Evaluation of near infrared spectroscopy for prediction of quality attributes and authentication of green coffee beans

Dissertation
to obtain the Ph. D. degree
in the International Ph. D. Program for Agricultural Sciences in Goettingen
(IPAG)
at the Faculty of Agricultural Sciences,
Georg-August-University Göttingen, Germany

presented by

Adnan

born in Jakarta, Indonesia

Göttingen, August 2020

D7

1. Name of supervisor: Prof. Dr. Elke Pawelzik

2. Name of co-supervisor: Dr. Christian Möllers

Date of dissertation: 23 November 2017

Acknowledgements

Göttingen, August 2020

Adnan

Table of Content

List of Tables

List of Figures

1. General introduction

1.1. The origin, species characterization, distribution and economic importance

Coffee is one of the most popular beverages in the world. It is easily found at groceries, coffee shops and at home. There is a variety of coffee products such as green beans, roasted beans, ground coffee or coffee drink on the market. Drinking coffee has become the world culture since centuries and it's still progressing (Adams, 2012; Kjeldgaard and Ostberg, 2007).

Coffee spread around the world starting from Ethiopia, which is known as the origin of the wild coffee species. Ethiopia has a great diversity of coffee species with regard to morphological, ethnobotany, and DNA based marker data. There are more than 5,000 accessions conserved in the main gene bank (Labouisse et al., 2008). Nevertheless, gene erosion can be observed in commercial coffee cultivars. Approximately 103 species and seven intraspecific coffee genera (*Rubiaceae*) were recorded in taxonomy, still, the coffee species traded most frequently are *Coffea arabica*, also known as Arabica and *Coffea canephora* which is known as Robusta (Davis et al., 2006).Arabica and Robusta have different species characterization. Arabica is tetraploid species ($2n = 4x = 44$ chromosomes) and Robusta is diploid with $2n = 2x = 22$ chromosomes. The leaves of the Arabica plant are slender and more delicate than Robusta. Arabica is usually cultivated above 1300 m above sea level (asl) and Robusta below 100 m asl (Wintgens, 2004).

Arabica and Robusta have distinct taxonomic classification (Davis et al., 2006); chemical composition (Alves et al., 2009; Carrera et al., 1998; Casal et al.,

2000; González et al., 2001; Ky et al., 2001; Martín et al., 1998, 2001); and sensory attributes (Gimase et al., 2014; Maeztu et al., 2001). There are several analytical methods to determine between species, such as physical, chemical, and biological approach. Nevertheless, it is still a challenge to discriminate green coffee beans species in terms of the accuracy of the method, time duration to perform analysis and the samples varieties (Toci et al., 2016).

As an alternative, NIR spectroscopy is potential to discriminate among green beans species. NIR spectroscopy discriminates between pure Arabica and Robusta and also blends of these species, with an accuracy of up to 96.2% (Downey et al., 1994). Another study also demonstrated NIR spectroscopy is reliable to identify Arabica and Robusta of green beans from different producing countries, up to 100% accuracy (Buratti et al., 2015).

Coffee plantation can be found in the regions located between latitudes 30°N and 30°S (Anderson and Smith, 2002). Hence, the origin of coffee refers to coffee growing region at the present time. Whereas, the origin related to a certain quality or characteristics, that can be used as a product differentiation (Teuber, 2010).

Ensuring the authenticity of the origin of the green coffee beans is still a challenge. Traders usually depend on the certificates of the origin to authenticate the beans. A certificate of origin contains information regarding the quality parameter of coffee, such as the origin, species, defects, beans size, color and sensory attributes (Feria-Morales, 2002; Van Hilten et al., 2011). This information is based on trust and needs validation.

Several chemical compositions of the green beans, such as chlorogenic acids, cinnamoyl amides, cinnamoyl glycosides, free phenolic acids, and

methylxanthines, are potential as a marker to determine the origin by using multivariate analysis, but it still need further studies with a larger representative sample set at different location (Alonso-Salces et al., 2009). Chemical composition in the beans depended on post-harvest treatment, which could lead to variations even in the same location (Duarte et al., 2010). Therefore, chemical analysis as a reference method to determine the origin needs to be taken with precaution.

Another promising method to identify the origin is stable isotope analysis (Rodrigues et al., 2009). There is still limited information regarding stable isotope analysis for the determination of the origin of the green beans. Stable isotope ^{2}H and ^{18}O are potential as marker with an error rate using linear discrimination analysis (LDA) of 5.1% and cross-validation error rate of 7.7%, proved qualified to determine whether coffee beans originated from Africa or Central-South America but could not reliably determine an Indonesian, Jamaican or Hawaiian origin (Weckerle et al., 2002). Other stable isotopes ^{11}B and ^{87}Sr were a feasible assessment for origin determination (Liu et al., 2014).

NIR spectroscopy showed as a potential tool to identify between Nicaraguan and Costa Rican coffee origins (Bertrand et al., 2005). However, it is unclear, whether NIR spectroscopy is a good tool for origin determination due to genotype differences or due to other factors in that study. There is an information gap regarding the origin determination by NIR spectroscopy and thus more experiments in this area are necessary. Indonesia, as one of coffee exporting country, is the third biggest coffee producing country after Brazil and Vietnam (Baroh et al., 2014). In 2014, Indonesia produces 643,857 ton of green beans (7.2% of world production) and most of the coffee plantation is owned by the small

farmer (96.19%). Robusta (73.57%) is produced more than Arabica (26.43%). From the total production, Indonesia export green coffee beans of 384,816 ton with trading value US$ 1,039,341, which 9.49% of this quantity is exported to Germany (Triyanti, 2016).

1.2. Quality parameters and authentication of green coffee beans

The perception of quality in coffee has a broad spectrum of understanding. It can be started from the quality of the green beans that include several parameters e.g. origin, moisture content, defects, bean size and shape, color, roasting characteristic and sensory attributes (Feria-Morales, 2002). The quality of coffee can also be related to how the beverage is served (Chen and Hu, 2010).

Species and origin authentication of green beans is also a concern in the coffee industry. Adulteration between Arabica and Robusta as the most traded coffee species is a problem (Toci et al., 2016). Consumer preference also varies on different coffee origin. Fraud can occur due to price variability on different species and origin (Sepúlveda et al., 2016). Therefore, authentication is important to avoid potential adulteration and fraud.

Table 1.1. Quality attributes and authentication parameter of green coffee beans.

Parameters	Attributes*	Author(s)/year
Quality		
Moisture content	The safety range is 8.0-12.5%. There are several methods to determine moisture content: ISO 1446, 1447 and 6673	(Reh et al., 2006)
Defects	Indonesia categories defect beans such as fully or partially black beans, faded beans, crushed beans, moldy beans, immature beans. Grade 1 considered as the best quality with a maximum score of 11 defect beans from 300 g sample of green coffee.	(Indonesian National Standard, 2008)
Bean size and shape	bean shape: round, long; bean size: small (<14 mm), medium (14 - 16 mm), bold (>17 mm)	(Kathurima et al., 2009)
Sensory test	Fragrance/aroma, flavor, aftertaste, balance, preference, acidity, body	(Gimase et al., 2014)
Authentication		
Species	Arabica and Robusta	(Górnaś et al., 2014)
Origin	Brazil, Kenya, Ethiopia, Indonesia	(Rodrigues et al., 2011b)

*Different countries have their own regulations on determination methods.

Despite the complexity of coffee quality and authentication issues under different parameters (Table 1.1.), it is still very important to have a method to assess the quality and authenticate that is fast, reliable and accurate for practical applications. One of the methods that meet this requirement is near infrared (NIR) spectroscopy (Toci et al., 2016). However, research on its applicability for green coffee beans analysis is still lacking. Therefore, this study is focused on this theme.

1.3. NIR spectroscopy as an alternative method to determine quality

The reference methods used to determine the quality of coffee underlie different regulations in each country. These differences in regulation can become a barrier for trading. Furthermore, the reference method usually consumes a lot of time, labor and chemicals. NIR spectroscopy, as an alternative to the reference method, has the advantages of saving time and resources. NIR spectroscopy requires minimal sample preparation and for a particular quality does not need

sample preparation at all. The whole green beans or roasted beans can be used as a sample without any beforehand treatment. The quality measurement using NIR can potentially be used as a non-destructive analysis and real time analysis (Barbin et al., 2014).

The implementation of NIR spectroscopy in coffee industry still faces the challenge of building a sufficient database for robust and reliable prediction. Coffee beans are organic matter and therefore influenced by several factors such as climate, genotype, agricultural practice and postharvest treatment (Decazy et al., 2003; Joët et al., 2010; Kitzberger et al., 2014; Läderach et al., 2011; Oberthür et al., 2011). This variability has to be covered in the prediction model. A wide variability of coffee samples is important in order to obtain a better calibration.

Another challenge is the requirement of specific skills in statistical data analysis. Chemometrics is employed in order to understand the correlation between NIR spectral data and the beans quality. Multivariate analysis is normally used to simplify the complex data (Barbin et al., 2014).

There are several steps to perform NIR spectroscopy. The first recommended step after acquiring the NIR wavelength data is an outlier detection. The identification of correct outlier is important to prevent an inaccurate model (Shabbak and Midi, 2012). Afterward, pre-processing data is needed to reduce unrelated data information (Esteban-Díez et al., 2004). The next step is to analyze data using multivariate analyses such as principal component analysis (PCA) and partial least square regression (PLS) through calibration and prediction model (Craig et al., 2012; Wold et al., 2001).

1.4. Objectives of the studies

NIR spectroscopy has been utilized to predict quality and authenticate green coffee beans. However, to what extent NIR spectroscopy examine green beans origin from Indonesia is still lacking. Indonesia has a lot of variety of coffee and needs to be evaluated in the term of quality and authentication. Therefore, the present study was performed to achieve the following objectives:

a. Predict moisture content in intact green coffee beans of Arabica and Robusta by NIR spectroscopy and chemometrics.

b. Evaluate the applicability of UV-Vis and NIR spectroscopy to discriminate between green coffee beans of Arabica and Robusta.

c. Examine the potential of NIR spectroscopy and stable isotope analysis of oxygen, hydrogen, and strontium to identify the origin of the green coffee beans from Java Island, Indonesia.

The result of the present studies will be presented in the following chapters. Rapid prediction of moisture content in intact green coffee beans will be investigated in chapter 2. Discrimination of Arabica and Robusta will be evaluated in chapter 3, while in chapter 4, the origin of Java green coffee beans will be identified. Finally, a general discussion and conclusion will be presented in chapter 5, followed by a summary in chapter 6.

2. Rapid prediction of moisture content in intact green coffee beans using near infrared spectroscopy

2.1. Abstract

Moisture content (MC) is one of the most important quality parameters of green coffee beans. Therefore, its fast and reliable measurement is necessary. This study evaluated the feasibility of near infrared (NIR) spectroscopy and chemometrics for rapid and non-destructive prediction of MC in intact green coffee beans of both *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) species. Diffuse reflectance (log 1/R) spectra of intact beans were acquired using a bench top Fourier transform NIR instrument. MC was determined gravimetrically according to The International Organization for Standardization (ISO) 6673. Samples were split into subsets for calibration (n = 64) and independent validation (n = 44). A three-component partial least squares regression (PLSR) model using raw NIR spectra yielded a root mean square error of prediction (RMSEP) of 0.80% MC; a four component PLSR model using scatter corrected spectra yielded a RMSEP of 0.57% MC. A simplified PLS model using seven selected wavelengths (1155, 1212, 1340, 1409, 1724, 1908, and 2249 nm) yielded a similar accuracy (RMSEP: 0.77% MC) which opens the possibility of creating cheaper NIR instruments. In conclusion, NIR diffuse reflectance spectroscopy appears to be suitable for rapid and reliable MC prediction in intact green coffee; no separate model for Arabica and Robusta species is needed.

Keywords: quality; rapid methods; infrared spectroscopy; Coffea arabica (Arabica); Coffea canephora (Robusta); chemometrics

2.2. Introduction

Moisture content (MC) is one of the most important quality parameters of green coffee beans. Most importing and exporting countries regulate MC as one of the quality standards for green coffee beans. The safety range for MC is 8.0–12.5%, based on fresh matter (International Coffee Organization, 2013; Pittia et al., 2007; Reh et al., 2006). MC outside the safety range impairs the bean quality and safety. Beans with a MC above 12.5% are not allowed to be shipped and traded (Van Hilten et al., 2011). MC below 8% causes shrunken beans and an

unwanted appearance (Gautz et al., 2008), whereas MC above 12.5% facilitates fungal growth and mycotoxin production (e.g., ochratoxin A) that are risks to human health (Palacios-Cabrera et al., 2004; Pardo et al., 2005).

Coffee is harvested in the form of ripe berries and has a MC of more than 60% (Finzer et al., 2003). These ripe berries are processed through several steps of (wet or dry) postharvest treatments resulting in green coffee beans. Farmers generally dry the beans under the sun. The dried beans often do not meet the standard requirements for MC, resulting in a lower price (Subedi, 2011). For example, green beans obtained in the Bengkulu Province of Indonesia had a MC of 10.1–18.6% (Yani, 2008) and those in West Nusa Tenggara Province had a MC of 11.0–14.1% (Aklimawati et al., 2014).

MC control is also important for the storability of the beans. An inappropriate storage environment (e.g., non-aerated silos and bag storage) affects MC fluctuation. The MC of green coffee beans stored in non-aerated silos increased up to 15.4% during rainy season. This moisture increase leads to the accumulation of glucose and an unpleasant taste in the beverage (Bucheli et al., 1998).

Furthermore, MC is crucial before the roasting process. The same roasting temperature and time with different MCs can result in different quality attributes—like color, density, and aroma—of the end product (Baggenstoss et al., 2008). Consequently, an identical MC of green coffee beans is important for the roasting procedure in order to produce a consistent quality of roasted beans.

Therefore, a fast and accurate determination of MC in green coffee beans is vital. Up to date, the standard method for determining MC is the gravimetric method, where a drying chamber with a certain temperature and time is used to

dry the beans and afterwards the mass loss is calculated. International standards for MC measurement of green coffee beans are The International Organization for Standardization (ISO) 1446, 1447, and 6673 (Mendonça et al., 2007; Reh et al., 2006). Thereof, ISO 6673 which requires less preparation and the shortest drying time (105 °C for 16 h) is widely accepted as a reference method among importing and exporting countries. Apparently these gravimetric methods do not suffice when the information on MC is needed instantly (Gautz et al., 2008) which is why we researched alternative methods.

Near infrared (NIR) spectroscopy has been widely investigated for rapid, often non-destructive, determination of the compositional and quality traits of agricultural products. For example, previous work in our group predicted rapid and non-destructive analysis of mango quality attributes using NIR spectroscopy and chemometrics (Munawar et al., 2016). NIR spectroscopy makes use of the fact that NIR radiation in the range of 780–2500 nm predominantly interacts with hydrogen bonds—e.g., O–H, C–H, N–H, S–H. NIR radiation that hits a sample may be transmitted, absorbed, or reflected, this depends on the chemical composition and physical factors of the sample. The intensity of transmitted, absorbed, or reflected radiation is then recorded by NIR spectroscopy (Blanco and Villarroya, 2002; Nicolaï et al., 2007).

Specific wavelengths (1450 and 1940 nm) were identified to be highly correlated with water content (Isengard, 1995; Reh et al., 2006). Predicting MC using NIR spectroscopy in any agricultural product is more complex and should not be based on wavelengths limited to 1450 and 1940 nm. MC does not only reflect water, but also loss of volatile compounds during drying (Reh et al., 2006). In fact, NIR has some disadvantages, e.g., overlapping of wavelengths that

correspond to specific organic compounds, and scattering problems (Barbin et al., 2014; Blanco and Villarroya, 2002). It is therefore necessary to carefully develop calibration models for NIR based predictions (Isengard, 2001, 1995).

Prediction of MC by NIR spectroscopy has been developed over years for many agricultural products (Büning-Pfaue, 2003). A regression model was developed to predict MC in (ground) green coffee bean (*Coffea arabica* from Brazil) based on NIR diffuse reflectance (log 1/R) spectra (Morgano et al., 2008). To the best of our knowledge, this is the first study investigating the prediction of moisture content in intact green coffee beans of both *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) species by NIR spectroscopy and chemometrics. The main goal of this paper was to study the feasibility of NIR spectroscopy to predict moisture content (MC) in intact green coffee beans. We developed and validated calibration models based on diffuse reflectance spectra which were obtained using a benchtop near infrared instrument. Our decision to involve both Arabica and Robusta species stems from the fact that both species are commercially important but vary in their chemical composition. Furthermore, we used intact green beans such as no sample preparation would be needed— yet such an approach has not been documented. The results are especially relevant for those involved in coffee trading, production, and quality control. We also demonstrate the possibility of creating a simple NIR instrument which only uses a few important wavelengths to predict MC, rather than employing the full NIR spectrum.

2.3. Materials and methods

2.3.1. Materials

Green Arabica and Robusta coffee beans that were harvested in 2013 were obtained from a local trading company in Indonesia. The materials were divided into separate sample sets for calibration and validation purposes (Table 2.1). The beans were placed in an open plastic box with the size of 15.5 × 11 × 6 cm, and were stored in a climatic chamber (Rumed® type 1301, Rubarth Apparate GmbH, Laatzen, Germany) at 25 °C and a relative humidity range of 30–85%, in order to obtain a broad range of MC within 6–22%. Upon equilibration, samples were removed from the climatic chamber to record diffuse reflectance (log 1/R) data. Immediately thereafter, MC of the beans was determined.

Table 2.1. Characteristics of the coffee samples including species and origin

No.	Purpose	Species	Origin
1			West Nusa Tenggara
2		Arabica	South Sulawesi
3			Aceh
4	Calibration		South Sumatera
5		Robusta	Bali
6			East Java
7			North Sumatera
8		Arabica	West Java
9			North Sumatera
10	Validation		South Sumatera
11		Robusta	East Java
12			Bengkulu

2.3.2. NIR spectroscopy

A bench top Fourier transform near infrared (FT-NIR) instrument with sample cup rotation (Thermo Nicolet Antaris MDS, Thermo Fisher, Waltham, MA, USA) was used to acquire diffuse reflectance spectra (log 1/R) of bulk samples of green coffee beans (40 g) on a Petri dish with a diameter of 7 cm.

Spectra were collected according to a workflow developed using the software Result Integration Software (Result[TM] version 3.0, Thermo Fisher, Waltham, MA, USA). Internal background spectra were collected once every hour. High resolution diffuse reflectance (log 1/R) spectra at a wavelength range of 1000 to 2500 nm with 2 nm intervals were recorded as the averages of 64 scans. Thus, the spectra consisted of 1557 data points. Three replicates were acquired per sample and the spectra were averaged before further calculations. In total, this resulted in 108 spectra of 12 samples differing in moisture content, species, and origin.

2.3.3. Moisture content determination

MC (% wet basis) was determined was based on ISO 6673 (Reh et al., 2006). A forced air electrical oven (Thermicon P[®] type UT6120, Heraeus Instruments GmbH, Hanau, Germany) was used to dry approximately 10 g whole green coffee beans in open glass petri dishes (diameter: 14 cm, height: 2.3 cm) at 105 ± 1 °C for 16 h. Samples were limited to six origins with two replications per drying cycle in order to maintain an equal amount of mass loss during drying. The petri dishes were closed with glass lids immediately after drying had completed, and then they were stored in desiccators for 1 h in order to cool down the samples to ambient temperature. Finally, samples were weighted (Type LP 620 S, Sartorius AG, Göttingen, Germany) to calculate MC based on weight loss; data are given as the average from two replications (Equation 2.1). Across all samples, average standard deviation of replicate MC determinations was 0.21% MC (Median: 0.08% MC).

$$MC = \frac{W_w - W_d}{W_w} \tag{2.1}$$

where MC is the moisture content (%) of green coffee beans (wet basis), W_W is the wet weight of the sample, and W_d is the weight of the sample after drying.

2.3.4. Data processing

The statistical software (The Unscrambler® X version 10.2 Network Client, CAMO software AS, Oslo, Norway) was used for further processing of the spectral data. Regression models to predict MC in green coffee beans were developed with a subset of calibration samples ($n = 64$), and then the models were tested using the subset of validation samples ($n = 44$) to evaluate the accuracy.

Firstly, spectral outliers were identified using Principal Component Analysis (PCA) and Hotelling's T^2 ellipse 5% plot, based on all samples' ($n = 108$) raw spectra. Afterwards, several pre-processing methods were applied to compensate the disadvantages of NIR, e.g., the scattering and material size (Blanco and Villarroya, 2002; Esteban-Díez et al., 2004). In detail, smoothing (moving average, Gaussian filter, median filter) window size of 3, 7, 11, 15, 19; Savitsky–Golay derivative (First derivative, two polynomial order; second derivative, two polynomial order; third derivative, three polynomial order) window size of 3, 7, 11, 15, 19; normalization (area, mean); baseline correction (baseline offset, linear baseline correction); standard normal variate (SNV); orthogonal signal correction (OSC) (non-linear iterative partial least squares algorithm, number of component 1); multiplicative scatter correction (MSC) (full MSC model); and extended multiplicative scatter correction (EMSC) were applied. Subsequently, the models were compared in terms of prediction accuracy and model robustness (number of latent variables). MSC and EMSC were applied to

the calibration data. Upon model validation, the processing was also applied to the validation data set prior to prediction.

Calibration models were developed using both partial least squares regression (PLSR) and multiple linear regression (MLR). For PLSR, the full spectra (1557 wave numbers, mean centered) were used. Full cross validation was applied to estimate calibration errors. Regression coefficients were obtained from PLSR to determine the important wavelengths, i.e., those that correlated most to MC. A subset of selected wavelengths was then used as an input for full rank MLR and PLS regression to identify the most parsimonious yet robust model. Leverage correction was applied with MLR to estimate calibration errors. The calibration models derived from PLSR and MLR were evaluated by the number of latent variables (LVs), R^2 of calibration, R^2 of cross validation, root mean square error of calibration (RMSEC), and root mean square error of cross validation (RMSECV). Finally, all models were validated in terms of their prediction accuracy using a separate validation data set. Parameters used were R^2 of prediction, root mean square error of prediction (RMSEP), standard error of prediction (SEP), bias, and residual predictive deviation (RPD) (Fearn, 2002; Morgano et al., 2008).

2.4. Results

2.4.1. Spectral properties, outliers, and effect of pre-processing

According to an initial PCA using all raw spectra and projection of the Hotelling's T^2 ellipse, four samples were suspected as spectral outliers (Figure 2.1). Subsequent modeling with and without these potential outliers, respectively revealed that model accuracy was not significantly affected. Thus, the suspected outliers were not excluded.

Inspection of the raw data also revealed that the NIR diffuse reflectance spectra of intact green coffee beans are influenced by scatter (Figure 2.2a). Several pre-processing methods were applied to eliminate the scatter. Application of EMSC proved to improve the prediction accuracy; the EMSC corrected spectra are shown in Figure 2.2b. Inspection of EMSC corrected spectra indicated that several wavelength regions reflect the chemical information regarding moisture content.
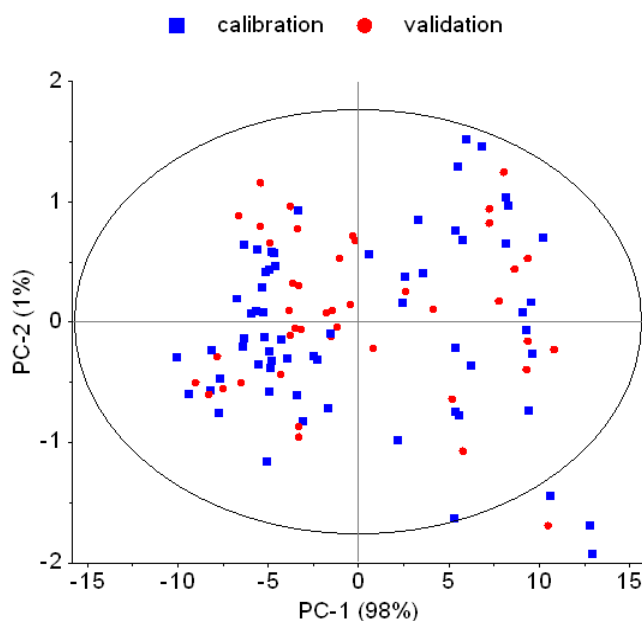
Figure 2.1. Score plot of principal component analysis (PCA) using raw NIR spectra (log 1/R) with Hotelling's T^2 ellipse for outlier inspection. Calibration samples (squares) and validation samples (circles) are marked accordingly. PC: principal component.
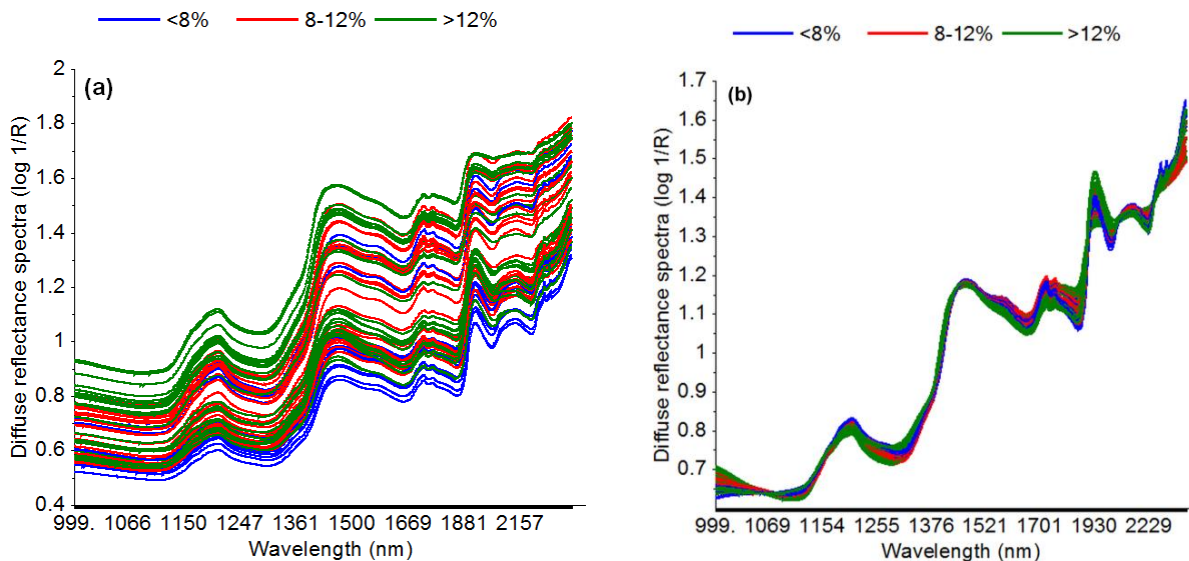
Figure 2.2. Diffuse reflectance spectra (log 1/R) of calibration model. (a) Raw spectra;);
(b) EMSC (extended multiplicative scatter) corrected spectra

2.4.2. Prediction of moisture content from NIR reflectance spectra

Several preprocessing methods were applied to build the model (see Section 2.3.4). Nevertheless, none of the preprocessing methods yielded a better accuracy than models using raw data (Appendix 1). Selected results of the various chemometric approaches to predict MC from NIR reflectance spectra are given in Table 2.2. The most parsimonious PLSR model on the full spectral range was achieved using raw spectra and three latent variables. Its prediction accuracy was, however, somewhat compromised when using the independent validation data set. Using the EMSC corrected spectra instead of the raw spectra yielded a similar R² while the prediction errors were comparably low both for the calibration and the validation data set. Yet, this model used four latent variables, e.g., it was less parsimonious compared to the model based on raw data.

Table 2.2. Statistical parameters of the developed prediction models for moisture content (MC) in green coffee beans using near infrared spectra.

Model	Parameter	Full Spectral Range PLSR		Spectral Subset	
		Raw	EMSC	Raw (MLR)	Raw (PLS)
	LVs	3	4	n/a	3
	R^2 calibration	0.9834	0.9850	0.9839	0.9743
Calibration	R^2 cross validation	0.9802	0.9811	0.9779	0.9698
	RMSEC (% MC)	0.52	0.49	0.51	0.65
	RMSECV (% MC)	0.58	0.56	0.60	0.71
	R^2 prediction	0.9641	0.9817	0.9632	0.9669
Prediction	RMSEP (% MC)	0.80	0.57	0.93	0.77
	Bias (% MC)	0.42	0.28	0.45	0.39
	RPD	6.21	8.53	3.47	6.39

PLSR: partial least squares regression using full spectral range (1000 to 2500 nm, 1557 data points); MLR/PLS: multiple linear and partial least squares regression using selected wavenumbers (1155, 1212, 1340, 1409, 1724, 1908, and 2249 nm); LVs: Latent variables (for PLS only); R^2: the coefficient of determination; RMSEC: root mean square error of valibration; RMSECV: root mean square error of cross validation; RMSEP: root mean square error of prediction; SEP: standard error of prediction; RPD: residual predictive deviation; n/a: not applicable; MC: moisture content.

Principal components (PC) 1 and 2 of the PLSR model based on raw spectra explain 99% of spectral data variance and 51% of MC variance; a clear separation of Arabica and Robusta species is to be seen (Figure 2.3a). PC 2 and 3 together explain 94% of MC variance (Figure 2.3b).
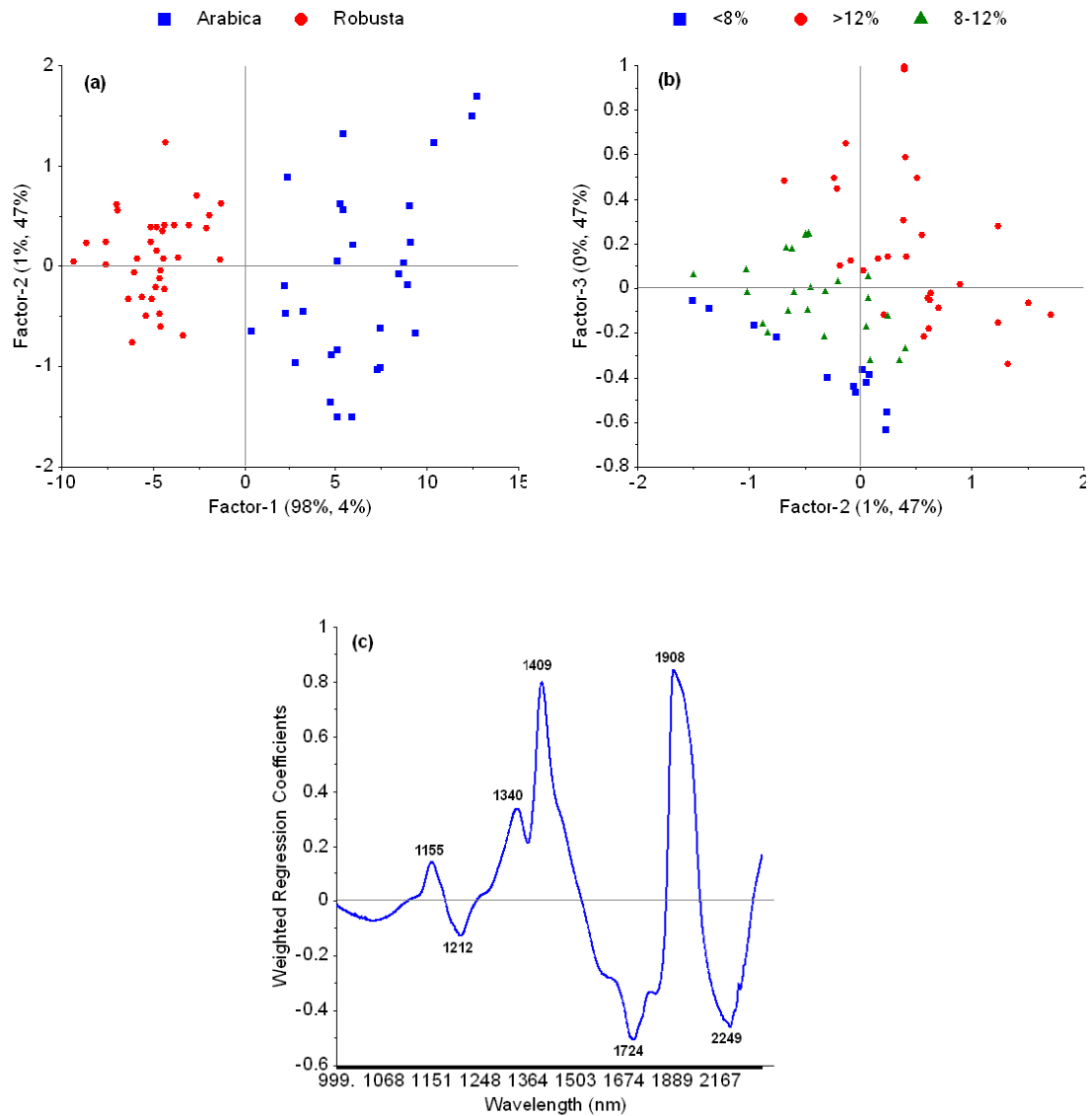
Figure 2.3. Score plots of PLSR for moisture content prediction based on raw diffuse reflectance (log 1/R) near infrared spectra. (a) A distinct clustering of Arabica and Robusta coffee samples is observed; (b) Sample allocation is following moisture content indicating the importance of PC 2 and 3 for moisture prediction; (c) Weighted regression coefficients obtained from PLSR using raw spectra.

Weighted regression coefficients obtained from PLSR on raw data (Figure 2.3c) were then used to study whether the model could be even simplified. Note that weighted and raw regression coefficients are the same as long as spectral data are not scaled but only mean centered; this was applied here. Seven wavelengths were selected due to their regression weights. That is, the intensities of 1155, 1212, 1340, 1409, 1724, 1908, and 2249 nm were used as input data to

develop a MLR calibration model. Thus, a similarly accurate model was obtained (Table 2.2); the prediction error for the validation test set was significantly lower ($p < 0.05$) for the MLR model (0.93% MC) as compared to the EMSC model using raw data (0.57% MC). The resulting MLR model is given in Equation 2.2.

$$MC\ (\%) = -4.20 + 115.02\ (V1) + 0.40\ (V2) - 116.18\ (V3) +$$
$$76.16\ (V4) - 97.72\ (V5) + 63.76\ (V6) - 17.59\ (V7)$$

(2.2)

where, V1 to V7 are the intensities of the wavelengths 1155, 1212, 1340, 1409, 1724, 1908, and 2249 nm, respectively. When subjecting this spectral subset to PLS, the predictive ability of a three LV model was even improved as compared to the full-rank MLR model (Table 2.2); its prediction error (0.77% MC) was significantly lower than the MLR model ($p = 0.015$). It is, however, not significantly different from the PLSR model using raw data ($p > 0.05$).
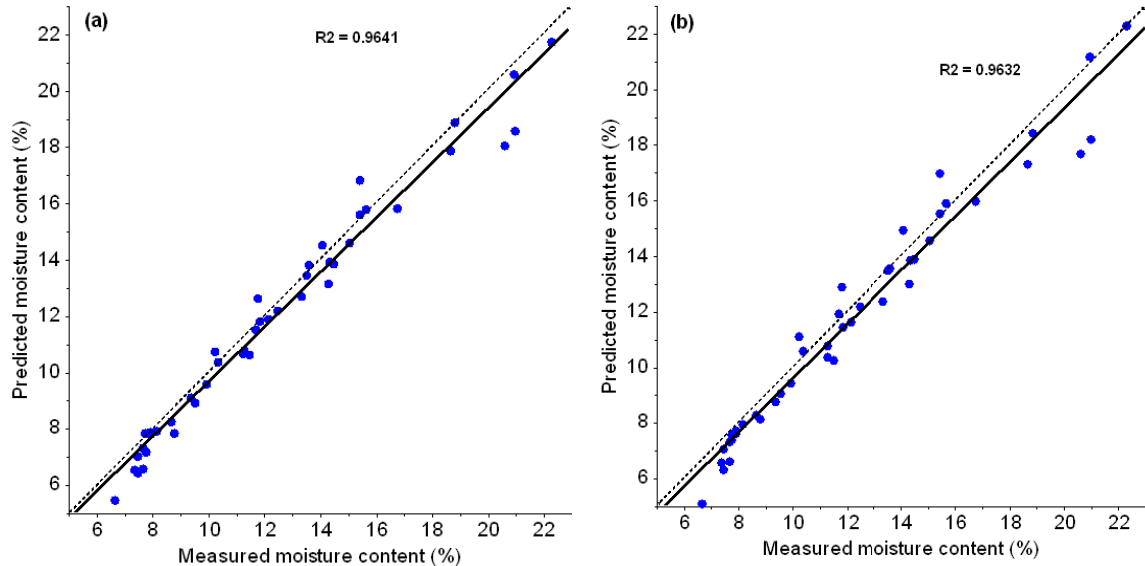
Figure 2.4. Predicted vs. measured moisture content of green coffee beans based on raw diffuse reflectance (log 1/R) near infrared spectra. (a) PLSR; (b) MLR.

PLSR and MLR using raw spectral data yielded a good correlation of reference versus predicted MC (Figure 2.4a,b). Also, the model's bias is close to the error of the reference method (0.21% MC, see 2.3.3).

2.5. Discussion

2.5.1. Outliers and effect of pre-processing

For outlier detection, PCA and subjection of the Hotelling's T^2 ellipse along with residuals and influence plot, and Q-residuals plot, were used which are common approaches in multivariate analysis. Identifying true outliers is important to prevent false inferences (Shabbak and Midi, 2012). In this experiment, four samples were suspected to be outliers (Figure 2.1). Explained spectral variance (PC1 + PC2) based on diffuse raw data reflectance (log 1/R) was 99%. Elimination of suspected outliers did not increase the explained variance. Further comparisons of PLSR with and without the suspected outliers yielded only very slight improvement in R^2 which indicates that the suspected outliers were no real outliers. Similarly, Morales-Medina and Guzmán (2012) examined multivariate data using Hotelling's T^2 ellipse. They also decided to not exclude the suspected outliers because they did not significantly affect the explained variance found through PCA.

Various pre-processing methods were applied to the raw spectra. This aims at reducing noise and improving the accuracy of the prediction model (Pizarro et al., 2004). EMSC was effective to remove scatter which was shown also in other studies (Sørensen et al., 2012). Accordingly, the prediction errors were the lowest when using EMSC corrected data for PLSR (Table 2.2). The resulting model, however, was surprisingly less parsimonious, i.e., it needed one more latent variable. Pizarro et al. (2004) also reported that none of the pre-processing

methods studied (first and second derivation, MSC, standard normal variate) improved the prediction for ash and lipid content in roasted coffee significantly as compared to using raw data; only OSC and direct orthogonal signal correction (DOSC) enhanced the model performance remarkably.

2.5.2. Prediction of moisture content using NIR spectra

Raw spectra were selected as an input to build the final PLSR model because this resulted in the lowest number of latent variables, the highest R^2 and lowest root mean square error compared to other pre-processing methods (Table 2.2). A model with these criteria is preferable. Kamruzzaman et al. (2012) also considered the number of latent variables together with R^2 and prediction errors to select the most appropriate model for prediction of water, fat, and protein content in lamb meat. Both the robustness and the predictive ability of a given model are of importance. If one considers only R^2, RMSEP, or RPD, which reflect the predictive ability, likely models using more latent variables would be preferred over models using less latent variables. In terms of robustness, however, a model using less latent variables is less prone to overfitting than a model using more latent variables.

Further examination of the PLSR score plots (based on raw spectra) revealed a distinct clustering of Arabica and Robusta samples on the first latent variable, explaining 98% in the spectral data variance but only 4% of moisture variance (Figure 2.3a). To understand this clustering, the loading weights of the first LV were inspected. As a result, important wavelengths are related to several chemical compounds, e.g., caffeine, chlorogenic acid, lipids, protein and amino acids, sucrose, carbohydrates, trigonelline and, of course, water (Ribeiro et al., 2011). These compounds were shown to vary between species which is why their

spectral contributions can be used to discriminate between species (Ky et al., 2001; Martín et al., 2001). Using PC 2 and 3 which together explain 94% of moisture variance, samples are allocated according to moisture content levels (Figure 2.3b). Thus, a three component PLSR model allows prediction of moisture content on both Arabica and Robusta species. The advantages of inputting raw spectra rather than pre-processed spectra, firstly reduces the complexity of calculations and therefore secondly reduces the computation time. These advantages will be useful for online and real time prediction in the future.

The statistical parameters of calibration and prediction accuracy were similar for the developed PLSR models, especially for the model based on EMSC corrected spectra. This indicates that the PLSR model is robust in terms of predicting unknown samples accurately. We also investigated PLSR models based on raw spectra within individual species. However, the results were not better than the PLSR model which was developed across species. The PLSR model obtained in this experiment resulted in a similar accuracy compared to what was reported by Morgano et al. (2008). That study predicted the MC of green Arabica coffee beans, based on smoothed spectra, which yielded an R^2 of calibration = 0.507, R^2 of validation = 0.669, and RMSEV of 0.55% MC (R^2 recalculated from r).

Even simplified MLR and PLS models were built using selected wavelengths based on their relative importance in the PLSR model. This experiment showed that near infrared diffuse reflectance intensities at 1155, 1212, 1340, 1409, 1724, 1908, and 2249 nm highly correspond to MC (Figure 2.3c). According to Ribeiro et al. (2011), these wavelengths are related to the absorbance of the second overtone of C–H, first combination overtone of C–H, first overtone of O–H and N–

H, second overtone of C=O, and combination of O–H and N–H, respectively. Obviously, these wavelengths are not exactly located at the water bands which indicate that it may well be useful to apply indirect relationships in prediction models. Plus, it was shown that the degradation of organic components during drying for MC determination needs to be considered. Reh et al. (2006) proved that, using ISO 6673, the beans lose 0.39% of their mass besides water. Thus, MC is calculated as a sum of extracted water and mass losses of other compounds. Similarly, Pan et al. (2015) found that MC in beet slices highly corresponded to spectral intensities at 968, 1078, and 1272 nm, i.e., not exactly located at the water bands.

The MLR model, as well as the PLS model based on the spectral subset, yielded a good accuracy both for calibration and validation thus proving their robustness (Figure 2.4b). The biases measured by PLSR and MLR were close to the method error of determining moisture content based on ISO 6673. Moreover, the ratio of the standard deviation of the target variable and the SEP of a given model, commonly referred to as RPD (residual predictive deviation), is often used to assess the performance of prediction models; higher RPD values indicate a better predictive performance (Fearn, 2002). Here, the models yielded RPD values of about 3 to 8 (Table 2.2) which is considered good (Williams, 2006). This shows the potential of near infrared spectroscopy to replace the reference method when a fast and non-destructive prediction is needed, e.g., when trading or for in-line process control.

Finally, the remarkable reduction of variables without a relevant loss of accuracy opens the possibility of creating a simple NIR instrument which only uses a few important wavelengths to predict MC, rather than employing the full

NIR spectrum. Specific LED light sources emitting only selected wavelengths can potentially reduce the costs of an NIR instrument.

2.6. Conclusion

The results indicate that a fast, non-destructive prediction of moisture content in intact green coffee beans is feasible using near infrared diffuse reflectance spectroscopy. EMSC effectively reduces scatter apparent in raw spectra. Thus, the prediction accuracy using EMSC corrected spectra is improved at the cost of a somewhat less parsimonious model. A simplified model based on only seven selected wavelengths points to the possibility of a cheaper instrumentation. The calibration model can be applied for both Arabica and Robusta species. In conclusion, NIR is deemed feasible to replace gravimetric methods for routine applications where a timely result may outweigh the loss of accuracy as compared to the drying methods.

3. Reliable Discrimination of Green Coffee Beans Species: A Comparison of UV-Vis-Based Determination of Caffeine and Chlorogenic Acid with Non-Targeted Near-Infrared Spectroscopy

3.1. Abstract

Species adulteration is a common problem in the coffee trade. Several attempts have been made to differentiate among species. However, finding an applicable methodology that would consider the various aspects of adulteration remains a challenge. This study investigated an ultraviolet–visible (UV-Vis) spectroscopy-based determination of caffeine and chlorogenic acid contents, as well as the applicability of non-targeted near-infrared (NIR) spectroscopy, to discriminate between green coffee beans of the *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) species from Java Island, Indonesia. The discrimination was conducted by measuring the caffeine and chlorogenic acid content in the beans using UV-Vis spectroscopy. The data related to both compounds was processed using linear discriminant analysis (LDA). Information about the diffuse reflectance (log 1/R) spectra of intact beans was determined by NIR spectroscopy and analyzed using multivariate analysis. UV-Vis spectroscopy attained an accuracy of 97% in comparison to NIR spectroscopy's accuracy by selected wavelengths of LDA (95%). The study suggests that both methods are applicable to discriminate reliably among species.

Keywords: Arabica; Robusta; caffeine; chlorogenic acid; linear discriminant analysis; food fraud

3.2. Introduction

The adulteration of *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) is a common problem in the coffee trade (Toci et al., 2016). It results, among others, from the price difference between the species. Arabica receive more than 50% higher market price compared with Robusta. From 1990 to 2017, the average annual price of green beans of Arabica (US $2.51 per kg) was higher than that of Robusta (US $1.63 per kg). Arabica takes up approximately 58% of the global production of coffee compared with Robusta's 42% global share. This

implies that the consumption of Arabica is more preferable than Robusta (International Coffee Organization., 2018). Consequently, trade fraud involving the substitution of Arabica with Robusta cannot be avoided. Such fraud includes the addition of low-cost materials like coffee beans from different geographical regions or species without stating that in the product label (Martins et al., 2018).

Arabica and Robusta differ in several aspects—for example, taxonomic classification, morphology, bean size and color, chemical compounds, and sensory evaluation (Davis et al., 2006; Feria-Morales, 2002; Keidel et al., 2010). For example, the mean liking scores in a consumer test based on aroma, flavor, and mouthfeel led to significantly higher values for Arabica (6.0) in comparison with Robusta (4.4) based on a nine-point category scale where "1" means "not at all vivid" and "9" means "very vivid" (Kim et al., 2016). The results of this study support the view that many consumers prefer Arabica to Robusta. For these reasons, Robusta is considered an adulterant for Arabica.

The conventional procedure to discriminate among species is based on a visual inspection of the size, shape, and color of the beans (Mendonça et al., 2009). The limitation of this approach is that the physical characteristics of the beans differ considerably between species and variety due to various genotypes and environmental factors (Keidel et al., 2010). Another common method for differentiating among species is sensory testing (Kim et al., 2016; Salamanca et al., 2017). The disadvantages of this approach are that trained panels are not always available and they are expensive (Di Donfrancesco et al., 2014). A third disadvantage is that certain varieties of Arabica have sensory properties similar to those of Robusta—in terms of mouthfeel and bitterness; this would distort the test results (Esteban-Díez et al., 2004).

Caffeine and chlorogenic acid play important roles for flavor formation and health effects on humans (Jeszka-Skowron et al., 2016). Caffeine consumption has well-known effects on the stimulation of brain functions and improvement in mood and physical performance; it is also associated with the reduction of the development of chronic degenerative diseases. However, excessive caffeine consumption may expose the drinker to cardiovascular disease and blood pressure problems (dePaula and Farah, 2019). Chlorogenic acid is a major component of green coffee beans and an important dietary polyphenol with potential health benefits, including anti-inflammatory, antidiabetic, anti-obesity, and anticarcinogenic effects (Tajik et al., 2017).

Chromatographic techniques (e.g., high-performance liquid chromatography (HPLC) and gas chromatography (GC)) are commonly applied successfully to discriminate between coffee beans species (Górnaś et al., 2014; Toci et al., 2016). Chemical compounds, for example, trigonelline, tocopherol, caffeine, and chlorogenic acid, are used to differentiate between Arabica and Robusta (Oestreich-Janzen, 2013). Trigonelline levels and the sum of α-, β-, and γ-tocopherols vary depending on the species. About 0.8% of the dry weight (dw) is composed of trigonelline and 0.028% of the sum of tocopherols in Arabica, which is higher than in Robusta (0.7% dw and 0.011%, respectively) (Górnaś et al., 2014; Oestreich-Janzen, 2013). Conversely, the caffeine (1.3% dw) and chlorogenic acid (8.1% dw) content in Arabica are lower than in Robusta (2.3% dw and 9.9% dw, respectively) (Oestreich-Janzen, 2013).

Despite their accuracy, these chromatographic techniques involve equipment-intensive and time-consuming processes. Out of the different chemical compounds present in coffee, caffeine and chlorogenic acid were

selected as key compounds in the present study; as mentioned, the concentrations of these compounds tend to differ across green coffee beans species.

An alternative to chromatographic techniques is ultraviolet–visible (UV-Vis) spectroscopy, which provides simplified measurement procedures that are time- and cost-effective (Tomaszewska et al., 2013). To date, UV-Vis spectroscopy has been applied to measure the caffeine and chlorogenic acid content of coffee beans (Belay et al., 2008; Belay and Gholap, 2009; Dankowska et al., 2017; Navarra et al., 2017), and it has also been used to discriminate species on ground and roasted coffee. However, no previous study has so far used these compounds measured by UV-Vis spectroscopy to discriminate among green coffee beans species. We, therefore, investigated this approach as an alternative method to discriminate between Arabica and Robusta to help prevent fraud within the global coffee beans trade.

Previous studies have reported the applicability of NIR spectroscopy to discriminate between species (Buratti et al., 2015; Esteban-Díez et al., 2007). However, they rarely covered aspects such as altitude or genotype that can cause considerable differences within species and variety (Kitzberger et al., 2014). In this study, we have therefore evaluated the capability of NIR spectroscopy to discriminate among the species of intact green beans from a different origin, variety, and altitude, in comparison with the UV-Vis-based determination of caffeine and chlorogenic acid. These methods allow high throughput and low involvement of labor—faster examination at a lower cost to discriminate species is preferable. Thus, the application of these methods may

help prevent fraud in a desirable manner for the coffee industry, as compared with other existing methods.

3.3. Materials and Methods

3.3.1. Samples

Seventy-four green coffee beans samples from various locations on Java Island, Indonesia, were used in this study. The samples set of green beans represented different environmental factors, agricultural practices, and genetic characteristics, and were sourced from 38 processing facilities on Java Island during the harvesting season from July to August 2014. Of these 74 samples, 32 samples belonged to Arabica and 42 to Robusta. To ensure the authenticity of the samples, the coffee species were validated by agricultural extension officers and farmers.

The first step to obtain green beans was the harvest of the red coffee cherries from the coffee plantation surrounding the processing facility. After harvesting, the red cherries were processed (e.g., pulping, washing, drying) into green beans and stored in 60 kg bags. From these bags, the samples under study were collected randomly at 250 g per sample and were then transported using double-sealed plastic bags for analysis.

3.3.2. Determination of Caffeine and Chlorogenic Acid Content by UV-Vis Spectroscopy

The samples ($n = 74$) were prepared in line with the procedures used by Belay et al. and Navarra et al. (Belay et al., 2008; Navarra et al., 2017). First, the beans were freeze-dried (Epsilon 2-40, Christ, Germany) and ground into a powder using a ball mill (Schwingmühle MM 400, Retsch, Germany). Next, the powder

was screened through a 0.355 mm sieve. A total of 10 mg of sieved coffee powder was dissolved in 10 mL of distilled water. The solutions were stirred (550 rpm, 35 °C, 1 h) using a stirrer (Eppendorf ThermoMixer® comfort, Eppendorf, Germany) and then passed through a paper filter (MN 615 1/4, Macherey-Nagel, Germany).

Caffeine extraction from the sample solution was performed by mixing 5 mL of filtrate with 5 mL of dichloromethane and stirring the liquid for one minute using a vortex mixer. Finally, the absorbance of caffeine in dichloromethane was measured by the UV-Vis spectrophotometer (HP 8453, Hewlett Packard, Germany) within the range of 200–500 nm against the corresponding blank reagent (dichloromethane) and—as per to the Beer–Lambert law—at a maximum wavelength of $\lambda = 276$ nm (Appendix 2). The standard solutions were prepared by dissolving caffeine (anhydrous ≥98.5%, Carl Roth, Germany) in dichloromethane (Rotipuran ≥99.5%, Carl Roth, Germany) ranging from 10 to 35 ppm; the calibration yielded an R^2 of 0.9974. The caffeine content was calculated in percent on a dry weight (dw) basis (Appendix 2). Each measurement was replicated twice. The average standard deviation of the replicated caffeine determination was 0.1% dw.

The chlorogenic acid content was determined using a method similar to the one outlined for caffeine measurement, excluding the extraction step using dichloromethane (Belay and Gholap, 2009; Navarra et al., 2017). The samples ($n = 74$) were prepared as follows: First, 1 mg of sieved coffee was dissolved in 10 mL of distilled water. The solutions were stirred for one minute and passed through a paper filter (MN 615 1/4, Macherey-Nagel, Germany). The chlorogenic acid was measured using UV-Vis spectroscopy against the corresponding blank reagent (distillate water), according to the Beer–Lambert law at a maximum

wavelength of λ = 324 nm (Appendix 2). Dissolved chlorogenic acid (≥97% Carl Roth, Germany) in distilled water was used to prepare standard solutions in the range of 20 to 150 ppm (R^2 of the calibration curve = 0.9999). The chlorogenic acid content was calculated in percent of dw based on two replications; the average standard deviation of duplicate measures was 0.9% dw (Appendix 2).

3.3.3. Determination of Species by Near-Infrared Spectroscopy

The samples were analyzed using the NIR spectroscopy method, as previously reported (Adnan et al., 2017). A bench-top Fourier transform (FT-) NIR instrument with a rotating sample cup (Thermo Nicolet Antaris MDS, Thermo Fisher, USA) was used to acquire the diffuse reflectance spectra (log 1/R) of bulk samples of green coffee beans (40 g per sample) (n = 74) in a rotating sample cup (petri dish made of Schott Duran glass). The internal background spectra were collected once every hour. High-resolution diffuse reflectance (log 1/R) spectra at a wavelength range of 1000–2500 nm at 2 nm intervals were recorded as the averages of 64 scans. Each sample was replicated three times. Before conducting further calculations, the spectra were averaged.

3.3.4. Statistical Procedures

Statistical analysis was performed using R software (R Foundation, Austria) for the UV-Vis spectroscopy method. The linear discriminant analysis (LDA) was carried out using Unscrambler® X version 10.2 Network Client (CAMO software AS, Norway). The box plot and Welch's unequal variances t-test were used to explore the variability of caffeine and chlorogenic acid content among species (Krzywinski and Altman, 2014; Puth et al., 2014). The data of both compounds were then analyzed using LDA to discriminate among species. Cross-validation was performed to validate the results and stated as classification accuracy (in

percent) (Dankowska et al., 2017). The correlation between the altitude and the chemical compound was tested using Pearson's product-moment correlation coefficient (r).

Multivariate analysis was carried out using the following procedure (Adnan et al., 2017). The first step in this analysis was to detect spectral outliers using principal component analysis (PCA) and Hotelling's T^2 ellipse 5% plot, based on the raw spectra of all samples ($n = 74$). Following the detection of spectral outliers, several preprocessing methods were applied (e.g., smoothing, the Savitsky–Golay derivative, normalization, baseline correction, orthogonal signal correction (OSC), multiplicative scatter correction (MSC), and extended multiplicative scatter correction (EMSC)). On the preprocessed spectral data, calibration ($n = 49$) and validation ($n = 23$) models were developed using partial least squares discriminant analysis (PLS-DA). Finally, all models were verified with regard to their prediction accuracy—that is, the number of latent variables (LVs), the coefficient of determination (R^2), the root mean square error of calibration (RMSEC), and the root mean square error of prediction (RMSEP) (Appendix 2) (Bassbasi et al., 2014).

LDA was applied to the selected wavelengths of raw spectra. These selected wavelengths were derived from PLS-DA. The accuracy (in percent) was calculated using a full cross-validation procedure.

3.4. Results

3.4.1. UV-Vis Spectroscopy

3.4.1.1. Caffeine Content

As shown in Figure 3.1a, the median and range values for caffeine content differed among species. The caffeine content in Arabica was significantly lower than that in Robusta (Welch's unequal variances $t(67.9) = -17.8$, p-value < 0.001). On average, Arabica had a caffeine content of 1.8% dw, while Robusta contained 2.9% dw. The 95% confidence interval for the differences in the caffeine content among the species was between −1.3% dw and −1.0% dw.
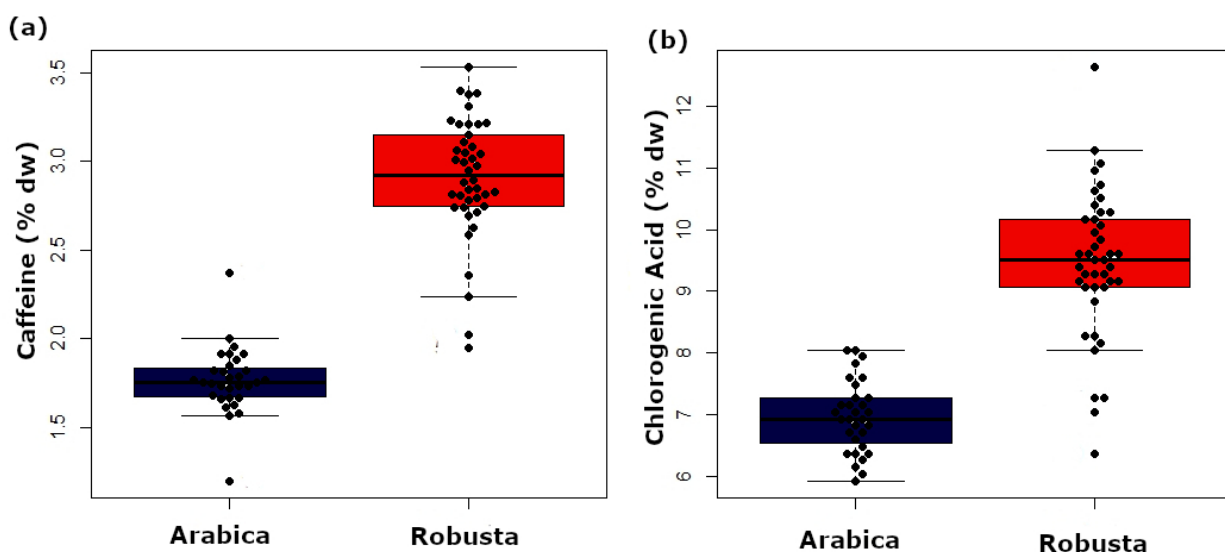
Figure 3.1. Caffeine content (a) and chlorogenic acid content (b) of green Arabica and Robusta beans obtained using UV-Vis spectroscopy ($n = 74$).

3.4.1.2. Chlorogenic Acid Content

The mean value of chlorogenic acid content in Arabica was 7.0% dw and in Robusta was 9.5% dw (Figure 3.1b). Arabica and Robusta differed significantly in their chlorogenic acid content according to Welch's t-test ($t(66.2) = -11.1$, $p <$

0.001; 95% CI of mean difference: −2.0% dw and −1.9% dw), although few samples (5.4%) displayed similar levels of chlorogenic acid for both species.

3.4.1.3. Discrimination among species on the basis of caffeine and chlorogenic acid content (by UV-Vis spectroscopy)

As shown in the previous sub-section, discrimination among species on the basis of only a single chemical component (caffeine or chlorogenic acid) is unreliable. To discriminate among species, both compounds were used for a linear discriminant analysis (LDA). As shown in Figure 3.2, LDA discriminates between Arabica and Robusta with an accuracy of 97.3%. Only two samples of Arabica were mistakenly identified as Robusta samples. This shows that using both caffeine and chlorogenic acid content is a reliable method for discriminating among species.
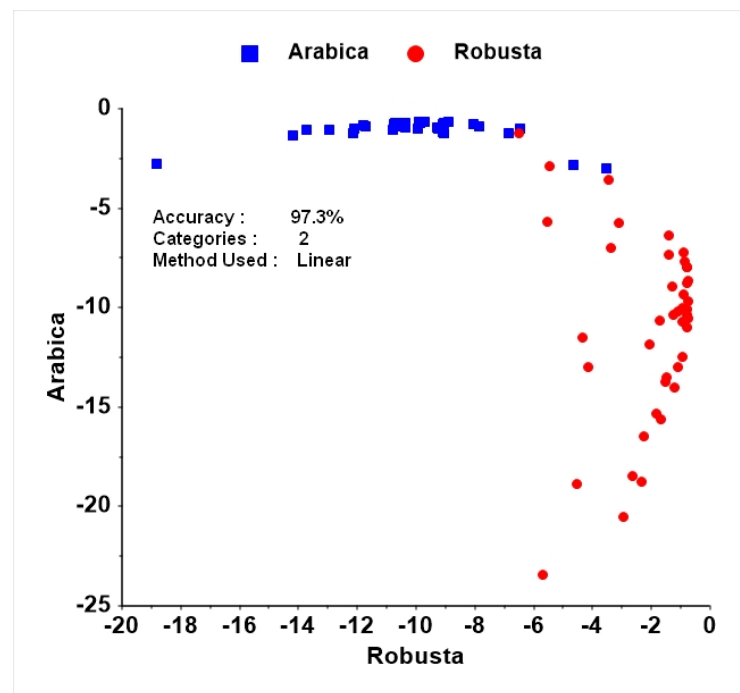
Figure 3.2. Discrimination among coffee species by linear discriminant analysis (LDA) using caffeine and chlorogenic acid derived from UV-Vis as predictor variables.

3.4.2. Discrimination among species using NIR spectroscopy

The following section examines the applicability of NIR spectroscopy. Raw diffuse reflectance (log 1/R) spectra ($n = 74$) were inspected to detect outlier data using principal component analysis (PCA) and projection of the Hotelling T^2 ellipse (Figure 3.3a,b). Data points located outside the ellipse were considered to be spectral outliers and were removed. Two samples were spoiled microbially. Hence, they were deleted due to their potential negative influence on the model, though they may have contained valuable information regarding species variation (Bylesjö et al., 2006).
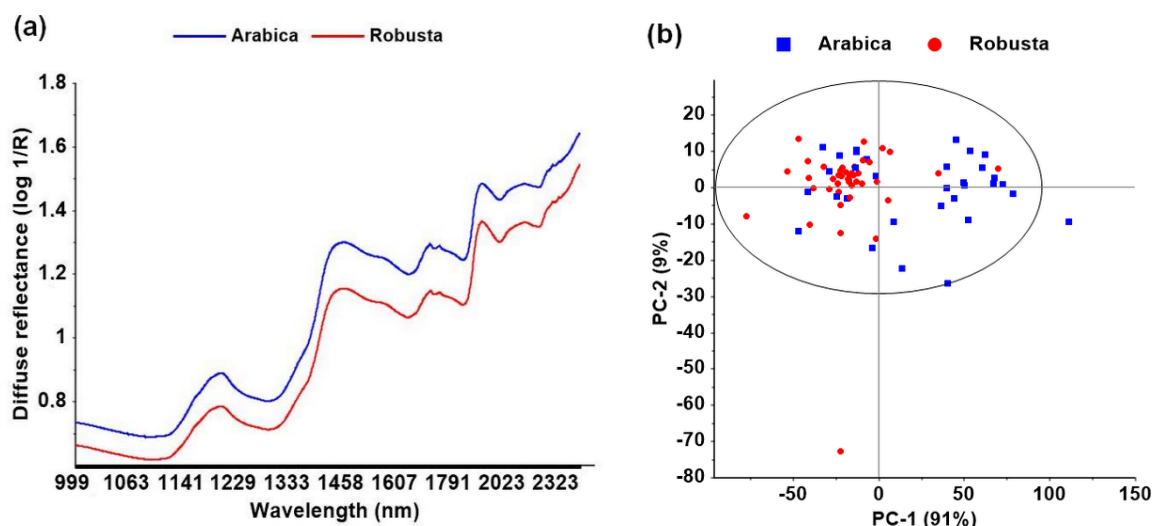
Figure 3.3. Mean diffuse reflectance (log 1/R) of raw spectra (a); score plot of principal component analysis using raw near-infrared spectra (log 1/R) with Hottelling's T^2 ellipse for outlier inspection. Samples outside the Hottelling's T^2 ellipse are considered spectral outliers (b).

Preprocessing methods were applied to raw spectra ($n = 72$) before using PLS-DA to discriminate among species. Table 3.1 shows that several PLS-DA models based on preprocessing methods produced fairly high R^2. The PLS-DA models based on EMSC, normalization (area and mean), smoothing (moving average, three segments), and MSC yielded R^2 of the validation model of

81.3−90.5%; however, using various numbers of latent variables (LVs). Those models yielded higher accuracy compared with the validation model based on raw spectra of 71.5%. The details pertaining to the other PLS-DA models are presented in Appendix 3.

Table 3.1. Statistical parameters of partial least squares discriminant analysis models discriminate green coffee beans species using near-infrared spectra

Preprocessing Method	LVs	R^2 of Calibration Model (%)	RMSEC	R^2 of Validation Model (%)	RMSEP
Raw	7	89.0	0.3266	71.5	0.6005
EMSC	6	91.4	0.2884	90.5	0.3641
Normalization (area)	7	93.2	0.2570	90.3	0.3745
Normalization (mean)	6	93.2	0.2570	90.3	0.3745
Smoothing (Moving average, 3 segments)	7	89.0	0.3266	88.9	0.3270
MSC	3	85.3	0.3774	81.3	0.4734

R^2: the coefficient of determination; LVs: latent variables; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction; MSC: multiplicative scatter correction; EMSC: extended multiplicative scatter correction. The results of the best functioning models are displayed in Appendix 3.

Another significant aspect of the multivariate analysis is the ability to disclose wavelengths that are causative for discrimination and considered to be related to chemical compounds present in the sample. The weighted regression coefficient plot shows that several wavelengths (1212 nm, 1342 nm, 1465 nm, 1674 nm, 1929 nm, 2021 nm, and 2227 nm) contributed to the discrimination among coffee species (Figure 3.4a,b). These wavelengths are related to caffeine, chlorogenic acid, carbohydrates, sugars, trigonelline, lipids, water, proteins, and amino acids (Ribeiro et al., 2011).
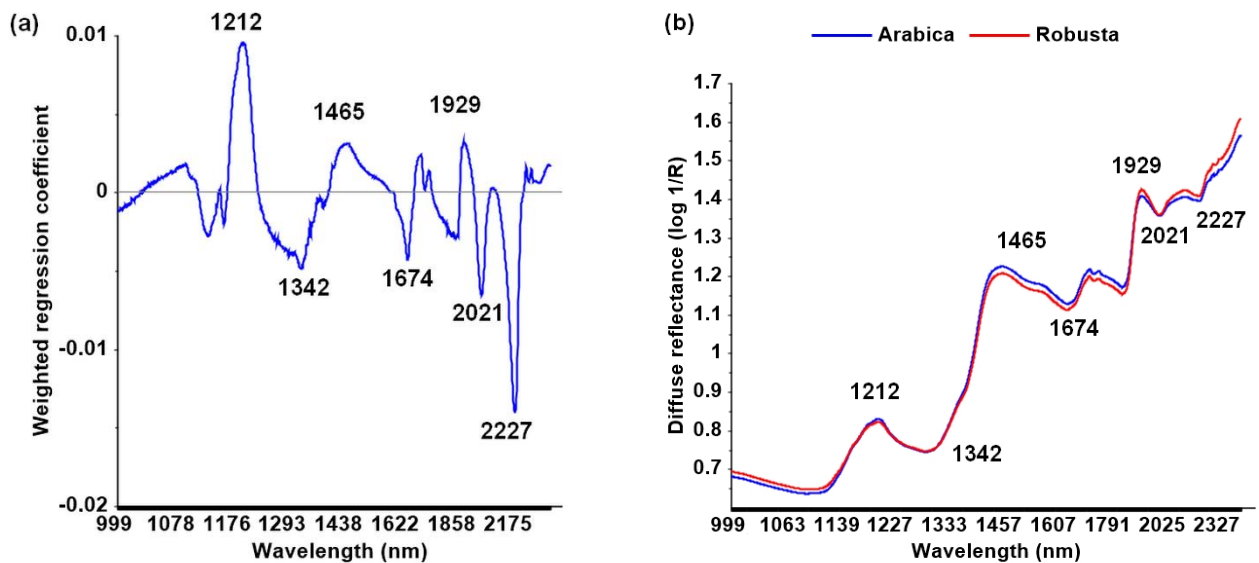
Figure 3.4. Weighted regression coefficient plot of the partial least squares discriminant analysis (PLS-DA) model based on multiplicative scatter correction (MSC) spectra of intact green beans (number of latent variables = 3) (a); mean diffuse reflectance (log 1/R) of MSC spectra (b).

Subsequently, wavelengths identified to contribute to the PLS-DA discrimination among coffee species (Figure 3.4a,b) were used for LDA, which resulted in a classification accuracy of 95% (95% CI (87%, 98%)) (Figure 3.5). The LDA method using selected wavelengths performs with higher accuracy than the full spectra of PLS-DA. This result suggested that the selected wavelengths of 1212 nm, 1342 nm, 1465 nm, 1674 nm, 1929 nm, 2021 nm, and 2227 nm are satisfactory to discriminate between coffee species. When Arabica is considered the positive case to be detected, sensitivity and specificity of the test were 100% and 91%, respectively. Using MSC corrected spectra and the above seven wavelengths, the classification accuracy improved slightly to 97%.
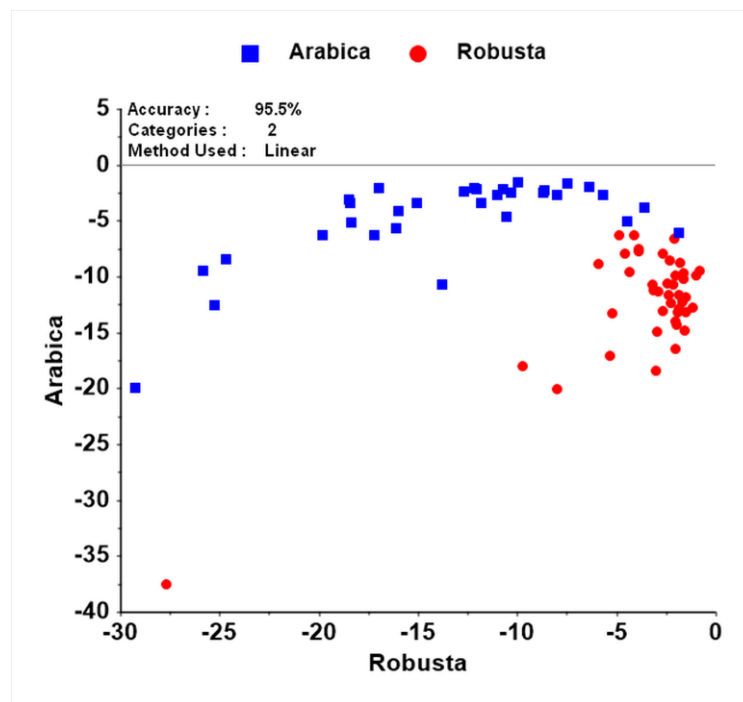
Figure 3.5. Discrimination among coffee species by linear discriminant analysis (LDA) using selected wavelengths derived from NIR spectroscopy as predictor variables ($n = 74$ samples).

3.5. Discussion

3.5.1. UV-Vis Spectroscopy

3.5.1.1. Caffeine Content

A closer inspection of Figure 3.1a shows that one sample of Arabica exhibited caffeine levels similar to those in Robusta. These species have an overlapping range of 4.1% samples. Thus, merely relying on caffeine content to discriminate among species can lead to false classifications.

In accordance with the present results, a previous study has demonstrated that the caffeine content ranged from 0.8% to 1.8% dw (mean = 1.3% dw) in Arabica as compared with a range of 1.2% to 2.5% dw in Robusta (mean = 1.8% dw) (Tran et al., 2017). This study also showed that the caffeine content values overlapped among species, which is consistent with other studies (Babova et al.,

2016; Couto et al., 2019). Taken together, this confirms that caffeine content is unsuitable as a single factor to discriminate between Arabica and Robusta.

A correlation test was conducted to obtain more insight whether the altitude as an environmental factor is responsible for the variability of caffeine content in the present study. The results showed a slightly negative but not significant correlation for Arabica of $r = -0.25$, ($p = 0.17$) and for Robusta of $r = -0.11$ ($p = 0.50$) (Appendix 4). Other studies confirmed that environmental factors such as total irradiance, rainfall, temperatures, and potential evapotranspiration at various altitudes do not affect the caffeine content (Barbosa et al., 2012; Joët et al., 2010).

3.5.1.2. Chlorogenic Acid Content

This result suggested that chlorogenic acid cannot be used as the only trait to discriminate among coffee species (Figure 3.1b). Accordingly, previous studies have demonstrated that the chlorogenic acid content values overlapped among species. In general, the chlorogenic acid content of green beans ranged between 4.0% dw and 8.4% dw in Arabica, and between 7.0% dw and 14.4% dw in Robusta (Farah and Donangelo, 2006).

Finding a correlation between chlorogenic acid and altitude is interesting because altitude may affect chlorogenic acid. A previous study showed lower chlorogenic acid on higher altitude (Barbosa et al., 2012). The results of the present study showed that the chlorogenic acid content did not correlate with the altitude for Arabica ($r = -0.25$, $p = 0.17$) and Robusta ($r < 0.01$, $p = 0.99$) (Appendix 5).

The multivariate statistical tools, that is, PCA and PLS-DA based on chlorogenic acid constituents (caffeoylquinic acid, feruloylquinic acid, and p-coumaroylquinic acid isomers), can be used for determining three coffee

agricultural practices (organic, conventional, and biodynamic), but not for geographical identification of roasted coffee. The 5-caffeoylquinic acid and 4-caffeoylquinic acid were higher than 3-caffeoylquinic acid in all agricultural practices. Chlorogenic acid constituents are assumed to result from the absence of pesticide and pest-defense compounds (Badmos et al., 2020).

3.5.1.3. Discrimination among Species on the basis of Caffeine and Chlorogenic Acid Content (by UV-Vis Spectroscopy)

Another advantage of measuring the caffeine and chlorogenic acid contents is that they are key compounds indicating the coffee flavor. Caffeine and chlorogenic acid are related to bitterness (Cheng et al., 2016). Thus, a higher level of these compounds may produce a more bitter coffee taste (Ribeiro et al., 2011) and, subsequently, contribute to lower cup quality as determined by trained panelists (Farah et al., 2006). This may explain why Arabica obtained a higher consumer liking than Robusta (Kim et al., 2016).

The present study indicates that UV-Vis spectroscopy can be used as an alternative approach to discriminate among species of green coffee beans on the basis of caffeine and chlorogenic acid content. The discrimination accuracy in our study is comparable to a previous study using caffeine, chlorogenic acid, trigonelline, total polyphenols, total free amino acids, and aqueous extract as determined by liquid chromatography (HPLC) for species discrimination (Martín et al., 1998). A K-nearest neighbors' classification resulted in an accuracy of 92.7–97.6%. The error rate occurred because some Arabica samples had an unusually high and some Robusta samples had an uncharacteristically low caffeine content (Martín et al., 1998). As the present study uses less compounds, it can be considered comparable.

Recent research has shown that PLS-DA differentiated 100% of ground Arabica and blends. Caffeine, total soluble solids, quercetin-3-rutinoside, the Folin–Ciocalteu reducing capacity, and antioxidant capacity (DPPH assay) were the most discriminating variables. The classification of ground Arabica and blends was also obtained with 100% accuracy using LDA. Total flavonoids, Fe2+ chelating ability, quercetin-3-rutinoside, and total phenolic content were the analytical responses that discriminated between groups (Monteiro et al., 2019).

3.5.2. Discrimination among Species Using NIR Spectroscopy

The results show that the PLS-DA model based on MSC preprocessed spectra needed the lowest number of LVs. While the number of LVs was the lowest, the calibration and validation models based on MSC obtained neither the highest R^2 nor the lowest RMSEC and RMSEP compared with the other preprocessing methods. Simultaneously, the EMSC and normalization of the area and mean method yielded the highest R^2 (>90%) and the lowest RMSEC and RMSEP (<0.40), but the number of LVs is high (Table 3.1).

Selecting the ideal model is challenging because none of the models fit the criteria for the best model—that is, highest R^2, lowest number of LVs, lowest RMSEC, and lowest RMSEP (Westad and Marini, 2015). However, a low number of LVs is considered to produce a more robust model (Munawar et al., 2016). If a model is selected on the basis of different LV numbers, the R^2 value should be examined closely to avoid an overoptimistic model (Brereton, 2006).

Thus, the PLS-DA model based on MSC processed spectra is selected here as the ideal model using full-range spectra for discriminating among green coffee beans species (Table 3.1). MSC (LVs = 3) reduces the number of LVs in the respective discrimination model compared with the model based on raw spectra

data (LVs = 7). This indicates the presence of scattering and simultaneously shows that MSC can reduce the noise (Esteban-Díez et al., 2004). Scattering in NIR spectroscopy can be influenced by the sample traits and the measuring conditions (Barbin et al., 2014). Species variety can lead to different sizes and shapes (Severa et al., 2012; Tran et al., 2017). As shown in Appendix 6, the present study used green coffee beans samples of different species, and variety may lead to scattering problems. The beans were not screened to be of the same size—the aim was to create the actual sample conditions that prevail during trading.

However, using only selected wavenumbers and an LDA approach yielded a superior classification performance. This model was built using authentic sample conditions including samples with different sizes and shapes. The percent of correctly classified samples by the validation model was 95.5%. The classification performance based on selected wavelengths is more accurate than those of a previous study. Downey et al. (Downey et al., 1994) reported that the accuracy of NIR spectroscopy in discriminating among green coffee beans species ranges from 86.5% to 88.6%. The validation models were built using a factorial discriminant analysis with eight LVs, which is potentially an overfit model.

Recent study has demonstrated NIRS data using the typical band of the spectra, and the PLS-DA classifier predicts farming system (organic and conventional) of roasted coffee and provides results with 89% accuracy. The NIRS classification model, which is much simpler to develop and deploy, can provide good prediction with less instrumentation complexity and at a lower cost than proton transfer reaction mass spectrometry (PTR-MS) does. However, geographic identification was somewhat complex. The PTR-MS models using

PLS-DA performed with slightly better accuracy than NIRS models (69% vs. 61%, respectively) (Monteiro et al., 2018).

According to the literature, wavelengths of 1209 nm, 1466 nm, 1726 nm, 1758 nm, 1904 nm, 2308 nm, and 2348 nm—relating to pure water and lipids—can be used to differentiate among green coffee beans species (Downey et al., 1994). Another study suggested different selected wavelengths to discriminate among species of green beans. The wavelengths of 1671 nm, 1673 nm, and 2154 nm are associated with caffeine, and the wavelengths of 1778 nm, 1834 nm, and 2251 nm are associated with cellulose (Buratti et al., 2015). Two previously identified wavelengths (i.e., 1671 nm and 1673 nm) [16] are close to the wavelength selection of the experiment resulting model (1674 nm), which are associated with caffeine.

Buratti et al. (Buratti et al., 2015) demonstrated that NIR spectroscopy offers 100% accuracy in discriminating between Arabica and Robusta. However, this study did not clarify the chemical composition of the beans. Hence, as shown in our study, accuracy may be compromised due to partly overlapping levels of caffeine and chlorogenic acid in both species.

Despite the potential of NIR spectroscopy, it is still a challenge to develop an applicable and sensitive method for discriminating among green coffee beans. As in the present study, the samples generally display variations among species and variety. Here, even samples from one particular location consist of several varieties because the farmers planted multiple batches at the same times. In addition to variety, models for species recognition could benefit from taking various environmental factors into account (Appendix 6).

3.6. Conclusions

This study evaluated the applicability of UV-Vis spectroscopy and NIR spectroscopy to discriminate between the green coffee beans of Arabica and Robusta from Java Island, Indonesia. The results showed that both approaches are acceptable in terms of their classification accuracy. UV-Vis spectroscopy-based determination of two important compounds—that is, caffeine and chlorogenic acid—attained a slightly higher classification accuracy of 97.3%. NIR spectroscopy using seven selected wavelengths and LDA yielded a similarly high classification accuracy (95.5%). The findings suggest that, given both the speed, nondestructiveness and low involvement of labor of NIR spectroscopy, it is superior for on-site species discrimination. This study was limited by the environmental conditions and varieties of the beans samples. Thus, further research should include samples of different species and varieties from various coffee-producing locations worldwide in order to evaluate the robustness of NIR-based species discrimination.

4. Identifying the origin of Java green coffee beans using near infrared spectroscopy and stable isotope analysis of oxygen, hydrogen, and strontium

4.1. Abstract

Java coffee is one of the famous single origins and has different coffee production areas on the west, central, and east part of the Island. No known empirical study has focused on identification of the green coffee beans from those locations using verifiable methodology. This present study aims to examine the potential of near infrared (NIR) spectroscopy and stable isotope analysis of oxygen ($\delta^{18}O$), hydrogen ($\delta^{2}H$), and strontium (δSr) to identify the origin of the green coffee beans from Java Island. The results indicate that NIR spectroscopy, as well as $\delta^{18}O$ and $\delta^{2}H$, are not suitable for identifying the sample's origin. The δSr in green beans, reflecting δSr of host rocks can potentially be used as a tracer to identify between two distinct locations of Java Island. This study implies that strontium analysis is prominent identifying the origin of Java green coffee beans.

Keywords: green coffee beans, origin, NIR spectroscopy, stable isotope, strontium.

4.2. Introduction

The issue of the green coffee (*Coffea* spp.) beans origin has received considerable an increasing attention recently, including in Indonesia (Durand and Fournier, 2017). The origin of the beans relates to several aspects such as specialty coffee and geographical indication. The first aspect is more to quality nuances and the second one is according to the law (Rahmah, 2017; Vellema et al., 2015). Eventually, those aspects determine green coffee price (Wilson and Wilson, 2014).

It has previously been observed that origin affects the beans quality. Each origin has their unique characteristics, e.g. annual precipitation, diurnal temperature range, number of dry months, altitude, latitude, soil type, and cultivated variety (Barbosa et al., 2012; Figueiredo et al., 2013; Oberthür et al.,

2011). The unique coffee quality of a certain origin may have caused it is favored than the others (Sepúlveda et al., 2016).

To protect and authenticating the origin, the certification has been applied on coffee trading (Barjolle et al., 2017; Offermans et al., 2015). However, If a dispute regarding origin validation occurs, the verifiable methodologies to identify green coffee beans are imperative (Burns et al., 2017). Previous studies have demonstrated several methodologies to identify the beans origin, e.g. near infrared (NIR) spectroscopy, stable isotope analysis, multi-element analysis, chromatographic techniques, and nuclear magnetic resonance (Arana et al., 2015; Liu et al., 2014; Marquetti et al., 2016; Mehari et al., 2016).

We are interested to identify the green coffee beans from Java Island, Indonesia by using NIR spectroscopy and stable isotope analysis. The advantages of NIR spectroscopy are for being rapid measurement, while stable isotope analysis can be related to the environmental coffee growing condition (Liu et al., 2014; Marquetti et al., 2016).

Java coffee is one of the famous single origins and it contributes to around 8% of the total national coffee production. Java origin has been referred to East Java production location (Wahyudi and Jati, 2012). However, there are other coffee production locations on West and Central Java and each location has been protected by regulation on geographical indication registration (Rahmah, 2017). This circumstance raises the intriguing issue regarding how to identify the green beans from each location using verifiable methodology.

No known empirical study has focused on identification of the green coffee beans originated from the west, central, and east part of Java Island. The aim of this study is to examine the potential of NIR spectroscopy and stable isotope

analysis of oxygen, hydrogen, and strontium to identify the origin of the green coffee beans from Java Island.

4.3. Materials and methods

4.3.1. Sites and sampling

The green beans samples ($n = 74$) of *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) were obtained from 38 coffee processing facilities on Java Island, Indonesia during the harvesting season from July to August 2014 (Appendix 6). The red cheery from surrounding coffee plantation were processed (e.g. wet/semi/dry processing, pulping, washing, and drying) into the green beans on these facilities. After drying, the beans were stored in 60 kg bags and the samples were collected randomly from these bags. Each sample was weighed 250 g and finally was placed in double sealed plastic bags for transportation. The processing facilities of sampling location were in the west, the central and the east area of Java Island. To date, the coffee beans from these areas have been regulated under geographical indication protection, indicate that each area has unique characteristic (Rahmah, 2017).

4.3.2. NIR spectroscopy

NIR spectroscopy procedures were done according to Adnan et al. (2017). Diffuse reflectance spectra (log 1/R) of bulk samples of green coffee beans (40 g) were acquired using a bench top Fourier transform near infrared (FT-NIR) instrument with rotating sample cup (Nicolet Antaris®, type Antaris MDS). High-resolution diffuse reflectance (log 1/R) spectra at a wavelength range of 1000 to 2500 nm with two nm intervals were recorded as an average calculated from the 64 scans. Due to the damage of one sample, only 73 samples could be acquired

for the spectra. Three replicates were acquired per sample and the spectra were averaged before conducting further calculations.

The spectra of the samples were analyzed by statistical software (The Unscrambler® X version 10.2 Network Client, CAMO software AS, Norway) to identify the origin of the beans. To begin this analysis, spectral outliers were detected using principal component analysis (PCA) and Hotelling's T^2 ellipse 5% plot, based on all samples' ($n = 73$) raw spectra. Three data were determined as outliers based on this technique. Afterwards, several pre-processing methods were applied to raw spectra ($n = 70$), i.e. smoothing, normalization, baseline correction, orthogonal signal correction (OSC), multiplicative scatter correction (MSC), and extended multiplicative scatter correction (EMSC). Finally, partial least square discriminant analysis (PLS-DA) was applied to the processed spectral data. Regression models were developed with a subset of calibration samples ($n = 52$) and thereafter the models were tested using the subset of validation samples ($n = 18$) to evaluate their accuracy. The number of latent variables (LVs), the coefficient of determination (R^2) of calibration, root mean square error of calibration (RMSEC), R^2 of prediction, and root mean square error of prediction (RMSEP) were functioned as parameters for the evaluation of accuracy.

4.3.3. Stable isotope

4.3.3.1. Determination of stable isotope ^{18}O and ^2H

Stable isotope ^{18}O and ^2H analysis was conducted at the Centre for Stable Isotope Research and Analysis, Georg August University, Göttingen. The first step in this analysis was to freeze dried the green coffee beans. The beans were then ground using a ball mill (Schwingmühle MM 400, Retsch, Germany), and

they were screened through a 0.355 mm sieve to obtain a uniform particle size. Subsequently, about 0.2 mg of powdered material was weighted in a silver tin cups for each sample. After this preparation, the samples were analyzed by using an isotope ratio mass spectrometer (Delta V Plus, Thermo Electron Cooperation, Germany) coupled to a Total Combustion Elemental Analyzer (TC/EA, Thermo Fisher Scientific, Germany) via a Conflo IV interface. Cellulose was used as reference material after every 10 samples to ensure the reproducibility of the measurement (SD_{mean} = 0.32‰ for delta (δ)^{18}O and 1.62‰ for δ^2H). Finally, isotope ratios ^{18}O and ^{2}H were calculated as delta-values (δ ‰) (Equations. 4.1 and 4.2). The Vienna-Standard Mean Ocean Water (V-SMOW) was applied as the international standard for oxygen and hydrogen isotopes.

$$\delta \ ^{18}O \ (‰) \ = \ \frac{R_{Sample} - R_{Standard}}{R_{Standard}} \times 1000 \qquad (4.1)$$

with $R_{Sample} = \frac{^{18}O}{^{16}O}$; $R_{standard}$ = the Vienna $-$ Standard Mean Ocean Water (V $-$ SMOW)

$$\delta \ ^2H \ (‰) \ = \ \frac{R_{Sample} - R_{Standard}}{R_{Standard}} \times 1000 \qquad (4.2)$$

with $R_{Sample} = \frac{^2H}{^1H}$; $R_{standard}$ = the Vienna $-$ Standard Mean Ocean Water (V $-$ SMOW)

In addition, the annual values of δ^{18}O and δ^2H in precipitation water from the location of the examined green beans samples were acquired from the online isotopes in precipitation calculator (Bowen, 2016). These data were needed to

find the relation between $\delta^{18}O$ and δ^2H in the green beans and environmental condition.

The calculation of means and standard deviations (SD) values were performed with Microsoft Excel 2007 (Microsoft, USA). The correlation between $\delta^{18}O$ and δ^2H in the green beans and precipitation water was tested using the Pearson's product-moment correlation coefficient (*r*) (R Software, Austria). Linear discriminant analysis (LDA) was performed to identify origin using the Unscrambler the Unscrambler® X version 10.2 Network Client (CAMO software AS, Norway).

4.3.3.2. Determination of stable isotope strontium ($^{87}Sr/^{86}Sr$)

Powdered samples (see 4.3.3.1) of the green coffee beans (*n* = 16) were analyzed to understand their isotopic composition of $^{87}Sr/^{86}Sr$. The strontium was separated using ion exchange chromatography with Sr spec resin (Eichrom Technologies). Nitric acid (3M) was used as eluent. All analyses were carried out with the Thermo-Finnigan Triton© thermal ionization mass spectrometry (TIMS) at the Isotope Geology Department of the University of Göttingen. The measurements were carried out using the re-double filament technique. Prior to digestion, all samples were mixed with a tracer solution enriched in ^{84}Sr. Concentrations were calculated using the isotope dilution technique. Reproducibility for NBS SRM 987 (*n* = 4) were at 0.71025±0.00003 and 0.05649±0.00001 for $^{87}Sr/^{86}Sr$ and $^{84}Sr/^{86}Sr$. The analytical mass bias was corrected with $^{88}Sr/^{86}Sr$ at 0.1194 using exponential law. The Sr isotope ratios were expressed as δ-values according to bulk earth ($^{87}Sr/^{86}Sr$) at 0.7047. The calculation of δ-values is presented in Equation 4.3.

$$\delta \, Sr \, (\‰) \; = \; \frac{R_{Sample} - R_{Standard}}{R_{Standard}} \times 1000 \qquad\qquad (4.3)$$

With $R_{Sample} \; = \; \frac{^{87}Sr}{^{86}Sr}$, $R_{standard} \; = \; \delta$ bulk earth $\frac{^{87}Sr}{^{86}Sr}$ at 0.7047

4.4. Results and discussion

4.4.1. Identifying the origins using NIR spectroscopy

The first step to analyze the diffuse reflectance (log 1/R) of intact beans spectra were outlier data detection. The outlier detection purpose was to locate the presence of outlier data as shown in the score plot using the principal component analysis (PCA) and projection of the Hotteling T^2 ellipse. The outlier detection was applied to raw spectra and the result was shown in Figure 4.1. Data points located outside the ellipse were considered outliers and were removed.

Following data outliers detection, pre-processing methods were applied to the raw spectra. The calibration model for the prediction of the origins of green coffee beans was developed using partial least square discriminant analysis (PLS-DA). From those pre-processing methods, the second derivative which applied the Savitsky-Golay algorithm (second polynomial order) achieved the lowest values in latent variables and R^2 calibration, thus this model was used to predict the origin of the green beans (Table 4.1). The other pre-processing methods did not achieve any better results (Appendix 7).
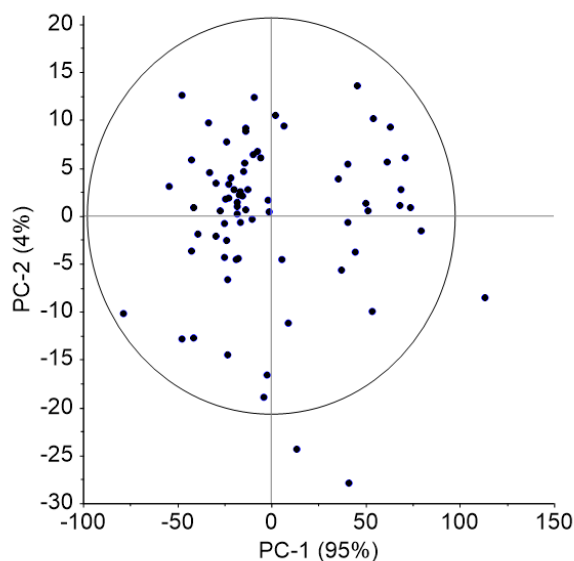
Figure 4.1. Score plot and the Hotteling T^2 ellipse showing raw infrared spectra (log 1/R) for outlier inspection. Data outside Hotelling's T^2 ellipse were considered outlier data.

The calibration model predicted origins, based on the second derivative using the Savitsky-Golay algorithm (second polynomial order), yielded a low LVs, low RMSEC and a high output in R^2. In contrast to these results, the prediction model yielded high RMSEP values and low R^2 (Table 4.1). Overall, the prediction model yielded low accuracy.

Table 4.1. Statistical parameters of PLS-DA models discriminate the origin of the green coffee beans using NIR spectroscopy

Pre-processing method	Origin	LVs	R^2 of the calibration model	RMSEC	R^2 of the prediction model	RMSEP
Raw spectra	West Java	7	0.6276	0.2903	0.1614	1.0633
	Central Java		0.3875	0.3723	0.0243	0.7488
	East Java		0.3930	0.3596	0.0182	0.6256
Savitsky-Golay (2nd derivative, 2nd Polynomial order)	West Java	3	0.8400	0.1903	0.0749	1.3266
	Central Java		0.8589	0.1787	0.7860	1.1282
	East Java		0.7458	0.2327	0.1276	1.1523
Baseline (Linear Baseline Correction)	West Java	4	0.6176	0.2941	0.238	0.464
	Central Java		0.4226	0.3615	0.0522	0.4845
	East Java		0.3204	0.3805	0.2417	0.4396

LVs: the number of latent variables, R^2: the coefficient of determination, RMSEC: root mean square error of calibration, RMSEP: root mean square error of prediction.

Further analysis showed that the principal components (PC)1 and PC2 of the PLS-DA model, based on the second derivative using the Savitsky-Golay algorithm (second polynomial order), explained only 12% of the spectral data variance as the independent variable. However, PC1 and PC2 of this model explained 70% of origin variance as the dependent variable (Figure 4.2). This result shows that the calibration model was neither sufficiently robust nor reliable to predict the origin of the examined samples.
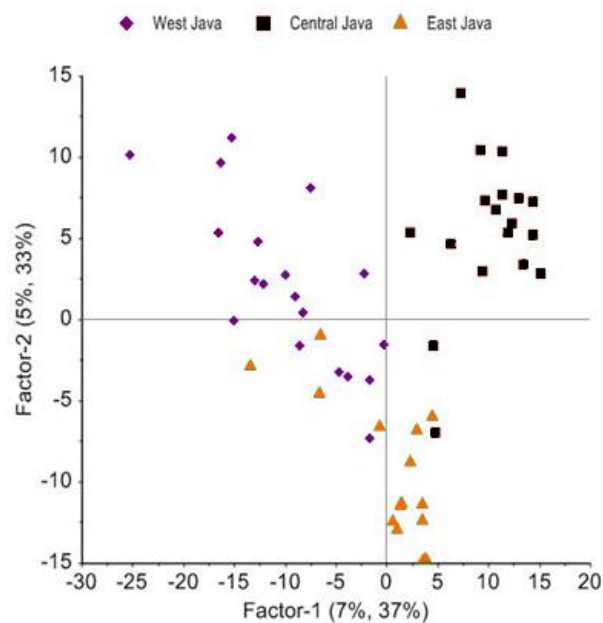
Figure 4.2. Score plot of the partial least square discriminant analysis calibration model for discriminating origins based on the 2nd derivative Savitsky-Golay method with two polynomial diffuse reflectances (log 1/R) spectra.

The robust model will be obtained if the covariance between the independent variables and dependent variable are maximized (Gromski et al., 2015). In other words, a few first PC of the model should explain a major proportion of the dependent and the independent variance. In addition, there should be only slightly different between the statistical parameter of the calibration with the prediction model (Pizarro et al., 2004). This was not the case for the present study.

The prediction model to identify the origin of this present study yielded a low accuracy of 32.95%. This low prediction accuracy is somewhat surprising, given the fact that other previous research shows near infrared discriminate the origin of the green coffee beans that cultivated in Brazil with high accuracy (Marquetti et al., 2016). This previous study presented that PLS-DA with pre-process method of MSC plus second-derivative and six LVs yielded an accuracy of 87.8%. One of the possible explanations for this might be that the differences of sample preparation. The present study used the whole beans while the previous study by Marquetti et al. (2016) used powdered samples. To date, there is little published information on identifying the origin of the green coffee beans using NIR spectroscopy on other specific regions (Burns et al., 2017). Future studies are needed to evaluate the sample preparation on different specific origin to investigate the ability and the limitation of NIR spectroscopy.

4.4.2 Stable isotope

4.4.2.1. Identifying origins using the stable isotopes ^{18}O and ^{2}H

Figure 4.3a displays an overview of δ^{18}O value of examined samples in West, Central, and East Java Island. The δ^{18}O values in the beans are 22.12±0.67‰ in West Java, 22.44±0.67‰ in Central Java and 23.35±0.88‰ in East Java. The range of δ^{18}O value between origins is narrow; indicate that this value is not reliable to serve as tracer to identify the origins of the examined samples. In comparison, the δ^{18}O value of the green beans from Sumatera Island, an Island near to Java Island, is 24.9‰, and are in the range of globally δ^{18}O values in the green beans from 18.7‰ to 34.8‰ (Rodrigues et al., 2011b).

These δ^{18}O values in the beans are likely to be related to δ^{18}O values in the precipitation water (Figure 4.3a.). The correlation between δ^{18}O values in

Java beans and precipitation water is positive and significantly different ($r= 0.44$, $p <0.001$). This finding supports the work of other studies that there is a positive correlation between $\delta^{18}O$ values in Hawaiian coffee and local precipitation ($r = 0.56$; $p < 0.05$) (Rodrigues et al., 2011a).This previous study has demonstrated that $\delta^{18}O$ values in green coffee beans from Hawaii Island is different than other four Island, which can be related to $\delta^{18}O$ values in local precipitation. Therefore, it seems that $\delta^{18}O$ values can be used to identify coffee origins if $\delta^{18}O$ values in precipitation water between those origins are significantly different.
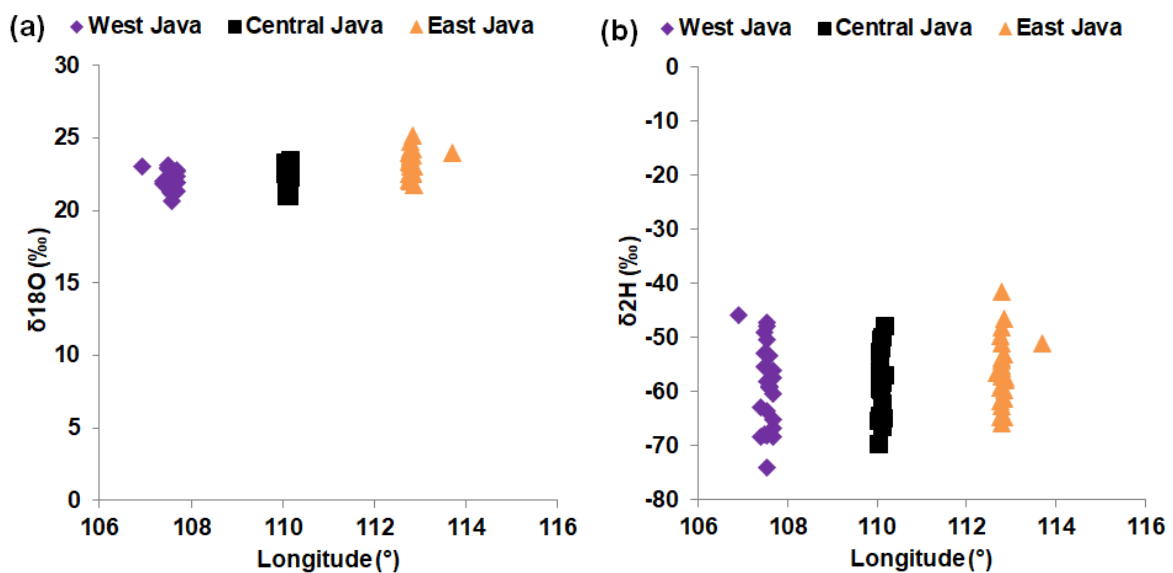
Figure 4.3. (a) $\delta^{18}O$ and (b) δ^2H value of green coffee beans of Java Island.

The differences of $\delta^{18}O$ value between the green beans and the precipitation water because of fractionation (Figure 4.3a and 4.4b). The soil water, contain the $\delta^{18}O$ from the precipitation water, is taken up through plant roots which lead to the fractionation of $\delta^{18}O$ in the leaf during photosynthesis and respiration processes. However, the $\delta^{18}O$ also can be also sourced from groundwater sources, atmospheric vapor and atmospheric CO_2 (Barbour, 2007; Yakir and Sternberg, 2000).
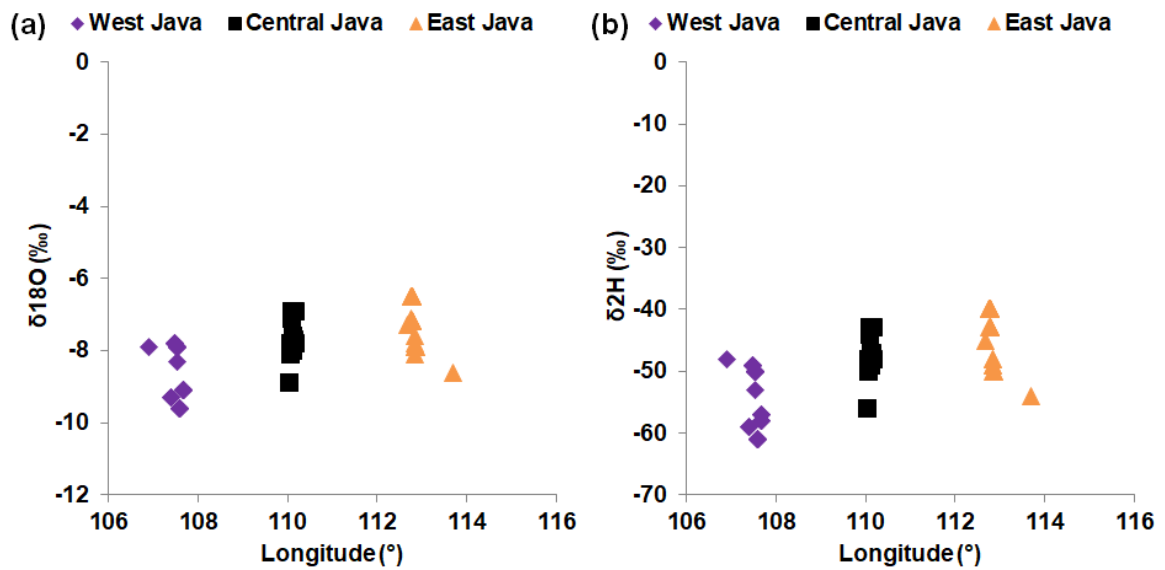
Figure 4.4. (a) $\delta^{18}O$ and (b) δ^2H value of precipitation water of Java Island. The $\delta^{18}O$ and δ^2H value of the precipitation water were obtained from (Bowen, 2016).

As shown in Figure 4.3b, a similar range of δ^2H value between regions indicates this value is not suitable to identify origins. The beans contain δ^2H value of -58.75±7.58‰, -58.46±5.84‰, and -56.07±6.08‰ in West, Central and East Java, respectively. In accordance with the present results, previous study has demonstrated that δ^2H value (range from -83‰ to -20‰) was not considered for predicting origins (Santato et al., 2012).

The present study shows a positive and significantly different correlation between δ^2H values in the beans and precipitation water (r = 0.29, p <0.05) (Figure 4.3b and 4.4b). However, little information is known about the correlation between δ^2H values in the green beans and precipitation water in the previous studies. The δ^2H is highly depleted during photosynthesis, which makes correlations between delta δ^2H in plant material and precipitation water more complex (Yakir, 1992).

Despite $\delta^{18}O$ and δ^2H seems not reliable as single tracer to identify origins, further examination of both tracers to identify origins was performed using LDA.

The LDA yielded an accuracy of 54.29% prediction of the examined samples. This result shows that the combination of $\delta^{18}O$ and δ^2H value is also not reliable to predict origin for Java green coffee beans. As can be seen from Figure 4.5, the beans from West, Central and East Java are not clearly divided into different groups on scatter plot.

The $\delta^{18}O$ and δ^2H analysis has a good potential to identify the origins of green coffee beans if $\delta^{18}O$ and δ^2H values in precipitation water vary among the origins and if there is a distinct correlation between the isotope ratios in precipitation water and its corresponding plant material; however, the uncertainty remains a challenge (Bowen et al., 2005). For example, the previous study analyzed $\delta^{18}O$ and δ^2H values in caffeine of green coffee beans to predict the origin (Weckerle et al., 2002). The LDA of $\delta^{18}O$ and δ^2H values in caffeine can differentiate whether coffee originates from Central/ South America or Africa at error rates of 5.7% when using calibration and of 7.7% when cross validation is applied.
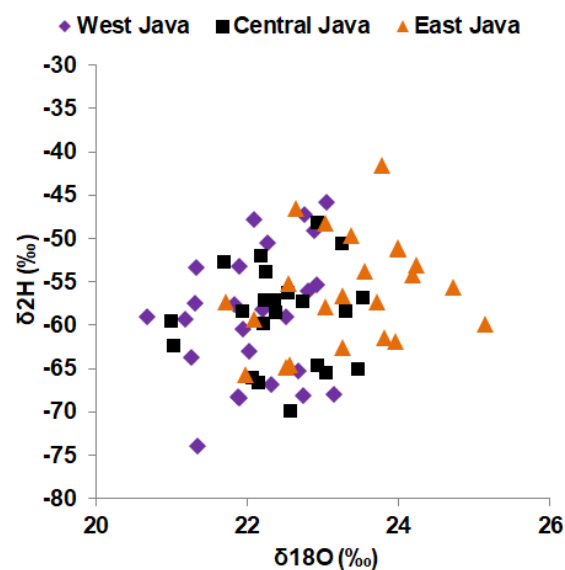
Figure 4.5. Scatter plot of $\delta^{18}O$ and δ^2H value in green coffee beans from West, Central, and East Java.

In addition, it seems that the potential of $\delta^{18}O$ and δ^2H value yield a high accuracy to predict the origin of the beans if those tracers are combined with other stable isotope and elemental analysis. For example, green beans from four continental areas (Central America, South America, Africa, and Asia) were identified using $\delta^{13}C$, $\delta^{15}N$, $\delta^{18}O$, Li, Mg, P, K, Mn, Co, Cu, Se, Y, Mo, Cd, La and Ce yielded an accuracy of 98% (Santato et al., 2012).

4.4.2.2. Identifying the origins of green coffee beans using the stable isotope strontium

The δSr value in green beans and the host rocks surrounding the sampling locations are presented in Figure 4.6. Generally, the δSr value of green beans is close to that of their host rocks. According to the literature, δSr value in green beans is either slightly lower or slightly higher than that of its corresponding host rock. Plant metabolism does not fractionate strontium isotopes significantly, thus δSr value in green beans is reflected by its host rocks and can potentially be used as an origin tracer (Rodrigues et al., 2011b).
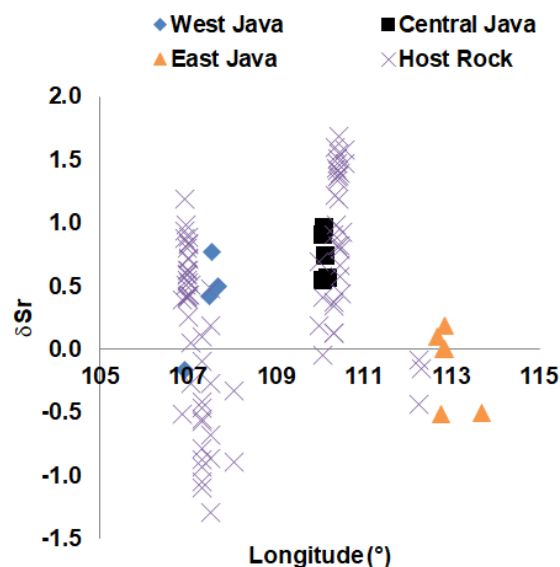
Figure 4.6. The δSr value in green coffee beans and their host rocks in Java Island. The δSr value of the host rocks was obtained from (EarthChem, 2011).

The examined green beans samples from East Java showed lower δSr value than those from West and Central Java (Figure 4.6). Thus, it seems more practical to divide coffee origins from Java Island into two rather than into three origins. Based on this finding, the green beans in western and eastern part of Java Island contain δSr value of 0.5887±0.3376 and -0.1204±0.3087, respectively. Additionally, Java is part of a volcanic island arc situated in the Indonesian archipelago at the southern margin of the Eurasian Plate along which the Muria-Progo lineament (Figure 4.7), a major fault system, divides it into two distinct structural regions (Clements et al., 2009). Both regions exhibit different characters along the deeper crusts resulting in varying contents of δSr in the respective rocks (Whitford, 1975).
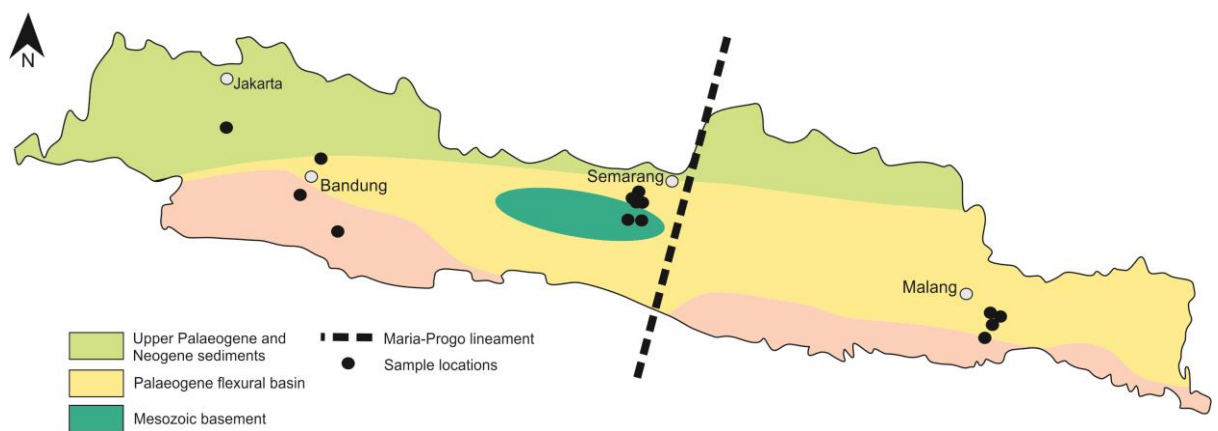
Figure 4.7. Simplified geological map of Java Island according to Clements et al. (2009). Sample locations represent δSr analysis of green coffee beans.

The ^{87}Sr/^{86}Sr isotope ratio is one of the most popular among the isotopic systems because it provides fingerprint data that can be traced back to its geographical origin. Strontium has four naturally occurring isotopes. The isotopes ^{84}Sr, ^{86}Sr, and ^{88}Sr are non-radiogenic, whereas ^{87}Sr develops, in part, out of the natural β-decay of ^{87}Rb. Its concentration in the minerals reflects the age of the

rock as well as the Rb/Sr ratio. This makes the $^{87}Sr/^{86}Sr$ isotope ratio an important parameter in geochronology. The $^{87}Sr/^{86}Sr$ isotope ratio was demonstrated not to change during biological processes and mirrors the growth environment of bedrock, soil, and soil–water in plants (Capo et al., 1998). It reflects the ranges of delta Sr for country rocks and their respective soils (Horn et al., 1998, 1993).

The Identification of origins using the analysis of stable isotope composition is dependent on environmental variables such as precipitation, altitude, distance from the coast, and host rock. Therefore, information on environmental factors that influence the isotopic signatures of plant materials is essential (Rodrigues et al., 2011b). Identifying the origins of green coffee beans origins from a limited geographical area is possible, however, if a combination of different tracers such as an isotope abundance of strontium ($^{87}Sr/^{86}Sr$), carbon ($\delta^{13}C$), nitrogen ($\delta^{15}N$), sulfur ($\delta^{34}S$), oxygen ($\delta^{18}O$), and the concentrations of 30 multi elements are utilized (Rodrigues et al., 2011a).

4.5. Conclusion

The purpose of the present study was to examine the potential of NIR spectroscopy and stable isotope analysis of $\delta^{18}O$, δ^2H, and δSr to identify the origin of the green coffee beans from Java Island. The results of this study show that NIR spectroscopy did not produce a precise prediction for the examined sample. The $\delta^{18}O$ and δ^2H also did not prove to be good tracers for origin distinction, as a single or in a combination. Analyzing the content of δSr in green beans turned out to be a promising tracer for identifying between origins of green coffee beans from Java Island and supports the evidence that Java Island is divided into two distinct structural regions. Although the present study is based on a limited examined sample, the findings suggest that identifying the origin of

the green beans should consider environmental factor. More research using a larger number of coffee production areas under various environmental conditions is needed to represent the origin of the green coffee beans.

5. General discussion and conclusion

5.1. General discussion

The present study focuses on evaluating near infrared (NIR) spectroscopy as a method for predicting quality attributes and authenticating the green coffee beans. The application of NIR spectroscopy in the coffee industry is still challenging (Barbin et al., 2014). The challenge occurs because of the variation in genetic and growing condition of the green beans samples (Toci et al., 2016). As consequences, the evaluation of NIR spectroscopy accuracy on the different origin of the beans is important in order to establish this method as a reference procedure in the coffee industry.

Several studies have been published regarding predicting quality attributes and authenticating the green coffee beans using NIR spectroscopy from several origins such as Brazil, Colombia, Costa Rica, Ethiopia, and Yemen (Bertrand et al., 2005; Morgano et al., 2008). Until now, not much is known about the reliability of NIR spectroscopy to assess the green beans quality from Indonesia. Therefore, green coffee beans from different origins in Indonesia were used as examined samples.

Prior to the experimental studies, the literature review in chapter 1 was performed to summarize the knowledge about the quality parameters of the green coffee beans, the standard methods for quality determination, and to review the literature about the reliability of NIR spectroscopy to measure the beans quality in comparison to the standard method. The previous studies have reported NIR spectroscopy has the potential to predict moisture content (MC), discriminate the beans species and identify the origin of the beans. Therefore, the experimental studies were conducted on those quality parameters.

In chapter 2, NIR spectroscopy was evaluated to predict MC in intact green beans of the *Coffea arabica* (Arabica) and *Coffea canephora* (Robusta) species. The results showed that a three-component partial least squares regression (PLSR) model using raw NIR spectra yielded a root mean square error of prediction (RMSEP) of 0.80% MC; a four component PLSR model using scatter corrected spectra yielded a RMSEP of 0.57% MC and a simplified PLS model using seven selected wavelengths (1155, 1212, 1340, 1409, 1724, 1908, and 2249 nm) yielded a RMSEP of 0.77% MC. These models were feasible to predict MC in Arabica and Robusta species. A previous study predicted MC only in intact the green beans of Arabica species (Morgano et al., 2008).

The reliability of NIR spectroscopy to predict MC is lower than the gravimetric method (ISO 6673) (SD_{mean} = 0.21% MC). However, NIR spectroscopy is feasible to replace the gravimetric method for a routine application if time is a limiting factor. The advantage of NIR spectroscopy is that MC can be obtained in few minutes whereas the determination with the gravimetric method needs 16 hours. A simplified model based on only seven selected wavelengths also opens the possibility of creating a simple NIR instrument for MC prediction.

The following study shown in chapter 3 aimed at comparing ultraviolet-visible (UV-Vis) and NIR spectroscopy analysis as methods for differentiating between Arabica and Robusta. The adulteration of Arabica and Robusta is a typical issue in coffee trading (Toci et al., 2016). The price differences among those species are a risk factor for the adulteration (International Coffee Organization., 2018).

UV-Vis spectroscopy has been utilized to measure the caffeine and chlorogenic acid contents in coffee beans (Belay et al., 2008; Belay and Gholap, 2009; Dankowska et al., 2017; Navarra et al., 2015). The present study examines the potential of the caffeine and chlorogenic acid contents in the green beans that were measured using UV-Vis spectroscopy to discriminate among coffee species. The result shows that UV-Vis spectroscopy-based determination of two important compounds—i.e. caffeine and chlorogenic acid—attained a slightly higher classification accuracy of 97.3%. Therefore, we propose UV-Vis spectroscopy as a new approach to discriminate between Arabica and Robusta species of green coffee beans.

NIR spectroscopy is a potential non-destructive analyzing method for the discrimination among species. NIR spectroscopy using 7 selected wavelengths and LDA yielded a similarly high classification accuracy (95.5%). A low number of LVs model is considered to produce a robust model and avoid an over-optimistic model (Brereton, 2006; Munawar et al., 2016).

The findings suggest that, given both the speed, non-destructiveness and low involvement of labor of NIR spectroscopy, it is superior for on-site species discrimination. This study was limited by the environmental conditions and varieties of the bean samples. Various methods have been proposed to differentiate among beans species (Toci et al., 2016). Each procedure has its own merits based on the accuracy, rapidity, simplicity, and cost.

Green beans from Java Island, Indonesia were identified using NIR spectroscopy and stable isotope analysis of oxygen ($\delta^{18}O$), hydrogen (δ^2H), and strontium (δSr) during the third study (chapter 4). The results indicate that NIR spectroscopy was not reliable to identify the origins of the green beans. The

calibration model predicted origins, based on the second derivative using the Savitsky-Golay algorithm (second polynomial order), yielded a low RMSEC and a high output in R^2. In contrast to this result, the prediction model yielded high RMSEP values and low R^2. Further analysis showed that the principal components (PC)1 and PC2 of the PLS-DA model, based on the second derivative using the Savitsky-Golay algorithm (second polynomial order), explained only 12% of the spectral data variance as the independent variable but explained 70% of origin variance as the dependent variable. This finding shows that the calibration model was neither sufficiently robust nor reliable to predict the origin of the examined samples.

The $\delta^{18}O$ and δ^2H value is also not suitable to determinate the origin of the examined samples. The $\delta^{18}O$ and δ^2H analysis has a good potential to identify the origins of green coffee beans if the $\delta^{18}O$ and δ^2H values in precipitation water vary between origins, which is not the case for the samples from Java Island (Bowen et al., 2005). The δSr value is a potential tracer because the δSr value in the green beans reflects its respective host rocks (Clements et al., 2009). The findings suggest that it seems more practical to divide coffee origins from Java Island into two rather than into three origins.

Further work is required to establish NIR spectroscopy as a standard method to examine quality parameters of the green coffee beans in the coffee industry. The NIR spectroscopy is able to predict quality and authenticate the green beans in the simultaneous times, which provide time and cost-effective procedures. Large databases consist of samples from different genetic characteristic and growing condition are required to overcome the complexity of

the issue. The appropriate multivariate analysis is also needed to be considered to build the model that yielded high prediction accuracy.

5.2. Conclusion

NIR spectroscopy has the potential to predict moisture contents and discriminate between Arabica and Robusta but it cannot be recommended as a tool to identify the origin of examined samples. It is possible that this conclusion is limited to the origin and number of examined samples. The green coffee beans from different origins may have differences in their quality attributes. Therefore, further studies are still necessary to evaluate the potential of NIR spectroscopy to predict quality and authentication of green coffee beans from other origins, in order to establish this rapid and non-destructive method as a routine analysis in the coffee industry.

6. Summary

Coffee is one of the most popular beverages in the world as well as an important commodity for several exporting and importing countries, including Indonesia. There are several quality parameters of the green beans that are generally used for trading, e.g., moisture content (MC), species, origin, and defect beans. There are no general agreements on the definitions and methods for the quality measurement. However, there is a variety of analytical methods for the determination of quality parameters—for example, physical, chemical, and biological approaches. Among these approaches, near infrared (NIR) spectroscopy has the potential to serve as an alternative method for the determination of green coffee beans quality because it is fast, reliable, and accurate.

Therefore, the main focus of the present study lay on the evaluation of NIR spectroscopy for its capacity to predict quality attributes and authenticity of green coffee beans. The study was divided into three parts: prediction of MC, discriminating among species, and identifying the origins of intact green coffee beans using NIR spectroscopy.

The green coffee bean samples used for prediction of MC were taken from different islands in Indonesia, while the samples used for discriminating among species and identifying origins were taken from Java Island to consider the variety of environmental factors, agricultural practices, and postharvest treatments.

The results of the first study showed that a three-component partial least squares regression (PLSR) model using raw spectra can fairly accurately predict MC in intact green coffee beans. It furthermore demonstrated that a simplified model based on only seven selected wavelengths opens the possibility of

creating a more affordable NIR instrumentation. The second study showed that UV-Vis spectroscopy-based determination of two important compounds—i.e. caffeine and chlorogenic acid and NIR spectroscopy using 7 selected wavelengths and LDA are applicable to discriminate reliably among species. The third study showed that neither NIR spectroscopy nor $\delta^{18}O$ and δ^2H values were suitable for origin determination. The δSr value in the green beans, however, can potentially be used as a tracer. It may, hence, be concluded that a combination of NIR spectroscopy and multivariate analysis can predict the moisture content in intact green beans and discriminate among coffee species but is not suitable for the identification of Java coffee origins in the examined samples.

7. References

Adams, J., 2012. Australia's American coffee culture. Australas. J. Pop. Cult. 2, 23–36. https://doi.org/10.1386/ajpc.2.1.23_1

Adnan, A., Hörsten, D. von, Pawelzik, E., Mörlein, and D., 2017. Rapid Prediction of Moisture Content in Intact Green Coffee Beans Using Near Infrared Spectroscopy. Foods 6, 38. https://doi.org/10.3390/foods6050038

Aklimawati, L., Yusianto, Mawardi, S., 2014. Characteristics of Quality Profile and Agribusiness of Robusta Coffee in Tambora Mountainside, Sumbawa. Pelita Perkeb. Coffee Cocoa Res. J. 30, 159–180.

Alonso-Salces, R.M., Serra, F., Reniero, F., Heberger, K., 2009. Botanical and Geographical Characterization of Green Coffee (Coffea arabica and Coffea canephora): Chemometric Evaluation of Phenolic and Methylxanthine Contents. J. Agric. Food Chem. 57, 4224–4235. https://doi.org/10.1021/jf8037117

Alves, R.C., Casal, S., Alves, M.R., Oliveira, M.B., 2009. Discrimination between arabica and robusta coffee species on the basis of their tocopherol profiles. Food Chem. 114, 295–299. https://doi.org/10.1016/j.foodchem.2008.08.093

Anderson, K.A., Smith, B.W., 2002. Chemical Profiling To Differentiate Geographic Growing Origins of Coffee. J. Agric. Food Chem. 50, 2068–2075. https://doi.org/10.1021/jf011056v

Arana, V.A., Medina, J., Alarcon, R., Moreno, E., Heintz, L., Schäfer, H., Wist, J., 2015. Coffee's country of origin determined by NMR: The Colombian case. Food Chem. 175, 500–506. https://doi.org/10.1016/j.foodchem.2014.11.160

Babova, O., Occhipinti, A., Maffei, M.E., 2016. Chemical partitioning and antioxidant capacity of green coffee (Coffea arabica and Coffea canephora) of different geographical origin. Phytochemistry 123, 33–39. https://doi.org/10.1016/j.phytochem.2016.01.016

Badmos, S., Fu, M., Granato, D., Kuhnert, N., 2020. Classification of Brazilian roasted coffees from different geographical origins and farming practices based on chlorogenic acid profiles. Food Res. Int. 134, 109218. https://doi.org/10.1016/j.foodres.2020.109218

Baggenstoss, J., Poisson, L., Kaegi, R., Perren, R., Escher, F., 2008. Roasting and Aroma Formation: Effect of Initial Moisture Content and Steam Treatment. J. Agric. Food Chem. 56, 5847–5851. https://doi.org/10.1021/jf8003288

Barbin, D.F., Felicio, A.L. de S.M., Sun, D.-W., Nixdorf, S.L., Hirooka, E.Y., 2014. Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. Food Res. Int. 61, 23–32. https://doi.org/10.1016/j.foodres.2014.01.005

Barbosa, J.N., Borem, F.M., Cirillo, M.A., Malta, M.R., Alvarenga, A.A., Alves, H.M.R., 2012. Coffee Quality and Its Interactions with Environmental Factors in Minas Gerais, Brazil. J. Agric. Sci. 4. https://doi.org/10.5539/jas.v4n5p181

Barbour, M.M., 2007. Stable oxygen isotope composition of plant tissue: a review. Funct. Plant Biol. 34, 83. https://doi.org/10.1071/FP06228

Barjolle, D., Quiñones-Ruiz, X.F., Bagal, M., Comoé, H., 2017. The Role of the State for Geographical Indications of Coffee: Case Studies from Colombia

and Kenya. World Dev. 98, 105–119. https://doi.org/10.1016/j.worlddev.2016.12.006

Baroh, I., Hanani, N., Setiawan, B., Koestiono, D., 2014. Indonesian Coffee Competitiveness in the International Market: Armington Model Application. Am. J. Econ. 4, 184–194.

Bassbasi, M., De Luca, M., Ioele, G., Oussama, A., Ragno, G., 2014. Prediction of the geographical origin of butters by partial least square discriminant analysis (PLS-DA) applied to infrared spectroscopy (FTIR) data. J. Food Compos. Anal. 33, 210–215. https://doi.org/10.1016/j.jfca.2013.11.010

Belay, A., Gholap, A.V., 2009. Characterization and determination of chlorogenic acids (CGA) in coffee beans by UV-Vis spectroscopy. Afr. J. Pure Appl. Chem. 3, 34–240.

Belay, A., Ture, K., Redi, M., Asfaw, A., 2008. Measurement of caffeine in coffee beans with UV/vis spectrometer. Food Chem. 108, 310–315. https://doi.org/10.1016/j.foodchem.2007.10.024

Bertrand, B., Etienne, H., Lashermes, P., Guyot, B., Davrieux, F., 2005. Can near-infrared reflectance of green coffee be used to detect introgression inCoffea arabica cultivars? J. Sci. Food Agric. 85, 955–962. https://doi.org/10.1002/jsfa.2049

Blanco, M., Villarroya, I., 2002. NIR spectroscopy: a rapid-response analytical tool. TrAC Trends Anal. Chem. 21, 240–250.

Bowen, G.J., 2016. The Online Isotopes in Precipitation Calculator, version 2.2. [WWW Document]. URL http://www.waterisotopes.org (accessed 10.1.16).

Bowen, G.J., Wassenaar, L.I., Hobson, K.A., 2005. Global application of stable hydrogen and oxygen isotopes to wildlife forensics. Oecologia 143, 337–348. https://doi.org/10.1007/s00442-004-1813-y

Brereton, R.G., 2006. Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. TrAC Trends Anal. Chem. 25, 1103–1111. https://doi.org/10.1016/j.trac.2006.10.005

Bucheli, P., Meyer, I., Pittet, A., Vuataz, G., Viani, R., 1998. Industrial storage of green Robusta coffee under tropical conditions and its impact on raw material quality and ochratoxin A content. J. Agric. Food Chem. 46, 4507–4511.

Büning-Pfaue, H., 2003. Analysis of water in food by near infrared spectroscopy. Food Chem. 82, 107–115. https://doi.org/10.1016/S0308-8146(02)00583-6

Buratti, S., Sinelli, N., Bertone, E., Venturello, A., Casiraghi, E., Geobaldo, F., 2015. Discrimination between washed Arabica, natural Arabica and Robusta coffees by using near infrared spectroscopy, electronic nose and electronic tongue analysis: NIR and e-devices for coffee classification. J. Sci. Food Agric. 95, 2192–2200. https://doi.org/10.1002/jsfa.6933

Burns, D.T., Tweed, L., Walker, M.J., 2017. Ground Roast Coffee: Review of Analytical Strategies to Estimate Geographic Origin, Species Authenticity and Adulteration by Dilution. Food Anal. Methods 10, 2302–2310. https://doi.org/10.1007/s12161-016-0756-3

Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E., Trygg, J., 2006. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J. Chemom. 20, 341–351. https://doi.org/10.1002/cem.1006

Capo, R.C., Stewart, B.W., Chadwick, O.A., 1998. Strontium isotopes as tracers of ecosystem processes: theory and methods. Geoderma 82, 197–225.

Carrera, F., León-Camacho, M., Pablos, F., González, A.G., 1998. Authentication of green coffee varieties according to their sterolic profile. Anal. Chim. Acta 370, 131–139.

Casal, S., Oliveira, M.B.P.P., Alves, M.R., Ferreira, M.A., 2000. Discriminate Analysis of Roasted Coffee Varieties for Trigonelline, Nicotinic Acid, and Caffeine Content. J. Agric. Food Chem. 48, 3420–3424. https://doi.org/10.1021/jf990702b

Chen, P., Hu, H., 2010. How determinant attributes of service quality influence customer-perceived value: An empirical investigation of the Australian coffee outlet industry. Int. J. Contemp. Hosp. Manag. 22, 535–551. https://doi.org/10.1108/09596111011042730

Cheng, B., Furtado, A., Smyth, H.E., Henry, R.J., 2016. Influence of genotype and environment on coffee quality. Trends Food Sci. Technol. 57, 20–30. https://doi.org/10.1016/j.tifs.2016.09.003

Clements, B., Hall, R., Smyth, H.R., Cottam, M.A., 2009. Thrusting of a volcanic arc: a new structural model for Java. Pet. Geosci. 15, 159–174. https://doi.org/10.1144/1354-079309-831

Couto, C.C., Santos, T.F., Mamede, A.M.G.N., Oliveira, T.C., Souza, A.M., Freitas-Silva, O., Oliveira, E.M.M., 2019. Coffea arabica and C. canephora discrimination in roasted and ground coffee from reference material candidates by real-time PCR. Food Res. Int. 115, 227–233. https://doi.org/10.1016/j.foodres.2018.08.086

Craig, A.P., Franca, A.S., Oliveira, L.S., 2012. Discrimination between defective and non-defective roasted coffees by diffuse reflectance infrared Fourier transform spectroscopy. LWT - Food Sci. Technol. 47, 505–511. https://doi.org/10.1016/j.lwt.2012.02.016

Dankowska, A., Domagała, A., Kowalewski, W., 2017. Quantification of Coffea arabica and Coffea canephora var. robusta concentration in blends by means of synchronous fluorescence and UV-Vis spectroscopies. Talanta 172, 215–220. https://doi.org/10.1016/j.talanta.2017.05.036

Davis, A.P., Govaerts, R., Bridson, D.M., Stoffelen, P., 2006. An annotated taxonomic conspectus of the genus Coffea (Rubiaceae). Bot. J. Linn. Soc. 152, 465–512.

Decazy, F., Avelino, J., Guyot, B., Perriot, J.-J., Pineda, C., Cilas, C., 2003. Quality of different Honduran coffees in relation to several environments. J. Food Sci. 68, 2356–2361.

dePaula, J., Farah, A., 2019. Caffeine Consumption through Coffee: Content in the Beverage, Metabolism, Health Benefits and Risks. Beverages 5, 37.

Di Donfrancesco, B., Gutierrez Guzman, N., Chambers, E., 2014. Comparison of Results from Cupping and Descriptive Sensory Analysis of Colombian Brewed Coffee: Coffee: Comparing Cupping and Descriptive Analysis. J. Sens. Stud. 29, 301–311. https://doi.org/10.1111/joss.12104

Downey, G., Boussion, J., Beauchêne, D., 1994. Authentication of whole and ground coffee beans by near infrared reflectance spectroscopy. J Infrared Spectrosc 2, 85–92.

Duarte, G.S., Pereira, A.A., Farah, A., 2010. Chlorogenic acids and other relevant compounds in Brazilian coffees processed by semi-dry and wet post-harvesting methods. Food Chem. 118, 851–855. https://doi.org/10.1016/j.foodchem.2009.05.042

Durand, C., Fournier, S., 2017. Can Geographical Indications Modernize Indonesian and Vietnamese Agriculture? Analyzing the Role of National and Local Governments and Producers' Strategies. World Dev. 98, 93–104. https://doi.org/10.1016/j.worlddev.2015.11.022

EarthChem, 2011. EarthChem KML for Google Earth [WWW Document]. URL http://www.earthchem.org/sites/earthchem.org/files/EarthChem.kml. (accessed 1.10.16).

Esteban-Díez, I., González-Sáiz, J.M., Pizarro, C., 2004. An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS. Anal. Chim. Acta 514, 57–67. https://doi.org/10.1016/j.aca.2004.03.022

Esteban-Díez, I., González-Sáiz, J.M., Sáenz-González, C., Pizarro, C., 2007. Coffee varietal differentiation based on near infrared spectroscopy. Talanta 71, 221–229. https://doi.org/10.1016/j.talanta.2006.03.052

Esteban-Díez, I., González-Sáiz, J.M., Pizarro, C., 2004. Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy. Anal. Chim. Acta 525, 171–182. https://doi.org/10.1016/j.aca.2004.08.057

Farah, A., Donangelo, C.M., 2006. Phenolic compounds in coffee. Braz. J. Plant Physiol. 18, 23–36.

Farah, A., Monteiro, M.C., Calado, V., Franca, A.S., Trugo, L.C., 2006. Correlation between cup quality and chemical attributes of Brazilian coffee. Food Chem. 98, 373–380. https://doi.org/10.1016/j.foodchem.2005.07.032

Fearn, T., 2002. Assessing Calibrations: SEP, RPD, RER and R 2. NIR News 13, 12–13.

Feria-Morales, A.M., 2002. Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. Food Qual. Prefer. 13, 355–367.

Figueiredo, L.P., Borém, F.M., Cirillo, M.Â., Ribeiro, F.C., Giomo, G.S., Salva, T.D.J.G., 2013. The Potential for High Quality Bourbon Coffees From Different Environments. J. Agric. Sci. 5. https://doi.org/10.5539/jas.v5n10p87

Finzer, J.R.D., Limaverde, J.R., Freitas, A.O., Júnior, J.L., Sfredo, M.A., 2003. Drying of coffee berries in a vibrated tray dryer operated with solids recycle and single-stage. J. Food Process Eng. 26, 207–222.

Gautz, L.D., Smith, V.E., Bittenbender, H.C., 2008. Measuring coffee bean moisture content. Eng. Noteb. 3, 1–3.

Gimase, J.M., Thagana, W.M., Kirubi, D.T., Gichuru, E.K., Kathurima, C.W., 2014. Beverage quality and biochemical attributes of arabusta coffee (C. arabica L. x C. canephora Pierre) and their parental genotypes. Afr. J. Food Sci. 8, 456–464.

González, A.G., Pablos, F., Martın, M.J., León-Camacho, M., Valdenebro, M.S., 2001. HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. Food Chem. 73, 93–101.

Górnaś, P., Siger, A., Pugajeva, I., Czubinski, J., Waśkiewicz, A., Polewski, K., 2014. New insights regarding tocopherols in Arabica and Robusta species coffee beans: RP-UPLC-ESI/MSn and NP-HPLC/FLD study. J. Food Compos. Anal. 36, 117–123. https://doi.org/10.1016/j.jfca.2014.08.005

Gromski, P.S., Muhamadali, H., Ellis, D.I., Xu, Y., Correa, E., Turner, M.L., Goodacre, R., 2015. A tutorial review: Metabolomics and partial least

squares-discriminant analysis – a marriage of convenience or a shotgun wedding. Anal. Chim. Acta 879, 10–23. https://doi.org/10.1016/j.aca.2015.02.012

Horn, P., Hölzl, S., Todt, W., Matthies, D., 1998. Isotope Abundance Ratios of Sr In Wine Provenance Determinations, in A Tree-Root Activity Study, and Of Pb in A Pollution Study on Tree-Rings. Isotopes Environ. Health Stud. 34, 31–42. https://doi.org/10.1080/10256019708036329

Horn, P., Schaaf, P., Holbach, B., Hölzl, S., Eschnauer, H., 1993. 87Sr/86Sr from rock and soil into vine and wine. Z. Für Lebensm.-Unters. Forsch. 196, 407–409.

Indonesian National Standard, 2008. Coffee Beans [Biji Kopi]. SNI 01-2907-2008.

International Coffee Organization., 2018. Historical Data on the Global Coffee Trade. [WWW Document]. URL http://www.ico.org/new_historical.asp (accessed 9.1.19).

International Coffee Organization, 2013. National quality standards. Promotion and Market Development Committee 6th meeting, 9 September 2013, Belo Horizonte, Brazil. [WWW Document]. URL http://www.ico.org/documents/cy2012-13/pm-29e-quality-standards.pdf (accessed 2.1.17).

Isengard, H.-D., 2001. Water content, one of the most important properties of food. Food Control 12, 395–400.

Isengard, H.D., 1995. Rapid water determination in foodstuffs. Trends Food Sci. Technol. 6, 155–162.

Jeszka-Skowron, M., Sentkowska, A., Pyrzyńska, K., De Peña, M.P., 2016. Chlorogenic acids, caffeine content and antioxidant properties of green coffee extracts: influence of green coffee bean preparation. Eur. Food Res. Technol. 242, 1403–1409. https://doi.org/10.1007/s00217-016-2643-y

Joët, T., Laffargue, A., Descroix, F., Doulbeau, S., Bertrand, B., kochko, A. de, Dussert, S., 2010. Influence of environmental factors, wet processing and their interactions on the biochemical composition of green Arabica coffee beans. Food Chem. 118, 693–701. https://doi.org/10.1016/j.foodchem.2009.05.048

Kamruzzaman, M., ElMasry, G., Sun, D.-W., Allen, P., 2012. Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression. Innov. Food Sci. Emerg. Technol. 16, 218–226. https://doi.org/10.1016/j.ifset.2012.06.003

Kathurima, C.W., Gichimu, B.M., Kenji, G.M., Muhoho, S.M., Boulanger, R., 2009. Evaluation of beverage quality and green bean physical characteristics of selected Arabica coffee genotypes in Kenya. Afr. J. Food Sci. 3, 365–371.

Keidel, A., von Stetten, D., Rodrigues, C., Máguas, C., Hildebrandt, P., 2010. Discrimination of Green Arabica and Robusta Coffee Beans by Raman Spectroscopy. J. Agric. Food Chem. 58, 11187–11192. https://doi.org/10.1021/jf101999c

Kim, S.-E., Lee, S.M., Kim, K.-O., 2016. Consumer acceptability of coffee as affected by situational conditions and involvement. Food Qual. Prefer. 52, 124–132. https://doi.org/10.1016/j.foodqual.2016.04.008

Kitzberger, C.S.G., Scholz, M.B. dos S., Benassi, M. de T., 2014. Bioactive compounds content in roasted coffee from traditional and modern Coffea

arabica cultivars grown under the same edapho-climatic conditions. Food Res. Int. 61, 61–66. https://doi.org/10.1016/j.foodres.2014.04.031

Kjeldgaard, D., Ostberg, J., 2007. Coffee Grounds and the Global Cup: Glocal Consumer Culture in Scandinavia. Consum. Mark. Cult. 10, 175–187. https://doi.org/10.1080/10253860701256281

Krzywinski, M., Altman, N., 2014. Points of significance: visualizing samples with box plots. Nat. Methods 11, 119–120.

Ky, C.-L., Louarn, J., Dussert, S., Guyot, B., Hamon, S., Noirot, M., 2001. Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild Coffea arabica L. and C. canephora P. accessions. Food Chem. 75, 223–230.

Labouisse, J.-P., Bellachew, B., Kotecha, S., Bertrand, B., 2008. Current status of coffee (Coffea arabica L.) genetic resources in Ethiopia: implications for conservation. Genet. Resour. Crop Evol. 55, 1079–1093. https://doi.org/10.1007/s10722-008-9361-7

Läderach, P., Oberthür, T., Cook, S., Estrada Iza, M., Pohlan, J.A., Fisher, M., Rosales Lechuga, R., 2011. Systematic agronomic farm management for improved coffee quality. Field Crops Res. 120, 321–329. https://doi.org/10.1016/j.fcr.2010.10.006

Liu, H.-C., You, C.-F., Chen, C.-Y., Liu, Y.-C., Chung, M.-T., 2014. Geographic determination of coffee beans using multi-element analysis and isotope ratios of boron and strontium. Food Chem. 142, 439–445. https://doi.org/10.1016/j.foodchem.2013.07.082

Maeztu, L., Andueza, S., Ibañez, C., Paz de Peña, M., Bello, J., Cid, C., 2001. Multivariate Methods for Characterization and Classification of Espresso Coffees from Different Botanical Varieties and Types of Roast by Foam, Taste, and Mouthfeel. J. Agric. Food Chem. 49, 4743–4747. https://doi.org/10.1021/jf010314l

Marquetti, I., Link, J.V., Lemes, A.L.G., Scholz, M.B. dos S., Valderrama, P., Bona, E., 2016. Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. Comput. Electron. Agric. 121, 313–319. https://doi.org/10.1016/j.compag.2015.12.018

Martins, V. de C., Godoy, R.L. de O., Gouvêa, A.C.M.S., Santiago, M.C.P. de A., Borguini, R.G., Braga, E.C. de O., Pacheco, S., Nascimento, L. da S. de M. do, 2018. Fraud investigation in commercial coffee by chromatography. Food Qual. Saf. 2, 121–133. https://doi.org/10.1093/fqsafe/fyy017

Martín, M.J., Pablos, F., González, A.G., 1998. Discrimination between arabica and robusta green coffee varieties according to their chemical composition. Talanta 46, 1259–1264.

Martín, M.J., Pablos, F., González, A.G., Valdenebro, M.S., León-Camacho, M., 2001. Fatty acid profiles as discriminant parameters for coffee varieties differentiation. Talanta 54, 291–297.

Mehari, B., Redi-Abshiro, M., Chandravanshi, B.S., Combrinck, S., Atlabachew, M., McCrindle, R., 2016. Profiling of phenolic compounds using UPLC–MS for determining the geographical origin of green coffee beans from Ethiopia. J. Food Compos. Anal. 45, 16–25. https://doi.org/10.1016/j.jfca.2015.09.006

Mendonça, J.C.F., Franca, A.S., Oliveira, L.S., 2009. Physical characterization of non-defective and defective Arabica and Robusta coffees before and after

roasting. J. Food Eng. 92, 474–479. https://doi.org/10.1016/j.jfoodeng.2008.12.023

Mendonça, J.C.F., Franca, A.S., Oliveira, L.S., 2007. A comparative evaluation of methodologies for water content determination in green coffee. LWT - Food Sci. Technol. 40, 1300–1303. https://doi.org/10.1016/j.lwt.2006.08.013

Monteiro, P.I., Santos, J.S., Alvarenga Brizola, V.R., Pasini Deolindo, C.T., Koot, A., Boerrigter-Eenling, R., van Ruth, S., Georgouli, K., Koidis, A., Granato, D., 2018. Comparison between proton transfer reaction mass spectrometry and near infrared spectroscopy for the authentication of Brazilian coffee: A preliminary chemometric study. Food Control 91, 276–283. https://doi.org/10.1016/j.foodcont.2018.04.009

Monteiro, P.I., Santos, J.S., Rodionova, O.Y., Pomerantsev, A., Chaves, E.S., Rosso, N.D., Granato, D., 2019. Chemometric Authentication of Brazilian Coffees Based on Chemical Profiling. J. Food Sci. 84, 3099–3108. https://doi.org/10.1111/1750-3841.14815

Morales-Medina, G., Guzmán, A., 2012. Prediction of density and viscosity of colombian crude oils from chromatographic data. CTF-Cienc. Tecnol. Futuro 4, 57–73.

Morgano, M.A., Faria, C.G., Ferrão, M.F., Bragagnolo, N., Ferreira, M.M. de C., 2008. Determination of moisture in raw coffee by near infra-red reflectance spectroscopy and multivariate regression. Food Sci. Technol. Camp. 28, 12–17.

Munawar, A.A., von Hörsten, D., Wegener, J.K., Pawelzik, E., Mörlein, D., 2016. Rapid and non-destructive prediction of mango quality attributes using Fourier transform near infrared spectroscopy and chemometrics. Eng. Agric. Environ. Food 9, 208–215. https://doi.org/10.1016/j.eaef.2015.12.004

Navarra, G., Moschetti, M., Guarrasi, V., Mangione, M.R., Militello, V., Leone, M., 2017. Simultaneous Determination of Caffeine and Chlorogenic Acids in Green Coffee by UV/Vis Spectroscopy. J. Chem. 2017, 1–8. https://doi.org/10.1155/2017/6435086

Navarra, G., Moschetti, M., Guarrasi, V., Mangione, M.R., Militello, V., Leone, M., 2015. Determination of caffeine and chlorogenic acids in raw coffee by absorption and excitation spectroscopy. Presented at the Third International Congress on Cocoa, Coffee and Tea, Aveiro, Portugal.

Nicolaï, B.M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K.I., Lammertyn, J., 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. Postharvest Biol. Technol. 46, 99–118. https://doi.org/10.1016/j.postharvbio.2007.06.024

Oberthür, T., Läderach, P., Posada, H., Fisher, M.J., Samper, L.F., Illera, J., Collet, L., Moreno, E., Alarcón, R., Villegas, A., Usma, H., Perez, C., Jarvis, A., 2011. Regional relationships between inherent coffee quality and growing environment for denomination of origin labels in Nariño and Cauca, Colombia. Food Policy 36, 783–794. https://doi.org/10.1016/j.foodpol.2011.07.005

Oestreich-Janzen, S., 2013. Chemistry of Coffee, in: Reference Module in Chemistry, Molecular Sciences and Chemical Engineering. Elsevier. https://doi.org/10.1016/B978-0-12-409547-2.02786-4

Offermans, E.S.A.S.A., Glasbergen, P., others, 2015. Sustainability Certification and Economic Performance An Analysis of Coffee Marketing Channels in Indonesia. J. Econ. Sustain. Dev. 6, 84–98.

Palacios-Cabrera, H., Taniwaki, M.H., Menezes, H.C., Iamanaka, B.T., 2004. The production of ochratoxin A by Aspergillus ochraceus in raw coffee at different equilibrium relative humidity and under alternating temperatures. Food Control 15, 531–535. https://doi.org/10.1016/j.foodcont.2003.08.006

Pan, L., Lu, R., Zhu, Q., McGrath, J.M., Tu, K., 2015. Measurement of moisture, soluble solids, sucrose content and mechanical properties in sugar beet using portable visible and near-infrared spectroscopy. Postharvest Biol. Technol. 102, 42–50. https://doi.org/10.1016/j.postharvbio.2015.02.005

Pardo, E., Ramos, A., Sanchis, V., Marin, S., 2005. Modelling of effects of water activity and temperature on germination and growth of ochratoxigenic isolates of on a green coffee-based medium. Int. J. Food Microbiol. 98, 1–9. https://doi.org/10.1016/j.ijfoodmicro.2004.05.003

Pittia, P., Nicoli, M.C., Sacchetti, G., 2007. Effect of moisture and water activity on textural properties of raw and roasted coffee beans. J. Texture Stud. 38, 116–134.

Pizarro, C., Esteban-Díez, I., Nistal, A.-J., González-Sáiz, J.-M., 2004. Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. Anal. Chim. Acta 509, 217–227. https://doi.org/10.1016/j.aca.2003.11.008

Puth, M.-T., Neuhäuser, M., Ruxton, G.D., 2014. Effective use of Pearson's product–moment correlation coefficient. Anim. Behav. 93, 183–189. https://doi.org/10.1016/j.anbehav.2014.05.003

Rahmah, M., 2017. The Protection of Agricultural Products under Geographical Indication: An Alternative Tool for Agricultural Development in Indonesia. J. Intellect. Prop. Rights 22, 90–103.

Reh, C., Gerber, A., Prodolliet, J., Vuataz, G., 2006. Water content determination in green coffee – Method comparison to study specificity and accuracy. Food Chem. 96, 423–430. https://doi.org/10.1016/j.foodchem.2005.02.055

Ribeiro, J.S., Ferreira, M.M.C., Salva, T.J.G., 2011. Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. Talanta 83, 1352–1358. https://doi.org/10.1016/j.talanta.2010.11.001

Rodrigues, C., Brunner, M., Steiman, S., Bowen, G.J., Nogueira, J.M.F., Gautz, L., Prohaska, T., Máguas, C., 2011a. Isotopes as Tracers of the Hawaiian Coffee-Producing Regions. J. Agric. Food Chem. 59, 10239–10246. https://doi.org/10.1021/jf200788p

Rodrigues, C., Máguas, C., Prohaska, T., 2011b. Strontium and oxygen isotope fingerprinting of green coffee beans and its potential to proof authenticity of coffee. Eur. Food Res. Technol. 232, 361–373. https://doi.org/10.1007/s00217-010-1362-z

Rodrigues, C.I., Maia, R., Miranda, M., Ribeirinho, M., Nogueira, J.M.F., Máguas, C., 2009. Stable isotope analysis for green coffee bean: A possible method for geographic origin discrimination. J. Food Compos. Anal. 22, 463–471. https://doi.org/10.1016/j.jfca.2008.06.010

Salamanca, C.A., Fiol, N., González, C., Saez, M., Villaescusa, I., 2017. Extraction of espresso coffee by using gradient of temperature. Effect on

physicochemical and sensorial characteristics of espresso. Food Chem. 214, 622–630. https://doi.org/10.1016/j.foodchem.2016.07.120

Santato, A., Bertoldi, D., Perini, M., Camin, F., Larcher, R., 2012. Using elemental profiles and stable isotopes to trace the origin of green coffee beans on the global market. J. Mass Spectrom. 47, 1132–1140.

Sepúlveda, W.S., Chekmam, L., Maza, M.T., Mancilla, N.O., 2016. Consumers' preference for the origin and quality attributes associated with production of specialty coffees: Results from a cross-cultural study. Food Res. Int. 89, 997–1003. https://doi.org/10.1016/j.foodres.2016.03.039

Severa, L., Buchar, J., Nedomová, Š., 2012. Shape and Size Variability of Roasted Arabica Coffee Beans. Int. J. Food Prop. 15, 426–437. https://doi.org/10.1080/10942912.2010.487967

Shabbak, A., Midi, H., 2012. An Improvement of the Hotelling T 2 Statistic in Monitoring Multivariate Quality Characteristics. Math. Probl. Eng. 2012, 1–15. https://doi.org/10.1155/2012/531864

Sørensen, K.M., Petersen, H., Engelsen, S.B., 2012. An On-Line Near-Infrared (NIR) Transmission Method for Determining Depth Profiles of Fatty Acid Composition and Iodine Value in Porcine Adipose Fat Tissue. Appl. Spectrosc. 66, 218–226. https://doi.org/10.1366/11-06396

Subedi, R.N., 2011. Comparative analysis of dry and wet processing of coffee with respect to quality and cost in Kavre District, Nepal: A case of Panchkhal Village. Int. Res. J. Appl. Basic Sci. 2, 181–193.

Tajik, N., Tajik, M., Mack, I., Enck, P., 2017. The potential effects of chlorogenic acid, the main phenolic components in coffee, on health: a comprehensive review of the literature. Eur. J. Nutr. 56, 2215–2244. https://doi.org/10.1007/s00394-017-1379-1

Teuber, R., 2010. Geographical Indications of Origin as a Tool of Product Differentiation: The Case of Coffee. J. Int. Food Agribus. Mark. 22, 277–298. https://doi.org/10.1080/08974431003641612

Toci, A.T., Farah, A., Pezza, H.R., Pezza, L., 2016. Coffee Adulteration: More than Two Decades of Research. Crit. Rev. Anal. Chem. 46, 83–92. https://doi.org/10.1080/10408347.2014.966185

Tomaszewska, E., Soliwoda, K., Kadziola, K., Tkacz-Szczesna, B., Celichowski, G., Cichomski, M., Szmaja, W., Grobelny, J., 2013. Detection Limits of DLS and UV-Vis Spectroscopy in Characterization of Polydisperse Nanoparticles Colloids. J. Nanomater. 2013, 1–10. https://doi.org/10.1155/2013/313081

Tran, H.T.M., Vargas, C.A.C., Slade Lee, L., Furtado, A., Smyth, H., Henry, R., 2017. Variation in bean morphology and biochemical composition measured in different genetic groups of arabica coffee (Coffea arabica L.). Tree Genet. Genomes 13. https://doi.org/10.1007/s11295-017-1138-8

Triyanti, D.R., 2016. Coffee Outlook. Center Data and Information System for Agriculture. General Secretariat. Ministry of Agriculture of The Republic of Indonesia, Jakarta, Indonesia.

Van Hilten, H.J., Fisher, P.J., Wheeler, M.A., 2011. The Coffee Exporter's Guide, 3rd ed. International Trade Centre UNCTAD/GATT.

Vellema, W., Buritica Casanova, A., Gonzalez, C., D'Haese, M., 2015. The effect of specialty coffee certification on household livelihood strategies and specialisation. Food Policy 57, 13–25. https://doi.org/10.1016/j.foodpol.2015.07.003

Wahyudi, T., Jati, M., 2012. Challenges of sustainable coffee certification in Indonesia, in: Seminar on the Economic, Social and Environmental Impact of Certification on the Coffee Supply Chain, International Coffee Council 109th Session, London, United Kingdom. Retrieved July. p. 2013.

Weckerle, B., Richling, E., Heinrich, S., Schreier, P., 2002. Origin assessment of green coffee (Coffea arabica) by multi-element stable isotope analysis of caffeine. Anal. Bioanal. Chem. 374, 886–890. https://doi.org/10.1007/s00216-002-1560-z

Westad, F., Marini, F., 2015. Validation of chemometric models – A tutorial. Anal. Chim. Acta 893, 14–24. https://doi.org/10.1016/j.aca.2015.06.056

Whitford, D.J., 1975. Strontium isotopic studies of the volcanic rocks of the Saunda arc, Indonesia, and their petrogenetic implications. Geochim. Cosmochim. Acta 39, 1287–1302.

Williams, P., 2006. Grains and Seeds, in: Ozaki, Y., McClure, W.F., Christy, A.A. (Eds.), Near-Infrared Spectroscopy in Food Science and Technology. John Wiley & Sons, pp. 165–217.

Wilson, A.P., Wilson, N.L.W., 2014. The economics of quality in the specialty coffee industry: insights from the Cup of Excellence auction programs. Agric. Econ. 45, 91–105. https://doi.org/10.1111/agec.12132

Wintgens, J.N., 2004. Coffee: growing, processing, sustainable production. A guidebook for growers, processors, traders, and researchers. WILEY-VCH Verlag GmbH & Co. KGaA.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

Yakir, D., 1992. Variations in the natural abundance of oxygen-18 and deuterium in plant carbohydrates. Plant Cell Environ. 15, 1005–1020.

Yakir, D., Sternberg, L. da S.L., 2000. The use of stable isotopes to study ecosystem gas exchange. Oecologia 123, 297–311.

Yani, A., 2008. Fungal infection at coffee beans during primary processing (case study in Bengkulu Province). Akta Agrosia 11, 87–95.

Appendix 1. Statistical parameters of calibration and prediction models for moisture content prediction

Step	Statistical Parameter	Pre-processing method				
		Raw	OSC	MSC	EMSC	SNV
	LV	3	2	3	4	3
Calibration	R^2 calibration (%)	98.34	98.17	98.25	98.51	98.43
	R^2 cross validation (%)	97.97	98	97.86	98.11	98.15
	RMSEC (%)	0.52	0.55	0.53	0.49	0.51
	RMSECV (%)	0.52	0.59	0.59	0.57	0.55
	R^2 prediction (%)	96.41	NA	97.78	98.18	97.79
	RMSEP (%)	0.8	5.53	0.63	0.57	0.63
Prediction	SEP (%)	0.69	0.65	0.57	0.5	0.55
	Bias (%)	0.42	5.5	0.28	0.28	0.31

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross validation; OSC: orthogonal signal correction; MSC: multiplicative scatter correction; EMSC: extended multiplicative scatter correction; SNV: standard normal variate.

Pre-processing method	Step	Statistical Parameter	3 segments	7 segments	11 segments	15 segments	19 segments
Smoothing Moving average	Calibration	LV	3	3	3	3	3
		R^2 calibration (%)	98.34	98.34	98.34	98.34	98.35
		R^2 cross validation (%)	97.93	98.06	98.01	97.73	98.12
		RMSEC (%)	0.52	0.52	0.52	0.52	0.52
		RMSECV (%)	0.59	0.58	0.57	0.63	0.56
	Prediction	R^2 prediction (%)	96.4	97.4	96.23	96.1	95.9
		RMSEP (%)	0.8	0.81	0.82	0.83	0.85
		SEP (%)	0.69	0.69	0.69	0.69	0.69
		Bias (%)	0.42	0.43	0.45	0.48	0.52
Smoothing GaussianFilter	Calibration	LV	3	3	3	3	3
		R^2 calibration (%)	98.34	98.34	97.88	98.34	98.35
		R^2 cross validation (%)	98.03	98.13	97.23	98.12	97.86
		RMSEC (%)	0.52	0.52	0.62	0.52	0.52
		RMSECV (%)	0.57	0.56	0.72	0.56	0.59
	Prediction	R^2 prediction (%)	96.41	96.38	0.98	96.26	96.16
		RMSEP (%)	0.8	0.8	0.62	0.82	0.83
		SEP (%)	0.69	0.69	0.62	0.69	0.69
		Bias (%)	0.42	0.43	0.01	0.45	0.47

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross validation.

Appendix 1. Continue

Pre-processing method	Step	Statistical Parameter	3 segments	7 segments	11 segments	15 segments	19 segments
Smoothing MedianFilter	Calibration	LV	3	3	3	3	3
		R^2 calibration (%)	98.34	98.34	98.34	98.35	98.35
		R^2 cross validation (%)	98.09	97.93	98.03	97.9	98.03
		RMSEC (%)	0.52	0.52	0.52	0.52	0.52
		RMSECV (%)	0.57	0.6	0.57	0.61	0.57
	Prediction	R^2 prediction (%)	98.34	96.41	96.41	96.4	96.41
		RMSEP (%)	0.52	0.8	0.8	0.8	0.8
		SEP (%)	0.52	0.69	0.69	0.69	0.69
		Bias (%)	0.42	0.42	0.41	0.41	0.42
SavitskyGolay 1st derivative, 2 Polinomial order	Calibration	LV	5	5	5	5	5
		R^2 calibration (%)	99.38	99.21	99.2	99.22	99.21
		R^2 cross validation (%)	98.31	98.77	98.83	98.78	98.7
		RMSEC (%)	0.32	0.36	0.36	0.36	0.36
		RMSECV (%)	0.53	0.46	0.43	0.45	0.46
	Prediction	R^2 prediction (%)	NA	NA	NA	NA	NA
		RMSEP (%)	6.2	6.09	5.62	5.15	4.78
		SEP (%)	4.05	4.01	4.02	4.04	4.05
		Bias (%)	4.7	4.63	3.97	3.25	2.6

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross validation.

Appendix 1. Continue

Pre-processing method	Step	Statistical Parameter	3 segments	7 segments	11 segments	15 segments	19 segments
SavitskyGolay 2st derivative, 2 Polinomial order	Calibration	LV	4	4	5	5	6
		R^2 calibration (%)	99.15	97.8	99.04	98.99	99.44
		R^2 cross validation (%)	81.6	91.5	96.9	98.4	99.02
		RMSEC (%)	0.37	0.6	0.4	0.41	0.3
		RMSECV (%)	1.74	1.19	0.72	0.53	0.41
	Prediction	R^2 prediction (%)	NA	NA	97.13	NA	NA
		RMSEP (%)	10.99	4.52	0.72	6.8	4.94
		SEP (%)	11.97	4.48	0.71	4.28	4.28
		Bias (%)	2.53	0.93	0.16	5.32	2.54
SavitskyGolay 3st derivative, 3 Polinomial order	Calibration	LV	NA	4	3	6	3
		R^2 calibration (%)	NA	97.97	92.97	99.33	92.98
		R^2 cross validation (%)	NA	74.37	82.3	90.51	82.3
		RMSEC (%)	NA	0.58	1.07	0.33	1.07
		RMSECV (%)	NA	2.06	1.72	1.29	1.72
	Prediction	R^2 prediction (%)	NA	NA	NA	NA	NA
		RMSEP (%)	NA	6.27	4.3	4.23	4.23
		SEP (%)	NA	10.16	4.62	4.33	4.28
		Bias (%)	NA	4.53	0.87	0.39	0.09

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross validation.

Appendix 1. Continue

Pre-processing method	Step	Statistical Parameter	area	mean
Normalization	Calibration	LV	4	4
		R^2 calibration (%)	98.25	98.25
		R^2 cross validation (%)	96.86	97.82
		RMSEC (%)	0.53	0.53
		RMSECV (%)	0.71	0.61
	Prediction	R^2 prediction (%)	96.3	96.3
		RMSEP (%)	0.81	0.81
		SEP (%)	0.74	0.75
		Bias (%)	0.34	0.34

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross validation.

Appendix 1. Continue

Pre-processing method	Step	Statistical Parameter	Baseline Offset	Linear Baseline Correction
		LV	4	3
	Calibration	R^2 calibration (%)	98.17	97.02
Baseline		R^2 cross validation (%)	97.58	96.56
		RMSEC (%)	0.55	0.70
		RMSECV (%)	0.63	0.75
		R^2 prediction (%)	96.31	97.29
		RMSEP (%)	0.81	0.70
	Prediction	SEP (%)	0.79	0.66
		Bias (%)	0.22	0.25

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSECV: root mean square error of cross validation.

Appendix 2. Equations

No	Parameter	Equation
1	Beer-Lambert	$A = \varepsilon * c * l$ Where: A = absorbance ε = the molar decadic absorption coefficient, c = the concentration of the absorbing compound. l = distance in the absorbing medium
2	Caffeine content (% dw)	$\dfrac{C * V}{cf * S} * 100\%$ Where: C = measured concentration (ppm) V = sample volume (ml) S = dry weight of sample (mg) cf = conversion factor (0.001 mg ml^{-1} ppm^{-1})

Appendix 2. Continue

No	Parameter	Equation
3	Chlorogenic content (% dw)	$\dfrac{C * V}{cf * S} * 100\%$ Where: C = measured concentration (ppm) V = sample volume (ml) S = dry weight of sample (mg) cf = conversion factor (0.001 mg ml^{-1} ppm^{-1})
4	Root mean squared error of prediction (RMSEP)	$\sqrt{\dfrac{1}{N}\sum (\tilde{y}_i - y_{i,ref})^2}$ where N= the size of the test set $\tilde{y}_i$ and $y_{i,ref}$ = the prediction and reference value for sample i, respectively

Appendix 3. Statistical parameters of several pre-processing method on diffuse reflectance (log 1/R) spectra by partial least squares discriminant analysis (PLS-DA) for species discrimination

No	Pre-processing method	LV	R^2 of the calibration model	RMSEC	R^2 of the prediction model	RMSEP
1	Raw	7	0.8896	0.3266	0.7151	0.6005
2	Smoothing (Moving average 3 segments)	7	0.8896	0.3266	0.8893	0.327
3	Smoothing (Moving average 7 segments)	7	0.8896	0.3265	0.7148	0.6162
4	Smoothing (Moving average 9 segments)	7	0.8897	0.3265	0.7146	0.6276
5	Smoothing (Moving average 11 segments)	7	0.8896	0.3266	0.7144	0.6426
6	Smoothing (Moving average 15 segments)	7	0.8895	0.3268	0.7137	0.6837
7	Smoothing (Moving average 19 segments)	7	0.8891	0.3273	0.7127	0.7375
8	Smoothing (Gaussian filter 3 segments)	7	0.8895	0.3268	0.7148	0.6014
9	Smoothing (Gaussian filter 7 segments)	7	0.8891	0.3273	0.7136	0.609
10	Smoothing (Gaussian filter 11 segments)	7	0.8889	0.3277	0.7121	0.623
11	Smoothing (Gaussian filter 15 segments)	7	0.8886	0.3281	0.7101	0.6441

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction.

Appendix 3. Continue

No	Pre-processing method	LV	R² of the calibration model	RMSEC	R² of the prediction model	RMSEP
12	Smoothing (Gaussian filter 19 segments)	7	0.8882	0.3286	0.7077	0.6729
13	Savitsky-Golay (1st derivative, 2 Polinomial order)	4	0.9727	0.8829	0.0129	1.0143
14	Savitsky-Golay (2st derivative, 2 Polinomial order)	2	0.85916	0.3689	0.0687	1.557
15	Savitsky-Golay (3st derivative, 3 Polinomial order)	2	0.8371	0.3967	0.1772	2.003
16	OSC	6	0.8877	0.3295	0.5615	8.305
17	MSC	3	0.8526	0.3774	0.8134	0.4734
18	EMSC	6	0.9139	0.2884	0.9049	0.3641
19	Normalization (area)	7	0.9317	0.257	0.903	0.3745
20	Normalization (mean)	6	0.9317	0.257	0.903	0.3745
21	Baseline (Baseline offset)	6	0.9422	0.2364	0.7071	0.5933
22	Baseline (Linear baseline correction)	6	0.8977	0.3144	0.5519	0.778
23	Baseline (Baseline offset + linear baseline correction)	7	0.894	0.32	0.7152	0.6338

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction; OSC: orthogonal signal correction; MSC: multiplicative scatter correction; EMSC: extended multiplicative scatter correction.

Appendix 4. The correlation between caffeine content and altitude. (a) Arabica. (b) Robusta.
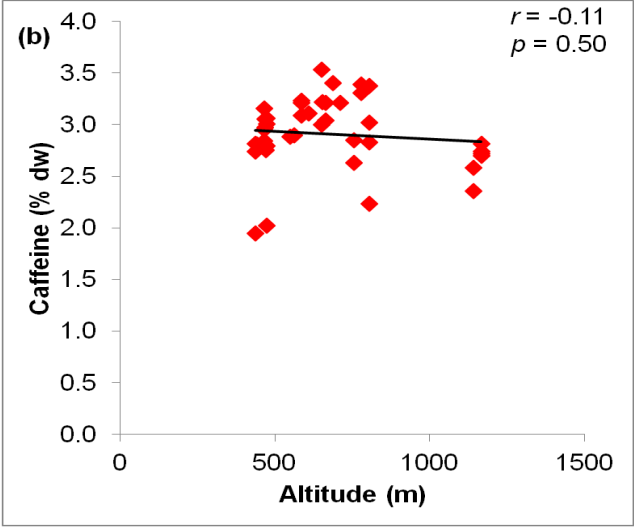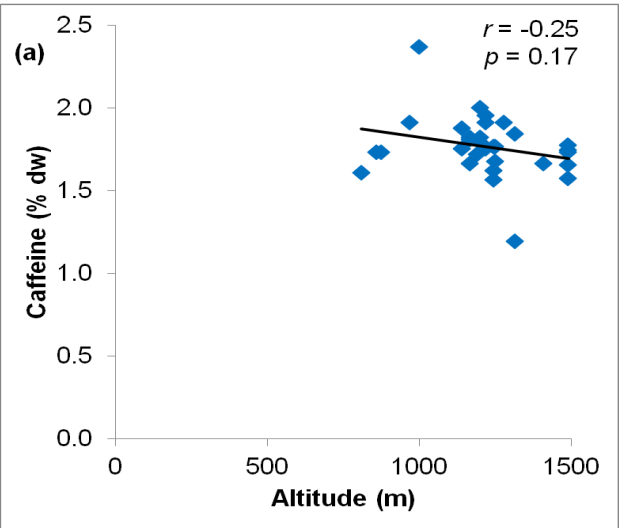
Appendix 5. The correlation between chlorogenic acid content and altitude. (a) Arabica. (b) Robusta.
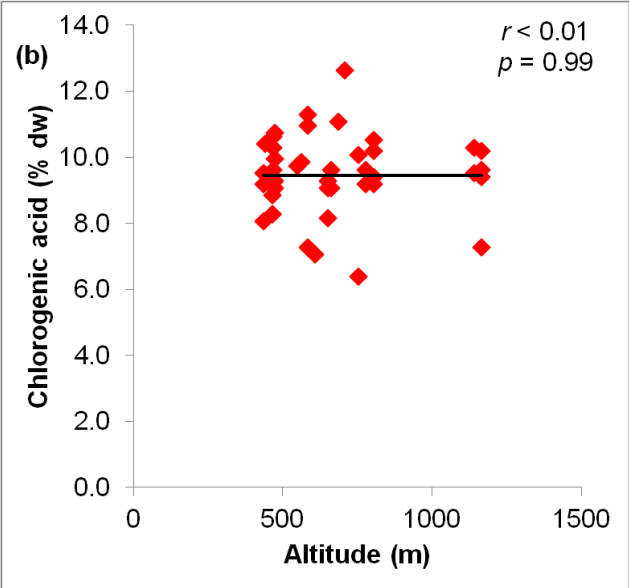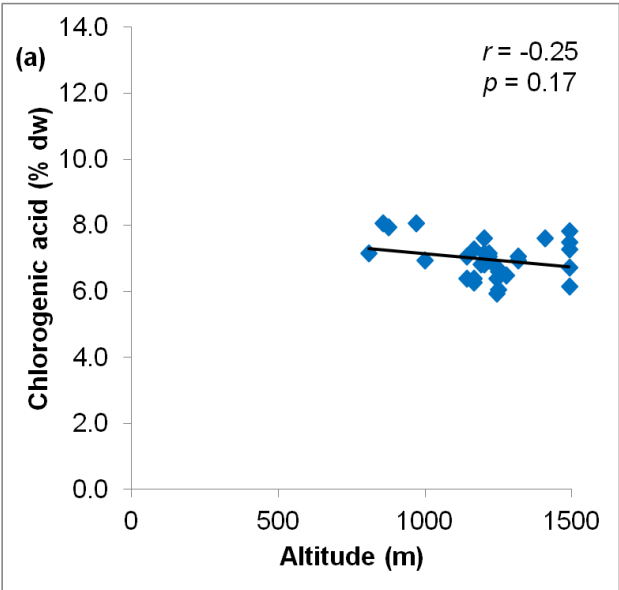
Appendix 6. Caffeine and chlorogenic acid content in different species and origin of green coffee beans samples.

Location	Origin	Species	Variety	Altitude (m)	Longitude	Latitude	Caffeine (% dw)	Chlorogenic acid (% dw)
1	West Java	Robusta	Wild	548	106.92	-6.59	2.88	9.73
2	West Java	Arabica	Java Preanger	1243	107.68	-6.82	1.62	5.93
	West Java	Arabica	Tim-tim	1243	107.68	-6.82	1.57	6.37
	West Java	Arabica	Linie S 795	1243	107.68	-6.82	1.77	6.71
3	West Java	Arabica	Sigararutang	1140	107.5	-7.11	2.36	10.29
	West Java	Robusta	Wild	1140	107.5	-7.11	2.58	9.50
	West Java	Robusta	Wild	1140	107.5	-7.11	1.88	6.37
	West Java	Arabica	Sigararutang	1140	107.5	-7.11	1.76	7.04
4	West Java	Arabica	Sigararutang + Linie S 795	1315	107.4	-7.06	1.20	7.04
	West Java	Arabica	Sigararutang + Linie S 795	1315	107.4	-7.06	1.85	6.93
5	West Java	Arabica	Tim-tim	1490	107.59	-7.17	1.58	7.27
	West Java	Arabica	Sigararutang	1490	107.59	-7.17	1.74	6.71
	West Java	Arabica	Linie S 795 + Sigararutang + Andungsari	1490	107.59	-7.17	1.78	7.49
	West Java	Arabica	Sigararutang	1490	107.59	-7.17	1.65	6.15
	West Java	Arabica	Sigararutang	1490	107.59	-7.17	1.73	7.83
6	West Java	Arabica	Linie S 795 + Sigararutang + Andungsari + Ateng	1409	107.56	-7.18	1.67	7.60

Appendix 6. Continue

Location	Origin	Species	Variety	Altitude (m)	Longitude	Latitude	Caffeine (% dw)	Chlorogenic acid (% dw)
7	West Java	Arabica	Linie S 795 + Ateng + Tim-tim + Preanger	1165	107.56	-7.11	2.72	10.18
	West Java	Arabica	Linie S 795 + Ateng + Tim-tim + Preanger	1165	107.56	-7.11	2.82	9.62
	West Java	Robusta	Wild	1165	107.56	-7.11	2.69	7.27
	West Java	Robusta	Wild	1165	107.56	-7.11	2.74	9.39
	West Java	Robusta	Wild	1165	107.56	-7.11	1.67	7.27
	West Java	Robusta	Wild	1165	107.56	-7.11	1.81	6.37
	West Java	Arabica	Wild	1165	107.56	-7.11	1.82	6.26
8	West Java	Arabica	Na	1217	107.69	-6.85	1.91	7.04
	West Java	Arabica	Na	1217	107.69	-6.85	1.76	7.16
	West Java	Arabica	Na	1217	107.69	-6.85	1.96	6.93
9	Central Java	Arabica	Na	1200	Na	Na	3.04	9.06
	Central Java	Arabica	Na	1200	Na	Na	3.21	9.62
10	Central Java	Robusta	Na	663	110.07	-7.18	1.82	7.16
	Central Java	Robusta	Na	663	110.07	-7.18	2.00	7.60
11	Central Java	Robusta	BP409	650	110.09	-7.14	3.00	9.28
	Central Java	Robusta	BP534	650	110.09	-7.14	3.53	8.16
12	Central Java	Robusta	BP534	804	110.07	-7.19	3.38	10.51
	Central Java	Robusta	BP409	804	110.07	-7.19	2.24	9.17
13	Central Java	Robusta	BP534	585	110.13	-7.18	3.21	7.27
	Central Java	Robusta	Wild	585	110.13	-7.18	3.23	11.29
	Central Java	Robusta	BP42	585	110.13	-7.18	3.09	10.96
14	Central Java	Robusta	Wild	609	110.16	-7.18	3.11	7.04
15	Central Java	Robusta	BP42	687	110.18	-7.18	3.40	11.07
16	Central Java	Robusta	BP534	652	110.17	-7.33	3.22	9.06
	Central Java	Arabica	Linie S 795	Na	Na	Na	2.23	8.83

Appendix 6. Continue

Location	Origin	Species	Variety	Altitude (m)	Longitude	Latitude	Caffeine (% dw)	Chlorogenic acid (% dw)
17	Central Java	Robusta	Wild	754	110.12	-7.22	2.85	10.06
	Central Java	Robusta	BP42 + BP534	754	110.12	-7.22	2.63	6.37
18	Central Java	Robusta	BP42	804	110.08	-7.21	2.83	10.18
	Central Java	Robusta	BGN371	804	110.08	-7.21	3.02	9.39
19	Central Java	Robusta	Wild	778	110.11	-7.21	3.31	9.62
	Central Java	Robusta	BGN371	778	110.11	-7.21	3.39	9.17
20	Central Java	Robusta	BP534	709	110.13	-7.23	3.21	12.64
21	Central Java	Arabica	Kartika	1248	110.05	-7.33	1.77	6.04
	Central Java	Arabica	Linie S 795	1248	110.05	-7.33	1.68	6.60
22	East Java	Robusta	BP534	473	112.77	-8.22	3.01	10.73
	East Java	Robusta	BP534	473	112.77	-8.22	2.02	9.06
	East Java	Robusta	BP534	473	112.77	-8.22	3.06	9.95
	East Java	Robusta	BP534	473	112.77	-8.22	2.80	9.28
23	East Java	Robusta	Umbulsari + Wild	466	112.77	-8.22	2.84	8.27
	East Java	Robusta	Malamsari	466	112.77	-8.22	2.78	9.28
	East Java	Robusta	BP534	466	112.77	-8.22	3.15	8.27
24	East Java	Robusta	BP534	436	112.77	-8.23	2.82	9.50
	East Java	Robusta	Excelsa	436	112.77	-8.23	1.95	8.05
25	East Java	Robusta	BP534	441	112.77	-8.23	2.81	10.40
26	East Java	Robusta	BP534	436	112.77	-8.23	2.74	9.17
27	East Java	Robusta	Grembyong + BP534	469	112.77	-8.22	2.75	10.62
28	East Java	Robusta	Umbulsari	464	112.77	-8.22	2.95	8.83

Appendix 6. Continue

Location	Origin	Species	Variety	Altitude (m)	Longitude	Latitude	Caffeine (% dw)	Chlorogenic acid (% dw)
29	East Java	Robusta	BP534	467	112.77	-8.22	2.97	9.62
	East Java	Robusta	BP534	467	112.77	-8.22	3.05	10.29
30	East Java	Arabica	Linie S 795	808	112.83	-8.17	1.61	7.16
31	East Java	Arabica	Linie S 795	874	112.83	-8.17	1.73	7.94
32	East Java	Arabica	Linie S 795	857	112.84	-8.17	1.73	8.05
33	East Java	Arabica	Linie S 795	1190	112.86	-8.14	1.72	6.82
34	East Java	Arabica	Linie S 795	1201	112.86	-8.14	1.79	6.82
35	East Java	Arabica	Na	968	112.84	-8.15	1.91	8.05
36	East Java	Arabica	Na	999	112.84	-8.15	2.37	6.93
37	East Java	Arabica	Linie S 795	1277	113.69	-7.93	1.91	6.49
38	East Java	Robusta	BP 42 + BP358	562	112.69	-8.28	2.90	9.84

+: there are several cultivated varieties in one plantation; Na: no available data; dw: dry weight.

Appendix 7. Statistical parameters of calibration and prediction models to identify the origin of Java green coffee beans

No	Origin	Pre-processing method	LV	R^2 of the calibration model	RMSEC	R^2 of the prediction model	RMSEP
1	West Java		7	0.6276	0.2903	0.1614	1.0633
	Central Java	Raw		0.3875	0.3723	0.0243	0.7488
	East Java			0.393	0.3596	0.0182	0.6256
2	West Java	Smoothing	7	0.6271	0.2905	0.1622	1.0676
	Central Java	(Moving average		0.387	0.3725	0.0245	0.7511
	East Java	3 segments)		0.3928	0.3596	0.019	0.6267
3	West Java	Smoothing	7	0.6267	0.2907	0.1616	1.0613
	Central Java	(Moving average		0.3863	0.3727	0.0247	0.7486
	East Java	7 segments)		0.3928	0.3596	0.0181	0.6251
4	West Java	Smoothing	7	0.6266	0.2907	0.1617	1.059
	Central Java	(Moving average		0.3861	0.3728	0.0248	0.7489
	East Java	11 segments)		0.3928	0.3596	0.0179	0.6243
5	West Java	Smoothing	7	0.6265	0.2908	0.1625	1.0615
	Central Java	(Moving average		0.3858	0.3728	0.025	0.7532
	East Java	15 segments)		0.3928	0.3596	0.0179	0.624
6	West Java	Smoothing	7	0.6263	0.2908	0.1621	1.0545
	Central Java	(Moving average		0.3855	0.3729	0.0252	0.7535
	East Java	19 segments)		0.3929	0.3596	0.0173	0.622
7	West Java	Smoothing	7	0.6274	0.2904	0.162	1.0669
	Central Java	(Gaussian filter		0.3873	0.3724	0.0245	0.7508
	East Java	3 segments)		0.3929	0.3596	0.01844	0.6265

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction.

No	Origin	Pre-processing method	LV	R² of the calibration model	RMSEC	R² of the prediction model	RMSEP
8	West Java	Smoothing	7	0.627	0.2906	0.1618	1.0633
	Central Java	(Gaussian filter		0.3866	0.3726	0.0246	0.7493
	East Java	7 segments)		0.3928	0.3597	0.0182	0.6255
9	West Java	Smoothing	7	0.6268	0.2906	0.1616	1.0605
	Central Java	(Gaussian filter		0.3864	0.3727	0.0247	0.7483
	East Java	11 segments)		0.3927	0.3597	0.0181	0.6245
10	West Java	Smoothing	7	0.6267	0.2907	0.1631	1.0682
	Central Java	(Gaussian filter		0.3863	0.3727	0.0248	0.7536
	East Java	15 segments)		0.3927	0.3597	0.0185	0.626
11	West Java	Smoothing	7	0.6265	0.2907	0.1621	1.059
	Central Java	(Gaussian filter		0.3862	0.3727	0.0249	0.7505
	East Java	19 segments)		0.3927	0.3527	0.0178	0.6237
12	West Java	Smoothing	7	0.6274	0.2904	0.1619	1.0665
	Central Java	(Median filter		0.3873	0.3724	0.0244	0.7504
	East Java	3 segments)		0.3929	0.3596	0.0184	0.6265
13	West Java	Smoothing	7	0.6273	0.2905	0.1618	1.0663
	Central Java	(Median filter		0.3868	0.3726	0.0244	0.7499
	East Java	7 segments)		0.3931	0.3596	0.0184	0.6268
14	West Java	Smoothing	7	0.6274	0.2904	0.1619	1.0644
	Central Java	(Median filter		0.3869	0.3725	0.0246	0.7494
	East Java	9 segments)		0.3931	0.3595	0.0183	0.626

LV: latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction.

Appendix 7. Continue

No	Origin	Pre-processing method	LV	R² of the calibration model	RMSEC	R² of the prediction model	RMSEP
15	West Java	Smoothing	7	0.6276	0.2903	0.1616	1.0617
	Central Java	(Median filter		0.3869	0.3725	0.0246	0.7479
	East Java	11 segments)		0.3933	0.3595	0.0181	0.6256
16	West Java	Smoothing	7	0.628	0.2902	0.1626	1.0718
	Central Java	(Median filter		0.3873	0.3724	0.0245	0.7538
	East Java	15 segments)		0.3934	0.3595	0.0185	0.6276
17	West Java	Smoothing	7	0.6282	0.2901	0.1621	1.0706
	Central Java	(Median filter		0.3874	0.3724	0.0244	0.7538
	East Java	19 segments)		0.393	0.3596	0.0183	0.6268
18	West Java	Savitsky-Golay	6	0.9161	0.1378	0.06	1.301
	Central Java	(1st derivative, 2		0.8679	0.1729	0.1665	0.6457
	East Java	Polynomial order)		0.7952	0.2088	0.1486	1.66
19	West Java	Savitsky-Golay	3	0.84	0.1903	0.0749	1.3266
	Central Java	(2st derivative, 2		0.8589	0.1787	0.786	1.1282
	East Java	Polynomial order)		0.7458	0.2327	0.1276	1.1523
20	West Java	Savitsky-Golay	3	0.8323	0.1948	0.1408	0.9867
	Central Java	(3st derivative, 3		0.8673	0.1733	0.0135	0.4796
	East Java	Polynomial order)		0.7736	0.2196	0.1405	0.9829
21	West Java		7	0.3067	0.3961	0.0195	4.24
	Central Java	OSC		0.4660	0.3477	0.0000	6.3691
	East Java			0.2061	0.4112	0.0665	4.26

LV: latent variables; R²: the coefficient of determination; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction; OSC: orthogonal signal correction.

Appendix 7. Continue

No	Origin	Pre-processing method	LV	R² of the calibration model		RMSEC	R² of the prediction model	RMSEP
22	West Java		4	0.4816	0.3425	0.1655		0.4734
	Central Java	MSC		0.3102	0.3951	0.0102		0.5413
	East Java			0.347	0.373	0.222		0.449
23	West Java		5	0.5567	0.3168	0.3372		0.4074
	Central Java	EMSC		0.3831	0.2244	0.0046		0.5269
	East Java			0.4183	0.352	0.2252		0.464
24	West Java		7	0.6468	0.2827	0.3186		0.4394
	Central Java	Normalization		0.3552	0.382	0.0094		0.553
	East Java	(area)		0.4207	0.3513	0.2595		0.4606
25	West Java		7	0.6468	0.2827	0.3186		0.4393
	Central Java	Normalization		0.3552	0.382	0.009		0.553
	East Java	(mean)		0.4207	0.3513	0.2595		0.4606
26	West Java		4	0.4354	0.3575	0.1147		0.468
	Central Java	Baseline		0.3269	0.3903	0.0016		0.5108
	East Java	(Baseline Offset)		0.2987	0.3865	0.1641		0.4553
27	West Java	Baseline	4	0.6176	0.2941	0.238		0.464
	Central Java	(Linear Baseline		0.4226	0.3615	0.0522		0.4845
	East Java	Correction)		0.3204	0.3805	0.2417		0.4396

LV: the number of latent variables; R^2: the coefficient of determination; RMSEC: root mean square error of calibration; RMSEP: root mean square error of prediction; MSC: multiplicative scatter correction; EMSC: extended multiplicative scatter correction.