Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

Erling Solberg

A Priori Spectroscopy of Coffee Roasts

Master's thesis in Electronic Systems Design and Innovation
Supervisor: Lise Lyngsnes Randeberg
June 2021

NTNU
Norwegian University of
Science and Technology

Erling Solberg

A Priori Spectroscopy of Coffee Roasts

Master's thesis in Electronic Systems Design and Innovation
Supervisor: Lise Lyngsnes Randeberg
June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU
Norwegian University of
Science and Technology

Abstract

Before brewing, coffee has to be roasted. Coffee roasters want as much control over the roasting process as possible. Today the coffee roasting process is mostly hand-tuned with little to no instrumentation. The chemical composition of coffee is crucial to its taste and smell. The concentration of important chemicals in coffee can be determined using hyperspectral imaging. We develop a roasting method to minimize variance in roast spectra and a hyperspectral imaging technique to minimize estimated roast spectrum variance. Roasting for a specific time gives a lower spectral variance than roasting until first crack, which is a common roasting technique. Based on our roasting method and randomized parameter search theory, we roast coffee to build a data set for prediction of roast spectra. We use this data set to train models that can predict the spectrum of coffee based on how it is roasted. Linear regression models can predict the roast spectra with $R^2 = 0.58 - 0.74$. These spectra can, in turn, be used to predict the chemical composition of the roast before roasting when combined with chemometric models.

Preface

This thesis is part of a masters project carried out by Erling Solberg at the Department of Electronic Systems at the Norwegian University of Science and Technology in colaboration with ROEST coffee AS. The project took place in spring 2021 and was supervised by Lise Lyngsnes Randeberg. Thanks to Trond Simonsen at ROEST coffee for provinding roasting equipment and supplies.

Erling Solberg
Trondheim, June 21, 2021

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Coffee is a product with great diversity, albeit good or bad. As with any product, producers want to make the product as good as possible. Today coffee roasting is a craft based on experience. To arrive at a desired roast, you have to experiment until you arrive at what you want, wasting both time and coffee. Predicting the properties of a roast before roasting would save some of these expenses. Spectroscopy is an established technique for monitoring roast quality and has already been used for chemometrics in green and roasted coffee. The next step in a priori chemometrics of coffee is to use the spectrum of green coffee to predict the effect of roasting on coffee roasts. It will enable a priori chemometrics of coffee roasts when combined with already existing chemometric models. Our first step towards a priori chemometrics of coffee roasts is reviewing the literature to get an overview of what chemical concentrations are measurable in coffee using spectroscopy and how coffee is best stored. Since we want to estimate the spectrum of coffee as accurately as possible, we want to keep the variance in the spectrum of each roast to a minimum. To achieve a low variance in the spectrum, we quantify the variance in the spectrum measured by hyperspectral imaging within and between roasts. Based on the variances, we choose an imaging and roasting method. Finally, we roast and image coffee based on our optimized technique using a ROEST sample roaster, shown in figure 1.2 and two hyperspectral cameras(SWIR320(962nm-2493nm) and VNIR1800(404nm-995nm), Norsk Elektro Optikk AS). A sample of beans captured with a VNIR camera through a microscope lens is shown in figure 1.1.

1.1 Background

The quality of coffee is affected by all steps in its processing and how it is treated in between. Coffee beans are actually seeds. When planted, they can grow into trees bearing fruit after three to four years. Since different places have different growth conditions, the origin of coffee affects its qualities. When the coffee cherries are deep red and ripe, they are ready for harvest. The cherries are either picked selectively or all at once. A larger variance in ripeness gives a larger variance in quality. After harvest, coffee is either first dried and then have its skin and pulp removed in what is called a *natural* process, or its skin and pulp is removed before drying in what is called a *washed* process. Finally, beans are sorted by size and put in jute bags or other containers before shipping to a supplier that distributes the coffee to roasters. To do a priori chemometrics of coffee both chemometrics and spectrum predictions are needed. Spectroscopy has been applied to coffee, measuring: caffeine (Nogales-Bueno et al. [2020],Huck et al. [2005]) theobromine [Huck et al., 2005], moisture [Davrieux et al., 2008], ash [Pizarro et al., 2004], sucrose [Santos et al., 2016], lipid [Pizarro et al., 2004], melanoid [Nogales-Bueno et al., 2020], phenolic [Nogales-Bueno

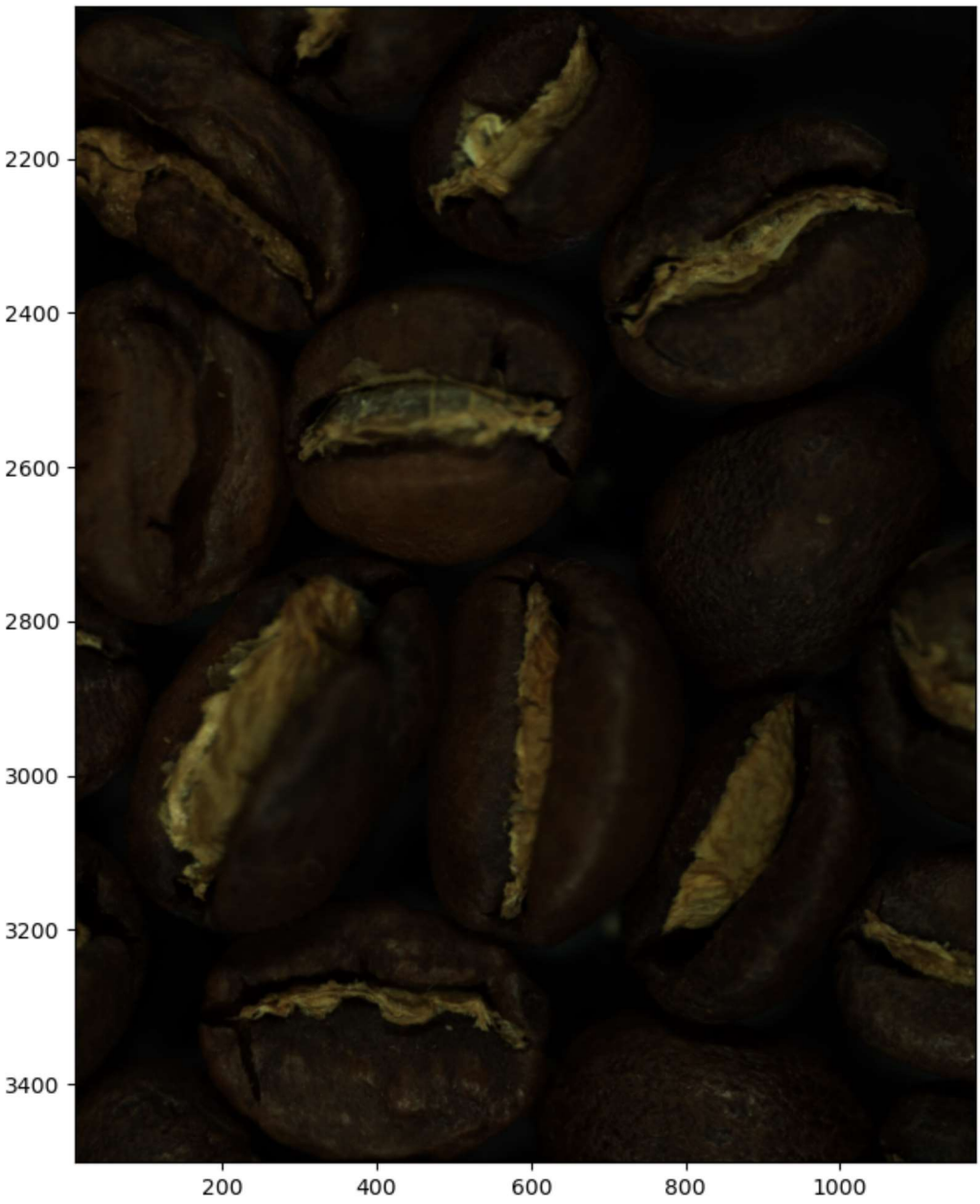


Figure 1.1: BLEND beans roasted for seven minutes captured through a microscope lens using a VNIR camera, the image is shown in standard RGB.

Figure 1.2: The ROEST sample roaster is a lab tool for coffee roaster made to roast 100g samples of coffee at a time.

et al., 2020], and chlorogenic [Nogales-Bueno et al., 2020] content. When analysing coffee samples roasted at different points in time it is important to have an estimate of how rapidly coffee spectra change before and after roasting. Abreu et al. [2019] have used Raman spectroscopy to compare coffee storage conditions. Smrke et al. [2017] have examined how roasted coffee degases over the course of 400 hours. These two studies are important in planing the timeline for our experiments.

1.2 Goals and Research Questions

What defines good coffee is not clear, but what is clear is that quality is chemistry-dependent. To give roasters more control of the quality of their roasts, we want to enable a priori chemometrics of coffee roasts.

Goal A priori chemometrics of coffee roasts.

Coffee beans have structure. The two most obvious parts of a coffee bean are the crack and the body. We do not know if the chemistry in the crack and the body is affected differently by roasting. We know that hyperspectral imaging can be used for chemometrics of coffee roasts. Can hyperspectral imaging be used with clustering to detect bean and background and to find structures of different materials on the surface of coffee beans?

Research question 1 Can clustering be used to detect coffee beans and structures of different materials on their surface?

For the prediction of roast spectra to make sense, the spectra of coffees with different roast profiles have to be separable using the hyperspectral cameras we have at our disposal. Their separability is dependent on their spectral variance.

Research question 2 How large is the within-roast variance of coffee roasts?

Before roasting coffee to build a data set for a priori spectroscopy, we want to know how we can roast coffee with minimal between-roast variance for roasts with identical roast profiles.

Research question 3 What roasting method minimizes the between-roast variance?

Finally, we want to know how accurately we can predict roast spectra based on the spectrum of the green coffee beans before roasting and the roast profile defined by the roast parameters.

Research question 4 How accurately can we predict roast spectra?

Chapter 2

Theory

There are three steps in coffee roasting analysis: roasting, imaging, and analyzing. First, we describe the roasting process and how we can control it. Next, we define the profiles we use in our experiments before defining a hyperspectral camera's field of view and resolution. Last but not least, we present the rationale behind our analysis and decision methods.

2.1 Coffee Roasting

After harvest, coffee berries have their pulp removed either through a *washed* or *natural* process and are dried until reaching around 10% humidity. Then it is time for roasting. We use the ROEST sample roaster, shown in figure 1.2, which is based on conductive heating of coffee beans and control heat transfer through temperature and airflow in the roasting chamber. It has a rotating roasting chamber, a fan that sucks out air from the roasting chamber, and a heating element. The three parameters: air temperature, fan rpm, chamber rpm as a function of time defines a roasting profile.

2.1.1 The Effect of Roasting on Coffee Beans

During roasting, it is evident that phase transitions occur. The two most apparent transitions are the first and second crack, which are both hearable. A little bit of gas is released from the coffee beans in a pop during the first crack (FC). The second crack is more violent than the first, with bursts of vapor erupting from the beans after releasing some of their oils. Roasters use these cracking sounds as indicators of where the beans are in the roasting process and typically decide to halt the roasting based on when the first crack happens. Roasting until first crack tends to produce pleasant roasts for most beans. In addition to the hearable phase transitions there are some visible ones. When coffee beans reach about 170 degrees they turn yellow, and when they reach 190 degrees they turn brown. After becoming brown they gradually become black.

2.2 Coffee Storage

From Abreu et al. [2019] we know that green beans change over time, but the change does not affect the quality of coffee within periods shorter than a month, and the change is smaller for coffee stored in plastic bags than for coffee stored in paper bags. Using a gravimetric method Smrke et al. [2017] have studied degassing of roasted coffee beans. Degassing was exponential, with a

time constant between 153 and 377 hours depending on the roast profile used. Smrke et al. [2017] found that a Weibull distribution with $k < 1$ was a better fit for their gravimetric timeseries, especially for quick dark roasts. This can be explained by two main exponential degassing processes, where one of them has a time constant of 37 hours. Dark and quick roasts have a short degassing time. When it comes to ground coffee, ROSS et al. [2006] did not find a significant quality difference in ground coffee two weeks after grinding.

2.3 Hyperspectral Imaging

Hyperspectral images (HSI) are images with a reflectance spectrum in each pixel. This spectrum can be used to analyze the composition of the material in each pixel, meanwhile having an image with many pixels shows us the spatial distribution of materials in the image.

2.3.1 Infrared Spectroscopy

When we image an object illuminated by a known light source, we get information about its composition through what light it absorbs. Molecules have different vibrational and rotational energy states, with energy gaps E_g between them matching the energy of infrared light. Even though there are many such vibrational states, only the ones where the electric dipole moment changes during vibration absorb infrared light. This is the *selection rule* for infrared spectroscopy. It is important to note that the energy gaps between the states always have some spread. There is a fundamental uncertainty in any energy state, and on top of that, the energy gap can be broadened by bonds with other molecules or other forces. The absorption spectrum, showing how much light of specific frequencies is absorbed in a medium, is characteristic for different molecules. Figure 2.1 shows the visible/near infrared(NIR) absorption bands of common molecules in coffee.

2.3.2 Absorbance

When trying to estimate the concentration of a chemical in a medium, we assume that the absorbance A of light is proportional to the concentration of the chemical ρ and the path length x of light in the medium. A= $\sigma\rho x$ so that the intensity of the light in the medium is given by Beer Lambert's law (2.1).

$$I(x) = I_0 e^{-\sigma\rho x} \tag{2.1}$$

2.3.3 Reflectance

In our hyperspectral images, we do not capture an absorption spectrum but a reflectance spectrum R. The reflectance spectrum consists of light being specularly reflected from the sample R_S and light transmitted through the sample's surface R_T, absorbing and scattering within the sample before ending up on the camera sensor. Specularly reflected light is light that is reflected from the surface, never entering the sample. When light is specularly reflected, it keeps its polarization, so we can use polarization filters to filter out most of it. If the sample is not flat, more specularly reflected light can get through the polarization filter. The absorption spectrum A and the reflectance spectrum without specular reflection R_T is related through (2.2)

$$A = -log_{10}(R_T) \tag{2.2}$$

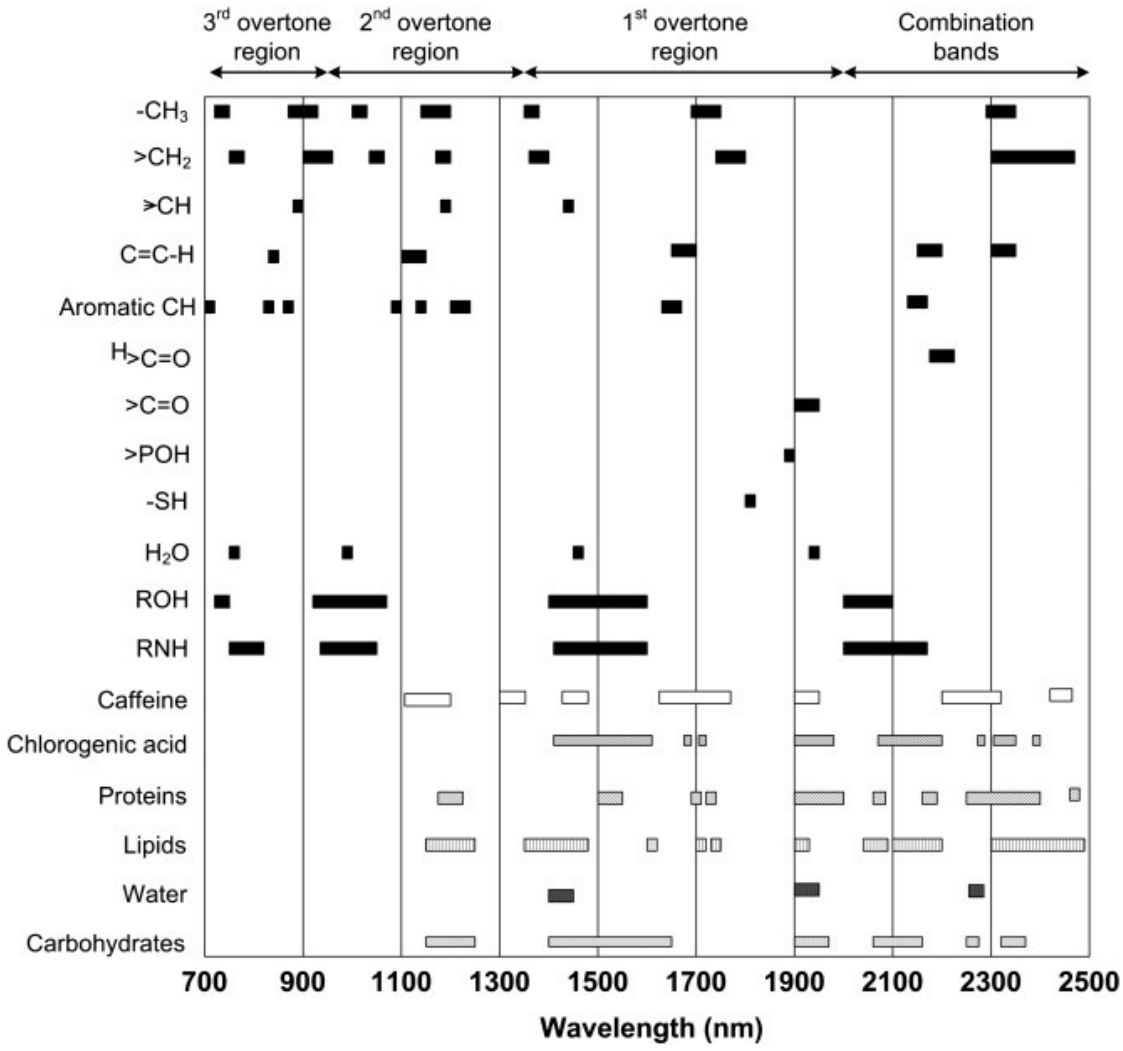


Figure 2.1: There are many molecules in coffee that have absorption peaks of different width in the visible/near infrared spectrum, the narrowest peaks are about 10-20nm wide. (adapted from Ribeiro et al. [2011] and B.G. Osborne [1993])

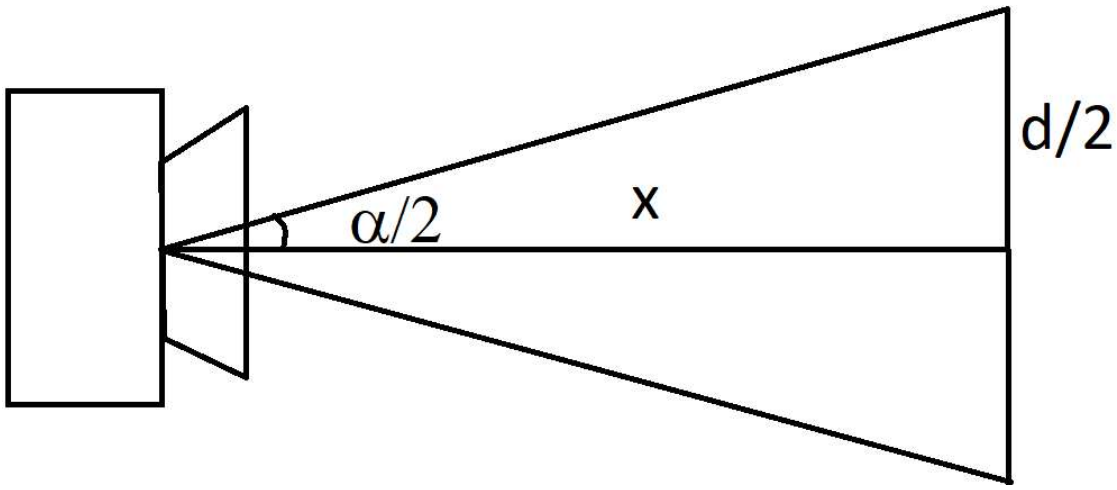


Figure 2.2: The relation between angle of view α and field of view d is given by (2.3)

2.4 Optics

Any spectroscopic system has some optical components. At the core of a dispersive hyperspectral imaging system, a detector array captures light to form an image. One of the spatial dimensions of the detector array represents a spatial direction in the focus plane of the camera, while the other one represents the spectrum of each point along the spatial direction in the focus plane. A line of light captured through the camera lens is spatially selected before it is dispersed into the orthogonal dimension through a dispersive element. The focus length of the camera lens determines the field of view and the depth of focus of the camera.

2.4.1 Angle and Field of View

Camera systems typically specify an angle of view, which describes how wide the camera can see. The angle of view α is illustrated in figure 2.2, where x is the distance from the lens and d is the field of view. If we know the angle of view α of a camera and want to find the field of view d in the focus plane when using a lens with focus length F we can use (2.3)

$$\mathrm{d} = 2\mathrm{F}\tan\alpha \tag{2.3}$$

2.4.2 Resolution

There are two types of resolution, optical resolution and digital resolution. The optical resolution is given by the point spread function of an optical system. When combined with digitalization in a camera, the coarsest resolution is the resolution of the system. Digital resolution r_d field of view d divided by number of pixels N as given by (2.4).

$$r_d = \frac{\mathrm{d}}{N} \tag{2.4}$$

2.5 Numerical Derivatives and Smoothing

The derivative $f'(x)$ of a function $f(x)$ with respect to a variable x is defined as the change in f for an infinitesimal change in x as given by (2.5). The best approximation to the derivative of a series is the difference between each step divided by the steplength l (2.6).

$$f'(x) = \lim_{dx \to 0} \frac{f(x + dx) - f(x)}{dx} \tag{2.5}$$

$$f'[i] = \frac{f[i] - f[i-1]}{l} \tag{2.6}$$

In series measurements, there is always some noise. If we know something about the signals we are looking for, we can filter out some of the noise. Common noise filtering techniques in digital signal processing include: infinite/finite impulse response filtering, discrete Fourier transform filtering, wavelet transform and empirical mode decomposition. A finite impulse response filter is defined by how the signal $\mathbf{s}$ affects the filtered signal $f(\mathbf{s})$ through the impulse response function $\mathbf{h}$ (2.7).

$$f_i = \sum_{j=-\infty}^{\infty} s_{i+j} h_j \tag{2.7}$$

Sometimes, it is convenient to define a function sparsely using a small set of points and interpolating between points that define the function. A much-used interpolation is the natural cubic spline. It has knots in all the points (x, y) in the data and is piece-wise cubic in between the knots. The function has to be linear before the first knot and after the last one while also being continuous up to the 2nd order derivative in all knots. This leaves only one parameter to be fit for each piece in between the knots. Mathematically the cubic spline with K knots at $x = \xi_k \forall k \in K$ is given by (2.8), where $b_1(x) = x$, $b_2(x) = x^2$, $b_3(x) = x^3$, $b_{k+3}(x) = (x - \xi_{k+3})^3 \forall x > \xi_{k+3} \forall k \in [1, K]$

$$y = \beta_0 + \beta_1 b_1(x) + \beta_2 b_2(x) + \beta_1 b_1(x) + ... + \beta_{K+3} b_{K+3}(x) \tag{2.8}$$

2.6 Error Metrics

To assess how good a model is in terms of predicting some quantity we need an error metric, for example the root mean square error (RMSE). If RMSEs are on different scales they are not comparable, to compare such models we use the coefficient of determination R^2

2.6.1 Root Mean Squared Error

The root mean square error is a common error metric in regression analysis. It compares an estimated vector $\hat{\mathbf{y}}$ with true values $\mathbf{y}$, where N is the length of the vectors.

$$\mathrm{RMSE}(\hat{\mathbf{y}}, \mathbf{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2} \tag{2.9}$$

2.6.2 Coefficient of Determination

The coefficient of determination R^2 is a measure of how much of the variation in the response a prediction model can explain. The definition of R^2 is given by (2.10), where y_{ij} is the true value of y_i at index j, $\hat{y}_{ij}$ is the predicted value of y_i at index j and $\bar{y}_i$ is the mean of y_i. When predicting roast frequencies, i indicates wavelength, j indicates roast, and y is the reflectance or 2nd derivative of reflectance.

$$R^2(\hat{\mathbf{y}}, \mathbf{y}) = 1 - \frac{\Sigma_{ij}(y_{ij} - \hat{y_{ij}})^2}{\Sigma_{ij}(y_{ij} - \bar{y}_i)^2} \tag{2.10}$$

2.6.3 Adjusted Coefficient of Determination

We define the adjusted coefficient of determination R^{2*} as (2.11), where σ_u^2 is variance in the response $\mathbf{y}$, which is known to be inexplainable based on the predictors. The adjusted coefficient of determination excludes errors known to be inexplainable so that variation in the response, which is impossible to predict, does not punish the prediction model.

$$R^{2*} = 1 - \frac{\Sigma_{ij}(y_{ij} - \hat{y}_{ij})^2 - \sigma_u^2}{\Sigma_{ij}(y_{ij} - \bar{y}_i)^2 - \sigma_u^2} \tag{2.11}$$

2.7 Statistical Learning

In statistical learning, we talk about predictors x and responses y. The goal of a prediction problem is to predict the response from the predictors. When trying to solve a prediction problem, it is smart to start with a simple model. Linear models are basic models that perform well for many applications. They assume that the response is linear in the predictors $Y = AX+b$. Sometimes we expect the response to be nonlinear in the predictors, and it is tempting to look for nonlinear models. However, nonlinear models suffer from the curse of dimensionality, the more complex the model, the more cursed. The number of coefficients of an ordinary least squares polynomial regression is polynomial so that a polynomial fit of degree D based on N predictors has $O(N^D)$ coefficients. To make a sensible regression, the number of coefficients should be smaller than the number of observations. Otherwise, the prediction model is free to choose coefficients to fit the signal and noise in the data perfectly, yielding a misleading model. In regression problems with many predictors, it is typically wise to transform the predictors and penalize the model's coefficients.

2.7.1 Standard Normal Variate Scaling

Standard normal variate(SNV) scaling is a much used technique to make variables comparable. It scales a variable $\mathbf{x}$ so that its standard deviation becomes 1, as in (2.12), where $\mathrm{sd}(\mathbf{x})$ is the standard deviation of $\mathbf{x}$.

$$\mathrm{SNV}(\mathbf{x}) = \frac{\mathbf{x}}{\mathrm{sd}(\mathbf{x})} \tag{2.12}$$

2.7.2 Evaluating Statistical Models

The empirical performance of a statistical model is stochastic. Therefore it is best practice to use at least two layers of acceptance in any model selection process. Both training and testing of a model should resemble its intended use case, and the steps should use non-overlapping data

sets. The first acceptance criterion for a model is based on cross-validation error, while the next one is based on an independent test error. This means that for a model to be accepted, it has to be selected through cross-validation and then accepted by testing.

2.7.3 Fisher Test

A fisher test (F test) compares two sets of samples and gives a p-value for how likely it is that all the samples originate from the same distribution.

2.7.4 K-Fold Cross-Validation

k-fold cross-validation is a common validation technique, where training data $\mathbf{x}$ is split in k folds. For $i \in [1, k]$ a model is trained on $x_j \forall (j \in k) \neq i$ and tested on x_i. The average validation error of all the validation errors is called the cross-validation error. The standard deviation of the cross-validation error using RMSE is the standard deviation of the validation errors divided by $\sqrt{k}$, assuming identically distributed independent variances in the validation errors.

2.7.5 Ordinary Least Squares Regression

Ordinary least squares regression(OLSR) tries to minimize $RMSE(\hat{\mathbf{y}}, \mathbf{y})$ of a regression model $\hat{\mathbf{y}} = \alpha + \beta \mathbf{x}$.

2.7.6 Principal Component Regression

Predictors in problems of high dimensionality are often correlated. If predictors are correlated enough, most of the variation in the predictors can be explained by a few linear combinations of the predictors called principal components. Selecting the r components that explain most of the variance in the predictors X is equivalent to finding $\tilde{V}$, the r most important eigenvectors of $X^T X$. The Echard-Young theorem states that the best rank r approximation to X is $X \approx \tilde{U}\tilde{D}\tilde{V}^T$. The columns of $\tilde{U}$ are the r first eigenvectors of X, $\tilde{D}$ is the $r \times r$ diagonal matrix of the first r singular values of X, and $\tilde{V}$ is the r most important eigenvectors of the correlation matrix $X^T X$. Using the principal components of X for regression is called principal component regression(PCR). In many applications, the predictors used in the principal component analysis have different units. Variances of variables with different units are typically incomparable, so the predictors are typically scaled to have unit variance. The first principal component is the unit vector in X, along which the variance is largest. Principal component z_j is the unit vector in the subspace of X orthogonal to the last principal components $z_1, ...z_{j-1}$ along which X varies the most.

2.7.7 Partial Least Squares Regression

Partial least squares regression(PLSR) is based on using the components of the predictors that are most correlated to the response to predict the response. The procedure to find the latent variables explaining most of the variance in the response starts with computing $\hat{\phi}_j =< x_j, y >$ for all predictors j. The first latent variable is $z_1 = \sum_j \hat{\phi}_j x_j$. y is regressed on z_1 to find z_1s regression coefficient β_1. Next $x_1 ... x_p$ are orthogonalized with respect to z_1, and so on.

2.7.8 Shrinkage Methods

The error $e(y) = var(y) + b^2(y)$ of a model comes from its variance $var(y) = \mathbb{E}[(\mathbb{E}(y) - \hat{y})^2]$ and its bias $b(y) = \mathbb{E}[\mathbb{E}(y) - \hat{y}]$. The Gauss-Markov theorem tells us that the mean square error estimator is the best unbiased estimator. However, if we introduce some bias, we might reduce the variance more than the square bias increases resulting in reduced error, this is called *regularization*. Having many predictors makes a model vulnerable to variance. If two variables are negatively correlated, one of the coefficients may be very large and positive while the other coefficient is very large and negative. Shrinkage methods regularize by penalizing the coefficients β. A common class of penalization is L_q penalization, where βs are chosen based on the criterion (2.13). L_0 is best subset selection, L_1 is lasso regression and L_2 is ridge regression. Just as in principal component analysis, predictors are SNV scaled before the coefficients are calculated.

$$\tilde{\beta} = \underset{\beta}{\operatorname{argmin}}(\sum_i (y_i - \beta_0 - \sum_j (x_{ij}\beta_{ij}))^2 + \lambda \sum_j (|\beta_j|^q)) \tag{2.13}$$

2.7.9 Neural Networks

The last decade has seen many applications of ever more advanced neural networks. The idea behind neural networks is to apply a neat property of nonlinear functions to approximate arbitrary functions. A neural network is just a set of matrix multiplications and transformations relating predictors $\mathbf{X}$ to responses $\mathbf{Y}$ through (2.14), where $\mathbf{C_i}$ are matricies and $_i$ are some functions. If the functions σ_i are locally bounded nonpolynomial piecewise continuous functions, the network can approximate arbitrary continuous functions with arbitrary accuracy.

$$\mathbf{Y} = \sigma_N(\mathbf{C_N}...\sigma_1(\mathbf{C_1}\sigma_0(\mathbf{C_0X}))) \tag{2.14}$$

2.7.10 Parameter Searching

When mapping a function in a parameter space it is tempting to do a grid search, which makes sense if the parameter space is one dimensional, but provides unsatisfactory little information if the parameter space is of higher dimensionality and the importance of the parameters are not equal. Random sampling is much better than grid search for sampling functions of high dimensionality, especially when the predictors are of unequal importance. This principle is explained by Bergstra and Bengio [2012], focusing on the application of random searching in hyper-parameter tuning for neural networks and is beautifully illustrated in 2.3.

2.7.11 Hierarchical Clustering of Spectra Using Complete Linkage

We cluster the spectra of pixels based on their scaled dot product. If two have an identical shape their dot product will be 1. A cluster is a set of one or more spectra. In hierarchical clustering all N spectra start out as leaf nodes in a forest of N trees. To connect trees we use the complete linkage criterion, which measures the longest distance between all pairs of spectra in a pair of trees. In each iteration the two trees with the largest overlap are connected in a new root node. All trees are connected in $N - 1$ iterations.

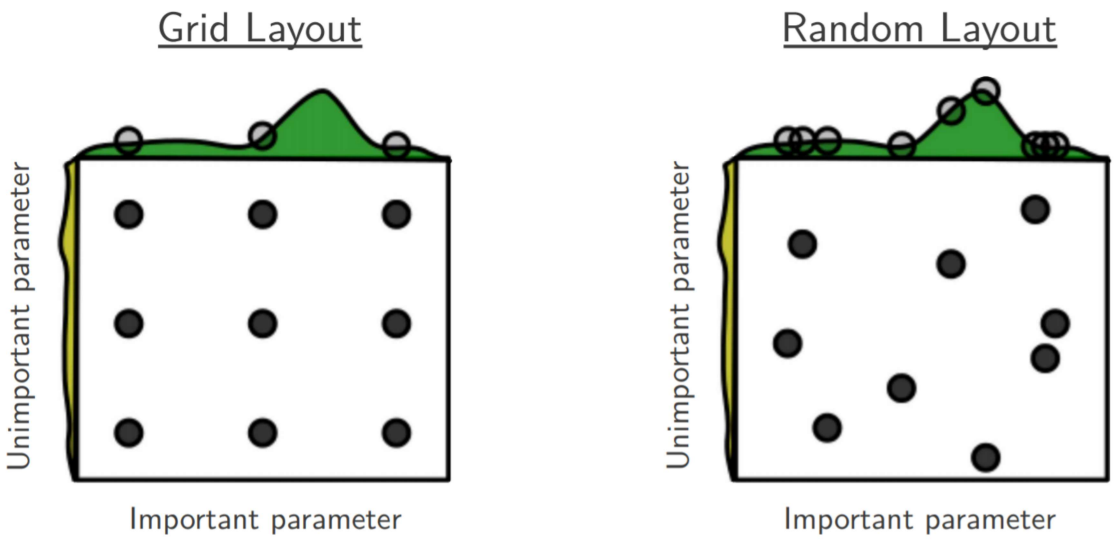


Figure 2.3: Grid and random sampling of nine points to estimate a function $f(x, y) = g(x) + h(y) \approx g(x)$ with low effetive dimensionality. Above each square $g(x)$ is shown in green, and left of each square $h(y)$ is shown in yellow. With grid search, nine samples only sample $g(x)$ in nine places. With random search, all nine samples explore $g(x)$ in distinct places. This failure of grid search is the rule rather than exception in high-dimensional spaces. Adapted from Bergstra and Bengio [2012]

Chapter 3

Method

Our study of coffee is an iterative process. Our end goal is a priori chemometrics of coffee roasts, and we know that Nogales-Bueno et al. [2020] have used hyperspectral cameras for estimation of extractable content of bioactive compounds in coffee beans. Using a hyperspectral camera with a spectral resolution of 3.25 nm, they estimate caffeine, chlorogenic, and phenolic content from the 950-1650 nm spectral band of roasted coffee beans. We intend to figure out how accurately we can estimate the post-roasting spectra of coffee beans $R_{roasted}$ based on their pre-roasting spectrum R_{green} and roast profile. In our work, we first image roasts of the same bean variety roasted for different amounts of time to check if we can see a difference between the roasts. We also use these images to check how large the spectral variance between different samples from the same roast is compared to the spectral variance between images of the same beans in different orientations. Then we check if we can find clusters of different materials on the surface of the coffee beans. Next, we choose a roasting method based on a variance analysis experiment before roasting coffee beans in different ways to produce a data set we can use to train prediction models. Finally, we test a range of prediction models to predict raw reflectance spectra, and 2nd derivative transformed spectra which are used for phenolic estimation by Nogales-Bueno et al. [2020]. In addition to the work we have done for spectral prediction, we have also imaged coffee after grinding and brewing. This enables an analysis of the relation between the spectra of whole bean, ground, and brewed coffee. This chapter first presents our hyperspectral imaging setups, our coffee roaster, and how it controls the roasting process. Then we present how we treat our samples before, during, and after imaging and roasting. We image all samples using two hyperspectral cameras.

3.1 Coffee Beans and Sample Roaster

We use a ROEST sample roaster for coffee roasting, always roasting 100g±1g of coffee at a time. We use five *varieties* of coffee for our roasting experiments: ETIOPIA, PERU, BLEND, PE-2020-038_ARS, which are all washed, and PSS-CO-2020-116, which is naturally processed.

3.2 Roasting Using the ROEST Sample Roaster

In our study of coffee, we look at 100g *samples* of coffee drawn from different *batches* of green coffee, shown in figure 3.1, where each batch consists of coffee beans of one specific variety. When roasting it is important that the temperature in the chamber *dropTemp* is stable before

Table 3.1: The ROEST standard eight minute profile with start temperature dropTemp and end temperature endTemp is defined by a natrual cubic spline (2.8) through these points in time t and temperature T.

t	**T**
0	dropTemp
30	dropTemp - 60
40	dropTemp - 60
90	dropTemp - 60 + $48.9\frac{\mathrm{dT}}{\mathrm{dt}}$
300	dropTemp - 60 + $366.7\frac{\mathrm{dT}}{\mathrm{dt}}$
520	endTemp

the beans are dropped into the chamber via the hatch shown in figure 3.2. We drop the beans into the chamber immediately after loading them into the hatch, so that they do not heat up before the roasting starts. The temperature in the ROEST sample roaster is preprogrammed and controlled automatically. When a sample is roasted, it becomes a *roast*. After *dropTime* seconds in the roasting chamber the roast is dropped into a ventilated aluminium traytray and is rapidly cooled, shown in figure 3.3. Beans typically lose about 7% of their mass during roasting, mostly water and their silverskin, collected in a chaff collector, shown in figure 3.4. Different samples roasted the same way are different roasts. We image either whole samples or *sub-samples*, shown in figure 3.5.

3.2.1 The ROEST Standard Profile

For roasting we have used ROESTs standard eight-minute temperature profile ROEST(dropTemp,endTemp,dropT which is defined by a natural cubic spline through the points in table 3.1, with time t, temperature T, dt = 480 and dT = endTemp - dropTemp + 60. dropTime is the time at which the beans are dropped into the aluminium tray to cool, it is measured in seconds. When roasting until first crack ROEST(dropTemp,endTemp,FC) dropTime is choosen dynamically and is set to the time at which first crack occurs + 50 seconds. Fan and chamber rpm are the same for all temperature profiles.

3.3 Coffee Storage

We have to keep an eye on what happens to coffee before roasting and between roasting and imaging. We know from Abreu et al. [2019] that coffee beans are best stored in plastic bags, so we use polyethylene bags for storage. The time it takes for coffee stored in plastic bags to change significantly in terms of sensory quality is longer than two months. We never store coffee for more than a week between imaging the green beans and roasting, keeping our storage time well within the bounds given by Abreu et al. [2019]. The time window between roasting and imaging is much shorter. Smrke et al. [2017] reports that degassing of some of the compounds in coffee is exponential with a time constant as low as 37 hours, which is why we are careful to image coffee beans we roast within 12 hours after roasting. When coffee is ground, it degases more rapidly.

Figure 3.1: Four batches of different coffee varieties, each in their own bag.

Figure 3.2: A sample is loaded in to the hatch of the ROEST sample roaster immediately before dropping into the roasting chamber.

Figure 3.3: A roast is cooled in an aluminium tray after roasting.

Figure 3.4: During roasting coffee beans loose their silverskin. The ROEST sample ROASTER collects the silverskin in a chaff collector.

Figure 3.5: Half of a green bean sample is loaded into a dish.

Table 3.2: Accessories we use with the SWIR camera in our imaging setup.

Lamps	HySpex "flicker-free" DC linear light source
Stage	Translationstage: 8MT195-540-10
Motor Controller	8SMC1-USBh Stepper Motor Controller
Background	Black Fabric

Table 3.3: Accessories we use with the VNIR camera in our imaging setup.

Lamps	Illumination technologies 3900
Stage	Standa 100833
Motor Controller	8SMC4-USB-B8-1 Motor Controller
Background	Black Cardboard

3.4 Coffee Bean Spectrum Measurements

Previous work (Nogales-Bueno et al. [2020],Zhang et al. [2018a]) have used averaged coffee bean spectra for estimation of chemical composition. We do the same to make our models compatible with the ones used in Nogales-Bueno et al. [2020] and limit our analysis scope. Zhang et al. [2018b] have found that image averaging yields better chemometric results than single pixel classification. To assess how much better image averaging is than point spectroscopy, we calculate the pixel-wise reflectance spectrum variance over all images.

3.5 Hyperspectral Imaging Setups

We image coffee using two different hyperspectral camera setups, one with a SWIR 320me(Norsk Elektrooptikk, 962nm-2493nm) camera, shown in figure 3.6, and one with a VNIR 1800(Norsk Elektrooptikk, 405nm-995nm) camera, shown in figure 3.7. The SWIR camera is mounted on a moving stage together with lamps over the sample. In contrast, the VNIR camera is mounted in a stationary configuration with samples mounted on a moving stage underneath. A reference reflectance sample (Spectralon Diffuse Reflectance Standard) of reflectance $R = 0.99$ is included in all images to enable calculation of the reflectance spectrum in all pixels of the images. The SWIR camera setup in the 30cm lens configuration is shown in figure 3.6 and the VNIR setup is shown in figure 3.7. Both setups use microscope lenses and no polarization filters to capture images for clustering analysis. For all other experiments, both cameras use 30cm lenses and polarization filters in cross-polarized configuration. We put the beans in NIR-UV transparent dishes for imaging, as shown in figure 3.8 and focus the camera on the bean surface. The accessories used for the cameras are listed in table 3.2 and 3.3, while the specs of the cameras are listed in table 3.4 and 3.5.

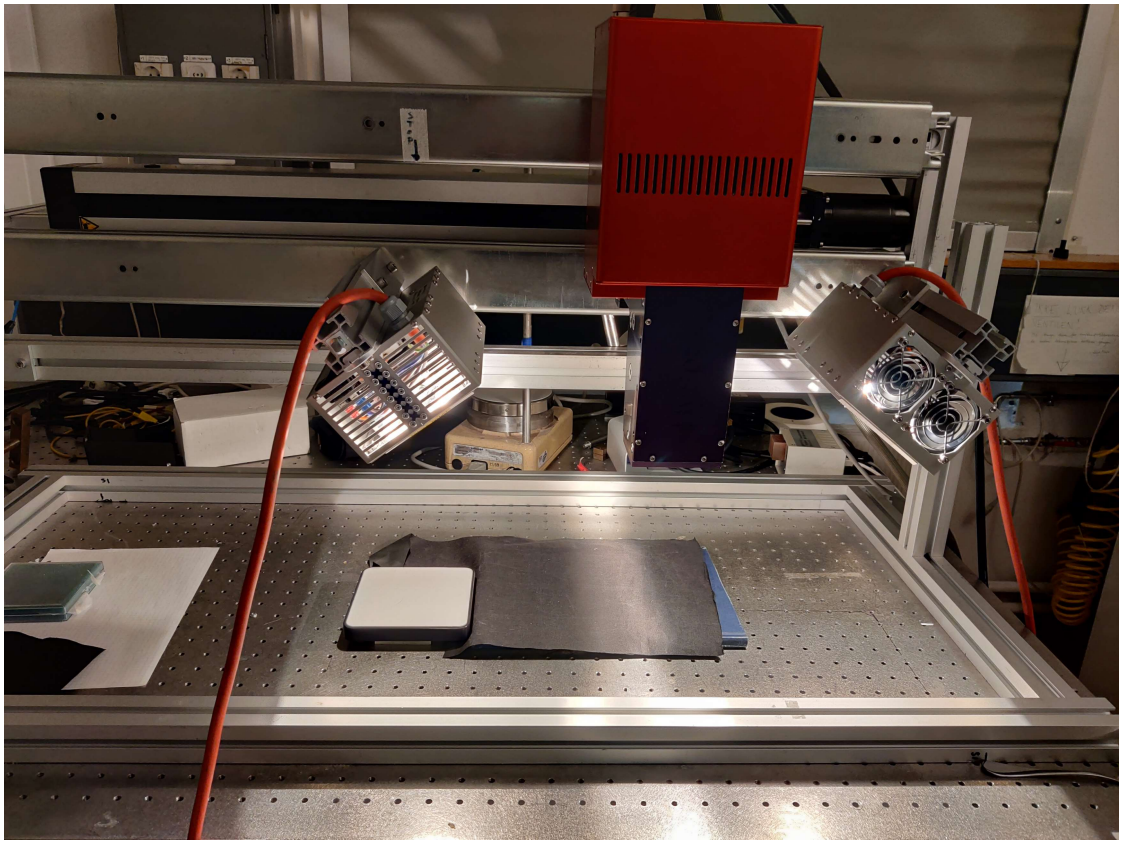


Figure 3.6: The SWIR320 962nm-2493nm hyperspectral camera is mounted over the sample with a 0.99 reflectance reference beside it.

Table 3.4: Specifications of the SWIR320 camera.

Spectrum	962nm-2493 nm
Spectral resolution	6.00 nm
AOV	16°
Bit resolution	16 bit
30 cm lens FOV	84 mm
30 cm lens resolution	263 μm
Microscope lens FOV	17 mm
Microscope lens resolution	53 μm

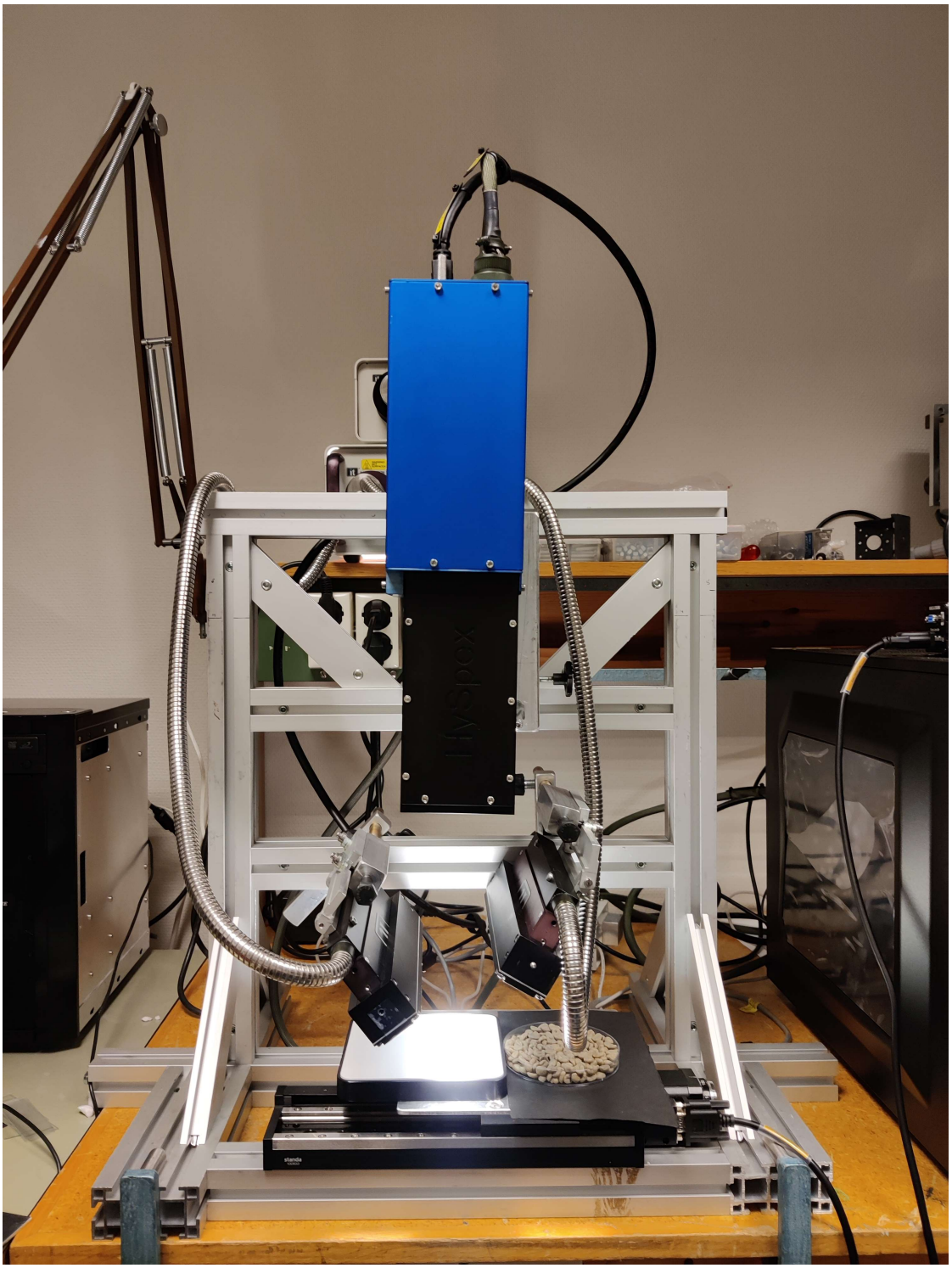


Figure 3.7: The VNIR1800 405nm-955nm hyperspectral camera is mounted over the sample with a 0.99 reflectance reference beside it.

Table 3.5: Specifications of the VNIR1800 camera.

Spectrum	405nm-995nm
Spectral resolution	3.26 nm
AOV Bit resolution	17° 16 bit
30 cm lens FOV	86 mm
30 cm lens resolution	168 μm
Microscope lens FOV	12 mm
Microscope lens resolution	24 μm

3.6 Using clustering to Find Regions of Interest

Zhang et al. [2018a] shows that the spatial distribution of the spectrum decays radially from the centre of coffee beans when their crack is facing away from the camera, we want to investigate if clustering can detect regions of interest on the surface of coffee beans in our images. We image eight different coffee roasts roasted from t = 0 − 7 minutes at one-minute intervals. All eight samples are of BLEND variety and are roasted using a ROEST(220,250,t) standard profile. We image the beans using a microscope lens to enable the clustering to find small-scale structures. We use hierarchical clustering with complete linkage.

3.7 Within-Roast Variance

Because we do not find prominent hyperspectral structures that are small in space, we decide to use a 30cm lens instead of the microscope lens to capture the rest of the images in our study. By increasing the field of view in this way, more of the variation in the beans is captured by having more beans in each image. Having switched lenses, we also add cross-polarized filters to the light sources and lenses before capturing images to look at the variance between images of the same coffee beans and between beans from the same roast.

3.7.1 Preparation and Imaging

In this experiment, we use the same eight samples of BLEND variety roasted for zero to seven minutes using a ROEST(220,250,t) standard profile as in the clustering experiment. We want to know how separable the spectra of our roasts are, which is determined by how large the variance in the spectrum of a roast is. We estimate the variance of the spectrum based on the orientation of the beans when the hyperspectral image is captured and based on what subset of beans from a sample we are imaging. To estimate the variance in the spectrum based on what subset of a roast we image, we draw three random samples from each roast. Each roast has between 90g and 100g of beans, and each sample is half of the beans from a roast drawn randomly amongst all beans in the roast. To estimate the variance in the spectrum based on the orientation of the beans, we capture three images of each sample, stirring in the dish of beans in between to randomize the orientation of the beans. We go through this process using both the SWIR and the VNIR camera. One sample of each roast is shown in figure 3.8.

Figure 3.8: Coffee beans roasted from zero to eight minutes, left to right, using the ROEST standard roasting profile defined by table 3.1 with dropTemp = 220 and endTemp = 250

3.7.2 Within-Roast Variance Analysis

Since clustering does not reliably separate bean and background or find meaningful patterns, we settle for a hand-tuned reflectance threshold for bean/background separation. For all 30cm lens images, we calculate the mean spectrum, the variance in the spectrum and make an image of the deviation from the mean spectrum of the image. From the images of the same beans in different orientations, we calculate the variance in the spectrum based on orientation, and from samples from the same roasts, we calculate the variance in the spectrum based on what sample of beans is randomly selected for imaging.

3.8 Between-Roast Variance

Before roasting beans to make prediction models, we want to know what roasting method is best in terms of minimal spectral variance and check whether the variance comes from the green beans or from roasting.

3.8.1 Roasting Beans for Variance Analysis

To test whether the variance in the spectrum of beans roasted using two different methods comes from the green beans or from roasting, we image twelve PE-2020-038_ARS bean samples. Then we roast the beans using two roasting methods, six using the first crack method ROEST(220,250,FC) and six for exactly six minutes ROEST(220,250,360), before imaging them again. In this way, we can compare the variances between the samples before and after roasting to determine if one roasting method causes more variance in the spectrum than the other one. We increase the size of a sample to include all beans in a roast so that all beans in each roast are imaged, eliminating the small variance arising from sample selection. Each sample is imaged three times with stirring in between. To investigate the effect of variety on the spectral variance caused by roasting, we also roast three additional samples of PE-2020-038_ARS, BLEND, ETIOPIA and PERU coffee

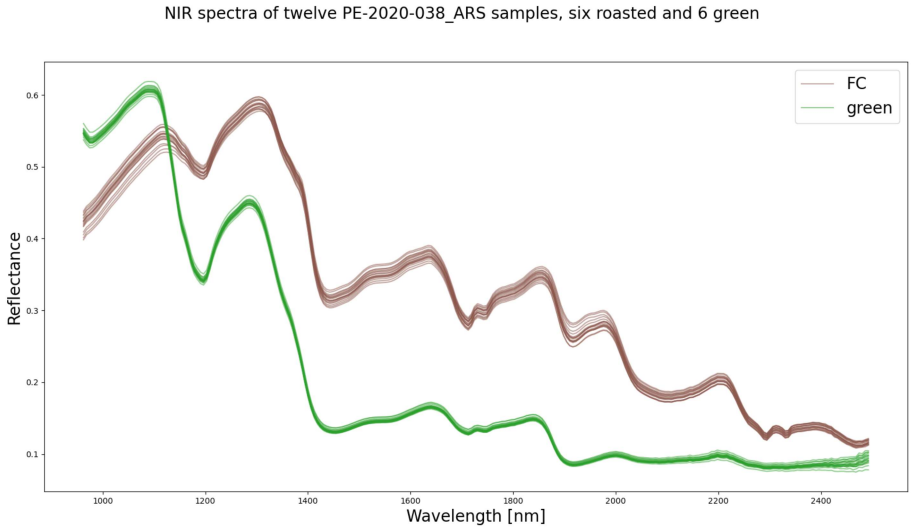


Figure 3.9: SWIR reflectance spectra of green and FC roasted PE-2020-038_ARS beans. Roasted beans clearly have a higher overall reflectance than green beans.

using the first crack method to investigate if there is a difference in spectral standard deviation based on variety.

3.8.2 Comparing Variances

Using a fisher test, we test whether the variances of the roasts are similar or not. We also use a fisher test to check if the variances can be explained by the green bean variance being scaled equally for all roasts. The scaling test is motivated by the fact that the mean of the SWIR spectrum increase when the beans are roasted, as shown in figure 3.9, however the mean of the VNIR spectrum decreases, as shown in figure 3.10.

3.9 Building a Dataset for A Priori Chemometrics of Coffee and Exploring Its Potential

We know that the variance based on what sample from a roast is selected for imaging is similar to the variance based on orientation. We know that roasting based on time yields lower spectral variance than roasting based on FC. We are ready to build a data set for a priori chemometrics of coffee. Our goal is to predict $R_{roasted}$ based on R_{green} and ROEST(dropTemp,endTemp,dropTime).

3.9.1 Building a Dataset for A Priori Chemometrics of Coffee

We want to enable a priori chemometrics of coffee to help make good coffee. This means we should limit the roasting of coffee beans to make roasts that are consumable. Since coffee needs to be heated to 190 degrees to be roasted, our lower bound for endTemp is 200 degrees. To stay within the operational temperature of the roaster, we limit endTemp to maximum 300

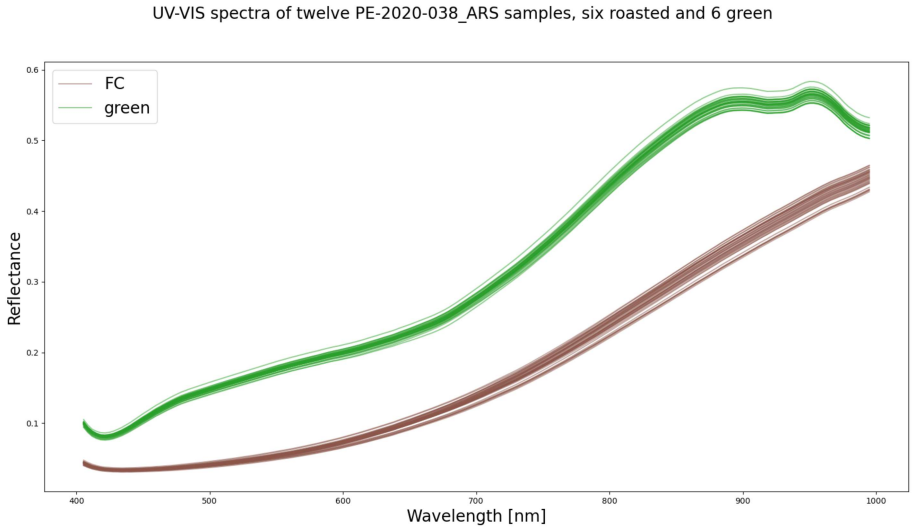


Figure 3.10: VNIR reflectance spectra of green and FC roasted PE-2020-038_ARS beans. Roasted beans clearly have a lower overall reflectance than green beans.

degrees. dropTemp is bounded to be minimum 200 degrees to ensure that the beans reach roasting temperature before dropping and maximum 240 degrees, since that is the maximal start temperature of the roasting machine. dropTime is kept within 360-480s, as this is the standard roast-time window for coffee. The beans we have available to build the data set are 25 ETIOPIA, 16 PERU, 12 BLEND, 1 PE-2020-038_ARS, and 1 PSS-CO-2020-116 samples. Since we expect that some of the roast parameters are more important than others, we randomly sample the roast parameters to produce roast profiles for all the bean samples. To image all the roasts within the lab hours we have available, we decide to image only one sample of green beans from each variety before roasting. All of the beans in the experiment are roasted and imaged within five days. The green beans are imaged at day three to make them as comparable to the other beans as possible. All images include all beans from the samples we image. Each sample is imaged three times with stirring in between.

3.9.2 Data Split

Before doing any analysis of the data whatsoever, we split the data into a training and a test set. To ensure the data sets are as independent as possible, we split them based on date while also ensuring that not all bean varieties are present in the training set. This is helpful for testing generalizability. The training set contains all roast spectra from days 1, 2, and 4, while the test set contains all roast spectra from days 3 and 5. In the training set, there are 10 BLEND roasts, 15 ETIOPIA roasts, and 15 PERU roasts. The test set has 2 BLEND roasts, 1 PERU roast, 10 ETIOPIA roasts, 1 PE-2020-038_ARS roast, and 1 PSS-CO-2020-116 roast.

3.9.3 Preprocessing of Data for Different Prediction Models

We now have a data set with predictors dropTemp, endTemp, dropTime and R_{green} and responses $R_{roasted}$. For each roast there are three different spectra $R_{roasted}$, all having identical predictors. To make linear models to predict $G(\cdot)$ we scale the response $\tilde{R}_{roasted} = \frac{R_{roasted}(\lambda)}{R_{green}(\lambda)} \forall \lambda$. So that $R_{green} \cdot G(\text{dropTemp}, \text{endTemp}, \text{dropTime}, R_{green}) \approx R_{green}\tilde{R}_{roasted} = R_{roasted}$. When calculating 2nd derivative spectra we choose to use length five sliding windows in the smoothing of the derivative with weights [0.05,0.15,0.60,0.15,0.05] to closely resemble the smoothing used by Nogales-Bueno et al. [2020] with a width of 30nm for the SWIR spectra and a 12.5nm width for the VNIR spectra. These widths are similar to narrow peaks in the visible/NIR absorption spectrum of common chemicals in coffee shown in figure 2.1. For most of our models we do no additional scaling, except for predictors for the lasso models, the 2nd derivative PCR and PLSR models and all neural networks, which are all standard normal variate scaled. The training set is always used as a scale for the test set, again, to test for generalizability.

3.9.4 Prediction Models

To do a priori chemometrics of coffee roasts, we first need to predict the spectrum of roasted coffee from the spectrum of green coffee and roast profile. Our approach to predicting $R_{roasted}(\text{dropTemp}, \text{endTemp}, \text{dropTime}, R_{green})$ is testing multiple common statistical learning models. We test linear regression(lin. reg.) of $R_{roasted}$ using only dropTemp, endTemp, and dropTime as predictors, with and without interaction(int.) between the roast parameters. We test linear regression of $R_{roasted}(\lambda)$ based on dropTemp, endTemp, dropTime, and $R_{green}(\lambda)$, with and without interaction between the roast parameters. We test principal component regression(PCR), partial least squares regression(PLSR), and lasso regression of $R_{roasted}$ based on dropTemp, endTemp, dropTime, and R_{green}, with interaction between the roast parameters for PCR and PLS, but not for lasso. Finally, we test a neural network regression of $R_{roasted}$ based on dropTemp, endTemp, dropTime, and R_{green}. The neural network takes the green bean spectrum and the roast parameters as input and outputs the roast spectrum. The hidden layers are two dense ReLU activated layers with dropout layers after each of them. For all models except the neural network models we try two kinds of models: models with an additive effect on the response $R_{roasted} = F(\cdot)$, and models with a multiplicative effect on the response $R_{roasted} = R_{green} \cdot G(\cdot)$. Additive models try to predict the roast spectrum directly, while multiplicative models try to predict how the green bean spectrum is scaled by roasting.

3.9.5 Model Training

When training statistical models, we use 10-fold cross-validation to calculate a cross-validation error and standard deviation so that we are less likely to choose using an overfitted model. We use the standard deviation of the cross-validated error to find its one standard deviation interval. After testing different models for a prediction problem, we choose the simplest model with one standard deviation cross-validation error interval overlapping with the model with the lowest cross-validation error. To find the best number of principal components at each frequency for the PCR and PLSR, as well as the best penalty λ for the lasso model, we use another layer of 9-fold cross-validation. More than one hyper-parameter needs to be tuned for the neural network, so we use randomized hyper-parameter tuning. We choose to tune the width of the dense layers, the dropout rate of the dropout layers, and the batch size used in training the network. These five parameters are random-uniformly sampled with widths between 16 and 256, dropout rates between 0 and 1, and batch size between 1 and 120.

3.9.6 Model Evaluation

We test models based on two kinds of data: Raw reflectance spectra and the 2nd derivative of reflectance spectra. To make fair comparisons between prediction models for raw spectra and prediction models for 2nd derivative spectra, we use the coefficient of determination R^2 (2.10), and the adjusted coefficient of determination R^{2*} (2.11). To evaluate R^{2*} we need to estimate the inexplainable variance σ_u^2. We set σ_u^2 to the maximum of within-roast variance in the test set, and the between-roast variance from the between-roast variance experiment. For each transformation, for each camera, we choose our preferred model through cross-validation. However, the highest error of the test and validation error is the error that best indicates how well the model performs.

3.10 Ground and Brewed Coffee

Both ground and brewed coffee were imaged using the 30 cm lens setup with polarisation filters. The coffee was ground using a handheld Bodum blade grinder, and the coffee was brewed on a Moccamaster filter coffee brewer. We never stored ground coffee for more than eight hours before imaging or brewing. We do not include these spectra in our analysis, but they are available for future work.

Chapter 4

Results

In total, we have captured over one thousand images of coffee, making up over 1TB of hyperspectral data. In this chapter we are going through the results from our experiments step by step. First, we look at the separability of beans of the same variety roasted for different amounts of time. Then we look at variances in the reflectance spectrum of the roasts $R_{roasted}$, between different samples from the same roast, between images with beans from the same sample in different orientations, and between different samples of beans from the same batch before and after roasting. We use these results to design an experiment to investigate how well we can estimate the spectra of roasted coffee beans as we take our first step toward a priori chemometrics of coffee roasts.

4.1 Clustering of Regions of Interest on the Surface of Coffee Beans

Clustering does not find regions of interest on the surface of coffee beans that we deem helpful for our analysis, indeed it does not even separate bean and background. The structures found are large enough to be captured using a 30cm lens. The topology of coffee beans rather than their structure dominates the spatial distribution of their spectrum.

4.2 Something Goes Wrong, But No Worries

For the rest of our images, we use polarization filters on both cameras. Unfortunately, we use a VNIR polarization filter on the SWIR camera, filtering the light in an unintentional way. On the flip side, the spectrum of the reflectance reference is filtered the same way as the coffee, and the normalization in the reflectance calculation undoes the unintended filtering.

4.3 Separability and Within-Roast Variance

After choosing to use the 30cm camera lens for imaging, we look at coffee roasts of beans from the same batch roasted using profiles defined by ROEST(220,250,t) with t spaced at one-minute intervals. The standard deviation in the spectrum between sub-samples from the same roast is 0.01 reflectance units and is much smaller than the gap between the roasts. The standard

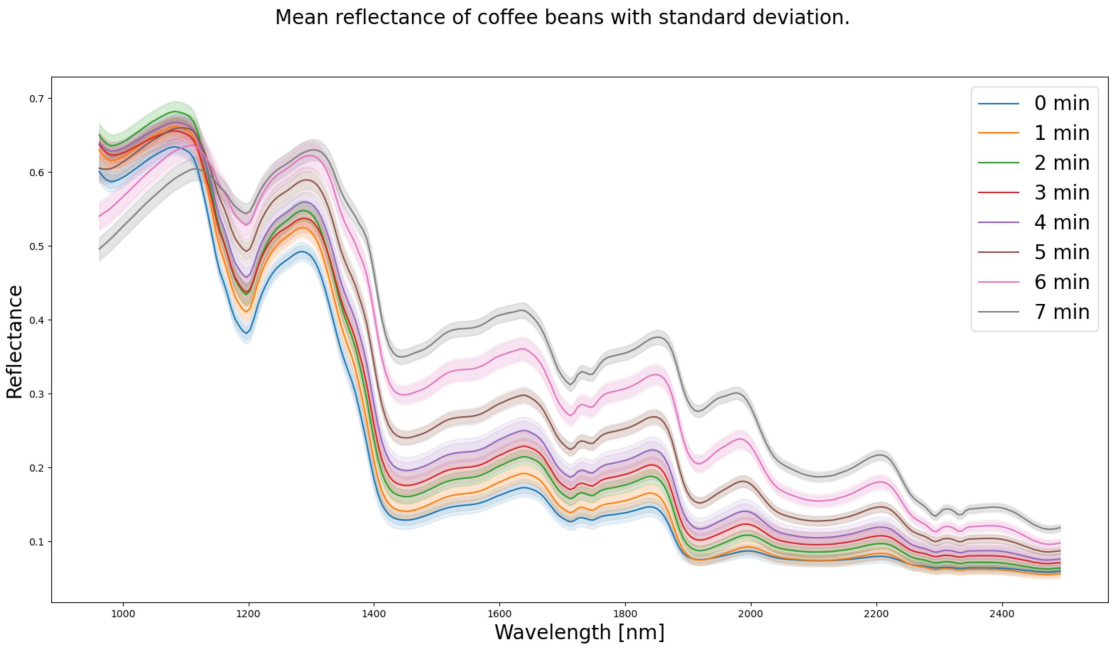


Figure 4.1: Mean spectra of BLEND beans roasted from 0 to 7 minutes at one minute intervals with inner standard deviation bars showing the within sample standard deviation and outer standard deviation bars showing between sample standard deviation, both within the same roast. The spectra are separable.

deviation in the spectrum of the same sub-samples in different orientations is 78% of the standard deviation between sub-samples from the same roast. The difference between the roasts is large enough to discriminate them from each other, especially when taking pictures of multiple orientations, as seen in figure 4.1. The figure shows the mean reflectance spectra of coffee roasted for one to seven minutes at one-minute intervals and the standard deviation of the reflectance at each frequency. There are two standard deviation bands, the inner one is calculated based on all images of the same sub-sample, and the outer one is based on all images of the same roast.

4.4 The Best Roasting Method

Roasting twelve samples of PE-2020-038_ARS coffee, six using the ROEST(220,250,360) method and six using the ROEST(220,250,FC) method, we find that the ROEST(220,250,360) method yields the lowest spectral variance in the raw spectrum while there is no difference in the 2nd derivative spectrum. All standard deviations within each sub-batch before and after roasting with 95% confidence intervals are listed in table 4.1 and table 4.2. The standard deviation of the six-minute roasts is 0.0056, and the standard deviation of the FC roasts is 0.0067. The p-value of an F test, testing if the variances are the same, is less than 2.2e-16. The F test of whether the increase in standard deviation arises from a scaling of the green bean standard deviations by the same factor is also less than 2.2e-16. The six-minute process also has a lower variance in the VNIR images with a standard deviation of 0.0051. The FC standard deviation is 0.0071; the

Table 4.1: This table shows standard deviations of the reflectance spectrum for the FC and time roasting methods. Standard deviations with 95% confidence interval(CI) upper and lower bounds(ub,lb) are listed in in units of reflectance units. The standard deviations are calculated for the FC and the time sub-batch before and after roasting, both between the samples in each sub-batch and within each sample within the sub-batch.

Camera	Set of samples	Sub-batch	roasted/green	Variation	CI lb	CI ub
SWIR	between-sample	FC	green	0.0034	0.0033	0.0035
SWIR	between-sample	FC	roasted	0.0067	0.0065	0.0068
SWIR	between-sample	time	green	0.0031	0.0030	0.0032
SWIR	between-sample	time	roasted	0.0056	0.0055	0.0057
SWIR	within-sample	FC	green	0.0027	0.0027	0.0027
SWIR	within-sample	FC	roasted	0.0044	0.0044	0.0045
SWIR	within-sample	time	green	0.0021	0.0020	0.0022
SWIR	within-sample	time	roasted	0.0036	0.0035	0.0036
VNIR	between-sample	FC	green	0.0046	0.0045	0.0046
VNIR	between-sample	FC	roasted	0.0071	0.0070	0.0073
VNIR	between-sample	time	green	0.0052	0.0051	0.0053
VNIR	between-sample	time	roasted	0.0051	0.0050	0.0052
VNIR	within-sample	FC	green	0.0041	0.0040	0.0042
VNIR	within-sample	FC	roasted	0.0036	0.0035	0.0037
VNIR	within-sample	time	green	0.0049	0.0048	0.0050
VNIR	within-sample	time	roasted	0.0047	0.0046	0.0048

difference between these standard deviations is even more significant than the difference between the SWIR standard deviations and can not be explained by scaling the spectrum. These results mean that roasting for six minutes using the ROEST standard profile is more reliable than roasting until first crack. The spectral standard deviation of different bean varieties roasted using the FC method is shown in table 4.3, we see that variety is more important than roast method in terms of variance with PERU roasts having a very small standard deviation.

4.5 Predicting Roast Spectra

Our goal is to predict the spectrum of roasted coffee $R_{roasted}$ based on the spectrum of green coffee R_{green} and the roast parameters dropTemp, endTemp and dropTime. When training models to predict the raw spectrum of roasted coffee, we found that multiplicative models $R_{roasted} = R_{green} \cdot G(\cdot)$ strictly outperformed additive models $R_{roasted} = F(\cdot)$, while opposite for models predicting the 2nd derivative spectrum. Therefore we only display the results of multiplicative models for the raw spectrum models and additive models for the 2nd derivative spectrum models. Table 4.5 and 4.6 show RMSE and R^2 for raw spectrum and 2nd derivative models, evaluated on the training set with cross-validation, evaluated on the test set, evaluated exclusively on known bean varieties in the test set, and exclusively on unknown varieties in the test set. The last two columns show the performance of models trained and evaluated on a single variety only. For the test error R^{2*} is also reported. The models are either with or without interaction(int.) between the roast parameters. All models except the ordinary linear(lin.) regression(reg.) models include R_{green} as predictor. Two of the ordinary linear regression models include $R_{green}(\lambda)$. For the SWIR raw spectrum data the between-roast variance from the between-roast experiment is

Table 4.2: This table shows standard deviations in the 2nd derivative of the reflectance spectrum with respect to frequency for the FC and time roasting methods. Standard deviations with 95% confidence interval(CI) upper and lower bounds(ub,lb) are listed in in units of reflectance units per square wavelength step. The standard deviations are calculated for the FC and the time sub-batch before and after roasting, both between the samples in each sub-batch and within each sample within the sub-batch.

Camera	**Set of samples**	**Sub-batch**	**roasted/green**	**Standard deviation**	**CI lb**	**CI ub**
SWIR	between-sample	FC	green	8.643e-5	7.659e-5	9.758e-5
SWIR	between-sample	FC	roasted	7.799e-5	7.138e-5	8.526e-5
SWIR	between-sample	time	green	7.499e-5	6.971e-5	8.072e-5
SWIR	between-sample	time	roasted	8.096e-5	7.255e-5	9.038e-5
SWIR	within-sample	FC	green	6.811e-5	5.678e-5	8.173e-5
SWIR	within-sample	FC	roasted	4.427e-5	3.846e-5	5.097e-5
SWIR	within-sample	time	green	5.538e-5	5.145e-5	5.963e-5
SWIR	within-sample	time	roasted	6.414e-5	5.483e-5	7.505e-5
VNIR	between-sample	FC	green	1.569e-5	1.424e-5	1.729e-5
VNIR	between-sample	FC	roasted	1.353e-5	1.218e-5	1.503e-5
VNIR	between-sample	time	green	1.610e-5	1.480e-5	1.752e-5
VNIR	between-sample	time	roasted	1.356e-5	1.161e-5	1.583e-5
VNIR	within-sample	FC	green	1.389e-5	1.278e-5	1.510e-5
VNIR	within-sample	FC	roasted	1.119e-5	1.007e-5	1.245e-5
VNIR	within-sample	time	green	1.502-5	1.372e-5	1.645e-5
VNIR	within-sample	time	roasted	1.236e-5	1.083e-5	1.412e-5

Table 4.3: The standard deviation in the spectra of roasts of the same variety roasted with the same profile using the FC method.

Bean Variety	**Standard deviation**
PE-2020-038$_A RS$	0.0055
BLEND	0.0057
PERU	0.0037
ETIOPIA	0.0047

Table 4.4: The performance of our preferred prediction models in terms of R^2

Model	R^2
SWIR raw spectrum	0.67
VNIR raw spectrum	0.74
SWIR 2nd derivative spectrum	0.61
VNIR 2nd derivative spevtrum	0.58

higher than the within-roast variance in the test set, so between-roast variance is used as the inexplainable error in the calculation of R^{2*}. In the VNIR raw spectrum data, the SWIR and the VNIR 2nd derivative spectrum data it is opposite. The R^2 of our preferred prediction models are listed in table 4.4

4.5.1 SWIR Spectrum Prediction Results

Simple linear regression of $R_{roasted}$ based solely on process parameters without interaction is the simplest model with an estimated cross-validation RMSE that overlaps with the lowest cross-validation RMSE of the raw spectrum SWIR models in table 4.5. It has a cross-validation RMSE of 0.043, with a one standard deviation interval of [0.040,0.045].

4.5.2 VNIR Spectrum Prediction Results

Simple linear regression of $R_{roasted}$ based solely on process parameters without interaction is the simplest model with an estimated coss validation RMSE that overlaps with the lowest cross-validation RMSE of our models in table 4.5. It has a cross-validation RMSE of 0.038 on the VNIR data, with a one standard deviation interval of [0.035,0.040].

4.5.3 SWIR 2nd Derivative Spectrum Prediction Results

Simple linear regression of $R_{roasted}$ based solely on process parameters without interaction is the simplest model with an estimated coss validation RMSE that overlaps with the lowest cross-validation RMSE of our models in table 4.6. It has a cross-validation RMSE of 3.21e-4 on the SWIR data, with a one standard deviation interval of [3.04e-4,3.37e-4]

4.5.4 VNIR 2nd Derivative Spectrum Prediction Results

Simple linear regression of $R_{roasted}$ based solely on process parameters without interaction is the simplest model with an estimated coss validation RMSE that overlaps with the lowest cross-validation RMSE of our models in table 4.6. It has a cross-validation RMSE of 4.09e-05 on the VNIR data, with a one standard deviation interval of [3.55e-5,4.57e-5].

Table 4.5: Raw roast spectrum prediction performance. The best models are highlighted in bold font. NAs indicate that the model is missing. (+) signs indicate that unseen R^2 is greater than seen R^2. Hyperparameters are listed on the form (width1,drop1,width2,drop2,batch_size).

Model	**Val. RMSE**	**Val. R^2**	**Test RMSE**	**Test R^2**	**Test R^{2*}**	**Seen RMSE**	**Seen R^2**	**Unseen RMSE**	**Unseen R^2**	**Single RMSE**	**Single R^2**
SWIR lin. reg. without int.	**0.0425**	**0.67**	**0.0431**	**0.71**	**0.71**	**0.0447**	**0.69**	**0.0301**	**0.86 +**	**0.0510**	**0.59**
SWIR lin. reg. with int.	0.0433	0.65	0.0483	0.64	0.64	0.0509	0.60	0.0248	0.90 +	0.0554	0.52
SWIR lin. reg. with $R_{green}(\lambda)$ without interaction	0.0391	0.72	0.0484	0.63	0.64	0.0471	0.65	0.0562	0.51 -	0.0510	0.59
SWIR lin. reg. with $R_{green}(\lambda)$ with int.	0.0381	0.73	0.0533	0.56	0.56	0.0508	0.60	0.0677	0.28 -	0.0554	0.52
SWIR PCR with int.	0.1195	-1.65	0.0437	0.70	0.70	0.0461	0.67	0.0225	0.92 +	0.0497	0.61
SWIR PLSR with int.	0.0382	0.73	0.0449	0.69	0.69	0.0474	0.65	0.0231	0.92 +	0.0498	0.61
SWIR LASSO reg. without int.	0.0408	0.69	0.0452	0.68	0.68	0.0466	0.66	0.0345	0.81 +	0.0530	0.56
SWIR Neural network (61,0.91, 70,0.35,7)	0.0758	-0.07	0.1144	-1.04	-1.05	0.0849	-0.12	0.2264	-7.00 -	0.0844	-0.11
VNIR lin. reg. without int.	**0.0376**	**0.74**	**0.0332**	**0.86**	**0.86**	**0.0349**	**0.85**	**0.0196**	**0.95 +**	**0.0321**	**0.87**
VNIR lin. reg. with int.	0.0407	0.69	0.0347	0.85	0.85	0.0365	0.83	0.0186	0.96 +	0.1106	-0.53
VNIR lin. reg. with $R_{green}(\lambda)$ without int.	0.0355	0.77	0.0352	0.84	0.85	0.0338	0.86	0.0431	0.77 -	0.0321	0.87
VNIR lin. reg. with $R_{green}(\lambda)$ with int.	0.0379	0.73	0.0362	0.84	0.84	0.0350	0.85	0.0432	0.77 -	0.1106	-0.53
VNIR PCR with int.	0.0410	0.69	0.0332	0.86	0.86	0.0347	0.85	0.0212	0.94 +	0.0607	0.54
VNIR PLSR with int.	0.0390	0.72	0.0325	0.87	0.87	0.0338	0.86	0.0226	0.94 +	0.0607	0.54
VNIR LASSO reg. without int.	0.0393	0.71	0.0382	0.82	0.82	0.0394	0.81	0.0295	0.89 +	0.0372	0.83
VNIR											

Table 4.6: 2nd derivative roast spectrum prediction performance. The best models are highlighted in bold font. NAs indicate that the model is missing. (+) signs indicate that unseen R^2 is greater than seen R^2. Hyperparameters are listed on the form (width1,drop1,width2,drop2,batch_size).

Model	**Val. RMSE**	**Val.** R^2	**Test RMSE**	**Test** R^2	**Test** R^{2*}	**Seen RMSE**	**Seen** R^2	**Unseen RMSE**	**Unseen** R^2	**Single RMSE**	**Single** R^2
SWIR lin. reg without int.	**3.21e-4**	**0.611**	**3.39e-4**	**0.64**	**0.67**	**3.51e-4**	**0.61**	**2.47e-4**	**0.81 +**	**3.75e-4**	**0.56**
SWIR lin. reg. with int.	3.38e-4	0.568	3.43e-4	0.63	0.66	3.60e-4	0.59	2.02e-4	0.87 +	8.68e-4	-1.38
SWIR lin. reg. with $R_{green}(\lambda)$ without int.	3.18e-4	0.617	3.52e-4	0.61	0.64	3.66e-4	0.58	2.37e-4	0.82 +	4.20e-4	0.44
SWIR lin. reg. with $R_{green}(\lambda)$ with int.	3.35e-4	0.575	3.52e-4	0.61	0.64	3.67e-4	0.58	2.37e-4	0.82 +	7.81e-4	-0.93
SWIR PCR	3.16e-4	0.623	3.64e-4	0.58	0.61	3.85e-4	0.53	1.86e-4	0.89 +	3.75	0.56
SWIR PLSR	3.08e-4	0.641	3.81e-4	0.54	0.57	4.04e-4	0.48	1.72e-4	0.91 +	3.76e-4	0.55
SWIR LASSO reg. without int.	3.16e-4	0.622	3.56e-4	0.60	0.63	3.71e-4	0.57	2.33e-4	0.83 +	8.24e-4	-1.15
SWIR Neural network (102,0.64, 41,0.38,19)	1.47e-3	-7.21	7.54e-2	-2e4	-2e4	1.45e-3	-5.61	0.206	-1e5 -	NA	NA
VNIR lin. reg. without int.	**4.09e-5**	**0.584**	**3.77e-5**	**0.71**	**0.73**	**3.98e-5**	**0.68**	**1.97e-5**	**0.92 +**	**3.21e-5**	**0.79**
VNIR lin. reg. with int.	4.25e-5	0.551	3.98e-5	0.68	0.70	4.23e-5	0.64	1.66e-5	0.94 +	7.73e-5	-0.21
VNIR lin. reg. with $R_{green}(\lambda)$ without int.	3.71e-5	0.658	3.96e-5	0.68	0.70	3.48e-5	0.75	6.22e-5	0.21 -	3.21e-5	0.79
VNIR lin. reg. with $R_{green}(\lambda)$ with int	3.79e-5	0.642	4.18e-5	0.65	0.66	3.67e-5	0.73	6.62e-5	0.11 -	7.73e-5	-0.21
VNIR PCR	3.68e-5	0.663	3.48e-5	0.76	0.78	3.44e-5	0.76	3.72e-5	0.72 -	4.35e-5	0.62
VNIR PLSR	3.73e-5	0.654	3.49e-5	0.75	0.77	3.52e-5	0.75	3.25e-5	0.79 +	3.88e-5	0.70
VNIR LASSO reg. without int.	4.08e-5	0.586	4.10e-5	0.66	0.68	4.31e-5	0.62	2.42e-5	0.88 +	4.64e-5	0.56
VNIR Neural network (51,0.49, 217,0.24,97)	1.00e-4	-1.48	1.37e-2	-4e4	-4e4	1.17e-4	-1.77	0.0374	-3e5 -	NA	NA

Chapter 5

Discussion and Conclusion

We go through how our results answer our research question and what future work is needed to reach our goal of a priori chemometrics of coffee roasts.

5.1 Using Clustering to Detect Beans and Cracks

Hierarchical clustering based on whole spectra with complete linkage does not separate bean and background or find the cracks of the coffee beans. Since the cracks in the coffee beans are easy to see by eye, they are probably easy to detect if looking at the correct wavelengths. However, since an intensity threshold can reliably separate beans and background, we decide that clustering is overkill. The radial spatial pattern found by clustering is reported in the literature Zhang et al. [2018a] and can also be seen in an image of the euclidian deviation from the mean spectrum in an image of coffee beans from one of our experiments shown in figure 5.1.

5.2 Within-Roast Variance Analysis

Our within-roast variance analysis finds that roasts roasted for zero to seven minutes at one-minute intervals are easily separable, as shown in figure 4.1. Relievingly we find that most of the variance within a sample of coffee can be captured within a smaller sample. At the same time, we notice the importance of imaging the coffee beans in multiple orientations.

5.3 Between-Roast Variance Analysis

Our between-roast variance analysis very clearly shows that roasting for a specific time is superior to roasting until first crack in terms of spectral variance. We therefore decide to parameterize profiles based on roast time.

5.4 Roast Spectrum Prediction Analysis

Our prediction models are capable of explaining much of the variance in the spectrum of roasted coffee. We show that roast spectrum prediction is possible and that even models trained on only three different R_{green} can make use of it in predictions of roasts with new R_{green}. Our discussion of neural network performance is very limited, as they performance is too bad to be

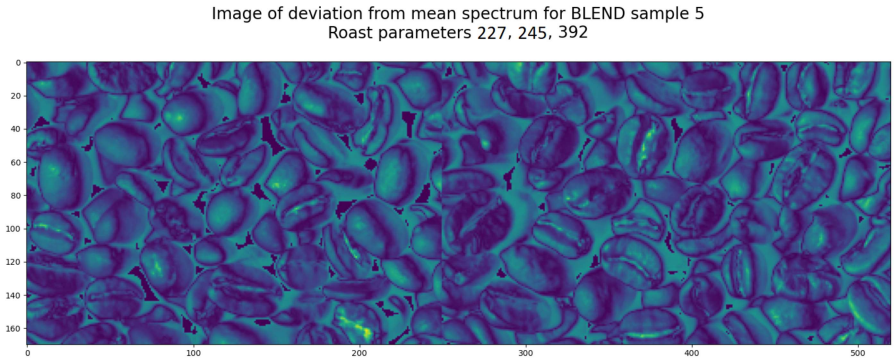


Figure 5.1: An image showing the euclidian distance from the mean spectrum to the spectru of each pixel over an image of coffee beans. The spectrum deviates most from the mean spectrum on the edges of the beans and in their centre.

worth our attention. Here we summarise the most important observations from the prediction model experiment.

- Making one model for each bean variety in the test set is mostly worse than making a joint model.
- Since $R^2 \approx R^{2*}$ the error in the predictions can not be attributed to inexplainable variance.
- We observe that the test error is mostly lower than the validation error for the linear models not knowing R_{green}, but greater for the basic linear regressions that include it, as indicated by the (+) signs in table 4.5 and 4.6.

5.4.1 The Effect of Only Having Three Distinct Values for R_{green}

Our regularized models can use more parameters than there are samples in our training data. When having more predictors than samples in a training set, we can typically choose coefficients to achieve a perfect fit. However, R_{green} only takes three distinct values, effectively reducing the number of degrees of freedom it introduces to three. Therefore the degrees of freedom are always less than the number of parameters in the models trained on the whole training set. Models trained on only one variety can still have more predictors than samples, but regularization with cross-validation helps choose models with low dimensionality. Since single-variety models can have too many predictors, we see that single-variety models with interaction between the roast parameters have a very low predictive power. The single-variety regularized models perform better than the non-regularized ones, and only the 2nd derivative SWIR single variety lasso model has no predictive power. Neural networks can benefit from more parameters than samples, so we are not concerned on their behalf.

5.4.2 The Effect of Including R_{green}

The fact that models not including R_{green} have better performance on the test set than similar models including R_{green} suggests that the models that include R_{green} are mislead by R_{green}. We do, however, see that PCR, PLSR and lasso are all able to predict $\tilde{R}_{roasted}$ quite well despite knowing R_{green}. This could be because the models learn that R_{green} is unimportant, but it is not. Not using R_{green} as predictors lowers the cross-validation R^2 of the raw SWIR PLSR model from 0.73 to 0.63. The fact that test R^2 and R^{2*} are close to identical indicates that our data has a high quality.

5.4.3 RMSE of Raw Spectrum Models Trained on Only a Single Variety

The RMSE of the basic SWIR linear regression model without interaction models trained on only a single variety of beans is high, compared to the normal one. This indicated that the information conveyed by bean variety is unimportant compared to having more training data. The same can not be said for the VNIR version of the basic linear regression, where the single variety model has the same performance as the normal one, indicating that the variation in the data is captured well enough in a smaller data set.

5.4.4 RMSE of 2nd Derivative Spectrum Models Trained on Only a Single Variety

The RMSE of the basic SWIR linear regression model without interaction models trained on only a single variety of beans is high, compared to the normal one. This indicated that the information conveyed by bean variety is unimportant compared to having more training data. The same can not be said for the VNIR version of the basic linear regression, where the single variety model has a lower error than the original one, indicating that variety is important.

5.4.5 Very Low Unseen Variety Error

The errors of our prediction models on unseen bean varieties are lower than the error on seen varieties for most models, as indicated by the (+) signs in table 4.5 and 4.6. All of the models taking $R_{green}(\lambda)$ as input without regularization, except the 2nd derivative SWIR models, have worse performance on unseen bean varieties than on seen bean varieties. Unregularized models are typically bad at generalization, so we are surprised that the 2nd derivative SWIR models perform better on unseen bean varieties than on seen ones. We can see that all the green beans have characteristic spectra in figure 5.6, 5.7 and 5.9, but the same is not true for 5.8, where information about bean variety is much less prominent. Even though the green bean spectra are characteristic, they are roughly the same size, as seen from figure 5.2, 5.3, 5.4 and 5.5. It seems like most of the regularized models have managed to pick up the most important information in the spectrum. Only neural networks are fooled by the discrepancy in the spectra, as seen by the abysmal prediction errors of the neural networks on unseen bean varieties in table 4.5 and 4.6.

5.4.6 The Spectral Distribution of Error

We see that the errors in our raw spectra are largest where the variance in the spectrum is largest in the data set, as seen when comparing figure 5.11 with 5.10 and figure 5.13 with 5.12. The same is true fro the 2nd derivative spectra, as seen when comparing figure 5.15 with 5.14

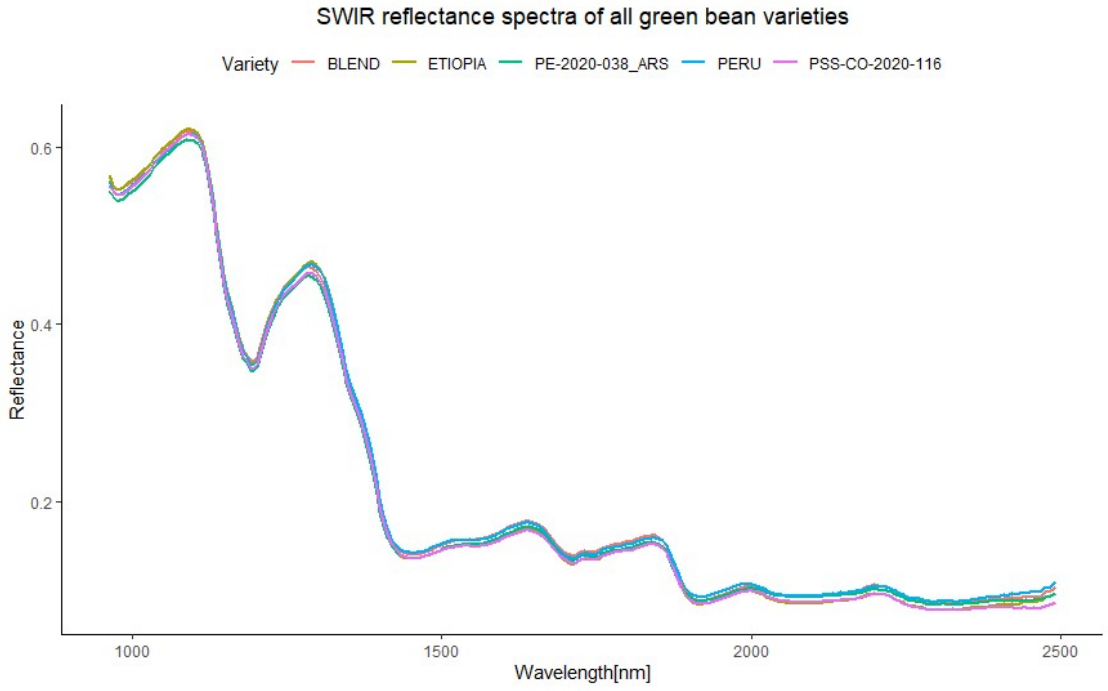


Figure 5.2: SWIR mean reflectance spectrum of all green bean varieties. The spectra are very similar.

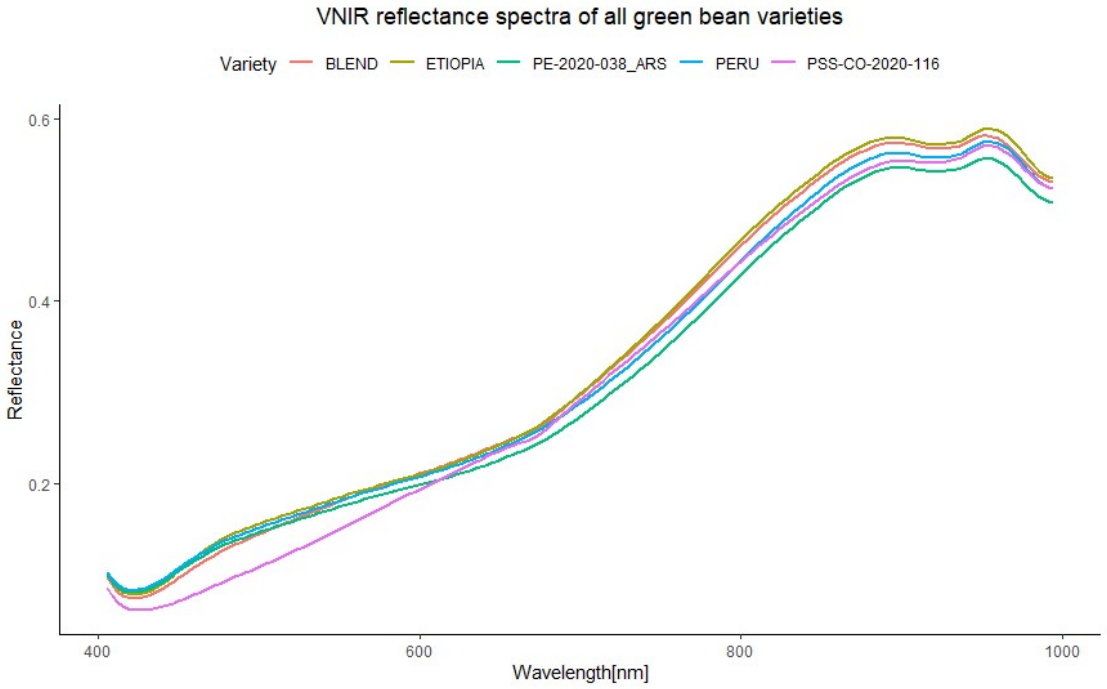


Figure 5.3: VNIR mean reflectance spectrum of all green bean varieties. The spectra are very similar.

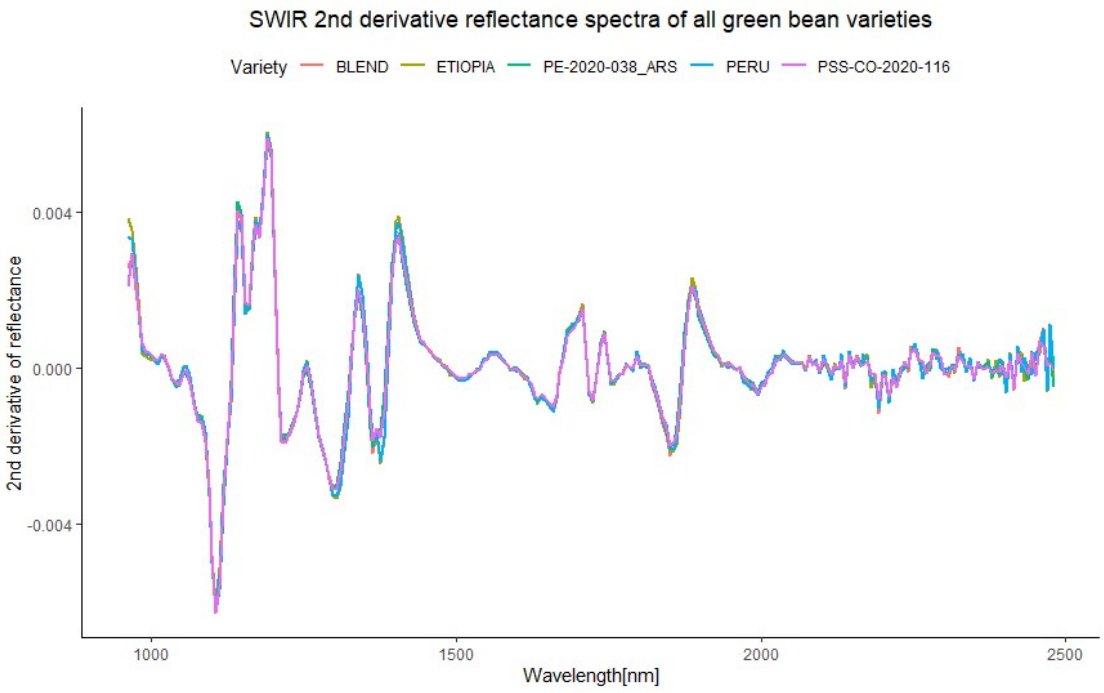


Figure 5.4: SWIR mean 2nd derivative reflectance spectrum of all green bean varieties. The spectra are very similar.

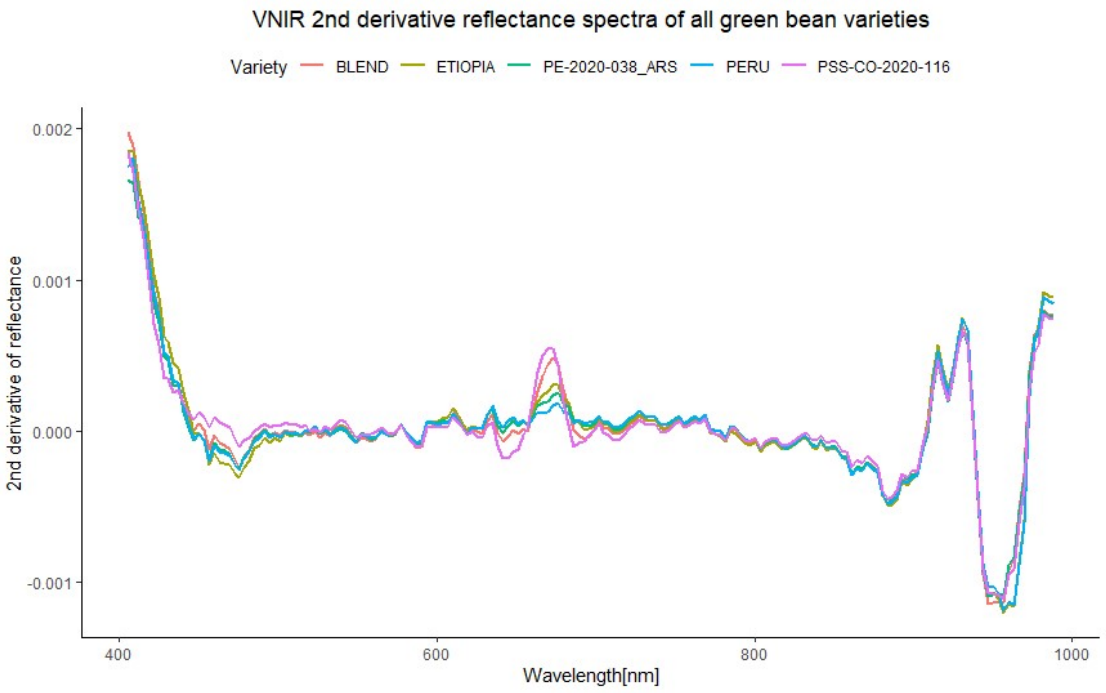


Figure 5.5: VNIR mean 2nd derivative reflectance spectrum of all green bean varieties. The spectra are very similar.

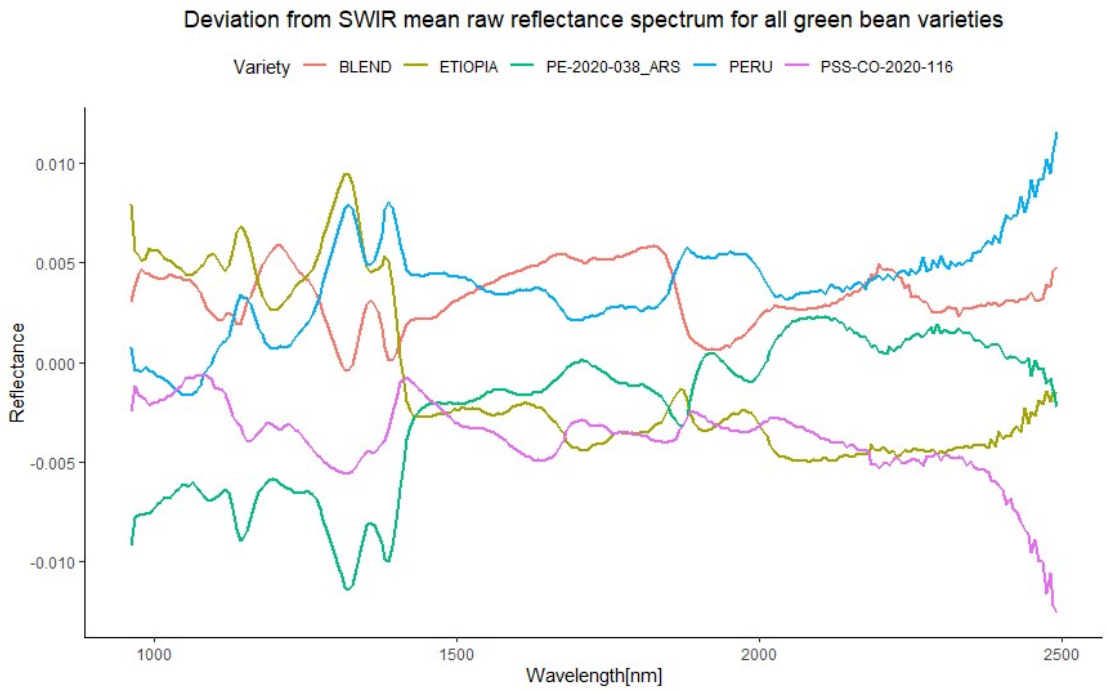


Figure 5.6: Deviation from the mean R_{green} SWIR raw reflectance spectrum for all green bean varieties. The spectra are characteristic for bean variety.

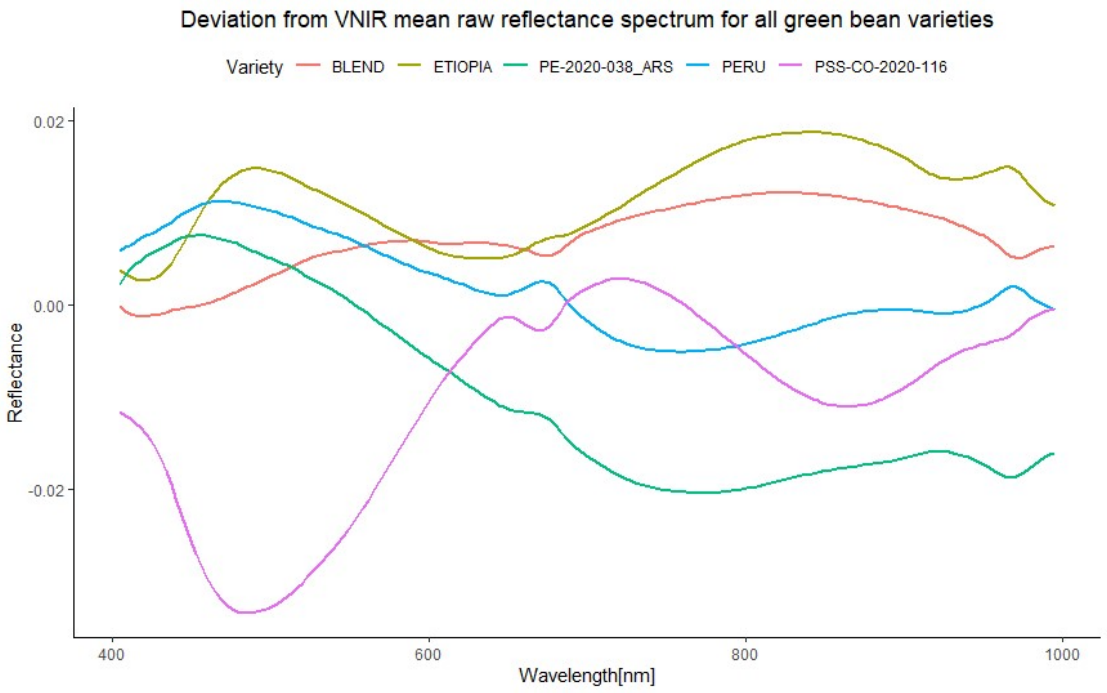


Figure 5.7: Deviation from the mean R_{green} VNIR raw reflectance spectrum for all green bean varieties. The spectra are characteristic for bean variety.

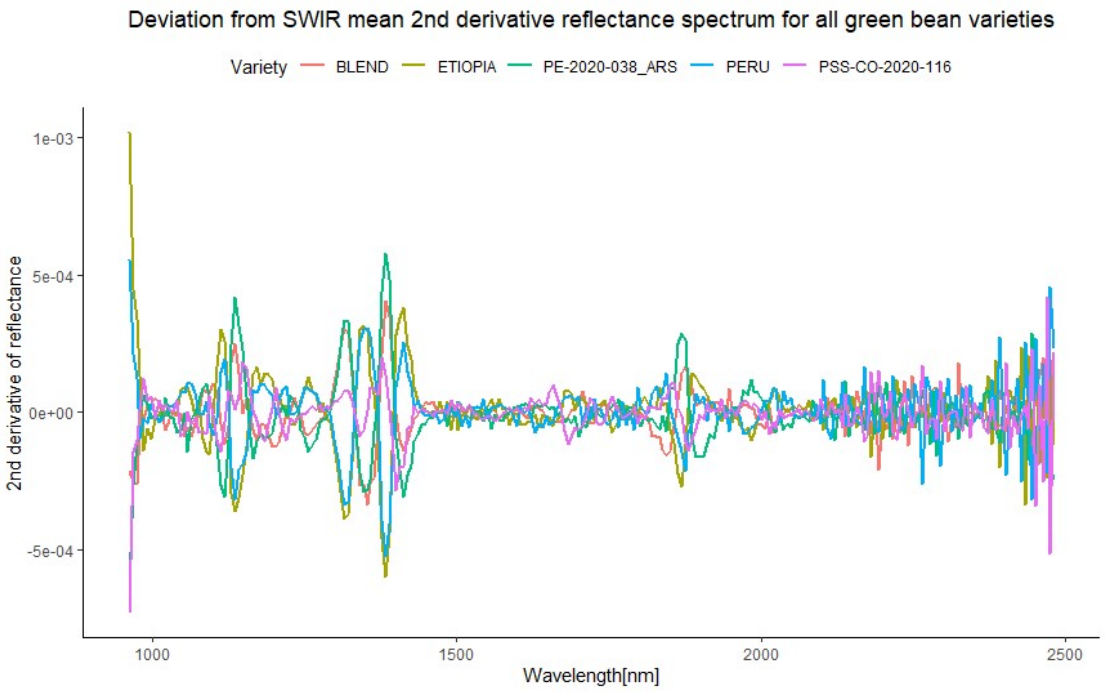


Figure 5.8: Deviation from the mean R_{green} SWIR 2nd derivative reflectance spectrum for all green bean varieties. The spectra are hard to distinguish.

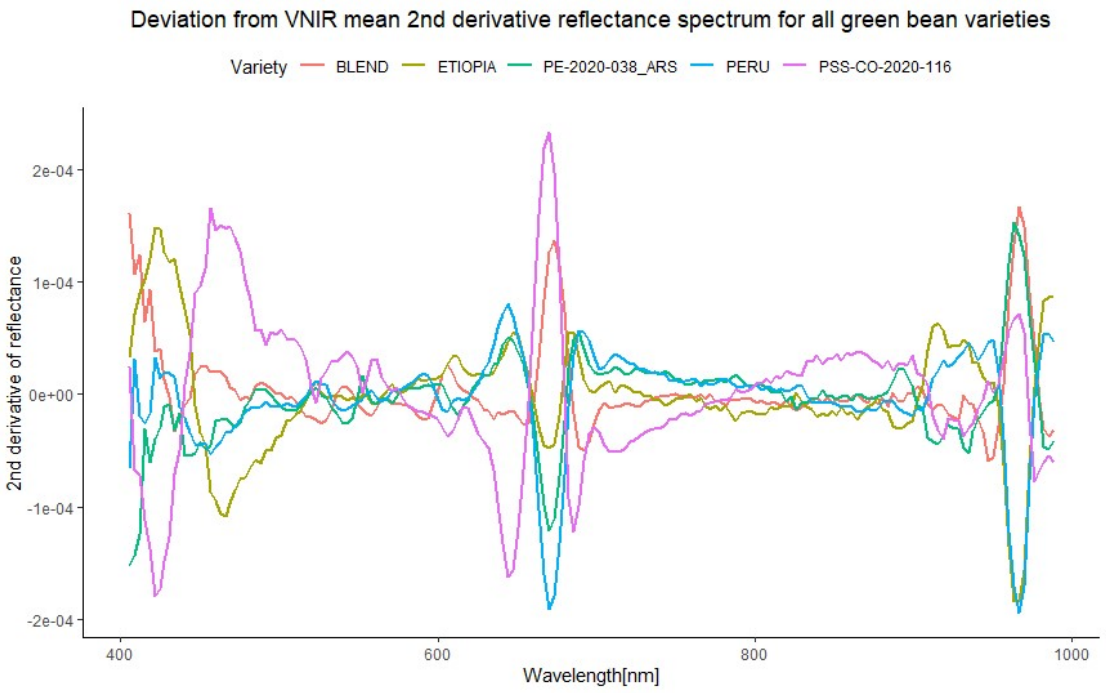


Figure 5.9: Deviation from the mean R_{green} VNIR 2nd derivative reflectance spectrum for all green bean varieties. The spectra are hard to distinguish, but easier than for figure 5.8.

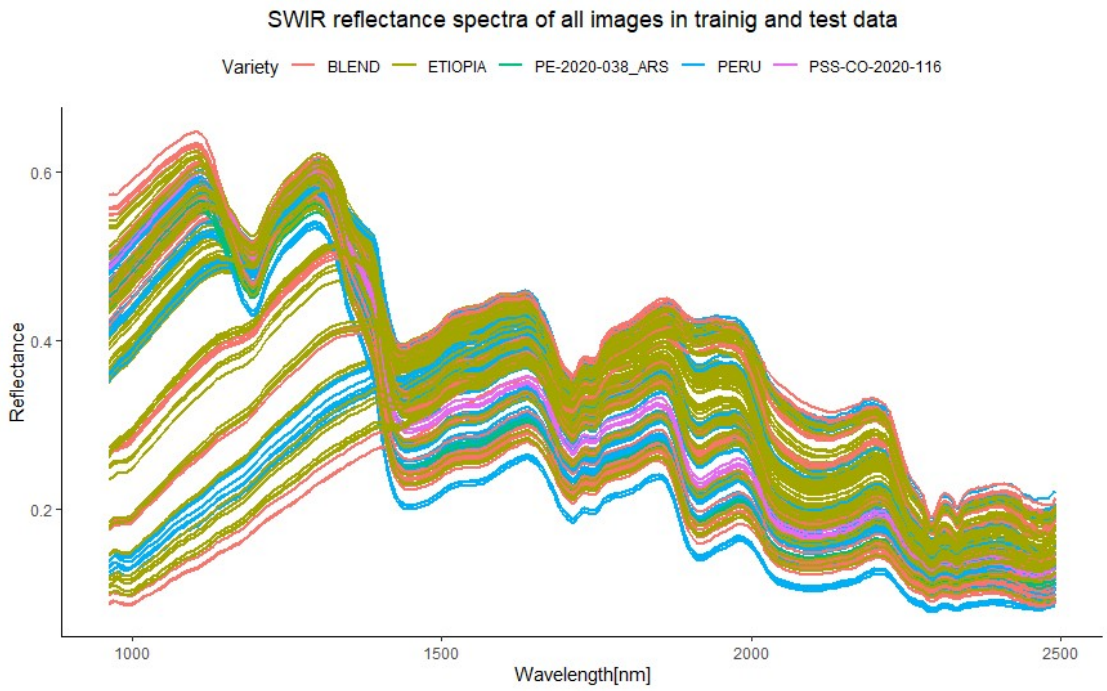


Figure 5.10: SWIR reflectance spectra of all images in the test and training set are plotted with varieties in different colours. The spectra have similar shape, but have a large variation.

and figure 5.17 with 5.16. This indicates that the models have not learned to pick up the most distinct differences between the roasts.

5.4.7 Applicability with Already Existing Models

We have made spectral prediction models predicting the spectral ranges 962-2493 nm and 405-995 nm. The error of our prediction models are not uniformly distributed over the spectral ranges they predict. To determine how well our models perform in synergy with the models of Nogales-Bueno et al. [2020] we need to look at how good they are at predicting the frequencies that are used by Nogales-Bueno et al. [2020], as shown in figure 5.18. Our errors are larger in the relevant region than otherwise, but so is the variance in the response, which means lowers the effect on R^2. To make models with synergy with Nogales-Bueno et al. [2020] one should make models using an error metric which punishes errors according to the weights of the latent vectors of the PLSR models of Nogales-Bueno et al. [2020]. However, our end goal is making a prediction model that can be used to predict all of the SWIR and VNIR spectrum, so that it may be used with any chemometric model. If the error is low enough it is no longer a point in training our prediction models for synergy with specific models.

5.4.8 Point Spectroscopy vs. Hyperspectral Imaging

From the standard deviation bands in figure 4.1 we see that very little extra information is captured by drawing more samples from a roast. The pixel-wise standard deviation is around

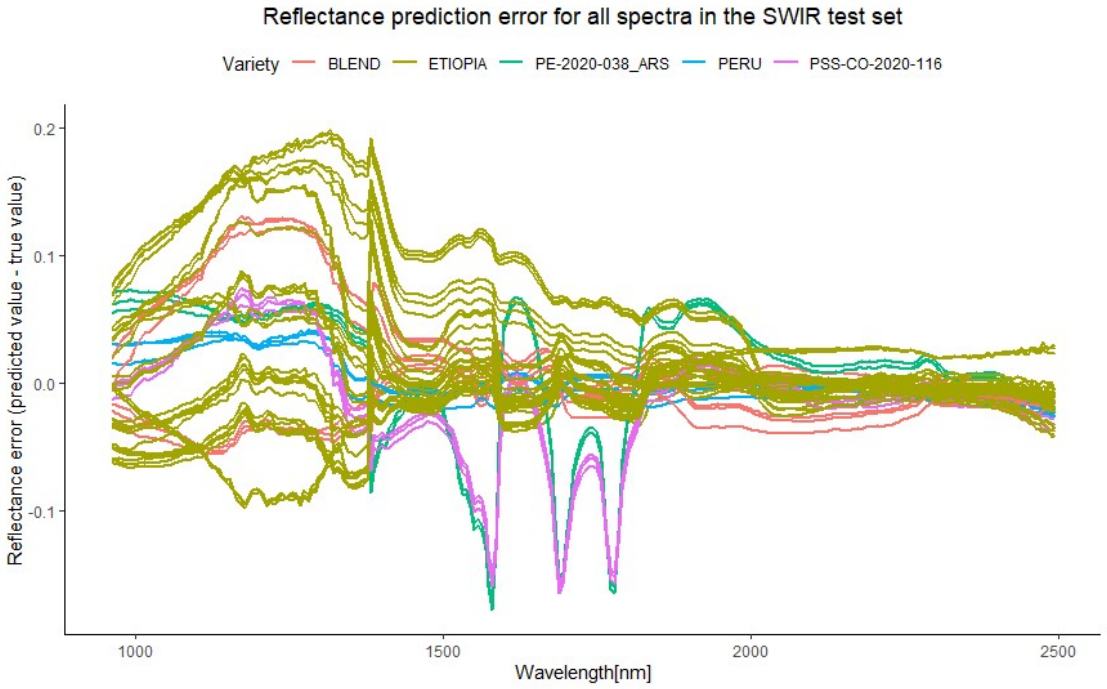


Figure 5.11: Reflectance prediction error for the basic linear regression model on the SWIR test data is plotted with different colours for each bean variety. We see that the error primarily comes from the 1000-1400nm region, which is also most important for chemometry.

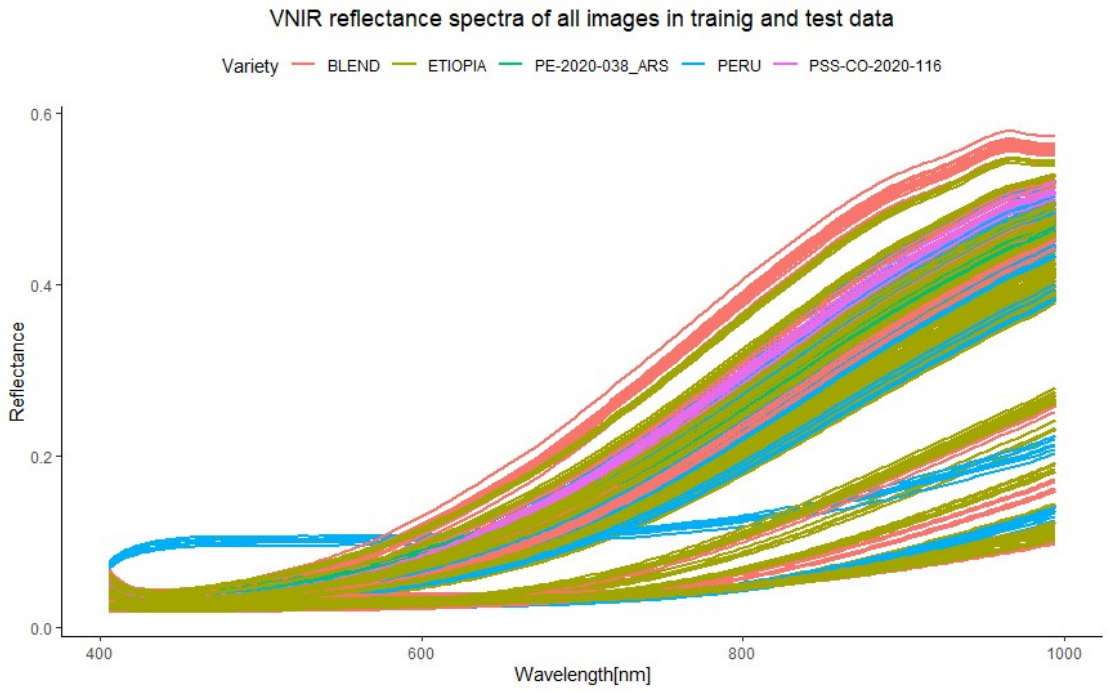


Figure 5.12: VNIR reflectance spectra of all images in the test and training set are plotted with varieties in different colours. We see that spetra of one of the PERU samples are consistently different from the rest of the spectra. We have inspected the regions of interest and the masks for the images used to calculate the mean spectrum of the beans without finding any errors. The spectra are similar in shape with large variation.

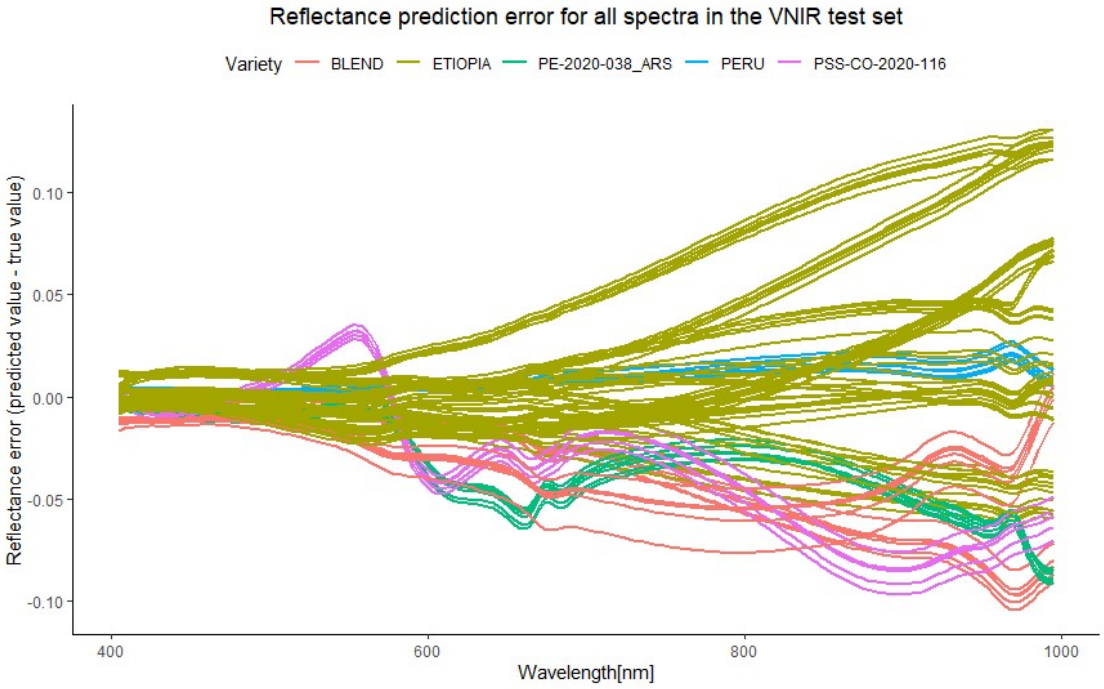


Figure 5.13: Reflectance prediction error for the basic linear regression model on the VNIR test data is plotted with different colours for each bean variety. We see that the error is larger for high frequencies, which makes sense, since the reflectance spectrum is highest in this area.

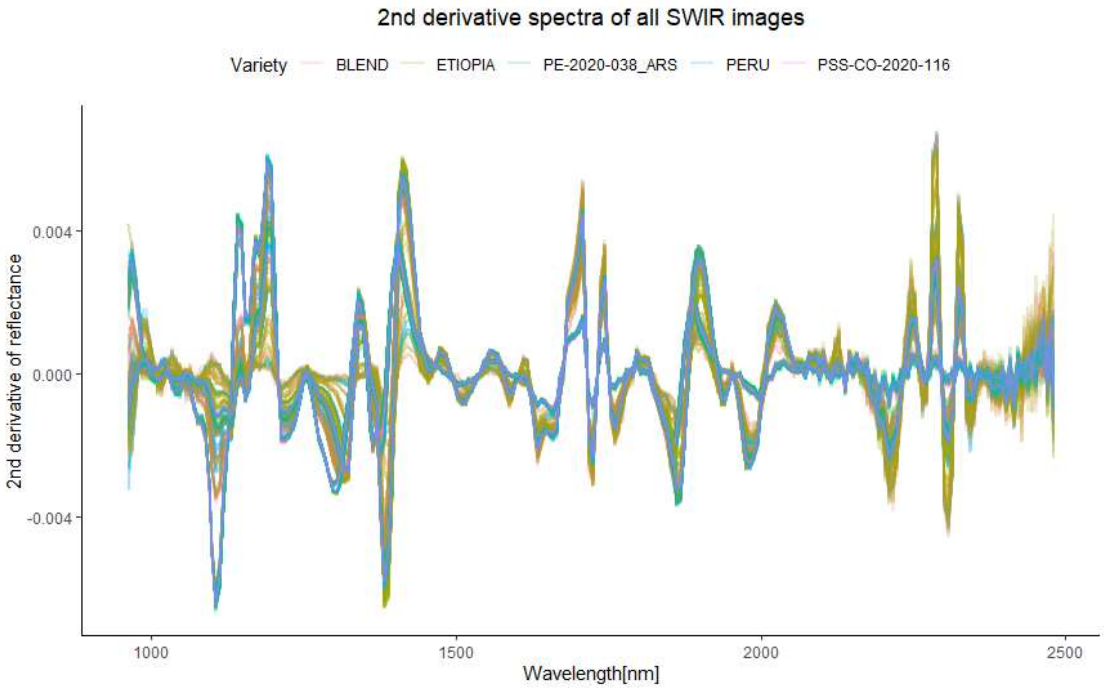


Figure 5.14: 2nd derivative reflectance spectra of all SWIR data. The spectra are similar in shape with a large variation.

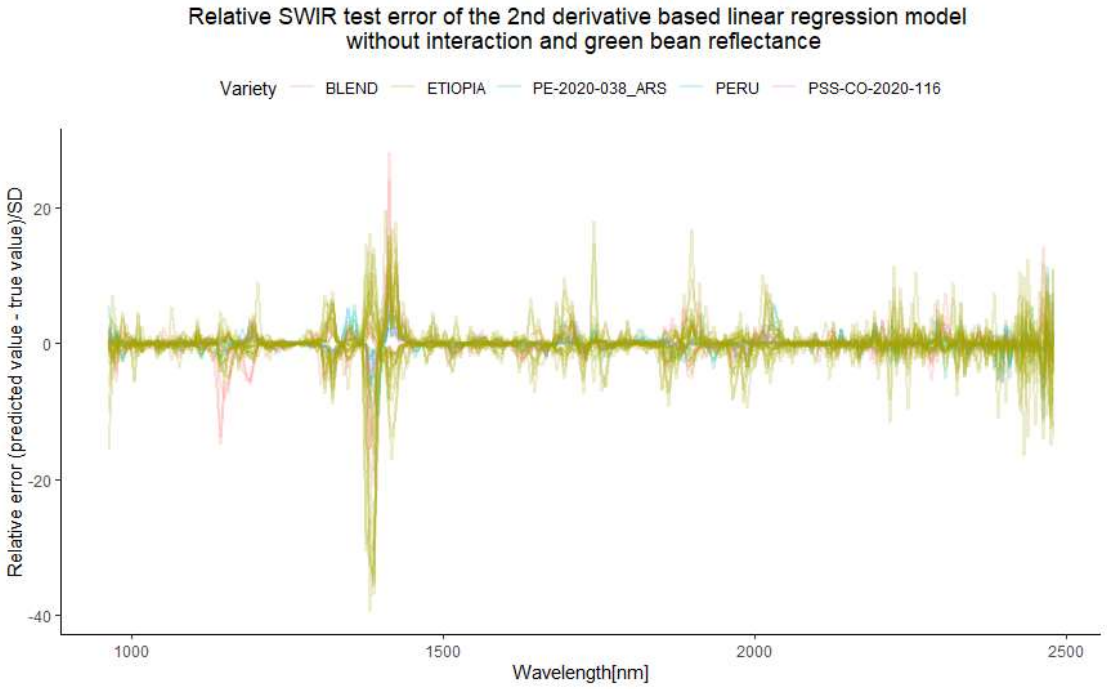


Figure 5.15: 2nd derivative reflectance spectrum prediction error of the basic linear regression model on the SWIR test data. The errors are largest where the variation on the 2nd derivative reflectance data is largest.

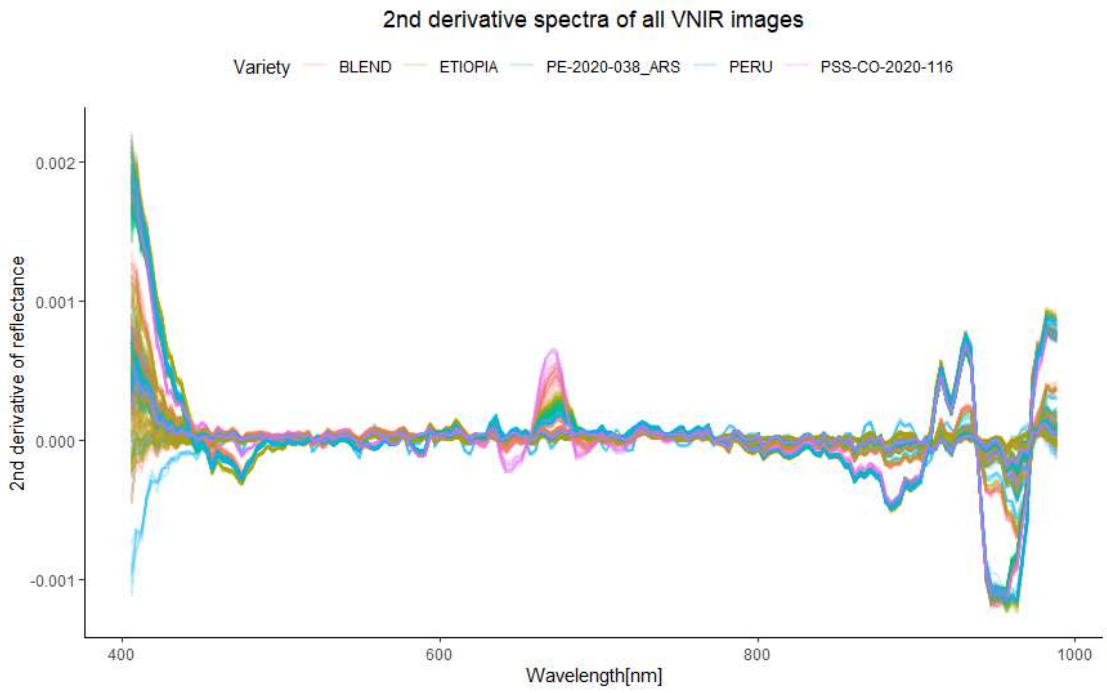


Figure 5.16: 2nd derivative reflectance spectra of all SWIR data. The spectra are similar in shape with a large variation.

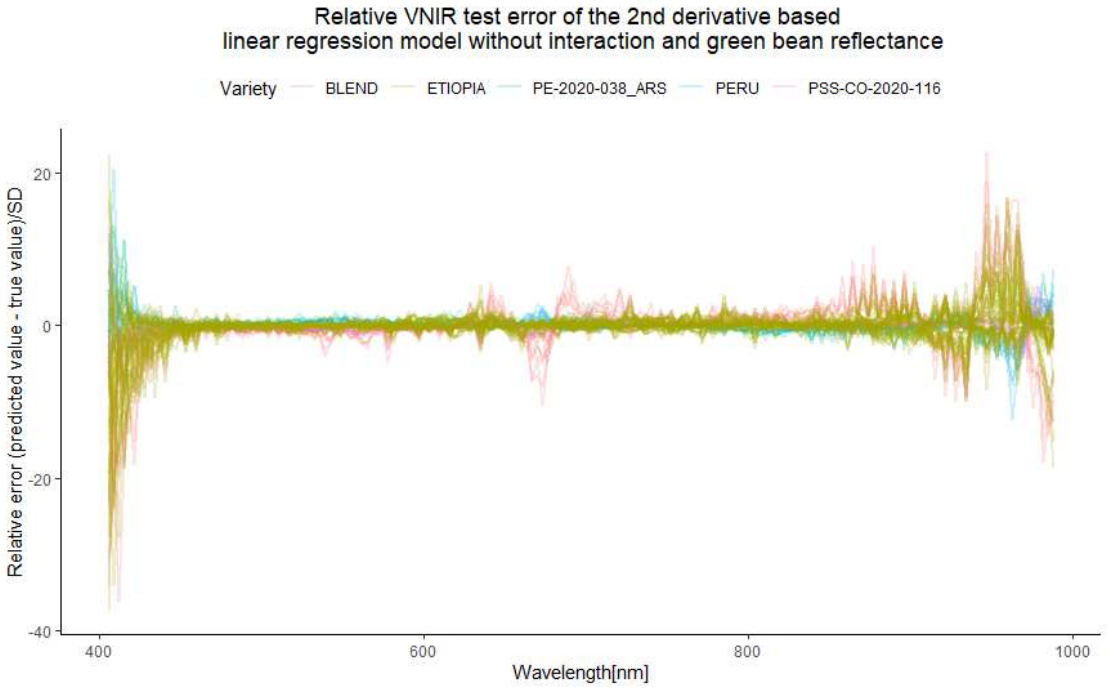
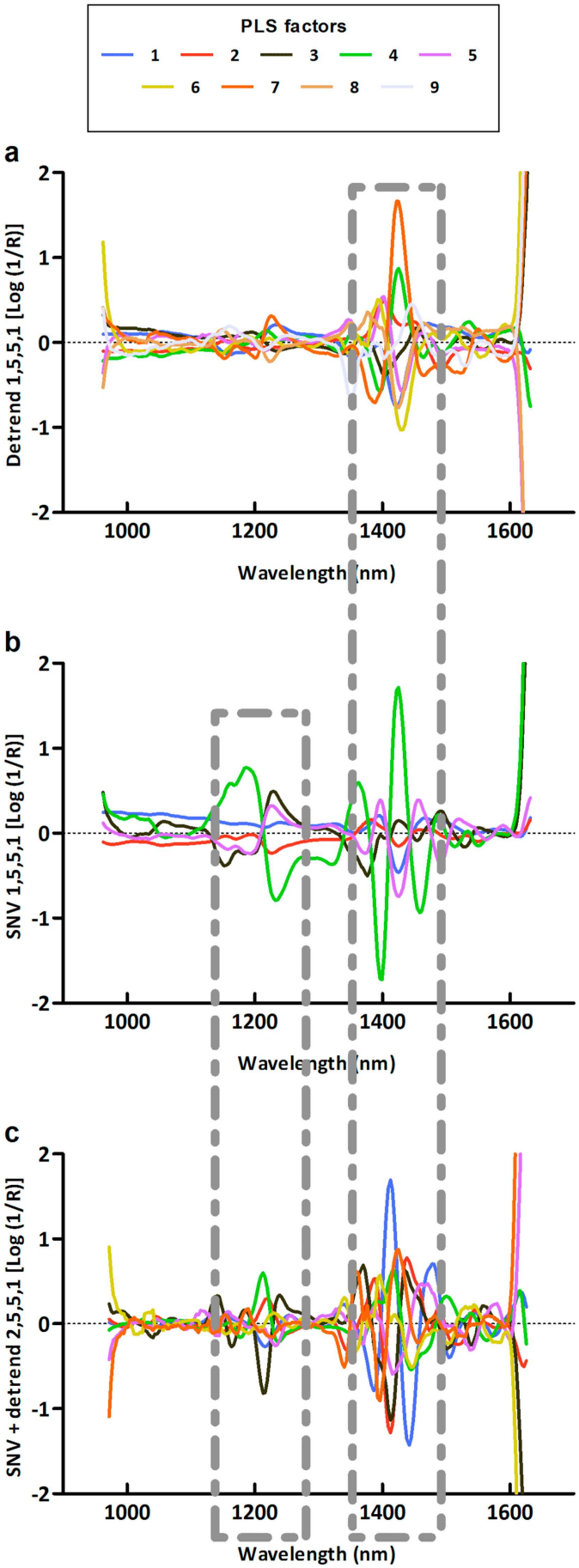


Figure 5.17: 2nd derivative reflectance spectrum prediction error of the basic linear regression model on the VNIR test data. The errors are largest where the variation on the 2nd derivative reflectance data is largest.


PLS factors
1
2
3
4
5
6
7
8
9
a
Detrend 1,5,5,1 [Log (1/R)]
Wavelength (nm)
b
SNV 1,5,5,1 [Log (1/R)]
Wavelength (nm)
c
SNV + detrend 2,5,5,1 [Log (1/R)]
Wavelength (nm)

0.07, so one could naively sample $\frac{0.07}{0.0051}^2 = 189$ pixels from the image to shrink the variance to be lower than the between-roast variance, however pixels close to each other are not independent, so to get a representative image the area covered by point measuring would have to be large to lower the variance sufficiently.

5.4.9 Sticky Beans

During roasting, we find that up to 0.3% of the beans from each roast stuck in the roasting chamber and are not released until after the next roast, as seen in figure 5.19. We do our best to get rid of double roasted beans, but for similar consecutive roasts, they might be too similar to discern. The double roasted beans we discover are from dark roasts prior to the roast we find them in. We suspect that the dark roasted beans get stuck in the roasting chamber because of the sticky oils on their surface, which are not present for light roasts.

5.4.10 Thoughts on Spectral Calculations

We started detecting regions of interest with the microscope setup, where we imaged a monolayer of beans. In later experiments, we increased the bean dose in each dish so that there are much smaller patches of not-bean pixels in each image. We kept finding regions of interest for comparability, even though it might be better not to. We had to adjust the intensity threshold based on the roast degree of the beans doing our best to optimize the trade-off between capturing the whole bean and leaving out the background. Averaging the whole image would lead to more reproducible results.

5.5 Conclusion

We have come closer to a priori chemometrics of coffee roasts. Most importantly, we have found that:

- The spectral standard deviation between different orientations is 78% of the spectral standard deviation between different samples of the same roast.
- Roasting for a specific time gives lower variance in the raw spectrum than roasting until first crack.
- The spectral variance between different roast with the same roast profile and bean variety is [list of variances]
- We can predict the 962-2493 nm reflectance spectrum of roasted coffee with RMSE = 0.039 and $R^2 = 0.67$
- We can predict the 405-995 nm reflectance spectrum of roasted coffee with RMSE = 0.038 and $R^2 = 0.74$
- We can predict the 2nd derivative of the 962-2493 nm reflectance spectrum of roasted coffee with RMSE = 3.39e-4 and $R^2 = 0.61$
- We can predict the 2nd derivative of the 405-995 nm reflectance spectrum of roasted coffee with RMSE = 4.09e-5 and $R^2 = 0.58$

Figure 5.19: Some beans get stuck in the roasting chamber and come out much darker than the other beans in the next roast.

5.6 Clustering

We did not separate material structures using clustering, indeed we did not separate bean and background. A reflectance threshold was able to separate beans and background.

5.7 Within-Roast and Between-Roast Variance

Both within-roast variance and between-roast variance are small enough to allow for accurate prediction of roast spectra when using averaged hyperspectral images. Using pixel sized point spectroscopy would have given an almost 200 times greater variance.

5.8 The Best Roasting Method

The most important conclusion we can make from our experiments is that roasting for a specific time gives more consistent results in terms of reflectance spectrum than roasting using the FC method. Roasting based on time should be used when trying to make reproducible results, while the FC method is useful to help find a good roast time.

5.9 A Priori Spectroscopy

Our prediction models can predict raw roast spectra in the near-infrared and UV-visible based on very few samples and only three roast parameters, generalizing to two unseen bean varieties.

5.10 Future Work

We captured a limited data set in our work, made no use of spatial information, did not compare our clustering results to a naive baseline, and did not draw samples of pixels from our images to test the performance of virtual point spectroscopy. To actually do a priori spectroscopy our models have to be combined with a chemometric model.

5.10.1 Combine our Model With Chemometric Models

The predictions from our models can be used as input to a vast range of chemometric models, this is what we intend our work to be used for.

5.10.2 Sample the Images

To get a proper measurement of how many pixels are needed to reduce the variance of an image to a desired threshold our pictures can be used to sample random pixels and check it empirically. If we want to evaluate a point spectroscopy method with points of a given size we can also use sampling.

5.10.3 Establish a Baseline

We should establish a baseline by making predictions based on averaging the whole image. Averaging the whole image might be better than regions of interest first if enough of the image is covered by beans. Predictions based on whole image spectra should be used as a baseline to compare any region of interest spectra to.

5.10.4 Make Use of Spatial Information

We have not made use of the spatial information in our hyperspectral images. Different parts of the beans are affected differently by roasting. For example, we know that the silverskin of the beans is shed during roasting. If we can classify different parts of the beans before and after roasting to compare how they are affected, we might make better predictions of roast spectra.

5.10.5 Gather More Data

We suspect that our data set is too limited to make much better predictions of roast spectra based on image averaging. More data should be collected to make better prediction models, especially if higher complexity is required. More roast profiles seem to be more important than more varieties, since our models generalized well to new varieties.

Bibliography

Abreu, G. F., Borém, F. M., Oliveira, L. F. C., Almeida, M. R., and Alves, A. P. C. (2019). Raman spectroscopy: A new strategy for monitoring the quality of green coffee beans during storage. *Food Chemistry*, 287:241–248.

Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(10):281–305.

B.G. Osborne, T. Fearn, P. H. (1993). *Practical NIR spectroscopy with applications in food and beverage analysis*. Number 2. Longman Scientific and Technical, Harlow, UK.

Davrieux, F., Berika, A., Raverdy, J.-L., and Excoffon, S. (2008). On-line roasted coffee quality control using nir spectroscopy. *International Conference on Coffee Science*, 22.

Huck, C., Guggenbichler, W., and Bonn, G. (2005). Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (nirs) compared to high-performance liquid chromatography (hplc) coupled to mass spectrometry. *Analytica Chimica Acta*, 538(1):195–203.

Nogales-Bueno, J., Baca-Bocanegra, B., Romero-Molina, L., Martínez-López, A., Rato, A. E., Heredia, F. J., Hernández-Hierro, J. M., Escudero-Gilete, M. L., and González-Miret, M. L. (2020). Control of the extractable content of bioactive compounds in coffee beans by near infrared hyperspectral imaging. *LWT*, 134:110201.

Pizarro, C., Esteban-Dıez, I., Nistal, A.-J., and 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. *Analytica Chimica Acta*, 509(2):217–227.

Ribeiro, J., Ferreira, M., and Salva, T. (2011). Chemometric models for the quantitative descriptive sensory analysis of arabica coffee beverages using near infrared spectroscopy. *Talanta*, 83(5):1352–1358.

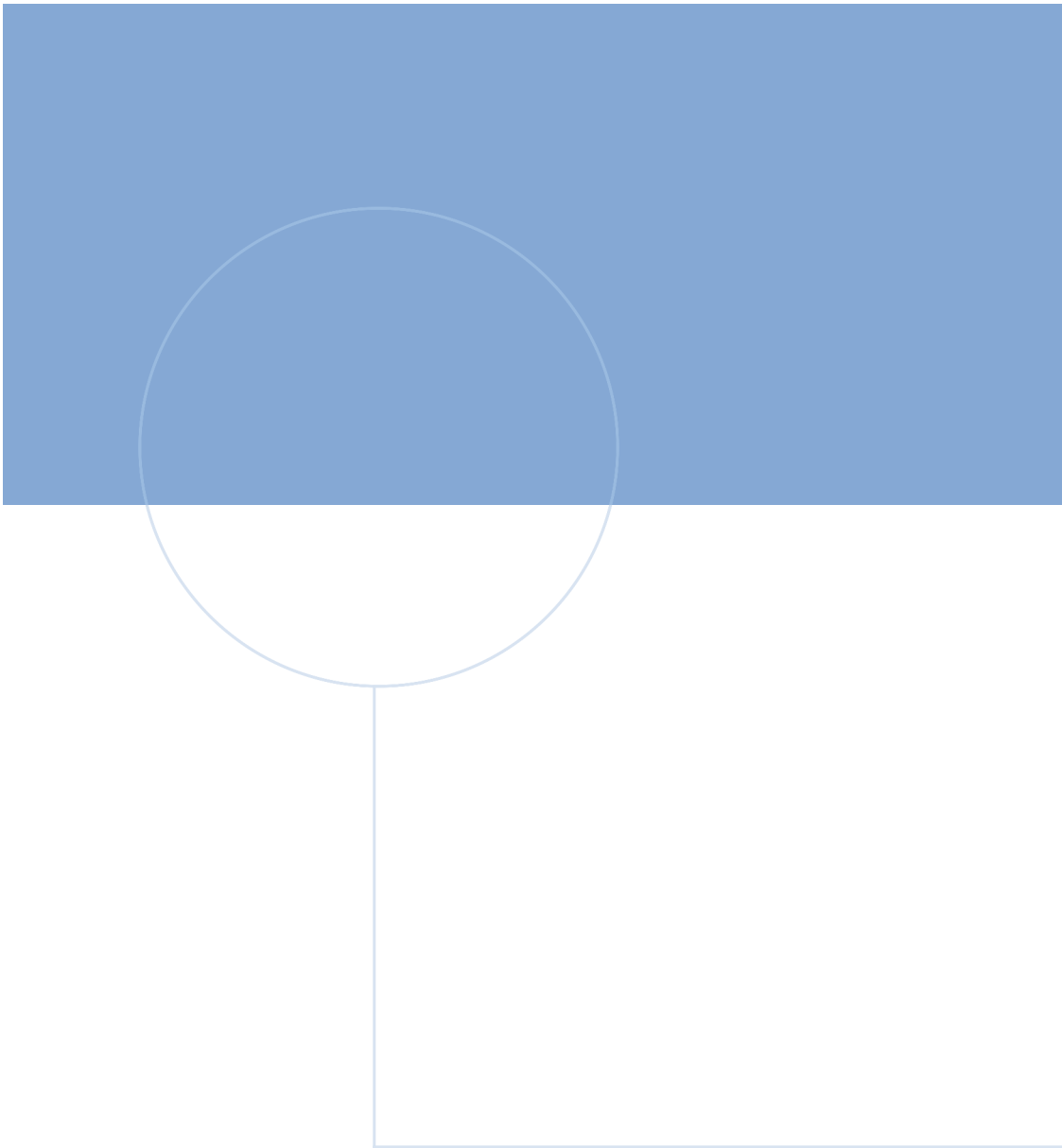
ROSS, C. F., PECKA, K., and WELLER, K. (2006). Effect of storage conditions on the sensory quality of ground arabica coffee. *Journal of Food Quality*, 29(6):596–606.

Santos, J. R., Viegas, O., Páscoa, R. N., Ferreira, I. M., Rangel, A. O., and 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.

Smrke, S., Wellinger, M., Suzuki, T., Balsiger, F., Opitz, S., and Yeretzian, C. (2017). A novel dynamic gravimetric method to assess degassing of roasted coffee. *Journal of Agricultural and Food Chemistry*, 66.

Zhang, C., Liu, F., and He, L. (2018a). Identification of coffee bean varieties using hyperspectral imaging: Influence of preprocessing methods and pixel-wise spectra analysis. *Scientific Reports*, 8.

Zhang, C., Liu, F., and He, Y. (2018b). Identification of coffee bean varieties using hyperspectral imaging: Influence of preprocessing methods and pixel-wise spectra analysis. *Scientific Reports*, 8.



NTNU

Norwegian University of
Science and Technology