https://doi.org/10.48196/016.02.2020.04

Prediction of Moisture and Caffeine Contents of the Roasted Coffee Beans (*Coffea liberica* Hiern) using NIRS and PLSR-MLR Models

Arnel R. Feraer[1], Engelbert K. Peralta[2], Erwin F. Quilloy[3], Delfin F. Suministrado[4], Paul R. Armstrong[5], and Kevin F. Yaptenco[6]

[1]MSAE Graduate, Agricultural, Food & Bioprocess Engineering Division, Institute of Agricultural and Biosystems Engineering , College of Engineering and Agro-Industrial Technology, University of the Philippines Los Baños, 4031 College, Laguna, Philippines (Author for correspondence email: arferaer@up.edu.ph)

[2]Professor 11, [6]Professor 9, Agricultural, Food & Bioprocess Engineering Division, Institute of Agricultural and Biosystems Engineering , College of Engineering and Agro-Industrial Technology, University of the Philippines Los Baños, 4031 College, Laguna, Philippines

[3]Assistant Professor 7, [4]Retired Professor 12, Agribiosystems Machinery and Power Engineering Division, Institute of Agricultural and Biosystems Engineering, College of Engineering and Agro-Industrial Technology, University of the Philippines Los Baños, 4031 College, Laguna, Philippines

[5]Research Agricultural Engineer, Center for Grain and Animal Health Research (CGAHR), United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Manhattan, Kansas, 66502, USA

ABSTRACT

Near-infrared spectroscopy was assessed for the prediction of moisture content (MC) and caffeine content (CC) of ground "Barako" roasted coffee. Individual models were developed using a chemometric analysis of the NIR spectra (900-1700 nm). Partial least squares regression (PLSR) cross-validation and validation results showed that the MC models could be used for at least quality assurance applications. However, the CC model for PLSR cross-validation can only be used for rough screening and approximate calibration applications due to low RPD (2.000) and R^2 (0.755) values. The results for the validation models of CC obtained lower RPD (0.220) and R^2 (0.136) that it did not pass for any use or application. The results of the PLSR modeling identified significant wavelengths based on the regression coefficient and variable importance of projection. These wavelengths were used to develop multiple linear regression (MLR) models. MC model, with 3 wavelengths, was suitable for most research applications with an RPD = 2.600 and R^2 = 0.851. CC model, with 8 wavelengths, did not pass for any use or application due to poor predictive performance (RPD = 1.378, R^2 = 0.471). The results showed that only the MC models can be used for quality assessment of roasted coffee.

Keywords: Near-infrared, coffee, Coffea liberica, coffee roasting, moisture content, caffeine content, PLSR, MLR

INTRODUCTION

Coffee roasting is an essential process for green beans to exhibit different flavors and aromas that can be categorized into several degrees of roasting. Each roasting condition would result in different kinds of taste and aroma profiles, as well as physical and chemical properties. Some of the determinants of roasting degree include color, aroma, sucrose content, acidity, moisture content, and caffeine content. However, there are still no reliable methods of roasted coffee authenticity, especially for the cv. Liberica variety except for manual inspection and coffee cupping which can only be done by coffee

experts. Obtaining the determinants of the roasting degree would also be time-consuming and will also require different kinds of standard tests. While it is true that coffee possesses a unique taste and aroma, it is also known to be associated with the chemical caffeine which stimulates the alertness of the brain and relieves drowsiness. The compound, 1,3,7-trimethylxanthine, or more commonly known as "caffeine" was originally coined by chemist F.F. Runge from the German word "kaffee" which directly translates to coffee (Tilling, 2001). This particular chemical is deteriorating as the degree of roasting is increased (Lokker, 2017), thus making it a good indicator of coffee quality. Caffeine content on the cv. Liberica has lower values compared to the cv. Arabica and cv. Canephora (Ling et al., 2000, Anthony et al. 1993, as cited by Amidou et al. 2007).

In the context of quality assurance of roasted coffee, near-infrared spectroscopy (NIRS) is a fast, simple and cheap non-destructive method that serves as an alternative tool for qualitative and quantitative analysis in different types of food. NIRS has been used to determine caffeine content and roasting color (Pizarro et al. 2007, Zhang et al, 2013, Ayu, Budiastra, & Rindang, 2020), sucrose (Santos et al, 2016), pH, and acidity (Araújo et al., 2020), trigonelline and chlorogenic acid (CGA) (Budiastra, 2020), blend ratio (Bertone et al., 2015) for both cv. Arabica and cv. Canephora samples. In-line monitoring of the roasting process has also been done with the use of this technique (Catelani, 2018). NIRS, paired with UV-VIS, was also used to discriminate green coffee bean species according to their caffeine content and amount of CGA (Adnan et al., 2020). Characterization of green coffee beans (GCB), whole roasted coffee bean, and ground coffee bean was also carried out using NIRS for real-time assessment of coffee matrices (Tugnolo et al., 2019).

Partial least squares (PLS) and variable selection and MLR can be used to relate matrix X to a vector y or a matrix Y. It is referred to as the projection of latent structures through pmLartial least squares. The statistical procedure is insensitive to collinear variables and can accept a large number of variables, such as NIR spectra. The resulting regression model predicts a property y from the original dependent variables (Bokobza, 1998 and Varmuza & Filzmoser, 2009).

MATERIALS AND METHODS

Coffee Processing and Sampling

Fresh coffee berries belonging to the cv. Liberica were purchased from two farms located in the mountainous lands bordering Tagaytay and Laguna. The first farm (Farm A) was located in Barangay Mabato, Calamba City, Laguna, and the second farm (Farm B) was located in Sitio Balagbag-Araw, Barangay Canlubang, Calamba City, Laguna. The wet processing method as suggested by Clarke & Macrae (1987) was employed to obtain the GCB starting with soaking the berries in water for 24 hours. The floaters were removed after this procedure. The electric coffee bean depulper machine (Model Number: VOS150, Zhengzhou VOS Machinery Equipment, Zhengzhou, China) was used to separate the pulp from the parchment. Then, the parchments were sun-dried from 7-10 days or until a maximum of 11% moisture content is reached. The dried parchments were processed with a dry coffee hulling machine (Model number: LG- QLG, Zhengzhou Longer Machinery, Zhengzhou, China) to obtain the green beans inside. Defective and infested green beans were manually removed from the samples. It was estimated that only 10% of the initial fresh weight of the berries yielded the GCB. The green beans were stored in a mason jar until the roasting process. Each green bean sample set weighed 100 g and a total of 11 sample sets were prepared for each harvest date and each farm. Ten sample sets were roasted and the remaining ones were the raw sample set.

Coffee Roasting

Results of the roasting profiling showed that slow and fast roasting can be achieved using temperatures 200 °C and 220 °C respectively. The roasting times used were 4, 7, 10, 13, and 16 mins for the fast roasting and 6, 12, 18, 24, and 30 mins for the slow roasting. Coffee roasting profiling was done by having a spoonful of the sample taken

every minute from the roasting chamber until the coffee beans were darkly roasted to nearly burnt. Agtron Gourmet Color Scale by SCAA (Specialty Coffee Association of America) was used to identify the degree of roasting from each of the samples with the help of a coffee roasting expert, Mr. Ronald Peña. The time and pre-set temperature at which the beans achieve Light, Medium, Medium -Dark, and Dark Roast were recorded and were used for the actual roasting operations. One hundred twenty grams (120 g) of GCB were weighed, 100 g of it were loaded into the coffee roaster once the temperature reading hit its set roasting temperature. The remaining 20 g, which was contained in a ziplock bag and stored in an empty mason jar, was used later for determining the initial GCB moisture content before roasting. The temperature of roasting was recorded every minute until the end of the roasting period. To facilitate smaller deviations in the set temperature, the air intake was changed accordingly. As observed, restricting air intake allowed a faster rise in drum temperature while opening the air intake allowed a faster falling rate. The roasted coffee was transferred immediately into a mason jar and was immediately put into an ice bath to halt the roasting and bring the internal temperature of the bean to room temperature. Once the bean temperature stabilized, the mason jar was wiped dry and stored in a styrofoam icebox to maintain a constant ambient temperature. This was repeated twice and the roasted coffee for the two trials was combined in a single mason jar. Roasting schedules were staggered every week to prevent long periods of storage before testing that may affect the parameter readings, such as the moisture content.

Moisture Content Determination

The moisture content determination was based on the routine method as provided by AOAC 979.11.1.2. Five grams (5 g) of ground coffee sample were weighed and was dried in a PEAK Carbolite natural convection oven (Derbyshire, United Kingdom) was set at 100 ± 2 ℃ for 5 to 6 hours. It was cooled down in a desiccator before weighing. It was dried again for 30 mins and again cooled in a desiccator. The process of heating and cooling was repeated until the difference in two successive weighings was less than 1 mg. The lowest reading was the final dry weight. The moisture content determination was carried out in triplicates.

Caffeine Content Determination

The reference method for caffeine content determination was adopted from the standards for reporting the determination of caffeine content using high-performance liquid chromatography by the Bureau of Indian Standards (BIS, 2012) and by the International Organization for Standardization (ISO, 2008). Half a gram (0.5 g) of ground coffee and 2.5 g of magnesium oxide (MgO, light grade) were weighed and placed in a 50 mL Erlenmeyer flask. The powder mixture was dissolved by adding at least 25 mL of distilled water and was stirred for 30 s. The solution was left to boil in a water bath mixture maintained at 90 ℃ with occasional stirring for 20 mins. The solution was immediately filtered using a gravity filtration setup with a Whatman paper filter #1. Additional hot distilled water (~80 ℃) was used to completely wash the remaining solids from the Erlenmeyer flask. The filtrate was transferred in a 50 mL volumetric flask and was filled with hot distilled water to the mark. The filtrate was allowed to cool down to room temperature. Three milliliters (3 mL) of the filtrate was filtered again with a syringe filter with 0.45 µm pores and was transferred in a vial before HPLC injection. The secondary filtration was done to ensure there were no solid particles in the liquid that might cause the HPLC column to clog. The sample preparation for the caffeine extract was done in duplicates. The caffeine extraction process was summarized in Figure 1.

Samples were analyzed using a Shimadzu Prominence UFLC system with a DAD detector. The separation was done using a Hypersil gold reverse phase C18 column. A 60% volume fraction methanol in water solution was used in the mobile phase with a flow rate of 1mL/min. Before testing, the mobile phase and corresponding pressure were allowed to stabilize. The UV detector was set to 273 nm, which was found to be the highest peak of the standard caffeine solution. A glass microliter syringe was used to inject 10 µL of the standard

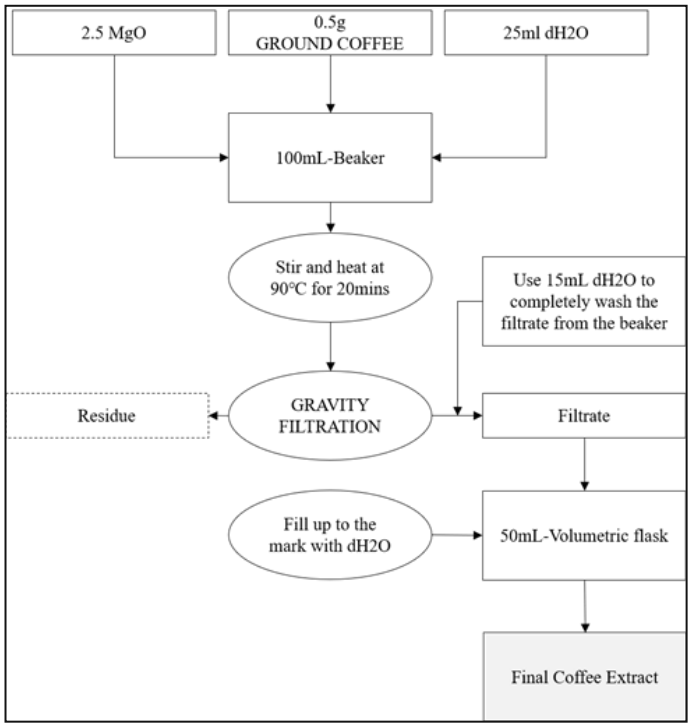

Figure 1. Caffeine extraction procedure from ground coffee samples

solution. Each run lasted 15 mins, with the caffeine peak starting around 6.8 mins and the highest peak appearing at 7.2 mins. After the injection of all the concentrations of the standard solution, it was followed by an equal volume of the sample solution. Each replicate of the sample solution was injected thrice, with a total of 6 injections for each sample. In between injections, a glass microliter syringe was washed with 60% methanol and distilled water to ensure there were no residues left in the needle and inside of the barrel. A clean run was also made by running the HPLC without injecting any solution to allow the mobile phase to remove any residues that may have been left in the column, which was also lasted 15mins. A clean run was made for every three runs of standard/sample solution. The area at the start to the end of the caffeine peak was used to calculate the caffeine content, given the Equation 1:

$$\%Caff = \frac{conc_{caff} \times V_{ce}}{w_{rc} \times \frac{1-mc_{rc}}{1-mc_{gcb}}} \times 100\%$$

Equation 1

where,

$$conc_{caff} = \frac{A-b}{m}$$

$\%Caff$ is the caffeine content, in mg/mg %, per dry matter of coffee

$conc_{caff}$ is the concentration, in mg/L, of caffeine in the coffee extract;

A is the area, in arbitrary units, of the HPLC caffeine peak of the sample coffee extract solution;

b is the y-intercept, in arbitrary units, of the standard curve generated from HPLC caffeine peak of the standard solution;

m is the slope, in arbitrary units over concentration, of the standard curve generated from HPLC caffeine peak of the standard solution

V_{ce} is the volume, in L, of sample coffee extract;

w_{rc} is the mass, in mg, of ground roasted coffee beans;

mc_{rc} is the moisture content, in decimal dry basis, of the roasted coffee beans

mc_{gcb} is the moisture content, in decimal dry basis, of the GCB

NIR Spectral Acquisition

The NIR measurement setup was composed of the following: NIR Quest 512 version 1.7 spectrometer equipped with a Hamamatsu G9204-512 InGaAs linear array as a detector with an operating range of 900-1700 nm, Tungsten Halogen HL-2000 light source, uninterrupted power supply (UPS), a spectral reflectance standard and a fiber optic reflectance probe. The NIR Quest 512 was connected to a laptop with SpectraSuite software installed. SpectraSuite® v.2.0 software was used to acquire reflection spectra. Before spectral collection, the following acquisition parameters were recorded from SpectraSuite® v.2.0: integration time (8ms), Boxcar width (30), and the number of scans-to-average A reference spectrum was obtained by scanning a white standard while the dark spectrum was obtained by blocking the light path of the spectrometer.

The Petri dishes containing ground coffee samples that were used previously in color determination were subjected to NIR scanning. The surface of the petri dish was divided into four quadrants,

representing four areas where spectral data was gathered using the spectrometer. The fiber optic probe was attached to an adaptor, with a 3mm clearance from the bottom surface, in a vertical alignment. The probe setup for the samples was visualized in Figure 2.

Statistical Analysis

Two-way analysis of variance (ANOVA) and Tukey's HSD test at a 5% level of significance was used to analyze the data to determine the significant differences in the moisture content and caffeine content across roasting times and roasting temperatures. For caffeine content, the values were also analyzed across roasting temperatures and farm sources. These statistical procedures were performed using IBM SPSS Statistics v.25.0 (USA).

NIR spectral data were analyzed using the chemometrics software, ParLeS version 3.1 which is capable of performing PLSR with leave-1-out cross-validation, PLSR modeling, and prediction (Viscarra-Rossel, 2008). Several pre-processing combinations were applied to NIR spectra and were evaluated using the PLSR Cross-Validation procedure. Two of the four statistical parameters that the PLSR Model function yielded were used to identify significant wavelengths relative to the selected parameters: regression coefficient (RC) and the variable importance of projection (VIP). A selection threshold of 0.7 was used on the normalized values of RC and VIP to narrow down the significant wavelengths impacting the PLSR model. The identified significant wavelengths were later used for multiple linear regression.

Multiple linear regression was performed on the reflectance values of the selected wavelengths from the PLSR modeling and the reference values of the moisture content, and caffeine content. The stepwise linear regression method was applied to regress multiple wavelengths while simultaneously removing those that are identified as not significant. The probability of the F-statistic was used to evaluate whether a wavelength was removed ($p>0.1$) or entered ($p<0.05$) in the model

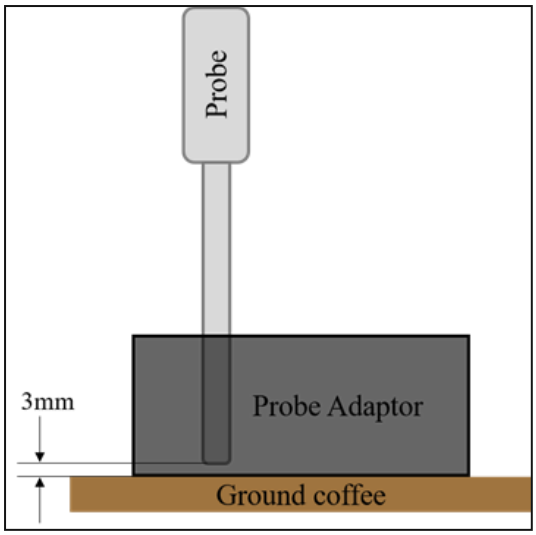

Figure 2. Probe setup for spectral acquisition with respect to ground coffee samples.

(Van den Berg, 2017). Wavelengths/variables were also analyzed for redundancy and were removed automatically by the statistical software. The spectra were preprocessed with mean centering only for all the parameters. The predictive performance of the MLR models was evaluated using the same statistics from PLSR models: R^2, R^2_{adj}, RMSE, ME, SDE, and RPD. The procedure was performed using IBM SPSS Statistics v.25.0. The performance of the PLSR and MLR models were assessed with their resulting RPD and R^2 values as summarized by Williams (2001) in Tables 1 and 2.

RESULTS AND DISCUSSION

Coffee Roasting

The drum temperature was recorded every minute during the whole duration of the roasting process as shown in Figure 3. Loading of the GCB was done once the set temperature was reached: 220°C for fast roasting and 200°C for slow roasting. A sudden drop in temperature reading was observed until the 2-minute mark, then it gradually rose again until it reached its respective set temperature in the 5-minute mark. From there, the temperature was seen fluctuating in a sinusoidal manner until the end of the roasting process. The first five

minutes of the roasting were characterized by rapid loss of moisture in the form of steam exiting from the roasting chamber. This was the drying stage of roasting, where the surface temperature of the bean reaches 100°C, and surface moisture was easily removed (Yeretzian et al, 2002). The first crack was observed uniformly for all the roasting trials during the 4 to 5-minute mark, indicating that the volume expansion of the beans had started already. Moisture in the cells turns from liquid to vapor causing high-pressure buildup. This high pressure, in combination with the CO_2 production inside, causes swelling once it exceeds the mechanical resistance of the bean. This expansion produces the first cracking sound (Fadai et al., 2017). However, the second crack was not observed for the slow roasting process, despite the prolonged roasting time, unlike with the fast roasting which exhibited the second crack as early as its 10-minute mark. Staub (1995), as mentioned by Songer (2012), indicated that the bean must reach an internal temperature of at least 230 °C for the second crack to occur, in which cell walls of the bean starts fracturing due to heating, hence the absence of the second crack in the slow roasting where the temperature is maintained only at 200°C. Kelly & Scott (2014) also attributed the second crack to gaseous build-up caused by pyrolytic reactions, a period when beans start to turn brown and sugar caramelization is occurring.

Table 1. Guidelines for the interpretation of coefficients of determination by Williams (2001).

R	R^2	INTERPRETATION
± 0.5	<0.25	Not usable in calibration
± 0.51 – 0.70	0.26 – 0.49	Poor correlation, needs further research to identify cause
± 0.71 – 0.80	0.50 – 0.64	Usable for rough screening
± 0.81 – 0.90	0.66 – 0.81	Suitable for screening and other approximate calibrations
± 0.91 – 0.95	0.83 – 0.90	Can be used with caution in most applications including research
± 0.96 – 0.98	0.92 – 0.96	Can be used in most applications, including quality assurance
± 0.99 or higher	Higher than 0.98	Excellent, can be used in any application

Table 2. Guidelines for interpreting RPD (Williams, 2001)

RPD	CLASSIFICATION	APPLICATION
0.0 – 2.3	Very poor	Not recommended for use
2.4 – 3.0	Poor	Very rough screening
3.1 – 4.9	Fair	Screening
5.0 – 6.4	Good	Quality Control
6.5 – 8.0	Very good	Process control
8.1 or higher	Excellent	Any application

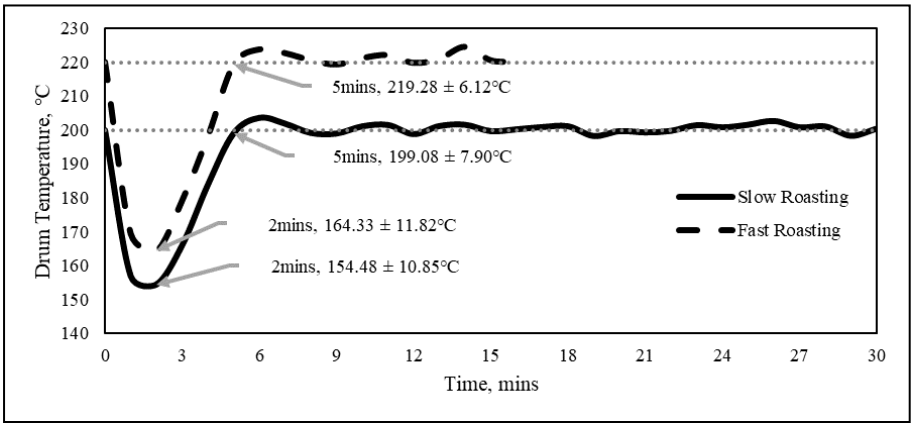

Figure 3. Mean drum temperature during slow and fast roasting.

Moisture Content

The moisture content of roasted coffee beans across different roasting temperatures and times are summarized in Table 3. Though the exact

temperature and time settings were set, there were still discrepancies in moisture content values for calibration and validation data sets. Environmental factors such as room temperature and relative humidity and the surface moisture content of the beans are some of the factors that could have caused the differences in moisture content values. Despite these discrepancies, both calibration and validation data sets showed an inverse correlation between roasting time and moisture content.

The mean moisture content (± S.D) of the GCB before roasting was 10.1 ± 0.7% (n=33) for samples from Farm A and 9.4 ± 0.45% (n=33) for

Table 3. Moisture content and caffeine content of roasted coffee beans across different roasting temperatures and times.

TEMPERATURE, (°C)	TIME (mins)	MOISTURE CONTENT, (d.b. %)		CAFFEINE CONTENT (g/g %)	
		calibration	validation	calibration	validation
25	0	9.00±0.15^{a}	10.36±0.05^{a}	0.79±0.05^{a}	0.82±0.02^{abc}
	6	3.85±0.14^{b}	2.68±0.04^{c}	0.89±0.12^{bcd}	0.84±0.02^{bc}
	12	2.01±0.56^{c}	1.10±0.05^{e}	0.87±0.07^{bc}	0.87±0.02^{bcde}
200	18	1.88±0.45^{c}	0.84±0.04^{fg}	0.81±0.11^{a}	0.75±0.01^{a}
	24	1.63±0.23^{d}	0.70±0.04^{hg}	0.85±0.09^{ab}	0.89±0.02^{cde}
	30	1.06±0.60^{e}	0.38±0.05^{i}	0.83±0.09^{ab}	0.89±0.02^{cde}
	4	5.87±0.15^{f}	4.18±0.04^{b}	0.92±0.06^{cde}	0.93±0.02^{e}
	7	2.31±0.22^{g}	1.43±0.04^{d}	0.84±0.10^{ab}	0.84±0.01^{bc}
220	10	1.53±0.57^{d}	0.95±0.05^{ef}	0.95±0.08^{e}	0.86±0.02^{bcd}
	13	1.10±0.42^{e}	0.86±0.04^{fg}	0.94±0.07^{de}	0.82±0.02^{ab}
	16	0.97±0.32^{e}	0.56±0.05^{hi}	0.95±0.05^{e}	0.93±0.02^{de}

* Values in the same column followed by different letters (a–i) differ significantly at $p < 0.05$ level.
* Values represent the mean ± SD

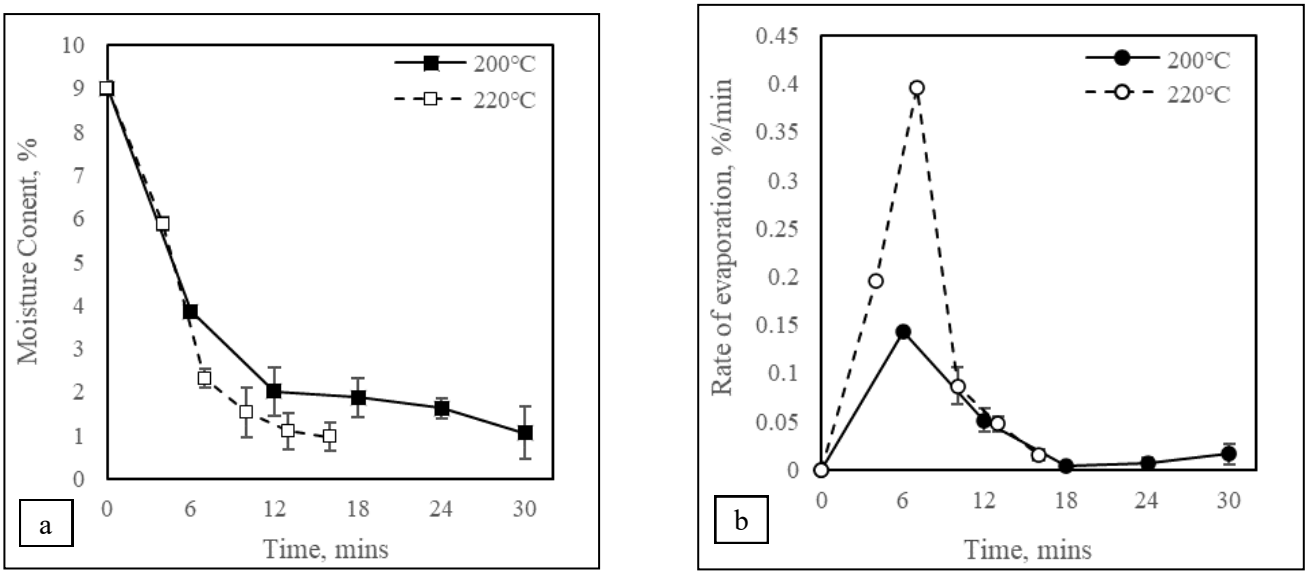

Figure 4. Time course of moisture content (a) and rate of evaporation (b) according to roasting temperature.

Farm B. The drying rate for the fast and slow-roasted coffees was almost equal during the drying stage as shown in Figure 4. Moisture loss decreased for the slow roasting, taking 30 minutes to reach 1.05% unlike for fast roasting that measured 0.96% moisture content after only 16 minutes. Evaporation of moisture is dependent on temperature; hence these results are also expected to follow this principle. The peak rate of evaporation was also higher for fast roasting compared to slow roasting.

Caffeine Content

Caffeine content determination was obtained using high-performance liquid chromatography (HPLC). Every time coffee sample extracts were tested, standard solutions at varying concentrations were injected in the HPLC to obtain a standard curve (Figure 5). These were used to interpolate the caffeine concentration of the coffee sample extract. The coefficient of determination for the standard curve generated was 0.9919, indicating a very high linear correlation between the area under the curve and the caffeine concentration.

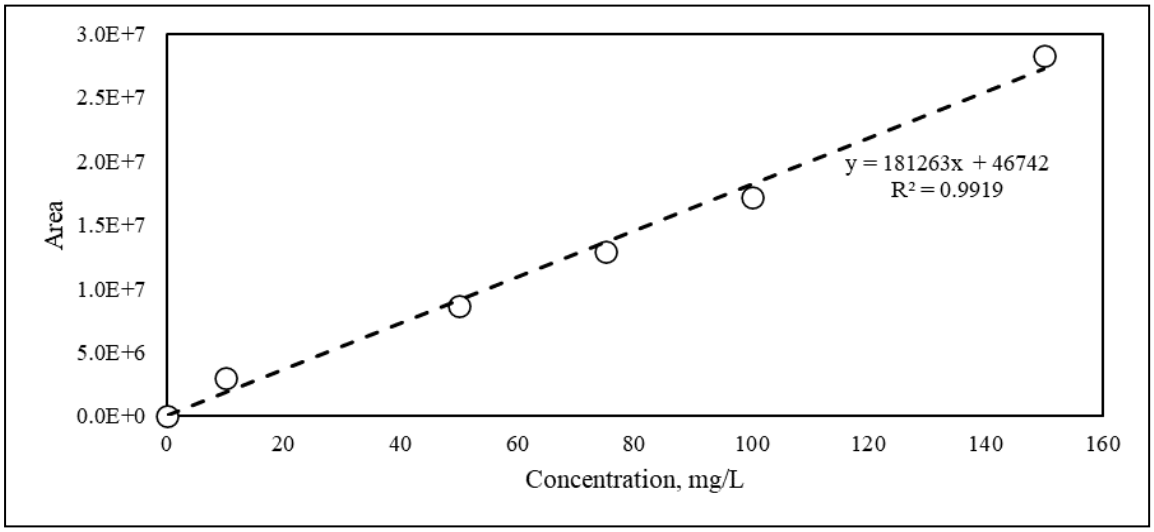

Figure 5. The standard curve generated using mean areas under the curve of the HPLC curve at 273nm and caffeine concentration.

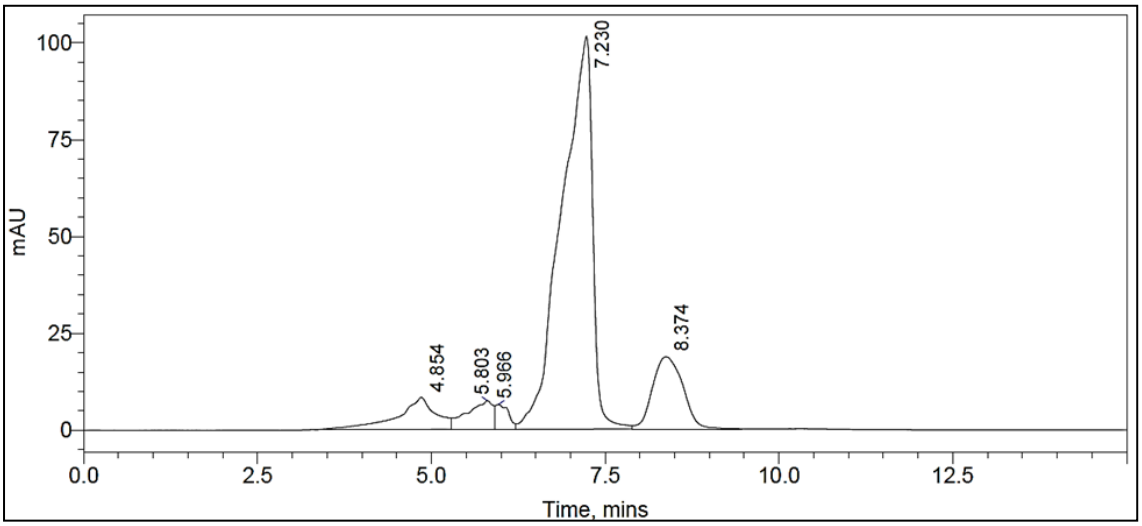

Figure 6. HPLC chromatogram for caffeine analysis of coffee extract samples.

The mean values for caffeine content across the calibration and validation data sets were also shown in Table 3. Caffeine content is also not consistent despite the same time and temperature for the two data sets. But unlike moisture content, caffeine content showed no pattern or correlation as roasting time is increased for both calibration and validation data sets.

The retention time of caffeine can be seen at the peak of the chromatogram which is 7.23 mins as illustrated in Figure 6. The area under the largest peak was used to calculate the caffeine concentration in Equation 1. The other small peaks adjacent to the caffeine are other coffee compounds that are also visible at 273 nm.

The summary of the caffeine content of coffee across different times, temperatures, and farm sources is shown in Figure 7. For both slow and fast roasting (Figure 7a), there was no clear indication of the correlation between the time and the caffeine content. However, when the means were grouped according to roasting temperature and farm source (Figure 7b), it was clear that the coffee coming from Farm A has a significantly higher caffeine content compared to Farm B with Sig < 0.05 for both Farm-Treatment and Farm-Temperature interactions. Fast-roasted coffee also obtained higher caffeine content compared to slow -roasted coffee with Sig < 0.05 for Farm-Temperature interaction. The caffeine content in GCB is affected by several factors such as genetic

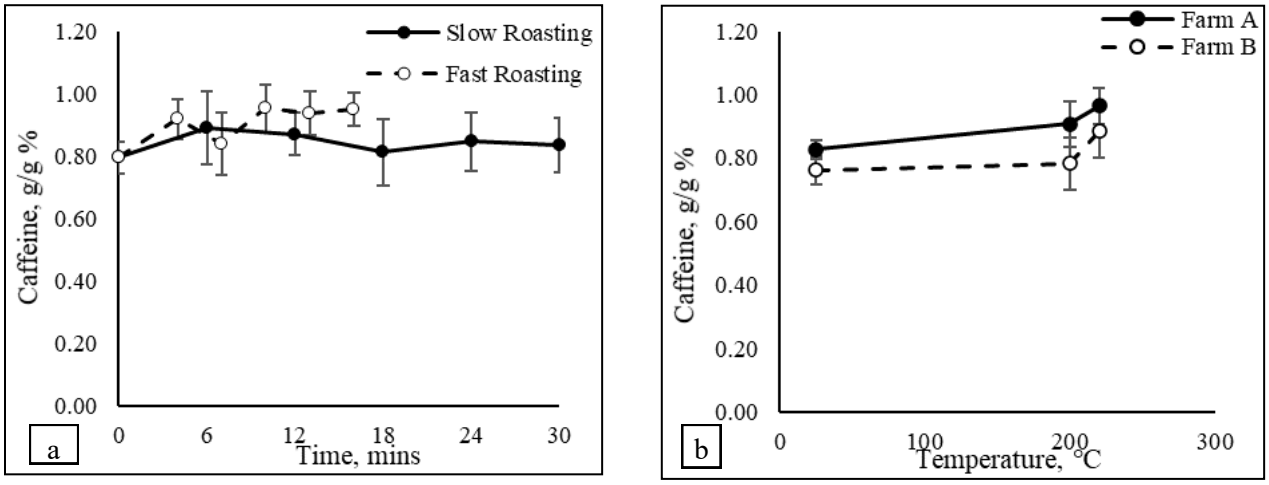

Figure 7. The caffeine content of coffee sample extracts relative to (a) time, and (b) temperature.

Table 4. Prediction statistics of the PLSR cross-validation models for moisture content and caffeine content of the roasted coffee beans.

PARAMETER	PRE-PROCESSING	CALIBRATION				VALIDATION		
		No. of PLSR Factors	R^2	RMSECV	RPD	R^2	RMSEP	RPD
Moisture Content	MC-MSC-MF-1D	9	0.970	0.0034	5.82	0.853	0.010	2.53
	MC-MSC-1D	9	0.968	0.0035	5.60	0.815	0.012	2.104
	MC-MSC	8	0.937	0.0049	4.00	0.837	0.012	2.090
Caffeine Content	MC-WD-MF-1D	17	0.755	0.0004	2.00	0.136	0.003	0.22
	MC-WD-1D	19	0.741	0.0004	1.95	0.160	0.001	0.556
	MC-WD	19	0.732	0.0004	1.92	0.153	0.001	1.042

diversity (Dessalegn, et al., 2008), geographical topology and climate (including rainfall, irrigation, and temperature) (Hameed, et al., 2018), and even agricultural practices. Coffees from Farm A and Farm B are classified as cv. Liberica varieties, but compounds such as caffeine can still vary widely due to the mentioned differences. Similar results were reported by (Lang, et al., 2013) as they have found that caffeine content was found to be significantly larger for coffee roasted at higher temperatures. The variability of caffeine content in the roasted coffee was also observed by having a large standard deviation

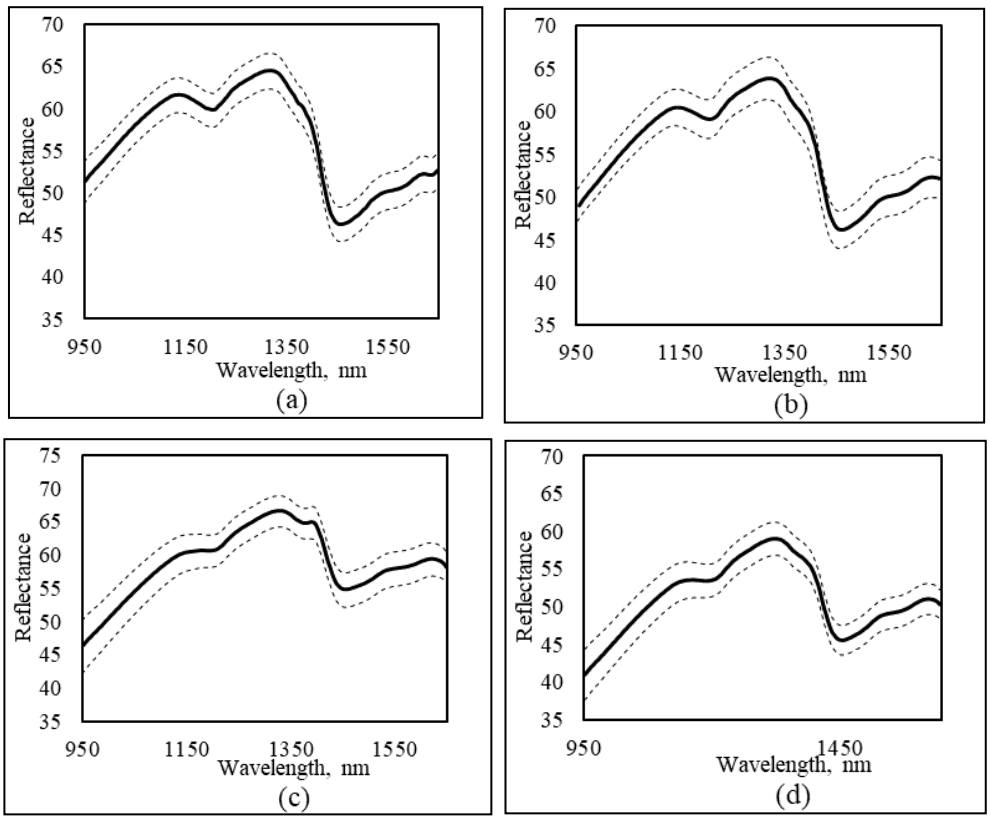

Figure 8. NIR spectra of ground roasted coffee beans from the calibration data set at 95% confidence interval from (a) Farm A – slow roasting (b) Farm B – slow roasting (c) Farm A – fast roasting; and (d) Farm B – fast roasting.

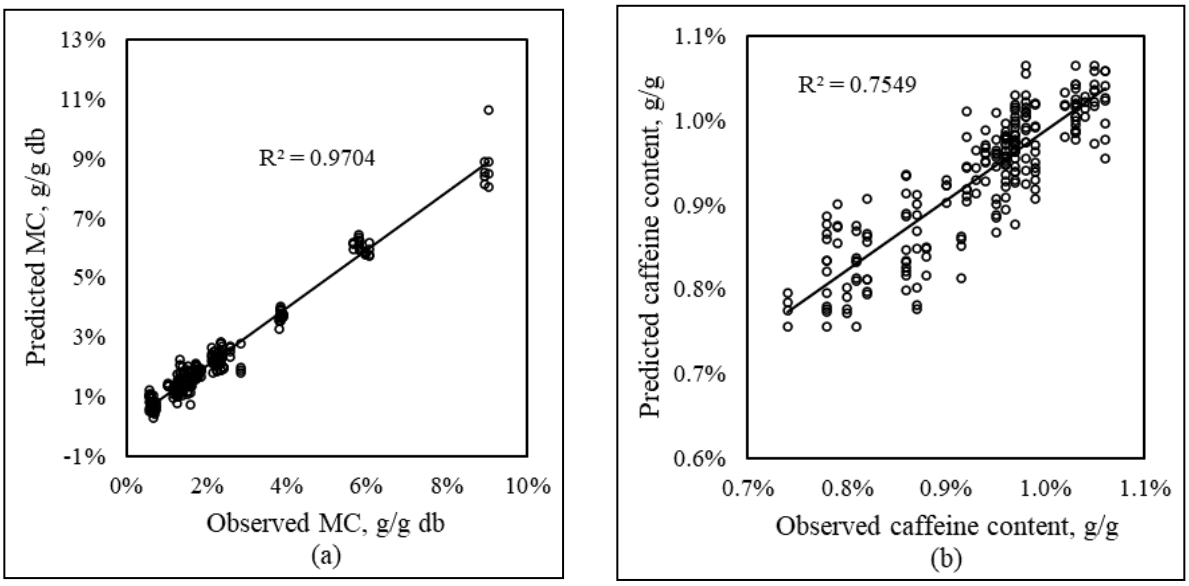

Figure 9. PLSR cross-validation model for (a) moisture content and (b) caffeine content using calibration data.

as the values converge in a range of 0.65 to 1.00 g/g % only.

Near-Infrared Spectra Analysis

NIR raw spectra of the ground roasted coffee beans from the calibration data set at 95% confidence interval are plotted in Figure 8. Each spectrum had its size reduced to half by increasing the wavelength increment to 3.3 nm and by rounding off the wavelength values to the nearest integer. The reflectance spectra were plotted from 950nm up to 1648nm. The spectra were grouped according to farm source and roasting temperature to identify if there were any notable differences when it comes to their recorded peaks and troughs.

Peak reflectance of the spectra was recorded at wavelength 1315-1322 nm (4.8a), 1312-1319 nm (4.8b), 1325-1332 nm (4.8c), and 1322-1328 nm (4.8d). Another notable depression in the spectra occurred at 1453-1460 nm [(4.8a) and (4.8c-d)] and 1450-1457 nm (4.8b). A smaller depression was also observed from 1199-1206 nm (4.8a-d). Based on these similarities, the NIR spectra for coffees sourced from both farms and roasted at different temperatures were almost the same, only the reflectance values were varied.

Reflectance at 1300-1350 nm is mainly attributed to the 1st overtone combinations of C–H stretch of methyl group, whereas 1450-1460 nm is for the 1st overtone of O–H and N–H of primary amides, 1200-1210 nm 2nd overtone of C-H stretch of methylene group (Otto et al., 2008 and Workman & Weyer, 2008).

Partial Least Squares Regression

The prediction statistics of the PLSR cross-validation models at different pre-processing combination was summarized in Table 4. The preprocessing combination used for moisture content was MC-MSC-MF-1D utilizing 9 PLSR factors, which also obtained the highest R^2 and RPD for both calibration and validation samples. Based on the R^2, the chosen model for calibration data can be used for most applications such as quality assurance (R^2: 0.92-0.96) whereas for the validation data it can only be used for research applications (R^2: 0.83-0.90). Despite high R^2, the model has a relatively poorer RPD value, limiting its recommended usage to quality control (calibration) and very rough screening (validation). The caffeine content model used four preprocessing methods (MC-WD-MF-1D) with 17 PLSR factors. The largest coefficients of

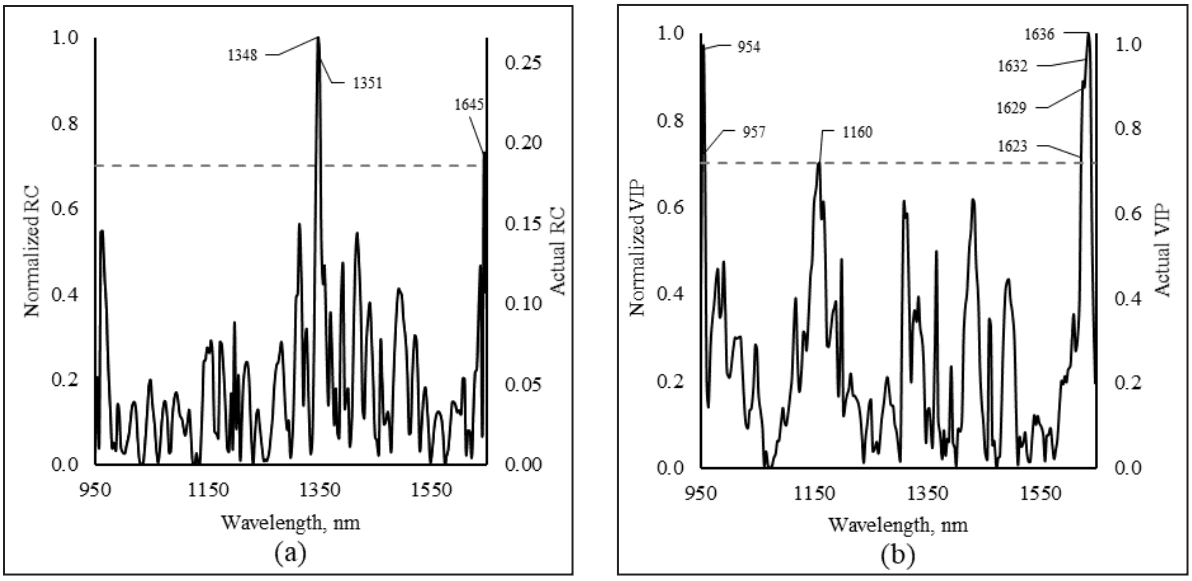

Figure 10. The absolute value of the (a) normalized regression coefficients plot and (b) normalized variable importance in projection generated from the PLSR model of the pre-processed spectral data and the moisture content.

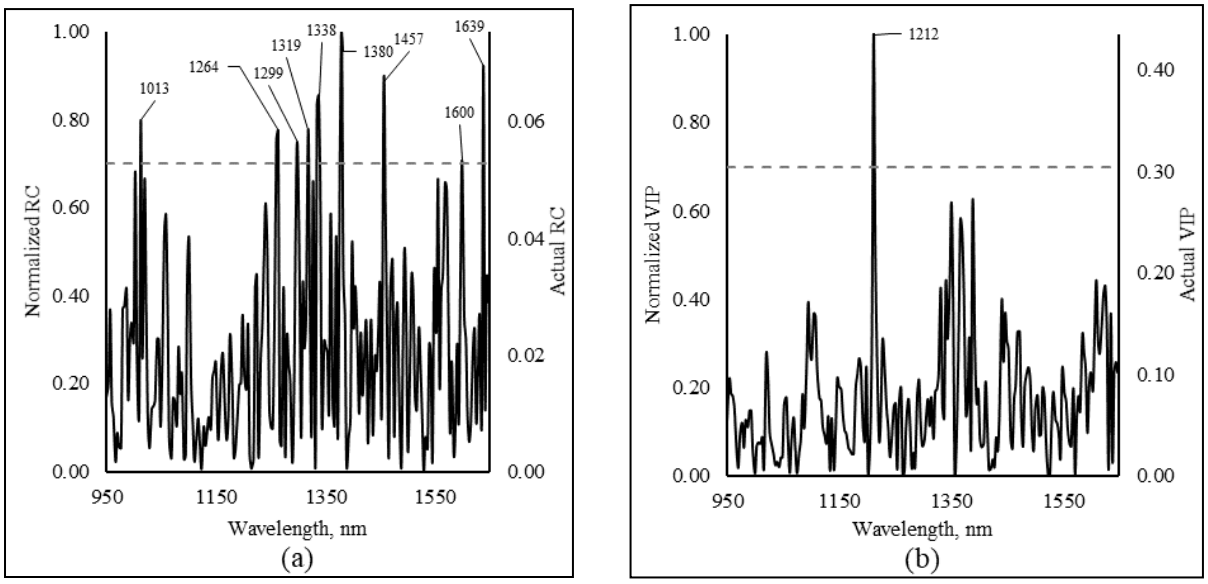

Figure 11. The absolute value of the (a) normalized regression coefficients plot and (b) normalized variable importance in projection generated from the PLSR model of the preprocessed spectral data and the caffeine content.

determination obtained from the three preprocessing combinations were R^2 = 0.755 (calibration) and 0.22 (validation). This is mainly because of the wide variability of the caffeine content as discussed in the earlier section. The lack of correlation between roasting time and caffeine content has caused the failure of the model to produce robust predictive statistics. The low RPD values for both calibration and validation data indicate that the caffeine content model is not recommended for any use. The predicted values are plotted against the observed values of the PLSR cross-validation model for both (a) moisture content and (b) caffeine content in Figure 9.

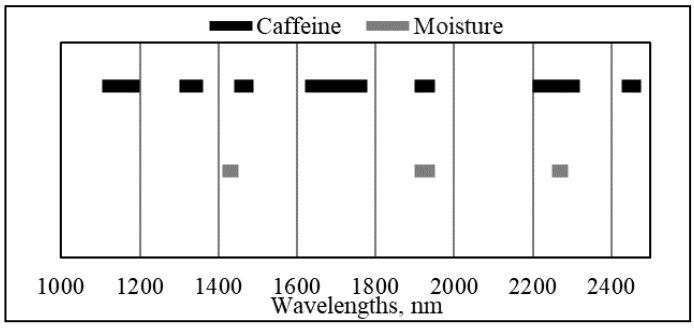

Figure 12. Absorption regions of caffeine and water found in coffee. Adapted from Ribeiro et al. (2010).

The normalized regression coefficient and variable importance of projection values of moisture content were plotted in Figure 10. For the normalized regression coefficient, there are only 3 wavelengths exceeding 0.7: 1348nm (0.27), 1351nm (0.25) and 1645nm (0.19). 1348-1351nm is attributed to the 1st overtone combination of C-H (CH_3). Davrieux et al. (2008) reported a PLS model to predict the moisture content of roasted coffee (cv. Arabica and cv. Canephora varieties). The reported wavelengths with large standard deviations of absorbance values are 1150nm and 1340nm, which they attributed to the first H-OH overtone absorption band (Otto et al., 2008 and Workman & Weyer, 2008). They mentioned that the high standard deviation values are attributed to the wide variation in spectral fingerprint due to water. An MLR model obtained by Adnan et al. (2017) in a similar experiment on moisture content prediction for GCB (cv. Arabica and cv. Canephora varieties) used the following wavelengths based on their weighted regression coefficient values: 1155, 1212, 1340, 1409, 1724, 1908, and 2249 nm. For the VIP of moisture content, the highest value was found at 1636 nm

with 1.03. Comparing to the previous parameters, the VIP value obtained for all wavelengths in moisture content was low. Nonetheless, sufficient information about the significant wavelengths was concluded since the VIP values were normalized. The selected wavelengths for moisture content based on VIP are as follows: 954-957, 1160, 1623-1639nm. The intensity at 954-957 nm was attributed to the 2nd overtone of O-H (H_2O, ROH, and ArOH), 1160nm to 2nd overtone of C-H (CH_3), and 1623-1639 nm to the 1st overtone of C-H (ArCH and CH_3) (Otto et al., 2008 and Workman & Weyer, 2008). In an experiment of moisture content prediction of peanuts, Kandala et al, (2008) used the following wavelengths in their MLR model: 1033, 1137, 1159, 1358, and 1393 nm. These are under the majority of the wavelengths selected for moisture content based on both the RC and the VIP, except those below 1000 nm and above 1600 nm. The wavelength bands 1160 nm, 1623-1639 nm, and 1645 nm are all located in NIR bands that do not correspond to water which contains O-H stretch and O-H bond vibrations (Luck, 1974 as mentioned by Büning-Pfaue, 2003). However, Reh et al. (2006) as mentioned by Adnan et al. (2017) have reported that coffee beans also lose 0.39% of their mass due to degradation reactions during the drying process; which is not accounted for as water. This means that the formula for

Table 5. Selected significant wavelengths based on partial least square regression (PLSR) models and their chemical assignments.

PARAMETER	WAVE-LENGTH	VIBRATION-AL MODE	CHARTS
Moisture Content	954-957	2nd overtone of O-H	H_2O, ROH, ArOH
	1160	2nd overtone of C-H	CH_3
	1348-1351	1st overtone combination of C-H	CH_3
	1623-1639, 1645	1st overtone of C-H	ArCH, CH_3
Caffeine Content	1013	2nd overtone of N-H	RNH2
	1212	2nd overtone of C-H	CH, CH_2
	1261-1264	2nd overtone region	n/a
	1299	2nd overtone region	n/a
	1319	2nd overtone region	n/a
	1335-1338	1st overtone combination of C-H	CH_3
	1377-1380	1st overtone combination of C-H	CH_3
	1457[b]	1st overtone combination of C-H, 1st over-tone of O-H	CH, H_2O, ROH, $CONH_2$, COHNR, RNH_2

Table 6. Model summary for the multiple linear regression of moisture content and selected wavelengths (selection threshold = 0.5) from the NIR spectra of roasted coffee.

MODEL	WAVE-LENGTHS, nm	R	R^2	R^2_{adj}	STD. ERROR OF THE ESTIMATE
0*	1623, 954, 1160, 1629	0.902	0.814	0.810	0.00852
1	1623	0.588	0.346	0.343	0.01586
2	1623, 954	0.839	0.703	0.700	0.01070
3	1623, 954, 963	0.922	0.850	0.848	0.00762
4	954, 963	0.921	0.849	0.847	0.00765
5	954, 963, 1354	0.923	0.851	0.849	0.00759

**Model was based on significant wavelengths at 0.7 selection threshold*

MC calculation does not exclusively pertain to moisture losses only, but also includes mass losses coming from other compounds. Thus, these wavelengths relating to vibrational modes of C-H can refer to the other compounds lost during the drying process.

Figure 11 summarizes the regression coefficients and variable importance in projection and the wavelengths plot for caffeine content. The highest regression coefficient was found at 1380nm with a value of 0.08. There was a total of 12 wavelengths with normalized regression coefficients exceeding 0.7: 1013, 1261-1264, 1299, 1319, 1335-1338, 1377-1380, 1457, 1600, and 1639nm. Among these, only 1013 and 1261-1264 were not included in the caffeine region as reported by Ribeiro et al. (2010) in their schematic representation of the absorption regions of main components in coffee (Figure 12). 1013nm is attributed to the 2nd overtone region of N-H (RNH_2) while 1261-1264nm falls under the 2nd overtone region only (Otto et al., 2008 and Workman & Weyer, 2008). For the variable importance in projection, the only wavelength selected was 1212 nm with a VIP of 0.04. Even lower than the VIPs obtained from the moisture content, the result is caused by the high variability of caffeine content observed from the roasted coffee across different time, temperature, farm source, and date of harvest. The intensity in 1212 nm is attributed to the 2nd overtone of C-H (CH and CH_2) (Otto et al., 2008 and Workman & Weyer, 2008). In a similar study conducted by Budiastra et al., (2018), they used the following wavelengths for the MLR model for caffeine prediction: 1128, 1298, 1672 nm plus several wavelengths which accommodate the scatter and intercorrelation effects between chemical compounds. The following wavelengths selected using the RC and VIP for caffeine content prediction were following the reported results of Budiastra et al., (2018): 1299nm and 1639nm.

The selected wavelengths based on the RC and VIP from the PLSR models are summarized with their respective chemical assignments in Table 5. A selection threshold equal to 0.7 for both RC and VIP yielded 4 wavelengths/wavelength ranges for moisture content and 8 wavelengths/wavelength ranges for caffeine content.

Multiple Linear Regression

The summary of the performance of the MLR model for moisture content and selected wavelengths from the NIR spectra of the roasted coffee is shown in Table 6. Initially, 12 wavelengths were identified as significant in the PLSR at the 0.7 selection threshold and were utilized for MLR regression and only 4 remained in the final model (Model 0: 1623, 954, 1160, and 1629nm). Model 0 has the highest R^2 of 0.814 and the smallest standard error of estimate of 0.0085. The total wavelengths used in the MLR model in Table 6 was 30 when the selection threshold was lowered to 0.5. Only 3 wavelengths were entered in the final model (Model 5) which had a higher R^2 (0.851) and lower standard error of estimate despite Model 0 using 4 wavelengths only.

Table 7 summarizes the performance of the MLR model of caffeine content and the selected wavelengths from the NIR spectra of roasted coffee. A total of 13 significant wavelengths was used in the MLR, 4 was entered in Model 6 and the rest were excluded as it was identified as insignificant. Model 6 obtained better performance in terms of R^2 (0.445) in comparison to Model 6 despite utilizing 1 less wavelength in the regression. To test if the model can be improved further, the selection threshold was lowered to 0.5 to increase the number of significant wavelengths that can be used in regression and the model summary was shown in Table 4.11b. Out of the initial 35 wavelengths, only 8 wavelengths were entered in Model 8 and obtained an insignificant improvement in terms of R^2 (0.471) and standard error of estimate (0.0062) in comparison to Model 0.

The stepwise multiple linear regression generated predicted values of the parameters using reflectance values of the selected wavelengths and the reference parameters. The predicted values of the moisture content and caffeine content were used to calculate the prediction statistics shown in Table 8. The moisture content model with R^2 =

0.851 and RPD = 2.6 was still acceptable but is only limited to research and rough screening applications. The caffeine content model had a poor performance with low R^2 = 0.471 and RPD = 1.378 despite having more wavelengths analyzed during regression. There was a poor correlation and the model needs more research to determine the cause, hence the model was not recommended for any use. The predicted values are plotted against the observed values of the MLR model for both (a) moisture content and (b) caffeine content in Figure 13.

The selected wavelengths based on the included wavelengths as selected by stepwise MLR are summarized with their respective chemical assignments in Table 9. Both MLR models used a 0.5 selection threshold and yielded 3 selected wavelengths for moisture content and 8 selected wavelengths for caffeine content. The 3 wavelengths for moisture content was also a subset of the selected wavelengths using the PLSR model. For the caffeine content, only 1572nm, 1639nm, and 1399nm are the only wavelengths that were not selected by the previous PLSR model.

SUMMARY AND CONCLUSION

Based on the R^2 (0.970), the PLSR model for moisture content can be used for quality assurance purposes whereas the RPD = 2.53 suggests that the model is only limited for rough screening purposes. For the PLSR model for the caffeine content, R^2 = 0.755 and indicates that it can only be used for screening and approximate calibrations but the RPD = 0.22 suggests that the model has no practical use yet and still needs further study to identify the cause.

MLR models were also developed using the significant wavelengths identified in PLSR. However, the performance of the models was almost similar to the PLSR models. The moisture

Table 7. Model summary for the multiple linear regression of caffeine content and selected wavelengths (selection threshold = 0.5) from the NIR spectra of roasted coffee.

MODEL	WAVELENGTHS, nm	R	R^2	R^2_{adj}	STD. ERROR OF THE ESTIMATE
0*	1418, 1409, 1389, 1386	0.667	0.445	0.429	0.00063
1	1457	0.236	0.056	0.051	0.00082
2	1457, 1572	0.454	0.206	0.199	0.00075
3	1457, 1572, 1639	0.473	0.224	0.213	0.00074
4	1457, 1572, 1639, 1215	0.602	0.363	0.351	0.00068
5	1457, 1572, 1639, 1215, 1338	0.655	0.429	0.415	0.00064
6	1457, 1572, 1639, 1215, 1338, 1013	0.665	0.443	0.427	0.00063
7	1457, 1572, 1639, 1215, 1338, 1013, 1399	0.676	0.457	0.439	0.00063
8	1457, 1572, 1639, 1215, 1338, 1013, 1399, 1377	0.686	0.471	0.450	0.00062

*Model was based on significant wavelengths at 0.7 selection threshold

Table 8. Regression statistics of the MLR model for moisture content, and caffeine content of the roasted coffee beans.

PARAMETER	PREDICTION STATISTICS							
	No. of Wavelengths excluded	included	R^2	Adj R^2	RMSEP	ME	SDE	RPD
Moisture Content	27	3	0.851	0.849	0.008	0.000	0.008	2.600
Caffeine Content	27	8	0.471	0.450	0.001	0.000	0.001	1.378

content model can be used for most applications including research based on the R^2 value (0.851), while the RPD (2.6) limits its usage for rough screening purposes only. For the caffeine content model, an improvement to the RPD with 1.378 was observed but is still not suggested for any practical use.

The findings indicate the potential use of NIR for the quality assessment using the moisture content of roasted coffee. This can be used as a quantitative indicator of roasting degree aside from the color of the beans during roasting. However, for the caffeine content, none of the chemometric techniques used were proven to be effective for determining the quality nor the roasting degree of

Table 9. Selected significant wavelengths based on stepwise multiple linear regression (MLR) models and their chemical assignments.

PARAME-TER	WAVELENGTH	VIBRATIONAL MODE	CHARTS
Moisture Content	954	2nd overtone of O-H	H2O, ROH, ArOH
	963	2nd overtone of O-H	H2O, ROH, ArOH
	1354	1st overtone combination of C-H	CH_3
CAFFEINE CONTENT	1457	1st overtone combination of C-H, 1st overtone of O-H	CH, H2O, ROH, $CONH_2$, COHNR, RNH_2
	1572	1st overtone region	n/a
	1639	1st overtone of C-H	ArCH, CH_3
	1215	2nd overtone of C-H	CH, CH_2
	1338	1st overtone combination of C-H	CH_3
	1013	2nd overtone of N-H	RNH_2
	1399	1st overtone combination of C-H	CH_3, CH2, ArOH
	1377	1st overtone combination of C-H	CH_3

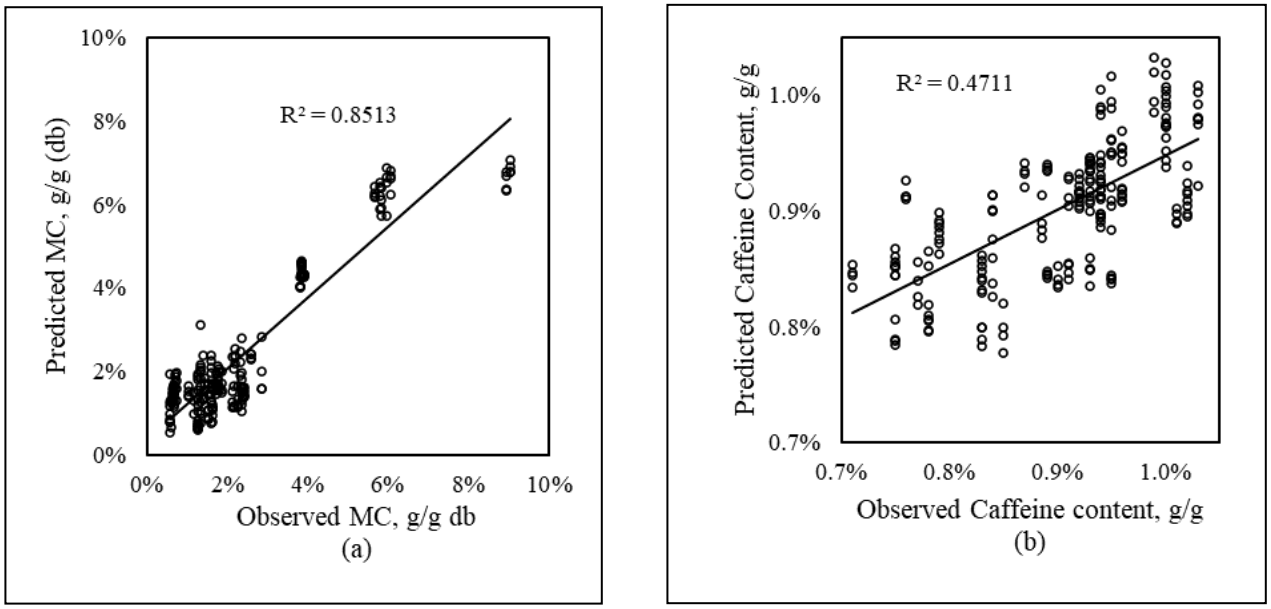

Figure 13. MLR model developed using calibration data for (a) moisture content and (b) caffeine content.

the roasted coffee beans. Roasting degrees as defined by the coffee industry are generally color-dependent, and the color tends to change easily with roasting time. Unlike color and moisture content, caffeine content does not follow a specific trend when plotted against roasting time. The development of a NIR model capitalizes on trends on the reference values of parameters as observed with the color parameters and the moisture content.

RECOMMENDATIONS

The study focused on determining moisture content and caffeine content on varying roasting degrees of cv. Liberica coffee using NIR spectroscopy. The performance of the majority of the models was satisfactory, but there were still limitations that call for improvement and further research. The following are the recommendations for further studies:

1. Increase of roasting temperature options to produce variation in roasting degree. Caffeine content was found to be dependent on the roasting temperature, hence this might lead to better-performing NIR models. This can also explain the conditions that can cause the maximum release of caffeine content in the coffee bean roasting.

2. Add more varieties and farm sources in the data set to improve sample composition for NIR models. The study attempted to eliminate bias caused by the use of a single species by having 2 different farm sources and 2 different harvest dates for each farm. Instead, one of the parameters, caffeine content, was found to be farm source dependent. Adding more varieties would increase the range of applicability of the models especially for caffeine content.

3. Obtain NIR data of roasted coffee samples that are being sold and used in the market with a clear identifier of roasting degree and variety to increase the robustness of the NIR caffeine model. This way, the information on how much caffeine content is present in a certain variety and roasting degree could add value to the coffee products.

ACKNOWLEDGMENT

The authors wish to acknowledge the funding support of the Philippine Department of Science and Technology – Engineering Research and Development for Technology (DOST-ERDT) scholarship program for the the Masteral thesis of the senior author. The authors also gratefully acknowledge the support provided by the Philippine Commission on Higher Education (CHED), Engr. Elizabeth B. Maghirang of the U.S. Department of Agriculture (USDA) – Stored Product Insect and Engineering Research Unit (SPIER) for the NIR instrumentation, and the technical assistance provided by Mr. Ronan Almeda and Dr. Milagros Peralta of the UPLB Nanotechnology Analytical and Instrumentation Service Laboratory for caffeine analysis. The expertise of Mr. Ronald Peña and Engr. Gerry Castillo, coffee experts of the National Coffee Research, Development and Extension Center of the Cavite State University (CvSU), and the use of the CvSU Coffee Processing Laboratory is also gratefully acknowledged.

LITERATURE CITED

ADNAN, A., VON HÖRSTEN, D., PAWELZIK, E., & MÖRLEIN, D. (2017). Rapid prediction of moisture content in intact green coffee beans using near infrared spectroscopy. *Foods*, *6*(38). doi: 10.3390/foods6050038

ADNAN, A., NAUMANN, M., MÖRLEIN, D., & PAWELZIK, E. (2020). 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. *Foods, 9*(6), 788. doi:10.3390/foods9060788

AMIDOU, N.D., MICHEL, N., SERGE, H., & VALÉRIE, P. (2007). Genetic basis of species differentiation between Coffea liberica Hiern and C. canephora Pierre: Analysis of an interspecific cross. *Genetic Resources and Crop Evolution,54*(5), 1011-1021. doi:10.1007/s10722-006-9195-0

ARAÚJO, C. D., MACEDO, L. L., VIMERCATI, W. C., FERREIRA, A., PREZOTTI, L. C., & SARAIVA, S. H. (2020). Determination of pH and acidity in green coffee using near-infrared spectroscopy and multivariate regression. *Journal of the Science of Food and Agriculture, 100* (6), 2488-2493. doi:10.1002/jsfa.10270

AYU, P. C., BUDIASTRA, I. W., & RINDANG, A. (2020). NIR spectroscopy application for determination caffeine content of Arabica green bean coffee. *IOP Conference Series: Earth and Environmental Science, 454*, 012049. doi:10.1088/1755-1315/454/1/012049

BERTONE, E., VENTURELLO, A., GIRAUDO, A., PELLEGRINO, G., & GEOBALDO, F. (2016). Simultaneous determination by NIR spectroscopy of the roasting degree and Arabica/Robusta ratio in roasted and ground coffee. *Food Control,59*, 683-689. doi:10.1016/j.foodcont.2015.06.055

BOKOBZA, L. (1998). Near Infrared Spectroscopy. Journal of Near Infrared Spectroscopy, 6(1), 3-17. doi:10.1255/jnirs.116

BUDIASTRA, I. W., SUTRISNO, S.M., WIDYOTOMO, S., & AYU, P. C. (2018). Prediction of Caffeine Content in Java Preanger Coffee Beans by NIR Spectroscopy Using PLS and MLR Method. *IOP Conference Series: Earth and Environmental Science*, *147*, 012004. doi: 10.1088/1755-1315/147/1/012004

BUDIASTRA, I. W., SUTRISNO, S.M., WIDYOTOMO, S., & AYU, P. C. (2020). Determination of trigonelline and chlorogenic acid (CGA) concentration in intact coffee beans by NIR spectroscopy. Agricultural Engineering International: The CIGR e-journal. 22. 162-168.

BÜNING-PFAUE, H. (2003). Analysis of water in food by near infrared spectroscopy. *Food Chemistry, 82*(1), 107-115. doi:10.1016/s0308-8146(02)00583-6

BUREAU OF INDIAN STANDARDS. (2012). Coffee and coffee products — Determination of the caffeine content using high performance liquid chromatography (HPLC) — Reference method (BIS Standard No. 16028). Retrieved from: https://www.services.bis.gov.in:8071/php/BIS/bisconnect/pow/ is_details?IDS=NTY3

CATELANI, T. A., SANTOS, J. R., PÁSCOA, R. N., PEZZA, L., PEZZA, H. R., & LOPES, J. A. (2018). Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. *Talanta, 179*, 292-299. doi:10.1016/j.talanta.2017.11.010

CLARKE, R. J., & MACRAE, R. (1987). Coffee (Vol. 2). Springer Netherlands. doi: 10.1007/978-94-009-3417-7

DAVRIEUX , F., BERKIA , A., RAVERDY , J. L., & EXCOFFON , S. (2008). pp 720-723. In *22nd International Conference on Coffee Science*. Retrieved from https://agritrop.cirad.fr/549606/

DESSALEGN, Y., LABUSCHAGNE, M. T., OSTHOFF, G., & HERSELMAN, L. (2008). Genetic diversity and correlation of bean caffeine content with cup quality and green bean physical characteristics in coffee (Coffea arabica L.). *Journal of the Science of Food and Agriculture*, *88*(10), 1726–1730. doi: 10.1002/jsfa.3271

FADAI, N. T., MELROSE, J., PLEASE, C. P., SCHULMAN, A., & GORDER, R. A. V. (2017). A heat and mass transfer study of coffee bean roasting. *International Journal of Heat and Mass Transfer*, *104*, 787–799. doi: 10.1016/ j.ijheatmasstransfer.2016.08.083

HAMEED, A., HUSSAIN, S. A., & SULERIA, H. A. R. (2018). "Coffee Bean-Related" Agroecological Factors Affecting the Coffee. *Bioactive Molecules in Food Reference Series in Phytochemistry*, 1–67. doi: 10.1007/978-3-319-76887-8_21-1

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. (2008). Coffee and coffee products — Determination of the caffeine content using high performance liquid chromatography (HPLC) — Reference method (ISO Standard No. 20481). Retrieved from: https://www.iso.org/standard/34185.html

KANDALA, C. V. K., NAGANATHAN, G. K., & SUBBIAH, J. (2008). NIR reflectance method to determine moisture content in food products. *Reflection, Scattering, and Diffraction from Surfaces*. doi: 10.1117/12.793267

KELLY, C. B. D., & SCOTT, J. B. (2014). Online Thermal Analysis of Batch Roasted Coffee Beans. Presented at the 21st Electronics New Zealand Conference (ENZCon), Hamilton, New Zealand, 20 - 21 Nov 2014, 2014.

LANG, R., YAGAR, E. F., WAHL, A., BEUSCH, A., DUNKEL, A., DIEMINGER, N., … HOFMANN, T. (2013). Quantitative studies on roast kinetics for bioactives in coffee. *Journal of Agricultural and Food Chemistry*, *61*(49), 12123–12128. doi: 10.1021/jf403846g

LING, L. S., NIK DAUD, N. I., & HASSAN, O. (2000). Determination of coffee content in coffee mixtures. *Malaysian Journal of Analytical Sciences*, *7*(2), 327–332. Retrieved from https://www.ukm.my/ mjas/v7_n2/7Liew.pdf

LOKKER, B. (2017, January 04). Coffee Roasts from Light to Dark. Retrieved from https://www.coffeecrossroads.com/coffee -101/coffee-roasts-from-light-to-dark

OTTO, T., SAUPE, R., WEISS, A., STOCK, V., WIESNER, K., LAMPE, U., FLEISCHER, M., GESSNER, T. (2008). Dual-detector optical MEMS spectrum analyzer: advances, applications, and prospects. *MOEMS and Miniaturized Systems VII*. doi: 10.1117/12.768356

PIZARRO, C., ESTEBAN-DÍEZ, I., GONZÁLEZ -SÁIZ, J., & FORINA, M. (2007). Use of near-infrared spectroscopy and feature selection techniques for predicting the caffeine content and roasting color in roasted coffees. *Journal of Agricultural and Food Chemistry,55*(18), 7477-7488. doi:10.1021/jf071139x

RIBEIRO, J. S., FERREIRA, M. M. C., & SALVA, T. J. G. (2010). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*, *83*(2011), 1352–1358. doi: 10.1016/j.talanta.2010.11.001

SANTOS, J. R., VIEGAS, O., PÁSCOA, R. N., FERREIRA, I. M., RANGEL, A. O., & LOPES, J. A. (2016). In-line monitoring of the coffee roasting process with near infrared spectroscopy: Measurement of sucrose and colour. *Food Chemistry,208*, 103-110. doi:10.1016/ j.foodchem.2016.03.114

SONGER, P. (2012). *A Thermodynamic Approach to Roasting Parameters and Stages of Roast*(pp. 1-3, Rep.). Songer and Associates, Inc.

TILLING, S. (2001). The Discovery and Isolation of Caffeine. Retrieved from http:// www.chm.bris.ac.uk/webprojects2001/ tilling/isolation.htm

TUGNOLO, A., BEGHI, R., GIOVENZANA, V., & GUIDETTI, R. (2019). Characterization of green, roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices. *Journal of Near Infrared Spectroscopy, 27*(1), 93-104. doi:10.1177/0967033519825665

VAN DEN BERG, R. G. (2017). *SPSS Stepwise Regression – Simple Tutorial*. SPSS tutorials SPSS Stepwise Regression Simple Tutorial Comments. https://www.spss-tutorials.com/stepwise-regression-in-spss-example/.

VARMUZA, K., & FILZMOSER, P. (2009). *Introduction to multivariate statistical analysis in chemometrics* (1st ed.). doi: 10.1201/9781420059496

VISCARRA ROSSEL, R. E. (2008). Chemometrics and Intelligent Laboratory Systems. *Chemometrics and Intelligent Laboratory Systems, 90*, 72-83. doi:10.1016/j.chemolab.2007.06.006

WILLIAMS, P. C., AND K. NORRIS. 2001. Implementation of near-infrared technology. In *Near-Infrared Technology in Agricultural and Food Industries*, eds. P. Williams, and K. Norris, 145–169. St. Paul, Minnesota: American Association of Cereal Chemists.

WORKMAN, J., & WEYER, L. (2008). *Practical guide to interpretive near-infrared spectroscopy*. Boca Raton: CRC Press.

YERETZIAN, C., JORDAN, A., BADOUD, R., & LINDINGER, W. (2002). From the green bean to the cup of coffee: Investigating coffee roasting by on-line monitoring of volatiles. *European Food Research and Technology,214*(2), 92-104. doi:10.1007/s00217-001-0424-7

ZHANG, X., LI, W., YIN, B., CHEN, W., KELLY, D. P., WANG, X., DU, Y. (2013). Improvement of near infrared spectroscopic (NIRS) analysis of caffeine in roasted Arabica coffee by variable selection method of stability competitive adaptive reweighted sampling (SCARS). *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,114*, 350-356. doi:10.1016/j.saa.2013.05.053 ■