerkhof, T. (2005). Implementation of a simple semi-quantitative near-infrared method for the classification of clinical trial tablets. *Journal of Pharmaceutical and Biomedical Analysis, 37*(1), 109-114.

Delgado-Ospina, J., Di Mattia, C. D., Paparella, A., Mastrocola, D., Martuscelli, M., & Chaves-Lopez, C. (2020). Effect of Fermentation, Drying and Roasting on Biogenic Amines and Other Biocompounds in Colombian Criollo Cocoa Beans and Shells. *Foods, 9*(4), 520.

Dhanoa, M., Lister, S., Sanderson, R., & Barnes, R. (1994). The link between multiplicative scatter correction (MSC) and standard normal variate (SNV) transformations of NIR spectra. *Journal of near infrared spectroscopy, 2*(1), 43-47.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Di Castelnuovo, A., Di Giuseppe, R., Iacoviello, L., & De Gaetano, G. (2012). Consumption of cocoa, tea and coffee and risk of cardiovascular disease. *European Journal of Internal Medicine, 23*(1), 15-25.

Dillinger, T. L., Barriga, P., Escárcega, S., Jimenez, M., Lowe, D. S., & Grivetti, L. E. (2000). Food of the gods: cure for humanity? A cultural history of the medicinal and ritual use of chocolate. *The Journal of nutrition, 130*(8), 2057S-2072S.

Eghbal, M (2018). Global Chocolate Industry: From Bean to Bar. Euromonitor International. Retrieved from https://blog.euromonitor.com/global-chocolate-industry.

García-Alamilla, P., González-Lauck, V., De La Cruz-Lázaro, E., Lagunes-Gálvez, L., & García-Alamilla, R. (2012). Description and physical properties of mexican criollo cacao during post-harvest processing. *Revista Iberoamericana de Tecnología Postcosecha, 13*(1), 58-65.

Ghana Cocoa Board (2018). Manual for cocoa extension in Ghana. pp. 53-61

Geladi, P. (2002). Some recent trends in the calibration literature. *Chemometrics and Intelligent Laboratory Systems, 60*(1-2), 211-224.

Geladi, P. (2003). Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B: Atomic Spectroscopy, 58*(5), 767-782.

Ghana Cocoa Board, https://cocobod.gh/news/cocobod-signs-us13-billion-syndicated-loan-facility-for-2020/2021-crop-season, accessed:17/12/ 2020.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Gilbert, C. L. (2009). Cocoa market liberalization in retrospect. *Review of Business and Economic Literature*, *54*(3), 294-313.

Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control*, *99*, 137-145.

Glin, L. C., Oosterveer, P., & Mol, A. P. (2015). Governing the Organic Cocoa Network from G hana: Towards Hybrid Governance Arrangements? *Journal of Agrarian Change*, *15*(1), 43-64.

Gourieva, K., & Tserevitinov, O. (1979). Method of evaluating the degree of fermentation of cocoa beans. *USSR patent*, *646*.

Grand View Research. (2019). Cocoa Beans Market Size, Share & Trends Analysis Report by application (Cosmetics, Confectionery, Pharmaceuticals), By Product (Butter, Powder, Liquor), By Distribution Channel (Online, Offline), And Segment Forecasts, 2019-2025.

Gu, F., Tan, L., Wu, H., Fang, Y., Xu, F., Chu, Z., & Wang, Q. (2013). Comparison of cocoa beans from china, indonesia and papua new guinea. *Foods*, *2*(2), 183-197.

Gu, L., House, S. E., Wu, X., Ou, B., & Prior, R. L. (2006). Procyanidin and catechin contents & antioxidant capacity of cocoa & chocolate products. *Journal of Agricultural & Food Chemistry*, *54*(11), 4057-4061.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Hadiwijaya, Y., & Putri, I. E. (2018). *Determination of Water Content of Intact Sapodilla Using Near-Infrared Spectroscopy*. Paper presented at the IOP Conference Series: Earth and Environmental Science.

Hart, J. R., Norris, K. H., & Golumbic, C. (1962). Determination of the moisture content of seeds by near-infrared spectrophotometry of their methanol extracts. *Cereal Chem, 39*(2), 94-99.

Hashimoto, J. C., Lima, J. C., Celeghini, R. M., Nogueira, A. B., Efraim, P., Poppi, R. J., & Pallone, J. A. (2018). Quality control of commercial cocoa beans (Theobroma cacao L.) by near-infrared spectroscopy. *Food analytical methods, 11*(5), 1510-1517.

Hii, C., Law, C., Suzannah, S., & Cloke, M. (2009). Polyphenols in cocoa (Theobroma cacao L.). *Asian Journal of Food and agro-industry, 2*(4), 702-722.

Huang, X., Teye, E., Owusu-Sekyere, J. D., Takrama, J., Sam-Amoah, L. K., Yao, L., & Firempong, C. K. (2014). Simultaneous measurement of titratable acidity and fermentation index in cocoa beans by electronic tongue together with linear and non-linear multivariate technique. *Food analytical methods, 7*(10), 2137-2144.

Hue, C., Gunata, Z., Bergounhou, A., Assemat, S., Boulanger, R., Sauvage, F.-X., & Davrieux, F. (2014a). Near-infrared spectroscopy as a new tool to determine cocoa fermentation levels through ammonia nitrogen quantification. *Food Chemistry, 148*, 240-245.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Hue, C., Brat, P., Gunata, Z., Samaniego, I., Servent, A., Morel, G., Davrieux, F. (2014b). Near infra-red characterization of changes in flavan-3-ol derivatives in cocoa (Theobroma cacao L.) as a function of fermentation temperature. *Journal of agricultural and food chemistry*, *62*(41), 10136-10142.

Hurtado-Barroso, S., Tresserra-Rimbau, A., Vallverdú-Queralt, A., & Lamuela-Raventós, R. M. (2019). Organic food and the impact on human health. *Critical reviews in food science and nutrition*, *59*(4), 704-714.

International Cocoa Organization, 2018. ICCO Quarterly Bulletin of Cocoa Statistics, Vol. XLIV, No.3 Cocoa year 2017/18.

International Office of Cocoa, C., & Confectionary, S. (1996). Determination of the Fat Content of Cocoa Powder by Soxhlet Extraction. In: IOCCC.

Javidnia, K., Parish, M., Karimi, S., & Hemmateenejad, B. (2013). Discrimination of edible oils and fats by combination of multivariate pattern recognition and FT-IR spectroscopy: A comparative study between different modeling methods. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *104*, 175-181.

Junfang, X., Xiaoyu, L., Peiwu, L., Wei, W., & Xiaoxia, D. (2007). Approach to nondestructive measurement of Vitamin C content of orange with near-infrared spectroscopy treated by wavelet transform [J]. *Transactions of the Chinese Society of Agricultural Engineering*, *6*.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Jha, S., & Garg, R. (2010). Non-destructive prediction of quality of intact apple using near infrared spectroscopy. *Journal of food science and technology, 47*(2), 207-213.

Jinap, S., & Dimick, P. S. (1990). Acidic characteristics of fermented and dried cocoa beans from different countries of origin. *Journal of Food Science, 55*(2), 547-550.

Jinap, S., Dimick, P., & Hollender, R. (1995). Flavour evaluation of chocolate formulated from cocoa beans from different countries. *Food Control, 6*(2), 105-110.

Jonfia-Essien, W., West, G., Alderson, P., & Tucker, G. (2008). Phenolic content and antioxidant capacity of hybrid variety cocoa beans. *Food chemistry, 108*(3), 1155-1159.

Kademi, H. I., Ulusoy, B. H., & Hecer, C. (2019). Applications of miniaturized and portable near infrared spectroscopy (NIRS) for inspection and control of meat and meat products. *Food Reviews International, 35*(3), 201-220.

Karimi, M., Kheiralipour, K., Tabatabaeefar, A., Khoubakht, G., Naderi, M., & Heidarbeigi, K. (2009). The effect of moisture content on physical properties of wheat. *Pakistan Journal of Nutrition, 8*(1), 90-95.

Kawata, S. (2009). New techniques in near-infrared spectroscopy. *Near-infrared Spectroscopy Principles, Instruments, Applications*.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Kays, S. E., Shimizu, N., Barton, F. E., & Ohtsubo, K. i. (2005). Near-infrared transmission and reflectance spectroscopy for the determination of dietary fiber in barley cultivars. *Crop science, 45*(6), 2307-2311.

Kongor, J. E., Hinneh, M., Van de Walle, D., Afoakwa, E. O., Boeckx, P., & Dewettinck, K. (2016). Factors influencing quality variation in cocoa (Theobroma cacao) bean flavour profile- A review. *Food research international, 82*, 44-52.

Kongor, J. E., Takrama, J. F., Budu, A. S., Mensah-Brown, H., & Afoakwa, E. O. (2013). Effects of fermentation and drying on the fermentation index and cut test of pulp pre-conditioned Ghanaian cocoa (Theobroma cacao) beans. *Journal of Food Science and Engineering, 3*(11), 625.

Kortüm, G. (2012). *Reflectance spectroscopy: principles, methods, applications*: Springer Science & Business Media.

Kosmowski, F., & Worku, T. (2018). Evaluation of a miniaturized NIR spectrometer for cultivar identification: The case of barley, chickpea and sorghum in Ethiopia. *PloS one, 13*(3).

Krähmer, A., Engel, A., Kadow, D., Ali, N., Umaharan, P., Kroh, L. W., & Schulz, H. (2015). Fast and neat–Determination of biochemical quality parameters in cocoa using near infrared spectroscopy. *Food chemistry, 181*, 152-159.

Kubelka, P., & Munk, F. (1931). An article on optics of paint layers. *Z. Tech. Phys, 12*(593-601), 259-274.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Kulkarni, A. D., & Lowe, B. (2016). Random forest algorithm for land cover classification.

Kutsanedzie, F. Y., Chen, Q., Hassan, M. M., Yang, M., Sun, H., & Rahman, M. H. (2018). Near-infrared system coupled chemometric algorithms for enumeration of total fungi count in cocoa beans neat solution. *Food Chemistry, 240*, 231-238.

Kutsanedzie, F. Y., Chen, Q., Sun, H., & Cheng, W. (2017). In situ cocoa beans quality grading by near-infrared-chemodyes systems. *Analytical Methods, 9*(37), 5455-5463.

Lam Thi, V. H., Nusantoro, B., Phillip, R. A., Huynh Anh, P., Thanh Toan, H., Messens, K., & Dewettinck, K. (2016). Physico-chemical properties of fourteen popular cocoa bean varieties in Dongnai-highland Vietnam. *Can Tho University Journal of Science, 4*, 81-86.

Lee, H. W., Bawn, A., & Yoon, S. (2012). Reproducibility, complementary measure of predictability for robustness improvement of multivariate calibration models via variable selections. *Analytica chimica acta, 757*, 11-18.

Lee, K. W., Kim, Y. J., Lee, H. J., & Lee, C. Y. (2003). Cocoa has more phenolic phytochemicals and a higher antioxidant capacity than teas and red wine. *Journal of Agricultural and Food Chemistry, 51*(25), 7292-7295.

Lee, L. C., Liong, C. Y., & Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

(HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst, 143*(15), 3526-3539.

Li, X., He, Y., & Fang, H. (2007). Non-destructive discrimination of Chinese bayberry varieties using Vis/NIR spectroscopy. *Journal of Food Engineering, 81*(2), 357-363.

Li, X., He, Y., & Wu, C. (2008). Non-destructive discrimination of paddy seeds of different storage age based on Vis/NIR spectroscopy. *Journal of Stored Products Research, 44*(3), 264-268.

Lin, H., Zhao, J., Sun, L., Chen, Q., & Zhou, F. (2011). Freshness measurement of eggs using near-infrared (NIR) spectroscopy and multivariate data analysis. *Innovative Food Science & Emerging Technologies, 12*(2), 182-186.

Lin, L., He, Y., Xiao, Z., Zhao, K., Dong, T., & Nie, P. (2019). Rapid-detection sensor for rice grain moisture based on NIR spectroscopy. *Applied Sciences, 9*(8), 1654.

Lu, C., Xiang, B., Hao, G., Xu, J., Wang, Z., & Chen, C. (2009). Rapid detection of melamine in milk powder by near-infrared spectroscopy. *Journal of near-infrared spectroscopy, 17*(2), 59-67.

Maalekuu, B. K., & Teye, N. (2017). Comparative evaluation of some physical and biochemical quality attributes of cocoa beans produced in Assin Fosu, Assin Bereku and Nkawie cocoa districts of Ghana. *Journal of Agricultural Science (Toronto), 9*(12), 169-177.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Markets & Markets. (2013). *MarketsandMarkets*: Global Chocolate Market worth $ 98.3 billion by 2016. Retrieved from http://www.markets andmarkets.com/PressReleases/global-chocolate-market.asp.

MarketWatch. (2019). Global Chocolate Market is Projected to Reach US$ 189.08 Billion by 2016. Retrieved from https://www.marketwatch.com /press-release/global-chocolate-market-is-projected-cited-to-reach-us-18908 -billion-by- 2026-2019-01-23.

Marques, E. J. N., de Freitas, S. T., Pimentel, M. F., & Pasquini, C. (2016). Rapid and non-destructive determination of quality parameters in the 'Tommy Atkins' mango using a novel handheld near infrared spectrometer. *Food chemistry*, *197*, 1207-1214.

Martens, H., Nielsen, J. P., & Engelsen, S. B. (2003). Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures. *Analytical Chemistry*, *75*(3), 394-404.

Miller, K. B., Stuart, D. A., Smith, N. L., Lee, C. Y., McHale, N. L., Flanagan, J. A., & Hurst, W. J. (2006). Antioxidant activity and polyphenol and procyanidin contents of selected commercially available cocoa-containing and chocolate products in the United States. *Journal of Agricultural and Food Chemistry*, *54*(11), 4062-4068.

Minifie, B. (2012). *Chocolate, cocoa and confectionery: Science and Technology*: Springer Science & Business Media.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Mohsenin, N.N. (1968). Physical Properties of Plant and Animal Materials, Part I, Vol. 1, First preliminary edition. *Department of Agricultural Engineering, Pennsylvania State University*.

Mohsenin, N. N. (1980). Thermal properties of foods and agricultural materials, Gordon Breach Sci. *Publ., New York, USA*, 198-224.

Moros, J., Iñón, F., Garrigues, S., & De la Guardia, M. (2007). Near-infrared diffuse reflectance spectroscopy and neural networks for measuring nutritional parameters in chocolate samples. *Analytica chimica acta, 584*(1), 215-222.

Mota, M. M., El Makhloufi, A., & Scala, P. (2019). On the logistics of cocoa supply chain in Côte d'Ivoire: Simulation-based analysis. *Computers & Industrial Engineering, 137*, 106034.

Musingarabwi, D. M., Nieuwoudt, H. H., Young, P. R., Eyéghè-Bickong, H. A., & Vivier, M. A. (2016). A rapid qualitative and quantitative evaluation of grape berries at various stages of development using Fourier-transform infrared spectroscopy and multivariate data analysis. *Food chemistry, 190*, 253-262.

Nazaruddin, R., Seng, L., Hassan, O., & Said, M. (2006). Effect of pulp preconditioning on the content of polyphenols in cocoa beans (Theobroma cacao) during fermentation. *Industrial Crops and Products, 24*(1), 87-94.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Nogales-Bueno, J., Baca-Bocanegra, B., Rodríguez-Pulido, F. J., Heredia, F. J., & Hernández-Hierro, J. M. (2015). Use of near infrared hyperspectral tools for the screening of extractable polyphenols in red grape skins. *Food chemistry, 172*, 559-564.

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (i PLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Applied spectroscopy, 54*(3), 413-419.

Oliveira, L. S., & Franca, A. S. (2011). Applications of Near Infrafred Spectroscopy (NIRS) in food quality evaluation. *Food Quality: Control, Analysis and Consumer Concerns*, 131-179.

Oosterveer, P., Van Hoi, P., & Glin, L. (2011). Governance and greening global agro-food chains: cases from Vietnam, Thailand and Benin.

Ordonez, A., Gomez, J., & Vattuone, M. (2006). Antioxidant activities of Sechium edule (Jacq.) Swartz extracts. *Food chemistry, 97*(3), 452-458.

Osborne, B. G. (2006). Near-infrared spectroscopy in food analysis. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*.

Othman, A., Ismail, A., Ghani, N. A., & Adenan, I. (2007). Antioxidant capacity and phenolic content of cocoa beans. *Food chemistry, 100*(4), 1523-1530.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Ozaki, Y., McClure, W. F., & Christy, A. A. (2006). *Near-infrared spectroscopy in food science and technology*: John Wiley & Sons.

Pasquini, C. (2003). Near-infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian chemical society*, *14*(2), 198-219.

Pay, E. (2009). Increasing incomes and food security of small farmers in West and Central Africa through exports of organic and fair-trade tropical products. *The Market for Organic and Fair-trade Mangoes and Pineapples, Study Prepared in the Framework of FAO Project GCP/RAF/404/GER, FAO of the UN, Trade & Markets Division, Rome*.

Phuangsombut, K., Suttiwijitpukdee, N., & Terdwongworakul, A. (2017). Nondestructive classification of mung bean seeds by single kernel near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences*, *10*(03), 1650053.

Pissard, A., Fernández Pierna, J. A., Baeten, V., Sinnaeve, G., Lognay, G., Mouteau, A., Lateur, M. (2013). Non-destructive measurement of vitamin C, total polyphenol and sugar content in apples using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*, *93*(2), 238-244.

Quampah, A., Huang, Z. R., Wu, J. G., Liu, H. Y., Li, J. R., Zhu, S. J., & Shi, C. H. (2012). Estimation of oil content and fatty acid composition in cottonseed kernel powder using near-infrared reflectance spectroscopy. *Journal of the American Oil Chemists' Society*, *89*(4), 567-575.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Quarmine, W., Haagsma, R., Sakyi-Dawson, O., Asante, F., Van Huis, A., & Obeng-Ofori, D. (2012). Incentives for cocoa bean production in Ghana: Does quality matter? *NJAS-Wageningen Journal of Life Sciences, 60*, 7-14.

Quelal-Vásconez, M. A., Lerma-García, M. J., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2019). Fast detection of cocoa shell in cocoa powders by near infrared spectroscopy and multivariate analysis. *Food Control, 99*, 68-72.

Quelal-Vásconez, M. A., Pérez-Esteve, É., Arnau-Bonachera, A., Barat, J. M., & Talens, P. (2018). Rapid fraud detection of cocoa powder with carob flour using near-infrared spectroscopy. *Food Control*, *92*, 183-189.

Ramli, N., Yatim, A. M., Said, M., & Hok, H. C. (2001). HPLC determination of methylxanthines and polyphenols levels in cocoa and chocolate products. *Malaysian Journal of Analytical Sciences, 7*(2), 377-386.

Rinnan, Å., Nørgaard, L., van den Berg, F., Thygesen, J., Bro, R., & Engelsen, S. B. (2009a). Data pre-processing. *Infrared spectroscopy for food quality analysis and control*, 29-50.

Rinnan, Å., Van Den Berg, F., & Engelsen, S. B. (2009b). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201-1222.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., & Jent, N. (2007). A review of near-infrared spectroscopy and chemometrics in

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis, 44*(3), 683-700.

Rout, S. P., Acharya, R., & Maji, J. K. (2017). Discriminant analysis of shodhana (processing) on baliospermum montanum muell (danti) root samples based on near infrared spectroscopy and multivariate chemometric technique.

Sadler, G. D., & Murphy, P. A. (2010). pH and titratable acidity. In *Food analysis* (pp. 219-238): Springer.

Saltini, R., Akkerman, R., & Frosch, S. (2013). Optimizing chocolate production through traceability: A review of the influence of farming practices on cocoa bean quality. *Food Control, 29*(1), 167-187.

Sampaio, P. S., Castanho, A., Almeida, A. S., Oliveira, J., & Brites, C. (2020). Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods. *European Food Research and Technology, 246*(3), 527-537.

Sandorfy, C., Buchet, R., & Lachenal, G. (2007). Principles of molecular vibrations for near-infrared spectroscopy. *Near-Infrared Spectroscopy in Food Science and Technology; Ozaki, Y., McClure, WF, Christy, AA, Eds*, 11-46.

Sandoval, A. J., Barreiro, J. A., De Sousa, A., Valera, D., & Müller, A. J. (2019). Determination of the physical properties of fermented and dried Venezuelan Trinitrio cocoa beans (Theobroma cacao L.)/Determinacion

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

de las Propiedades Fisicas de Habas Fermentadas y Secas de Cacao Trinitario Venezolano (Theobroma cacao L.). *Revista Tecnica, 42*(2), 50-57.

Santander Muñoz, M., Rodríguez Cortina, J., Vaillant, F. E., & Escobar Parra, S. (2020). An overview of the physical and biochemical transformation of cocoa seeds to beans and to chocolate: Flavor formation. *Critical reviews in food science and nutrition, 60*(10), 1593-1613.

Seregely, Z., Deak, T., & Bisztray, G. D. (2004). Distinguishing melon genotypes using NIR spectroscopy. *Chemometrics and Intelligent Laboratory Systems, 72*(2), 195-203.

Shahbandeh, M., 2021. World cocoa production by country in 2018/2019 to 2019/2020 (in 1000 metric tons). https://www.statista.com/statistics/. coom/statistics/263855/cocoa-bean-production.

Shao, Y., He, Y., Bao, Y., & Mao, J. (2009). Near-infrared spectroscopy for classification of oranges and prediction of the sugar content. *International Journal of Food Properties, 12*(3), 644-658.

Siesler, H. W., Ozaki, Y., Kawata, S., & Heise, H. M. (2008). *Near-infrared spectroscopy: principles, instruments, applications*: John Wiley & Sons.

Stohner, J., Lukas, B., Suter, M., Zucchetti, B., Deuber, F., & Hobi, F. (2012). NIRS of chocolate and its chemometric analysis. *Newfood, 15*(6), 21-28.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Sun, J., Jiang, S., Mao, H., Wu, X., & Li, Q. (2016). Classification of black beans using visible and near-infrared hyperspectral imaging. *International Journal of Food Properties*, *19*(8), 1687-1695.

Sun, M., Zhang, D., Liu, L., & Wang, Z. (2017). How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method. *Food chemistry*, *218*, 413-421.

Sun, T., Tang, J., & Powers, J. R. (2005). Effect of pectolytic enzyme preparations on the phenolic composition and antioxidant activity of asparagus juice. *Journal of Agricultural and Food Chemistry*, *53*(1), 42-48.

Sunoj, S., Igathinathane, C., & Visvanathan, R. (2016). Nondestructive determination of cocoa bean quality using FT-NIR spectroscopy. *Computers and Electronics in Agriculture*, *124*, 234-242.

Svetnik, V., Liaw, A., Tong, C., Culberson, J. C., Sheridan, R. P., & Feuston, B. P. (2003). Random forest: a classification and regression tool for compound classification and QSAR modeling. *Journal of chemical information and computer sciences*, *43*(6), 1947-1958.

Tarkošová, J., & Čopíková, J. (2000). Fourier transform near-infrared spectroscopy applied to analysis of chocolate. *Journal of Near Infrared Spectroscopy*, *8*(4), 251-257.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Teye, E., & Huang, X. (2015). Novel prediction of total fat content in cocoa beans by FT-NIR spectroscopy based on effective spectral selection multivariate regression. *Food Analytical Methods*, *8*(4), 945-953.

Teye, E., Amuah, C. L., McGrath, T., & Elliott, C. (2019a). Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *217*, 147-154.

Teye, E., Anyidoho, E., Agbemafle, R., Sam-Amoah, L. K., & Elliott, C. (2019b). Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A Review. *Infrared Physics & Technology*, 103127.

Teye, E., Anyidoho, E., Agbemafle, R., Sam-Amoah, L. K., & Elliott, C. (2020). Cocoa bean and cocoa bean products quality evaluation by NIR spectroscopy and chemometrics: A review. *Infrared Physics & Technology*, *104*, 103127.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *114*, 183-189.

Teye, E., Huang, X., Sam-Amoah, L. K., Takrama, J., Boison, D., Botchway, F., & Kumi, F. (2015). Estimating cocoa bean parameters by FT-NIRS and chemometrics analysis. *Food chemistry*, *176*, 403-410.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Teye, E., Huang, X., Takrama, J., & Haiyang, G. (2014a). Integrating NIR spectroscopy and electronic tongue together with chemometric analysis for accurate classification of cocoa bean varieties. *Journal of Food Process Engineering*, *37*(6), 560-566.

Teye, E., Huang, X. Y., Lei, W., & Dai, H. (2014b). Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans. *Food research international*, *55*, 288-293.

Teye, E., Uhomoibhi, J., & Wang, H. (2016). Nondestructive authentication of cocoa bean cultivars by FT-NIR spectroscopy and multivariate techniques. *Focus On Medical Sciences Journal*, *2*(3).

Thanh Noi, P., & Kappas, M. (2018). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, *18*(1), 18.

Tran, P. D., Van de Walle, D., De Clercq, N., De Winne, A., Kadow, D., Lieberei, R., Van Durme, J. (2015). Assessing cocoa aroma quality by multiple analytical approaches. *Food research international*, *77*, 657-669.

Veselá, A., Barros, A. S., Synytsya, A., Delgadillo, I., Čopíková, J., & Coimbra, M. A. (2007). Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Analytica chimica acta*, *601*(1), 77-86.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Voora, V., Bermúdez, S., & Larrea, C. (2019). *Global Market Report: Cocoa*: JSTOR

Wang, H., Peng, J., Xie, C., Bao, Y., & He, Y. (2015). Fruit quality evaluation using spectroscopy technology: a review. *Sensors*, *15*(5), 11889-11927.

Wang, Q. Q., Huang, H. Y., & Wang, Y. Z. (2019). Geographical authentication of Macrohyporia cocos by a data fusion method combining ultra-fast liquid chromatography and Fourier transform infrared spectroscopy. *Molecules*, *24*(7), 1320.

Whitacre, E., Liver, J., Van Den Broek, R., Van Engelen, P., Remers, B., Van Der Horst, B., Jansen-Beuvink, A. (2003). Predictive analysis of cocoa procyanidins using near-infrared spectroscopy techniques. *Journal of Food Science*, *68*(9), 2618-2622.

Whitefield, R. (2008). *Making chocolates in the factory*: Kennedyþs Books Ltd.

Wood, G. A. R., & Lass, R. (2008). *Cocoa*: John Wiley & Sons.

Workman Jr, J., & Weyer, L. (2012). *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*: CRC press.

World Bank, T. (2012). Supply Chain Risk Assessment Cocoa in Ghana.

Xie, C., Chu, B., & He, Y. (2018). Prediction of banana color and firmness using a novel wavelengths selection method of hyperspectral imaging. *Food chemistry*, *245*, 132-140.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

Xie, L., Ye, X., Liu, D., & Ying, Y. (2011). Prediction of titratable acidity, malic acid, and citric acid in bayberry fruit by near-infrared spectroscopy. *Food research international*, *44*(7), 2198-2204.

Xie, L., Ying, Y., Ying, T., Yu, H., & Fu, X. (2007). Discrimination of transgenic tomatoes based on visible/near-infrared spectra. *Analytica chimica acta*, *584*(2), 379-384.

Yu, H., & Kim, S. (2012). SVM Tutorial-Classification, Regression and Ranking. *Handbook of Natural computing*, *1*, 479-506.

Zhang, C., Liu, F., & He, Y. (2018). Identification of coffee bean varieties using hyperspectral imaging: Influence of preprocessing methods and pixel-wise spectra analysis. *Scientific reports*, *8*(1), 1-11.

Zhang, M., Luypaert, J., Pierna, J. F., Xu, Q., & Massart, D. (2004). Determination of total antioxidant capacity in green tea by near-infrared spectroscopy and multivariate calibration. *Talanta*, *62*(1), 25-35.

Zhao, J., Chen, Q., Huang, X., & Fang, C. (2006). Qualitative identification of tea categories by near-infrared spectroscopy and support vector machine. *Journal of Pharmaceutical and Biomedical Analysis*, *41*(4), 1198-1204.

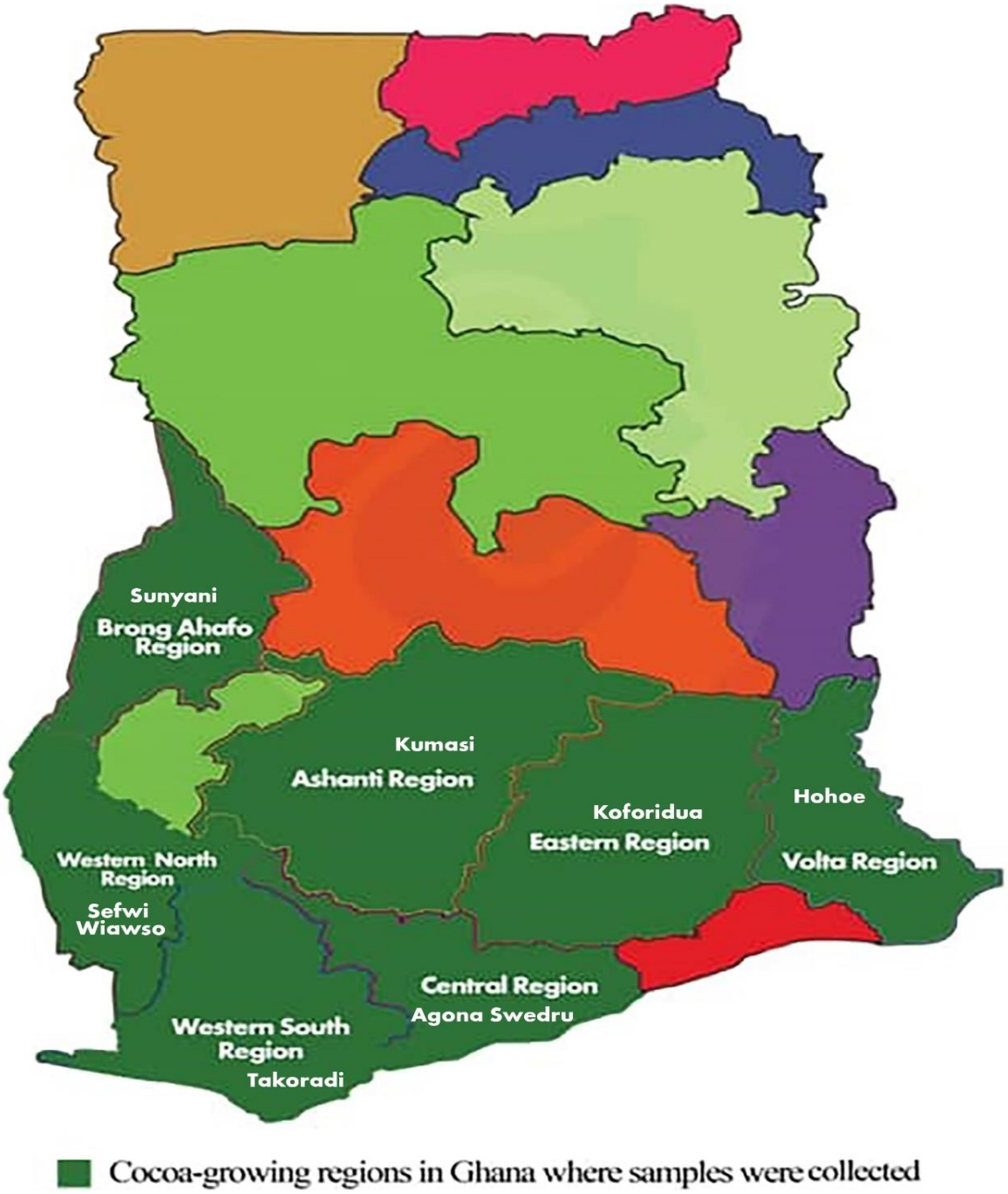
Zhu, Z., Chen, S., Wu, X., Xing, C., & Yuan, J. (2018). Determination of soybean routine quality parameters using near-infrared spectroscopy. *Food science & nutrition*, *6*(4), 1109-1118.

Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

APPENDICES

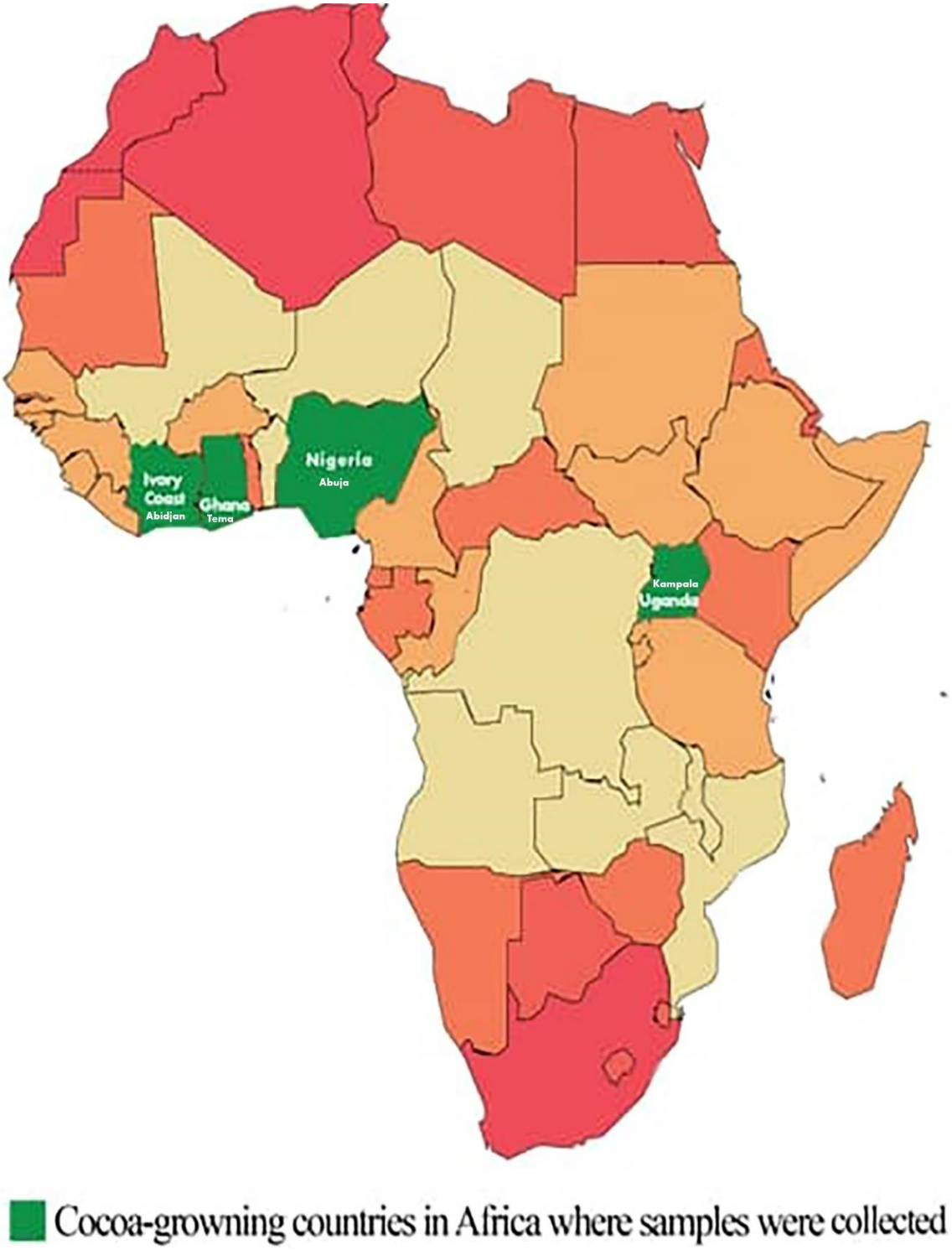
APPENDIX 1: COCOA-GROWING REGIONS IN GHANA WHERE SAMPLES WERE COLLECTED.



Digitized by Sam Jonah Library

University of Cape Coast https://ir.ucc.edu.gh/xmlui

APPENDIX 2: COCOA-GROWING COUNTRIES IN AFRICA WHERE SAMPLES WERE COLLECTED.



Digitized by Sam Jonah Library