

ARSE:
ADAPTIVE REGRESSION VIA SUBSPACE ELIMINATION
A NOVEL ALGORITHM FOR ELIMINATING THE CONTRIBUTION OF
UNCALIBRATED INTERFERENTS

by

Joshua Michael Ottaway

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemistry and Biochemistry

Winter 2017

© 2017 Joshua M Ottaway
All Rights Reserved

ProQuest Number: 10255104

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10255104

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

ARSE:

ADAPTIVE REGRESSION VIA SUBSPACE ELIMINATION

**A NOVEL ALGORITHM FOR ELIMINATING THE CONTRIBUTION OF
UNCALIBRATED INTERFERENTS**

by

Joshua Michael Ottaway

Approved: _____
Murray Johnston, Ph.D.
Chair of the Department of Chemistry and Biochemistry

Approved: _____
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Karl S Booksh, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Steven Brown, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Sandeep Patel, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed:

Katherine Bakeev, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I would like to extend my thanks to all those who have helped and encouraged me as I pursued my PhD. To start at the beginning I like to thank my high school chemistry, Ms. Hancock, because without her I would never have had any interest in chemistry. Next I must thank my undergraduate research advisor, Dr. John Kalivas, not only for his excellent advisement, mentorship and friendship but also for introducing me to the field of chemometrics. Last, but certainly not least, I must thank my graduate advisor, Dr. Karl Booksh, because without his mentorship, knowledge, or banter I would never survived this long. Also I would be remiss if I didn't thank the Booksh group as a whole for making sure the lab was never dull. In particular I would like to thank Dr. Devon Boyne for countless discussions, both work related and not, over the years, as well as Marcie Wiggins, William Gilbraith, Zachary Voras, and Christopher Goodwin (both Zach and Chris are honorary Booksh group members) for their willingness to have spirited discussion about every topic under the sun. Finally, a large portion of this work wouldn't have been possible without the help of my army of undergraduate: Annette Brocks, Ilkin Mahmudov, Collin Maunz, Dejuanah Collymore, Adriana Paldino, Daniel Kraiter, Clarke Snell, Tian Wei Sun, and Nicholas Favate, and high school researchers: Arelia Tapia, Paulina Cortez, Tahj Boston, Riise Taylor, Alex Muspratt, Breanna Johnson, and Luis Sedano.

I would also like to extend my thanks to both the University of Delaware Department of Chemistry and Biochemistry as well as Lawrence Livermore National

Lab (LLNL). The chemistry department has provided an extremely inviting environment for study, collaboration, and research. Additionally, the administrative staff within the department were invaluable to productivity through their knowledge, resources, and kindness. I would like to thank my committee members for donating their time, as well as their wisdom and expertise to help me through the doctoral process. Lastly, I'd like to thank LLNL for the opportunity to assist on several interesting projects as well as the ability to continue my work with them through a postdoctoral position.

My family has been a constant source of both encouragement and motivation throughout my education. I would like to express my heartfelt thanks to both of my parents, Thomas and Kimberly Ottaway, for always being there with a kind word or a kick in the ass depending on the situation. I would also like to thank my brother and sister, Matthew and Jennifer Ottaway, for their continued support and entertainment. Finally, and most importantly, I would like to thank my fiancée, Anna Murphy, for her unwavering support and insight, as well as her assistance in taking care of our beautiful child, Carlos the Turtle (who is actually a tortoise).

TABLE OF CONTENTS

| | |
|-----------------------|-----|
| LIST OF TABLES | ix |
| LIST OF FIGURES | x |
| ABSTRACT | xii |

Chapter

| | | |
|---------|--|----|
| 1 | INTRODUCTION | 1 |
| 1.1 | Motivation | 1 |
| 1.2 | Multivariate Calibration | 2 |
| 1.3 | Spectroscopy..... | 5 |
| 1.3.1 | Raman Spectroscopy | 5 |
| 1.3.2 | Near and Mid Infrared Spectroscopy | 7 |
| 1.4 | Overview of Dissertation..... | 8 |
| 2 | ADAPTIVE REGRESSION VIA SUB SPACE ELIMINATION..... | 10 |
| 2.1 | Introduction | 10 |
| 2.2 | Mathematics and the Approach | 12 |
| 2.3 | Experimental..... | 17 |
| 2.3.1 | Software..... | 17 |
| 2.3.2 | Data Sets | 17 |
| 2.3.2.1 | Data Set 1 | 17 |
| 2.3.2.2 | Data Set 2 | 18 |
| 2.4 | Data preprocessing and algorithm parameters | 19 |
| 2.5 | Results and Discussion | 20 |
| 2.5.1 | Data Set 1 | 20 |
| 2.5.2 | Data Set 2 | 29 |
| 2.6 | Conclusions | 34 |

| | | |
|---------|---|----|
| 3 | EDIBLE OIL DATA COLLECTION | 36 |
| 3.1 | Introduction | 36 |
| 3.2 | Data Set Descriptions and Motivations | 37 |
| 3.3 | Spectroscopic Techniques | 39 |
| 3.3.1 | Raman Spectroscopy | 39 |
| 3.3.1.1 | Ocean Optics Raman | 39 |
| 3.3.1.2 | Lawrence Livermore National Lab Raman | 40 |
| 3.3.2 | Mid Infrared Spectroscopy | 43 |
| 3.3.3 | Near Infrared Spectroscopy | 44 |
| 3.4 | Peroxide Value Determinations..... | 46 |
| 3.5 | Peroxide Value and Classification Data Tables | 47 |
| 3.6 | Data Collection Conclusions | 50 |
| 4 | EDIBLE OILS DATA ANALYSIS | 51 |
| 4.1 | Introduction | 51 |
| 4.2 | Materials and Methods | 52 |
| 4.2.1 | Data Sets | 52 |
| 4.2.2 | Software..... | 53 |
| 4.3 | Results and Discussion | 53 |
| 4.3.1 | Data Set 1 | 53 |
| 4.3.1.1 | Classification | 53 |
| 4.3.1.2 | Predicting Peroxide Values | 56 |
| 4.3.2 | Data Set 2 | 57 |
| 4.3.2.1 | Classification | 57 |
| 4.3.2.2 | Predicting Peroxide Values | 58 |
| 4.4 | Conclusions | 61 |
| 4.4.1 | Data Set 1 | 61 |
| 4.4.2 | Data Set 2 | 61 |
| 5 | CONCLUSIONS | 63 |

| | | |
|--------------------|---------------------------------------|----|
| 5.1 | ARSE Conclusions | 63 |
| 5.2 | Edible Oil Conclusions..... | 64 |
| 5.3 | Future Directions | 64 |
| 5.3.1 | ARSE..... | 64 |
| 5.3.2 | Edible Oils..... | 65 |
| BIBLIOGRAPHY | | 66 |
| Appendix | | |
| A | PERMISSION FOR MATERIAL REPRINT | 83 |

LIST OF TABLES

| | |
|--|----|
| Table 1: Wold's PLS Algorithm..... | 4 |
| Table 2: Results for Data Set 1 with no noise in wavelength space..... | 22 |
| Table 3: Results for Data Set 1 with no noise in wavelet space..... | 22 |
| Table 4: Results for Data set 1 with noise added in wavelet space..... | 25 |
| Table 5: Data Set 2 with Quinaldine Red as uncalibrated interferent in wavelet space | 31 |
| Table 6: Data set 2 with Quinaldine Red as uncalibrated interferent in wavelet space with added noise | 31 |
| Table 7: Data set 2 with Methyl Red as uncalibrated interferent in wavelet space..... | 32 |
| Table 8: Data set 2 with Methyl Red as uncalibrated interferent in wavelet space with added noise..... | 33 |
| Table 9: Sample IDs ,Peroxide Values, and Class IDs for University of Delaware Data Set (Data Set 1)..... | 48 |
| Table 10: Sample IDs, Peroxide Value, and Class IDs for Lawrence Livermore National Lab Data Set (Data Set 2)..... | 49 |
| Table 11: Class Identifiers..... | 50 |
| Table 12: PLS Peroxide Value Prediction..... | 57 |
| Table 13 PLS Peroxide Value Prediction (milliequivalents/kq) | 60 |

LIST OF FIGURES

| | |
|---|----|
| Figure 1: Raman Scattering Jablonski Diagram(78) | 6 |
| Figure 2: Ratio of analyte of interest pure component to uncalibrated interferent pure component in both wavelength and wavelet space | 14 |
| Figure 3: Diagram of the ARSE algorithm..... | 16 |
| Figure 4: Pure component spectra for Data Set 1 | 18 |
| Figure 5: Pure component spectra for Data Set 2..... | 19 |
| Figure 6: Histogram of errors for noiseless data set..... | 22 |
| Figure 7: Histogram of errors for 1% noise data set | 24 |
| Figure 8: Histogram of errors for 5% noise data set | 25 |
| Figure 9: RMSEP as a function of number of variables used. Inset histogram of the percentage of times a variable is chosen | 28 |
| Figure 10: Predicted vs True Y values for repeated 1% noise samples | 29 |
| Figure 11: Histogram of prediction errors with Quinaldine Red as uncalibrated interferent and no noise | 32 |
| Figure 12: Histogram of prediction error with Methyl Red as uncalibrated interferent and no noise | 34 |
| Figure 13: Raman spectra collected on Ocean Optics Spectrometer before any preprocessing..... | 39 |
| Figure 14: Ocean Optics Raman spectra after it has been baseline corrected, normalized to unit area and reduced to the area of interest..... | 40 |
| Figure 15: LLNL Raman spectra of 99 samples | 42 |
| Figure 16: LLNL Raman spectra after baseline correction, normalization and reduction to area of interest. | 43 |

| | |
|--|----|
| Figure 17: Mid IR spectra of 100 edible oil samples. | 44 |
| Figure 18: 24 mm path length NIR spectra | 45 |
| Figure 19: 8 mm path length NIR spectra (with baseline correction) | 46 |
| Figure 20: PLSDA results for Data Set 1, the black line represents the difference between the calibration set and the test set..... | 55 |
| Figure 21: Comparison between 24mm and 8 mm path length NIR spectra | 59 |

ABSTRACT

Chemometrics represents an extremely effective data analysis tool. Through its judicious application, information that would otherwise be obscured within a data set can be discovered. This dissertation investigates the development of novel chemometric algorithms for accommodating the presence of uncalibrated spectral components. It will also discuss building models for both the classification of edible oils as well as predicting the peroxide value of those edible oils.

Adaptive Regression via Subspace Elimination, ARSE, is demonstrated to be able to effectively handle the presences of uncalibrated chemical components within the prediction set. This is demonstrated first with a model system consisting of Gaussian “model spectra”. It is then expand to two different artificial data sets based upon actual pure component spectra. Across all the data sets a maximum of 4.2x improvement in prediction compared to just PLS is observed.

Also shown within this dissertation is the ability to build models to accurately predict the type of edible oil as well as the peroxide value. The classification models demonstrate an overall 92.75% accuracy. The error for predicting peroxide value depends on both the type of spectroscopy used as well as the composition of the data set. The prediction error varies from 3.60 to 8.72 on the same data set for Near IR and Raman spectroscopy, respectively.

Chapter 1

INTRODUCTION

1.1 Motivation

The field of chemometrics has been extensively studied and utilized for data analysis. Chemometrics utilizes patterns and relations in chemical data that may or may not be visible upon a simple visible inspection of the data. Chemometrics is employed in pharmaceuticals(1-10), the agricultural industry(11-19) and the technology sector(20-24) to help analyze a wide varieties of samples. This dissertation focuses on the development of a novel algorithm for chemometric modeling in the presence of uncalibrated chemical interferents.

Chemometric models are only as good as the data that is used to construct them. If a model is introduced to a sample that contains a chemical component not present in the calibration set that was used to construct that model the prediction will likely fail. There are three broad categories of approaches to account for an uncalibrated chemical intereferent. One can construct a model that contains all possible future chemical components. This is doable with very predictable systems but can quickly become if not impossible at least impractical with complicated systems. The next approach would be to construct a new model every time an additional chemical component is introduced into the system. This approach will work, however it requires extensive wet lab time for every new model, which is extremely costly and time consuming. Finally, the original model can be reduced, through some variable

elimination scheme, such that the variables contaminated by the uncalibrated chemical component are no longer present. This approach is purely data preprocessing, and therefore requires no additional spectra be collected. This final approach is the basis on which the novel algorithm Adaptive Regression via Subspace Elimination was developed, which will be a primary focus of this dissertation.

This dissertation also investigates the collection and analysis of various edible oil samples. Edible oils, e.g. Olive oil, Vegetable oil, Canola Oil, etc. have been the subject of numerous research papers (25-35). Two of the primary points of interest within the literature are the ability to differentiate different types of edible oils and determining the peroxide value of the edible oils. The identity of the edible oils is of interest due to fraud, primarily the dilution of Extra Virgin Olive Oil with lesser quality oils. The peroxide value is an indication of the rancidity of the oil, therefore a rapid determination of peroxide value via spectroscopy is a hotly researched topic(36-47). Within this dissertation techniques for building models to predict both type of oil as well as peroxide value will be discussed, with an addition study comparing the ability of various types of spectroscopy to accurately predict peroxide value.

1.2 **Multivariate Calibration**

The term chemometrics was first used by Svante Wold in a grant application in 1971(48). Since then several very detailed explanations of multivariate calibration have been developed (49-56), and a brief overview is presented here. Multivariate calibration involves relating an independent variable, such as concentration, to a set of dependent variables, such as a spectrum. This relationship is described by Equation 1-1 below.

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (1-1)$$

where \mathbf{y} represents a vector of independent variables, \mathbf{X} represents a matrix of dependent variables, \mathbf{b} represents the regression vector and \mathbf{e} represents the error.

The crucial portion of any multivariate calibration approach is the regression vector, this vector represents the mathematical relationship between your independent and dependent variables. Calculating the regression vector requires calculating an inverse for the \mathbf{X} matrix such that Equation 1-2 below can be reached.

$$\hat{\mathbf{b}} = \mathbf{X}^+ \mathbf{y} \quad (1-2)$$

where \mathbf{X}^+ represents the pseudoinverse of the \mathbf{X} matrix. Once the regression vector has been calculated it can then be applied to new unknown samples, as seen in equation 1-3, to obtain an estimate of the property of interest.

$$\mathbf{y}_{unk} = \mathbf{X}_{unk} \hat{\mathbf{b}} \quad (1-3)$$

where \mathbf{X}_{unk} represents the dependent variable of a new unknown sample and \mathbf{y}_{unk} represents the estimate of that samples independent variable.

The major difference between all multivariate modelling approaches is the determination of the pseudoinverse of \mathbf{X} . Multiple Linear Regression (MLR) requires eliminating enough variables such that there are more samples than variables, then a traditional inverse can be taken(57, 58). Partial least squares (PLS) and Principal component regression (PCR) require determining the number of latent variables or principal components to retain respectively (59-61). In Tikhonov Regularization, also known as Ridge Regression (RR), the appropriate weighting factor must be determined to stabilize the inverse (62). PLS is the approach that is used within this dissertation and therefore will be discussed in more detail

Partial least squares was first introduced by Herman Wold in 1979(63) and has been heavily used in the chemometrics community (64-73). Table 1 below shows the basic breakdown of Wold's PLS algorithm (59).

Table 1: Wold's PLS Algorithm

| |
|---|
| 1) Initialize: $y_0 = y$; $x_0 = x$; $\hat{y}_0 = 0$ |
| 2) For $k = 1$ to p do: |
| 2.1) $w_k = cov(y_{k-1}x_{k-1})$ |
| 2.2) $z_k = w_k^t x_{k-1}$ |
| 2.3) $r_k = \left[\frac{cov(y_{k-1}z_k)}{cov(z_k^2)} \right] z_k$ |
| 2.4) $\hat{y}_k = \hat{y}_{k-1} + r_k$ |
| 2.5) $y_k = y_{k-1} - r_k$ |
| 2.6) $x_k = x_{k-1} - \left[\frac{cov(z_k x_{k-1})}{cov(z_k^2)} \right] z_k$ |
| 2.7) if $ave(x_k^t x_k) = 0$ then exit |
| 3) End For |

Step 1 in the above Table serves to initialize the algorithm, where x_0 is the dependent variable, y_0 represents the independent variable, \hat{y}_0 represents the model. Step 2.1 computes the covariance w_k between y_{k-1} and x_{k-1} and then w_k is then used to form a linear combination of the x residuals in step 2.2. Then the y residuals are regressed against this linear combination in step 2.3. The results of this regression are then added to the model in step 2.4. Steps 2.5 and 2.6 represent creating the next set of y and x residuals respectively to be used for the next pass through the loop. Step 2.7 is a check to see if the residual of the x matrix has reached zero, once this happens this step terminates the algorithm. The algorithm will stop after k steps, where k is equal to the rank of the covariance matrix of xx^t (59). Once the algorithm has run its course it is up to the investigator to choose the best model.

1.3 Spectroscopy

1.3.1 Raman Spectroscopy

Raman spectroscopy is based around the principle of the inelastic scattering of light that was first proposed by Adolf Smekal in 1923(74). However it was first observed by Sir C.V. Raman in February 1928, and C.V. Raman was awarded the Nobel Prize in physics in 1930 for his experimental observation (75). Raman was further developed into a spectroscopy technique by George Placzek, a Czechoslovakian – physicist, between 1930 and 1934(76).

Raman spectroscopy allows for the probing of the vibrational modes of a sample. This is accomplished by illuminating a sample with a monochromatic light source, typically a laser, which temporarily elevates the molecule to a higher virtual energy state. If the molecule then relaxes to the same initial energy state this is called Rayleigh scatter. Rayleigh scattering must be filtered out due to its uninformative nature and the fact that it is several orders of magnitude more intense than the useful Raman scattering (77). If the molecule relaxes to a higher energy state than it was originally at this is referred to as Stokes Raman scattering, and if the molecule relaxes to a lower energy level than the original it is referred to as anti-Stokes Raman scattering. This can be seen visually in the Jablonski diagram below.

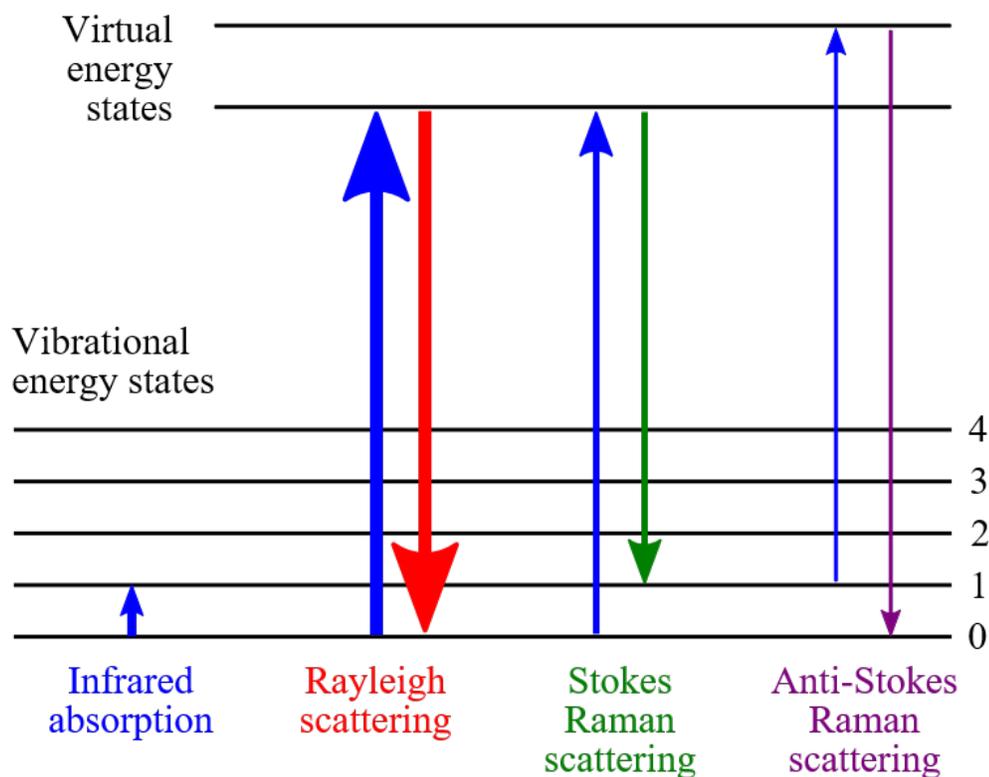


Figure 1: Raman Scattering Jablonski Diagram(78)

Raman Spectroscopy relies on the ability of a molecule to have a change in polarizability. This change in polarizability is an effect of the absorbed light causing a vibrational change with the molecule that may or may not produce a Raman active band. For example in carbon disulphide, the symmetric stretch is Raman active because it results in a shift in the polarizability, however that same stretch is not Infrared active because there is no change in the dipole moment. Conversely the asymmetric and bend of carbon disulphide are Infrared active due to change in the dipole by Raman inactive due to no change in the polarizability. This rule of mutual exclusion, i.e. within symmetric molecules no normal modes can be both Raman and

IR active (77), is what leads to Raman and Infrared spectroscopies being considered complimentary approaches.

1.3.2 Near and Mid Infrared Spectroscopy

Infrared radiation (IR) was discovered by Sir William Herschel a British Astronomer in 1800, through its effect on a thermometer (79). Mid-infrared Spectroscopy, or Mid-IR, encompasses the portion of the electromagnetic spectrum from approximately 4000 to 400 cm^{-1} . Near Infrared spectroscopy or NIR encompasses the region of the electromagnetic spectrum from 14000 to 4000 cm^{-1} . Due to this wide difference in the energy ranges covered by the two techniques they excite significantly different processes with a molecule.

Mid-IR spectroscopy is used to study the fundamental vibrations of a molecule, allowing for an understanding of the rotational-vibration structure of a molecule complementary of that of Raman spectroscopy. Mid-IR spectroscopy is accomplished by either sweeping the Mid-IR range of infrared wavelengths one at a time across the sample or passing the entire Mid-IR range at once over a sample and using a Fourier transform to obtain the spectrum. With either approach, when a molecule is hit with the appropriate wavelength to excite a change in the molecule's dipole moment radiation with that wavelength will be absorbed; and a dip or peak will be observed in the transmission or absorption spectrum respectively. Mid-IR is most commonly used for the identification of organic compounds(80-84); however, in recent years it has also been applied to a wide range of samples including paintings and other art objects(85-87) semiconductors(20-24), and food products(88-94).

NIR absorption, unlike Mid-IR, is a result of overtones and combination bands of those fundamental vibrations observed in the Mid-IR region. The first overtone bands occurs when a molecule is excited from vibrational mode $\nu = 0$ to $\nu = 2$. Subsequent overtone bands can occur but the probability isn't high for a first overtone excitation and is even lower for higher order overtones. Despite the slight difference in what is being measured NIR spectra are collected in a similar manner to that of Mid-IR spectra, just within a different wavelength range and a change of detector.

However, NIR bands resulting from overtones and combination bands are typically 10 to 100 times weaker than those in the Mid-IR region (95). NIR bands are also significantly more broad than those in the Mid-IR region (95) leading to the technique being much less popular than Mid-IR, until the widespread use of multivariate analysis lead to an increase in its usage. NIR also has some major advantages over MIR: Water is nearly transparent in the NIR, allowing for much easier sample prep in most cases and glass and quartz are transparent in the NIR, allowing for much easier sampling. Also the quantum yield of the NIR detectors is much better than those used in IR, resulting in much less noise than in MIR.

1.4 Overview of Dissertation

This dissertation will have two primary focuses, chemometric method development and spectroscopy. Chapter 2 is focused on the development of a novel algorithm allowing for prediction in the presence of uncalibrated chemical constituents. Chapter 3 focuses on the collection of two separate data sets of edible oil spectra. Chapter 4 focuses on the analysis of these two data sets. The first data set collected at the University of Delaware that will be used for both classification

modeling, determining the type of edible oil, as well as building models to predict the peroxide value of the edible oils. Whereas the second data set, collected at Lawrence Livermore National Lab, will be used not only for building models predicting the peroxide value of edible oils, but also to compare the effectiveness of different spectroscopic techniques in building those models. Finally Chapter 5 provides a conclusion as well as ideas for future directions.

Chapter 2

ADAPTIVE REGRESSION VIA SUB SPACE ELIMINATION

2.1 Introduction

Multivariate calibration models are a widely accepted means of quantitatively determining analyte properties, such as pH, oxidation state, and, most commonly, concentration (96, 97). In spectroscopy, calibration models are developed from a set of observed spectra with known reference values. These observed spectra describe the calibration space, the set of all instrumental, environmental, and chemical effects captured by the spectra. These calibration spectra and associated reference values often require extensive laboratory time and expense to collect. Due to this time and expense, many methods for updating an existing model to predict future spectra with uncalibrated interferences have been developed (98-106).

Calibration maintenance studies have been the subject of review articles (107). These methods typically fall into two broad categories: robust model building and model updating (108-117). The goal of any robust modeling approach is to construct the calibration set such that it spans all possible future chemical, environment, and instrumental interferences that may appear in future samples. This proves impractical in the majority of situations for two reasons. First it is extremely difficult to predict all the possible interferences that may appear in future samples. Second, even if one could predict all the possible future interferences one then must

build their calibration set containing all interferences in a sufficiently robust experimental design. The size of such a calibration set could quickly expand well past a practical size.

The other broad category of calibration maintenance is to update the calibration set. This can be accomplished by collecting or creating a large number of new samples containing the uncalibrated interference and augmenting the original calibration set or collecting a few samples and weighting those appropriately when augmenting the calibration set. These approaches have been well reviewed in literature but suffer the drawback of requiring reference values for those new samples (107).

Additionally, both strategies suffer from an expanding ‘interference space.’ Including more interferences in the experimental design decreases the net analyte signal (NAS). From the standpoint of NAS, an optimal experimental design would include only the interferences present in a future sample. Decreasing the NAS degrades the noise handling properties of the multivariate model. When the NAS becomes sufficiently small, the ability to reliably estimate properties of future samples is lost.

The goal of this new calibration maintenance process, Adaptive Regression via Subspace Elimination (ARSE), is not to update a calibration set but rather eliminate the contribution of the uncalibrated interference. By determining the set of variables in the test spectrum that have a contribution from the uncalibrated interference, those variables can be eliminated, and the original calibration set, minus the contaminated variables, can be reanalyzed to construct a new calibration model. Effectively, ARSE

is trading bias for an increase in variance. The variables most biased by the uncalibrated interferent are identified and eliminated. The remaining subset of variables, consequently, has diminished capacity to average the effects of random errors.

2.2 Mathematics and the Approach

The ARSE algorithm is based on the assumption that any uncalibrated interferent will contaminate a subset of the variables in a future sample. For instance, if S_c represents the multivariate space described by the variables in the calibration set, then a future contaminated sample, x_f , can be described as the set of variables that lie within S_c , x_{fa} , and the set of variables that would lie with in S_c if not for the contribution of the uncalibrated interferent (UI), x_{fb} , as seen in equation 2-1.

$$x_f = x_{fa} \subseteq S_c + (x_{fb} \subseteq S_c + UI) \quad (2-1)$$

This then becomes a combinatorics problem. If for example, the calibration set is described by 40 variables and a new sample has an uncalibrated interferent contaminating 30 of those variables, there consists one 10-variable subspace that is not contaminated by the interferent out of $8 \cdot 10^8$ possible 10-variable subspaces. Not only is this computationally impractical for an exhaustive search but it is also dependent on knowing exactly how many contaminated variables are present.

However, in many spectroscopic applications, observed interferent spectra have contributions at (most) all spectroscopically informative wavelengths in the calibration model. That is to say, the probability of finding a range of wavelengths that

are uncontaminated is low. For this reason, methods like secured-PCR (118) have not seen much practical utility.

To overcome the problems of a large search space with too few uncontaminated variables, ARSE was performed following transformation of all spectra into the wavelet domain. As described in reference 118, this is a frequency and phase transformation that preserves all the information contained within the spectra (119). This transform significantly increases the number of variables unique to our analyte of interest with respect to the uncalibrated interferent. (Figure 2) ARSE was then performed on a finite number of random subsets of variables. This allowed the computational issues to be avoided while also evaluating the usefulness of each variable in combination with many different variables.

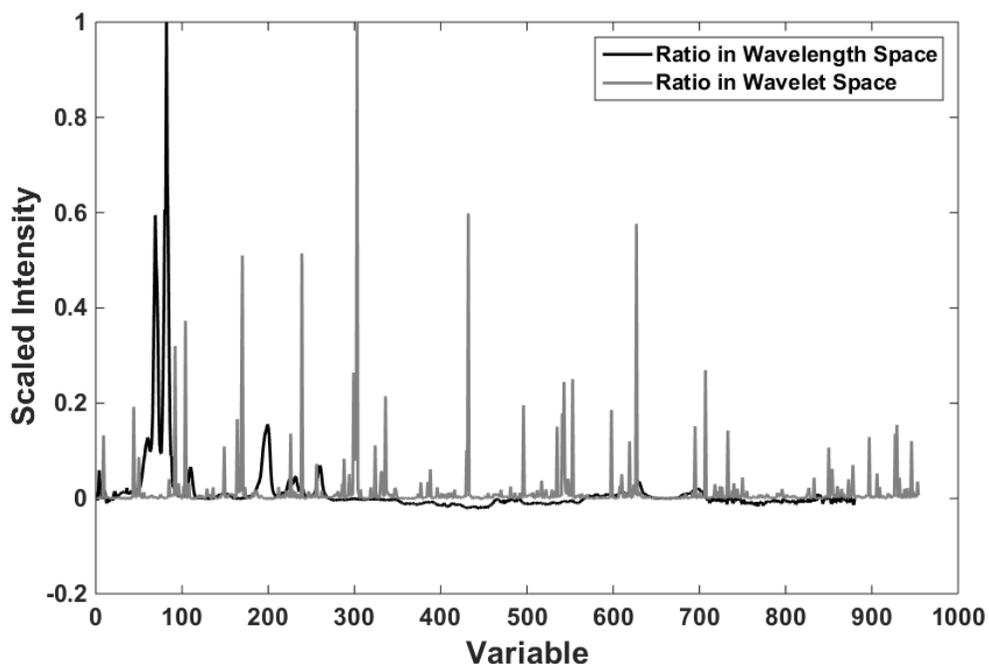


Figure 2: Ratio of analyte of interest pure component to uncalibrated interferent pure component in both wavelength and wavelet space

The goal of ARSE is determine those variables that are described by the calibration set and have predictive ability while having minimal contribution from the uncalibrated interferent. The presences of an uncalibrated interferent can be identified through the projection error,

$$ProjE = x * (I - V * V')^2 \quad (2-2)$$

In Equation 2-2, x represents a future, possibly contaminated, sample with m variables. , I represents an $m \times m$ identity matrix and V represent the eigenvectors of the calibration space. The projection error for any k -variable subspace of the calibration space can be equivalently determined by selecting any k of the m variables of x . The identity matrix becomes $k \times k$, and the eigenvector matrix, V , must be

recalculated from just the k variables in the calibration set. This distance outside of the calibration space is then a measure of the contamination within each set of k variables.

The net predictive ability of a k-variable subspace is determined by calculating the prediction error for a set of k variables

$$PredE = \sqrt{\frac{\sum(y-\hat{y})^2}{n}} \quad (2-3)$$

where y represent the calibration reference values, \hat{y} represent the PLS calculated estimate of the reference values and n represents the number of samples in the calibration set. A separate multivariate model is generated for each unique subspace analogous to moving window partial least squares regression (120) except the set of variables employed is not contiguous – and is in the wavelet domain.

After many k-variable subspaces are analyzed, the average projection and prediction errors can be calculated for each variable based on the observed projection and prediction errors of subspaces when each variable is employed. This then gives a measure of how contaminated each variable is by an uncalibrated interferent and how informative each variable is with respect to determining the analyte of interest in the calibration space.

After a large set of potential subspaces have been analyzed, the average projection and prediction errors are used to establish the best subset of variable to use to predict the sample containing the uncalibrated interferent. This is accomplished by rank ordering these two errors and then examining the union of those two ordered sets.

The union of the two sets can be expanded until the desired number of samples are present, i.e. both errors may need to be expanded until the best performing 30 variables, by each metric, are present before the union of the two sets contains five variables. Once the desired number of variables are present in the union, those variables can be selected to build a model and predict the possibly contaminated test sample. The algorithm in its entirety can be seen below in Figure 3.

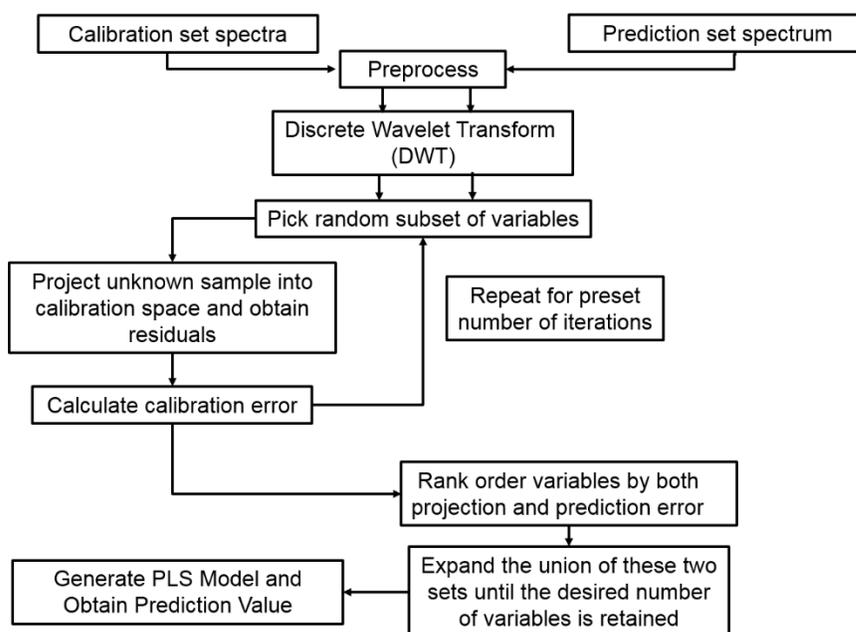


Figure 3: Diagram of the ARSE algorithm

This approach can then be repeated for each sample in the set of samples with uncalibrated interferent. By doing each uncalibrated sample independently the calibration model can be rebuilt to fit the needs of each specific sample. This allows

the algorithm to compensate for a set of samples that may have different interferences in different samples.

2.3 Experimental

2.3.1 Software

Programs for new methods were written in Matlab 8.4 (The Mathworks, Natick, MA). PLS programs were used from the PLS toolbox version 7.95 (Eigenvector Research, Inc., Manson, WA).

2.3.2 Data Sets

2.3.2.1 Data Set 1

Synthetic mixture spectra were made from pure component spectra obtained from the EPA Vapor-Phase IR Library. The pure component spectra were measured at 4 cm^{-1} resolution from 450 to 4000 cm^{-1} . The pure component spectra (Figure 4) were then used to create 40 calibration samples containing three species and 25 test samples containing the original three calibration species plus an uncalibrated interferent. Concentrations were randomly determined values between 0 and 1 and the concentration of the uncalibrated interferent in the test samples was entirely independent of the concentration of the analyte of interest.

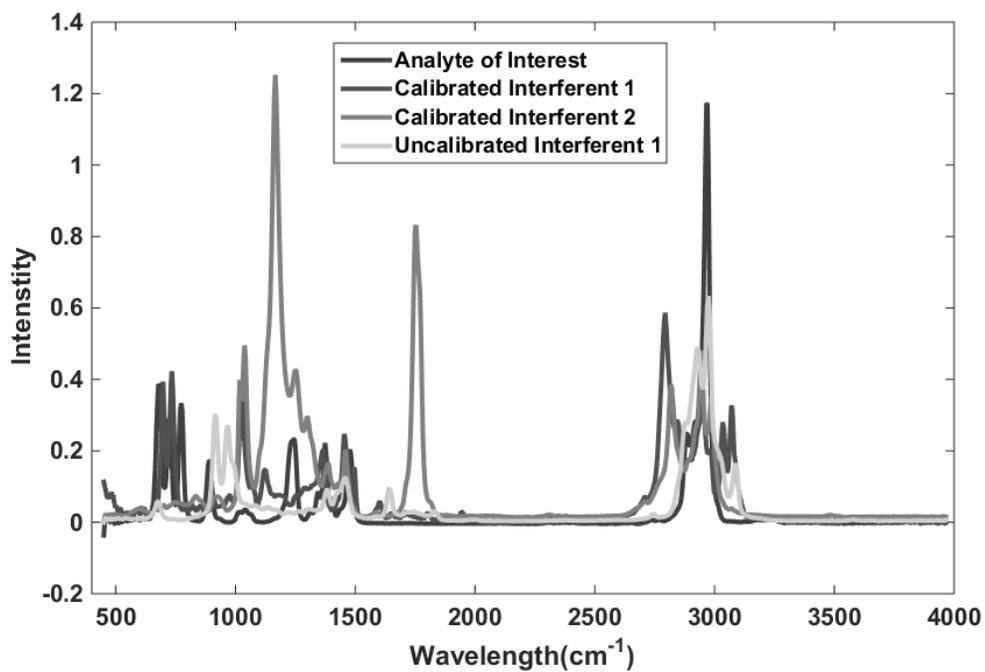


Figure 4: Pure component spectra for Data Set 1

2.3.2.2 Data Set 2

Synthetic mixture spectra using pure component UV-VIS spectra were measured in-house on a HP 8452a UV-VIS with 2 nm resolution from 190 to 820 nm. The pure components (figure 5) consist of three dyes, Eosin Y, green food coloring (Blue 1 and Yellow 5) and Rhodamine B, for the calibration set, where Eosin Y was treated as the analyte of interest. The test set consisted of two dyes, Methyl Red and Quinaldine Red, to act as different uncalibrated interferents. These were then used to create a calibration set containing 60 samples and two different 40 sample test sets with each test set containing one of the uncalibrated interferents. As with the previous set, concentrations were randomly distributed values from 0 to 1, $U(0,1)$, and the

concentration of the uncalibrated intereferents were independent of those of the analyte of interest and also followed a $U(0,1)$ distribution.

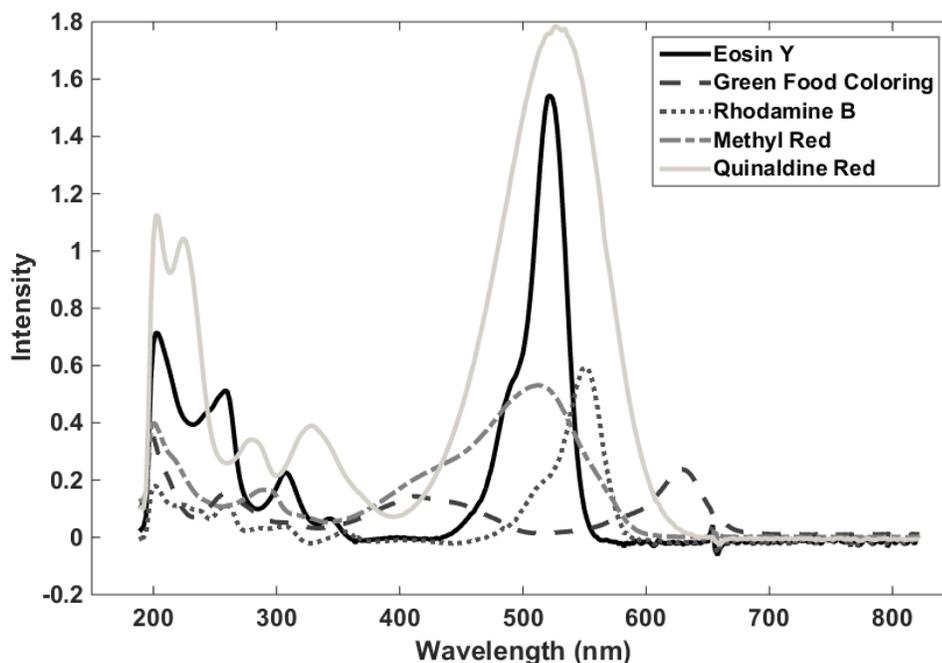


Figure 5: Pure component spectra for Data Set 2

2.4 Data preprocessing and algorithm parameters

All spectra were wavelet transformed using a symlet 8 wavelet to potentially increase the number of variables that are unique to the calibration set. All spectra and corresponding reference values were mean centered before any PLS model was built. The algorithm was allowed to run for 1 million iterations while selecting $k = 6$ random variables for each iteration. The ProjE (Eq. 2-2) and PredE (Eq. 2-3) were calculated for the 6-variable subspace. Preliminary results show that these parameters work well,

but may not be optimized for general use nor optimal for any particular data set. Further optimization and validation of the algorithm parameters is an ongoing project.

2.5 Results and Discussion

2.5.1 Data Set 1

The synthetic IR data are perfectly estimated by a three factor PLS model when no noise or uncalibrated interferent are present for calibration model construction in either the wavelength or wavelet space. When estimating concentrations in the presence of uncalibrated interferents, analysis in the wavelet and wavelength spaces perform equivalently. In the wavelength space, estimating the analyte concentration from the 25 spectra containing an uncalibrated interferent presents a root mean squared error of prediction (RMSEP) of 0.2738 for samples with a mean nominal concentration of 0.5 (Table 2). These 25 samples have a mean error of -0.2406 with a standard deviation of 0.1227 based solely on the distribution of added uncalibrated interferent. In the wavelet space, estimating the analyte concentration from the 25 spectra containing an uncalibrated interferent presents a RMSEP of 0.2786 for samples with a mean nominal concentration of 0.5 (Table 3). These 25 samples have a mean error of -0.2503 with a standard deviation of 0.1248 based solely on the distribution of added uncalibrated interferent. The distribution of prediction errors is shown in Figure 6.

However, application of ARSE performs significantly better on the wavelet transformed data than on the wavelength-space data. This is due to lack of interferent-

free variables in the wavelength space. Applying ARSE in the wavelength space does not improve the RMSEP with either a 6-variable or 12-variable model (Table 2). The RMSEP converges to the RMSEP without ARSE as more variables are included.

Applying ARSE to the noiseless data reduces the RMSEP by a factor of 5.3 for both the best 6-variable and best 12-variable ARSE models (Compare rows 2 and 3 to row 1 in Table 3). The absolute mean error and the standard deviation of observed errors are also reduced by a factor of 5.3. That is to say, the PLS model is 530% more biased without ARSE than following application of ARSE. The distribution of errors following ARSE are virtually identical for the 6-variable and 12-variable ARSE treatments (Table 3). That the bias shifts from negative to positive is a consequence of which subspace is kept. This should not be interpreted as an ‘overcorrection’ by ARSE. It happened that upon selection of variables with minimal uncalibrated interferent contribution, there was a net greater overlap with positive weighted variables in the regression vector than negatively weighted variables in the regression vector when the PLS model was rebuilt within the retained subspace.

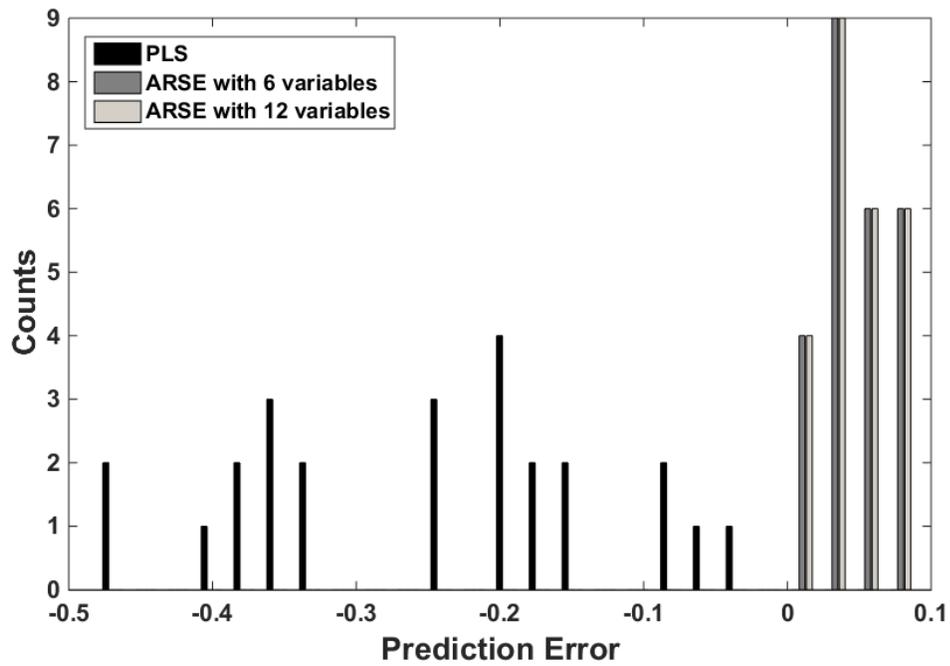


Figure 6: Histogram of errors for noiseless data set

Table 2: Results for Data Set 1 with no noise in wavelength space

| Method | Mean Error | STD | RMSEP |
|----------------------------|------------|--------|--------|
| PLS | -0.2460 | 0.1227 | 0.2738 |
| ARSE plus PLS 6 variables | -1.3785 | 0.6875 | 1.5342 |
| ARSE plus PLS 12 variables | -0.7455 | 0.3718 | 0.8298 |

Table 3: Results for Data Set 1 with no noise in wavelet space

| Method | Mean Error | STD | RMSEP |
|----------------------------|------------|--------|--------|
| PLS | -0.2503 | 0.1248 | 0.2786 |
| ARSE plus PLS 6 variables | 0.0470 | 0.0234 | 0.0523 |
| ARSE plus PLS 12 variables | 0.0470 | 0.0234 | 0.0523 |

To further demonstrate the efficacy of this new algorithm, two levels of normally distributed noise, $N(0, 1)$ were added to the calibration and test data sets. The noise levels were scaled to be 1 percent and 5 percent of the net spectral intensity of each variable in the wavelength space. The spectra were then converted from wavelength space to wavelet space prior to ARSE application.

The addition of 1% and 5% noise does not significantly impact the performance of the PLS model prior to treatment by ARSE. The RMSEP, mean bias, and standard deviation of observed biases are all within 0.2% of the values obtained by PLS analysis of the noiseless data. (Compare Table 4, first row to Table 3, first row) The errors of analysis are dominated by the bias derived from the uncalibrated interferent, not the added random spectral noise.

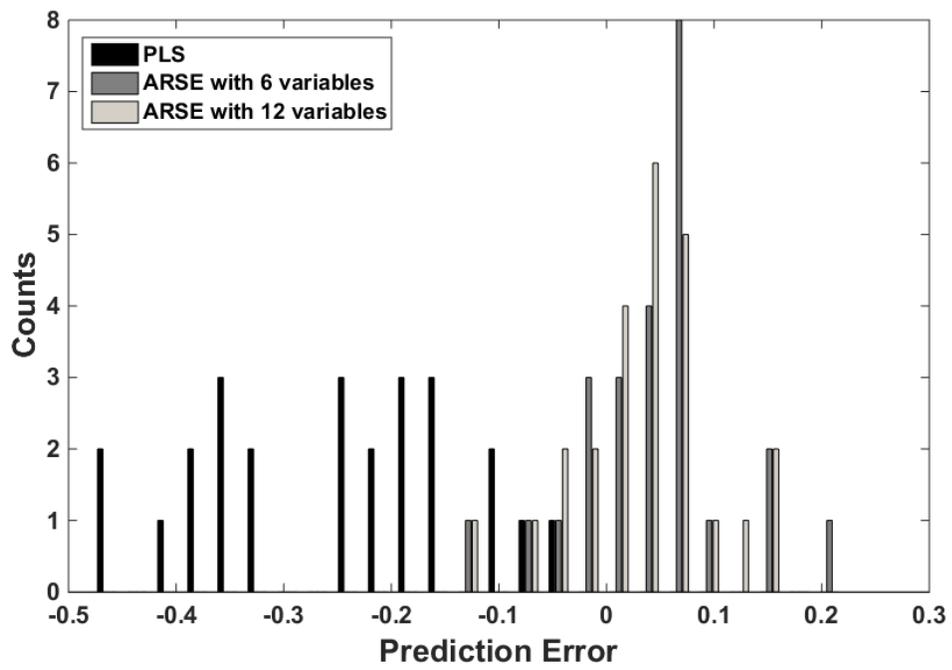


Figure 7: Histogram of errors for 1% noise data set

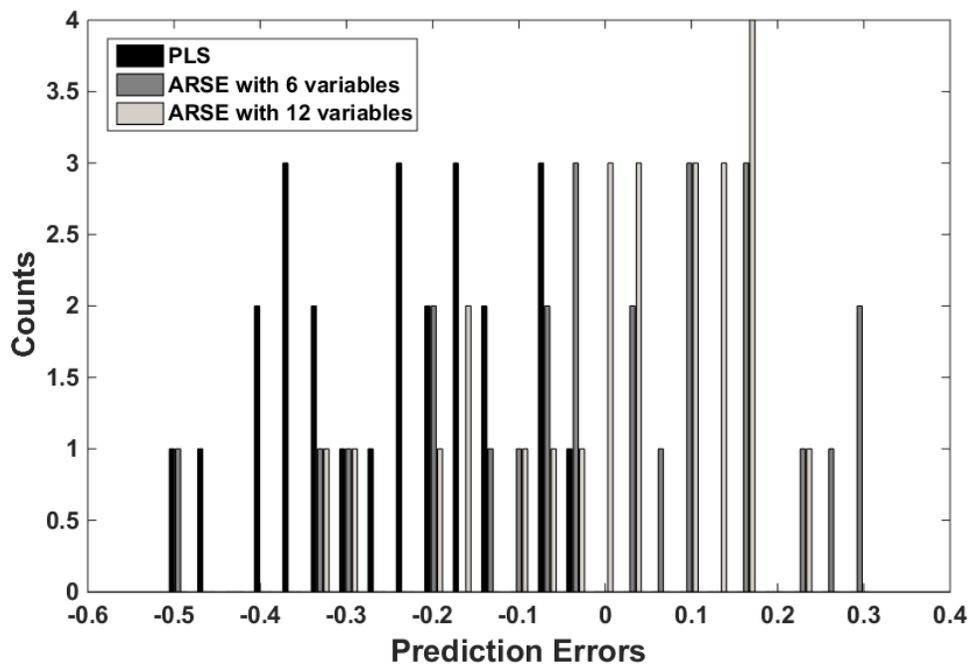


Figure 8: Histogram of errors for 5% noise data set

Table 4: Results for Data set 1 with noise added in wavelet space

| Method | 1 Percent Noise | | | 5 Percent Noise | | |
|------------------------------------|-----------------|--------|--------|-----------------|--------|--------|
| | Mean Error | STD | RMSEP | Mean Error | STD | RMSEP |
| PLS | -0.2497 | 0.1247 | 0.2780 | -0.2458 | 0.1268 | 0.2754 |
| ARSE plus PLS 6 variable | 0.0448 | 0.0733 | 0.0846 | 0.0001 | 0.2028 | 0.1987 |
| ARSE plus PLS 12 variable | 0.0332 | 0.0604 | 0.0679 | 0.0111 | 0.1518 | 0.1491 |

A Monte Carlo noise sensitivity analysis highlights the impact of random errors on the ability of ARSE to determine robust calibration models. With 1% normally distributed noise, both the 6 variable and 12 variable ARSE models show improvement of both accuracy (~3x-4x) and precision (~2x) (Table 3, Figure 6). With 5% normally distributed noise the tradeoff between accuracy and precision using ARSE becomes evident (Figure 8). While a greater than 10x improvement in accuracy is realized, the precision is degraded substantially. In this example the 12 variable model yields a better precision without loss of accuracy than the 6 variable model due to signal averaging because there are at least 12 uncontaminated variables in the wavelet space. In addition to signal averaging across more variables, where appropriate, loss in precision can be recovered by averaging replicate samples.

Ultimately the accuracy and precision of ARSE is based on which variables are selected. Of interest is the effect on variable selection by the noise present. Each test set sample was replicated 20 times with different realizations of 1% and 5% noise. Due to similarities in results, only the 1% noise will be discussed. Interestingly, although each sample was analyzed independently of all other samples, the ARSE algorithm under corrected samples with low analyte concentration and over corrected samples with high analyte concentration (Figure 9). Observing the frequency usage of each variable (Figure 10) in the 12-variable model shows that 5 variables were employed at least 90 percent of the time and 11 variables were employed at least 50 percent of the time. When the 12 most commonly employed variables for the 5 lowest concentrations are used to form a model that is applied to all samples, a similar under-

correction is observed; when the 12 most commonly employed variables for the 5 highest concentration are used to form a model that is applied to all samples, a similar over-correction is observed. Consequently it is concluded that the under-/over-correction problem is a function of the variables chosen, and is not intrinsic to the PLS model or the constructed data. Disaggregating the data shows that the major difference among the variables selected for the high concentration and low concentration samples is a simple shift in the variables being chosen, i.e. for high concentration variable 937 is chosen and for the low concentration variable 940 is chosen.

A more accurate and precise model can be realized with a different selection of variables. From the histogram of employed variables (Figure 9) models were constructed using the k most frequently selected variables and applied to all the samples, each with 20 different realizations of noise.(Figure 10) The RMSEP was determined for the 20 replicates of all 25 samples (from Figure 10) as a function of number of variables used (Figure 9). Clearly, more accurate and precise models can be obtained by choosing variables in a manner that is more robust to random noise. Based on the observed RMSEP for different calibration subspaces in Figure 9, judicious selection of variables through an improved ARSE algorithm could lead to a further 50% improvement in robustness against uncalibrated interferences.

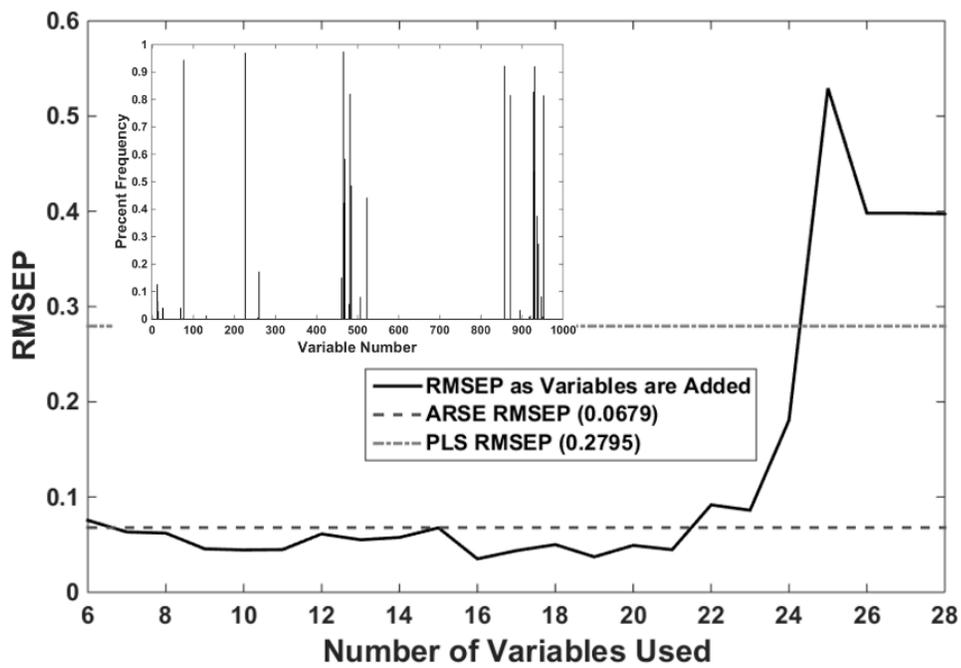


Figure 9: RMSEP as a function of number of variables used. Inset histogram of the percentage of times a variable is chosen

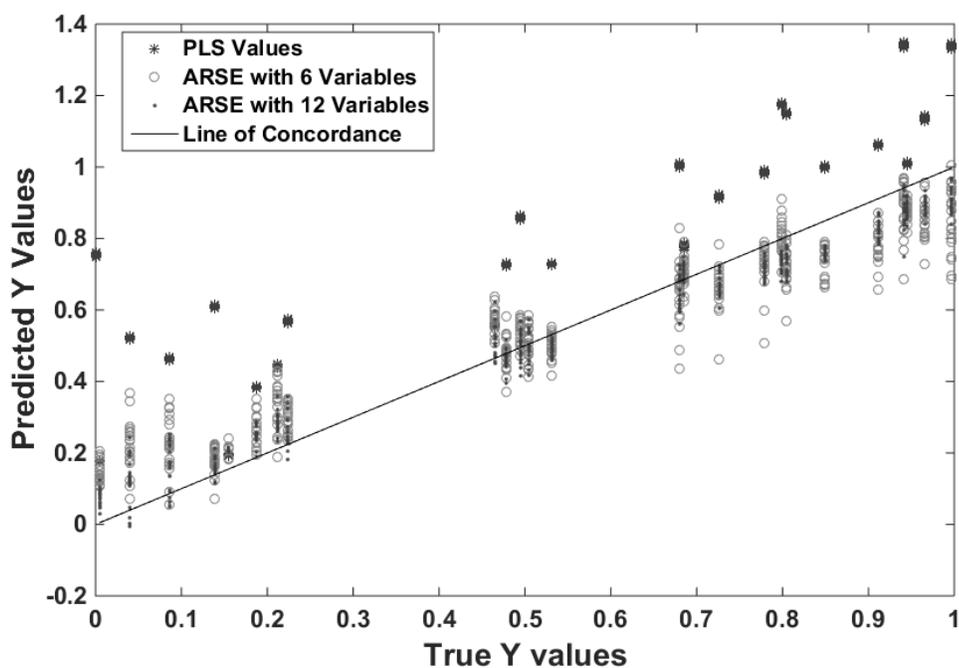


Figure 10: Predicted vs True Y values for repeated 1% noise samples

2.5.2 Data Set 2

The UV-Vis data set represents a more challenging scenario for ARSE. Not only are there no analytically useful uncontaminated variables in the wavelength space, there are few analytically useful, uncontaminated variables in the wavelet space. When building PLS models across all variables in the presence of Quinaldine Red, RMSEPs of 0.4872 and 0.4953 in the wavelet and wavelength space, respectively, were obtained; when in the presence of Methyl Red RMSEPs of 0.1841 and 0.1887 in the wavelet and wavelength space, respectively, were obtained.

Employing ARSE on the data with Quinaldine Red as the uncalibrated interferent, a 4.2x improvement is realized for the 6-variable model and 3.8x improvement is realized for the 12-variable model (Table 5, Figure 11). The ARSE models also present a 1.9x improvement in model accuracy and 2.4x improvement in model precision for the 6-variable model and 3.8x improvement in model accuracy and 4.0x improvement in model precision for the 12-variable model. When examining the data with Methyl Red as the uncalibrated interferent a slight increase in RMSEP (1.14x) is observed for the 6-variable model however a 1.8x improvement is observed for the 12-variable model (Table 6, Figure 12). The accuracy improves for both the 6-variable and 12-variable model, whereas the precision only improves for the 12-variable model. This degradation of the precision for the 6-variable model results in the higher overall prediction error when compared to the PLS model.

As with the IR data, two levels of normally distributed noise, $N(0, 1)$ were added to the calibration and both test data sets. The noise levels were scaled to be 1 percent and 5 percent of the net spectral intensity of each variable in the wavelength space. The spectra were then converted from wavelength space to wavelet space prior to ARSE application.

Again as in the IR data, the addition of noise had no significant effect on the PLS model. The RMSEP is within 0.1% of the noiseless value for the 1% noise data set and is exactly the same to 4 decimal places for the 5% noise data set (Compare row 1 in Tables 5 and 6, and 7 and 8). The model precision and accuracy also vary by, at most, 0.1% when comparing the noiseless data to the data sets with additional noise.

This again demonstrates that the error in the model is inherent to the presence of the uncalibrated interferent rather than from any additional spectral noise.

Table 5: Data Set 2 with Quinaldine Red as uncalibrated interferent in wavelet space

| Method | Mean | STD | RMSEP |
|---------------------------------|---------|--------|--------|
| PLS | -0.3976 | 0.2852 | 0.4872 |
| ARSE plus PLS 6 variable | 0.2139 | 0.1184 | 0.1184 |
| ARSE plus PLS 12 variable | 0.1039 | 0.0712 | 0.1254 |

Table 6: Data set 2 with Quinaldine Red as uncalibrated interferent in wavelet space with added noise

| Method | 1 Percent Noise | | | 5 Percent Noise | | |
|------------------------------------|-----------------|--------|--------|-----------------|--------|--------|
| | Mean Error | STD | RMSEP | Mean Error | STD | RMSEP |
| PLS | -0.3969 | 0.2885 | 0.4869 | -0.3988 | 0.2834 | 0.4872 |
| ARSE plus PLS 6 variable | 0.1303 | 0.2481 | 0.2774 | 1.1586 | 1.0178 | 1.5337 |
| ARSE plus PLS 12 variable | 0.9202 | 0.7289 | 1.1683 | 1.2792 | 1.1098 | 1.6844 |

When Quinaldine Red is the uncalibrated interferent, ARSE is able to improve the RMSEP by a factor of 1.75x for a 6-variable model on the 1% noise data (Table 5). However, for both the 12-variable model on the 1% noise data and both models on the 5% noise data there is a significant increase in all three factors of merit. This increase

is due to the inherent difficulty of this data set. The difficulty of the small number of analytically useful uncontaminated variables in this data set is further complicated by the introduction of any noise.

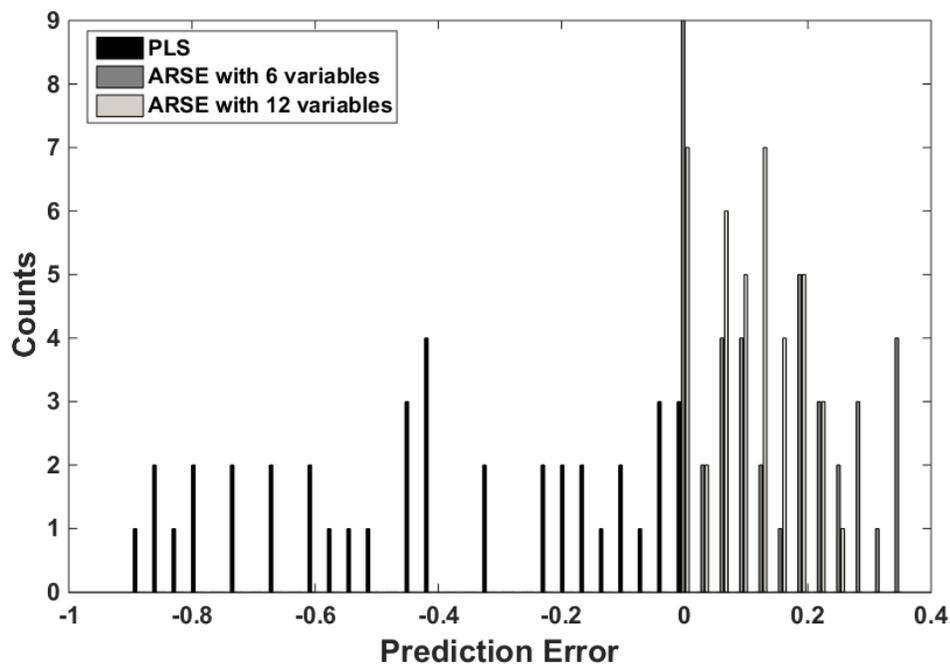


Figure 11: Histogram of prediction errors with Quinaldine Red as uncalibrated interferent and no noise

Table 7: Data set 2 with Methyl Red as uncalibrated interferent in wavelet space

| Method | Mean | STD | RMSEP |
|---------------------------------|---------|--------|--------|
| PLS | -0.1570 | 0.0973 | 0.1841 |
| ARSE plus PLS 6 variable | 0.1460 | 0.1583 | 0.2139 |
| ARSE plus PLS 12 variable | 0.0685 | 0.0761 | 0.1016 |

Table 8: Data set 2 with Methyl Red as uncalibrated interferent in wavelet space with added noise

| Method | 1 Percent Noise | | | 5 Percent Noise | | |
|------------------------------------|-----------------|--------|--------|-----------------|--------|--------|
| | Mean Error | STD | RMSEP | Mean Error | STD | RMSEP |
| PLS | -0.1568 | 0.0971 | 0.1838 | -0.1596 | 0.0992 | 0.1873 |
| ARSE plus PLS 6 variable | 0.0848 | 0.1055 | 0.1344 | 0.2465 | 0.2648 | 0.3593 |
| ARSE plus PLS 12 variable | 0.1917 | 0.1455 | 0.2396 | 0.2573 | 0.2703 | 0.3707 |

Similar to Quinaldine Red, when Methyl Red is the uncalibrated interferent ARSE is able to improve the RMSEP by a factor of 1.4x for the 6-variable model in the 1% noise data set (Table 7, 8). Again however for the 12-variable model for the 1% noise data set and both models for the 5% noise data set there is a significant degradation in the prediction error, though less than is seen in the case of Quinaldine Red.

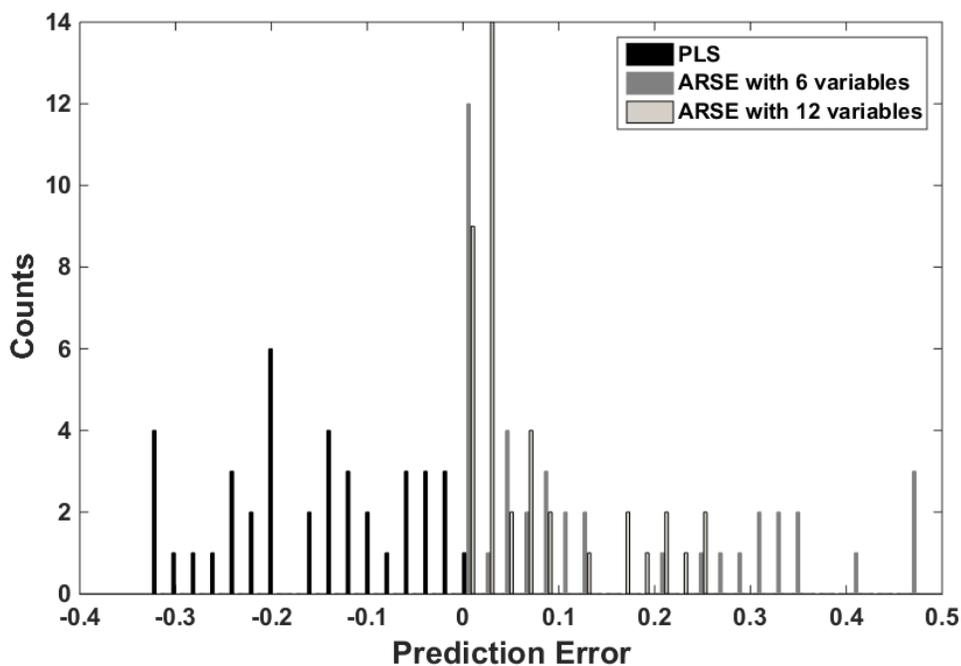


Figure 12: Histogram of prediction error with Methyl Red as uncalibrated interferent and no noise

2.6 Conclusions

The results in this chapter show that a solution to the problem of uncalibrated interferences in future samples exists in the form of determining uncontaminated variables and then re-building a model with just those variables. This approach is not without obstacles that must be overcome; first is the creation of variables that are both analytically relevant to the analyte of interest and uncontaminated by interferences. Within this chapter that was accomplished via a symlet based wavelet transform. Future work will focus on analyzing other possible wavelet families as well as wavelet preprocessing to eliminate irrelevant variables prior to the application of ARSE.

Once appropriate variables are created, the obstacle becomes selecting those variables. This work has shown that with an ARSE like algorithm it is possible to select variables and build a model that improves model accuracy with minimal increase in model precision. However the selection process must be further modified to be more robust to the effect of sample to sample variation.

Chapter 3

EDIBLE OIL DATA COLLECTION

3.1 Introduction

The health benefits of edible oils have caused them to frequently be recommended for daily consumption. One of the largest health benefits is their ability to help prevent heart disease (121). This benefit has been linked to the presence of unsaturated fatty acids in the oils (121). This has led to an increased interest in determining both the purity and level of unsaturation, or the degree of saturation, of edible oils.

Determining the purity of an edible oil first relies on being able to accurately identify an oil. This has been done in the past through various spectroscopic techniques, including MIR, Raman, and NIR(122-136); however, to our knowledge, no one has compared classification capabilities of all of these techniques on the same data set. Another gap in the classification literature is consideration of brand and co-packer differences. The standard procedure when performing an edible oils experiment for classification is to procure samples directly from one source, typically directly from the distributor, and use aliquots from those samples to make up the data set. One of the goals of this dissertation is to build classification models using samples procured from various grocery stores so as to include realistic variation from brand to brand and supplier to supplier.

Once the identity of an oil sample is established, the next step is to determine the degree of saturation, or peroxide value, of the oil. Traditionally the peroxide value of an oil is determined via the American Oil Chemists Society's procedure Cd-8b (137). This method will be discussed in greater detail later in this chapter. Another goal of this dissertation is to demonstrate the first steps in replacing this involved chemical procedure with a spectroscopic technique paired with chemometrics to greatly simplify the determination of the peroxide value.

3.2 Data Set Descriptions and Motivations

Literature has demonstrated that multiple different spectroscopic techniques are viable for both the classification of edible oils(138-144) as well as determining the peroxide value of a given edible oil(145-150). However, Raman spectroscopy proves to be the most popular for classification. This is likely due to the abundance of Raman active bonds in edible oils as well as easy of data collection with a typical dip probe Raman set up.

This ease of useful data collection led to the creation of Data Set 1, or the University of Delaware Data Set. This data set was collected via an Ocean Optics dip probe instrument detailed in the section below. The goal of this data set was to show both the classification and quantification power of dip probe Raman combined with the appropriate chemometric data pretreatment and model building which will be discussed in Chapter 4.

The goals of Data Set 2, or the Lawrence Livermore National Lab (LLNL) Data Set, was to compare both the classification and quantification power of various

spectroscopic techniques. To accomplish this, 100 samples of edible oils were titrated to determine their respective peroxide values and then were measured using Raman, NIR at 2 different path lengths, and MIR. The details of this data collection are described in the remaining sections of this chapter and the results of the comparison are laid out in Chapter 4.

Both of these data sets consist of edible oils purchased at grocery stores in and around Newark, DE. The bottles of oils have been stored at room temperature since they were purchased. The samples vary in purchase date from the summer of 2014 up to the spring of 2016. All samples were prepared for measurement by first removing an aliquot and placing it in a 20 mL scintillation vial until the measurement was made. The samples for Data Set 2 were aliquots removed from bottles in Newark, DE and then shipped overnight to Lawrence Livermore National Lab (Livermore, CA), where they were then stored in a room temperature lab until measurements were made. Peroxide value determination was done as close as possible in time to the spectroscopic measurements. For Data Set 1, all peroxide value determinations were done within 24 hours of any Raman collection. For Data Set 2, due to the logistics of the additional spectroscopic techniques, all data collection was done within a 6-day window.

3.3 Spectroscopic Techniques

3.3.1 Raman Spectroscopy

3.3.1.1 Ocean Optics Raman

This data set was collected on an Ocean Optics QE65000 Raman Spectrometer with a 785nm laser. The spectra were collected via a sapphire ball lens immersion probe. The samples were collected with a 90 second integration time with 2 cm^{-1} spectral resolution. Shown in the figures below is both the raw Raman data as well as the data once it has been reduced to the area of interest, then baseline corrected, and finally normalized to unit area. The spectra were baseline corrected by taking the reduced set of wavelengths and fitting a second order polynomial to the baseline. This polynomial was then subtracted from each spectra resulting in a flat baseline for the spectra.

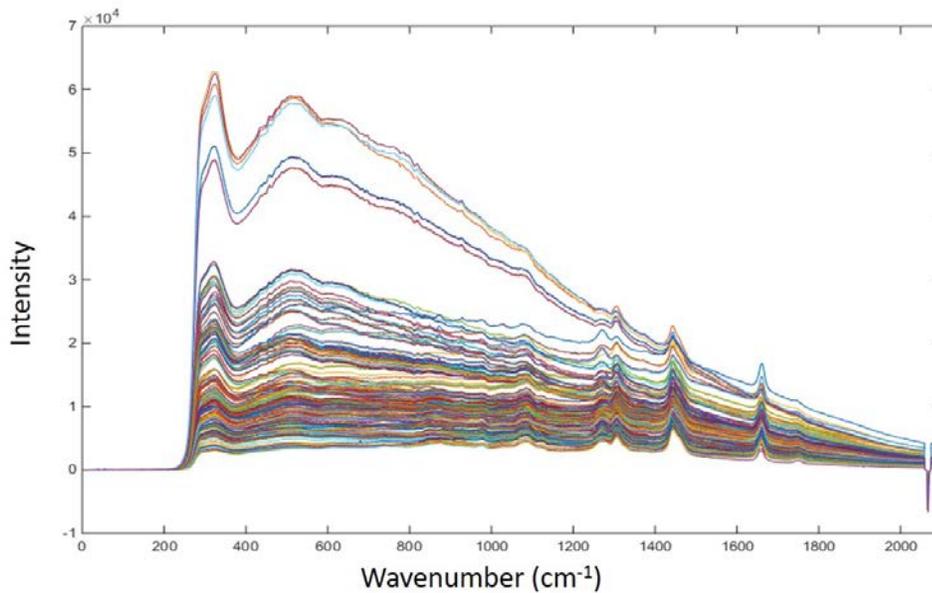


Figure 13: Raman spectra collected on Ocean Optics Spectrometer before any preprocessing

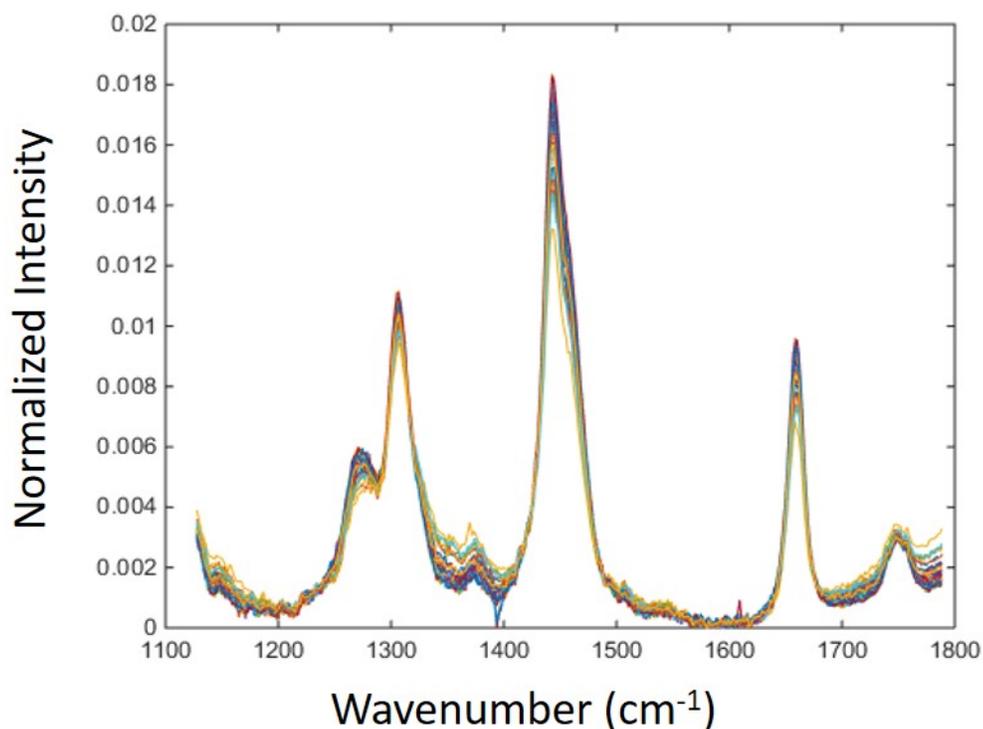


Figure 14: Ocean Optics Raman spectra after it has been baseline corrected, normalized to unit area and reduced to the area of interest.

3.3.1.2 Lawrence Livermore National Lab Raman

Raman spectra were measured with a Holospec *f*/1.8i spectrograph (Kaiser Optical Systems, Ann Arbor, MI) equipped with a liquid nitrogen-cooled charge-coupled device photodetector (Princeton Instruments Model LN/CCD-1340 BUV, Trenton, NJ). The 785 nm radiation from a diode laser (Kaiser Optical, Envictus) was used for excitation, with a maximum output power of 175 mW. A commercially available filtered fiber-optic Raman probe (Kaiser Optical Systems, Mark II probe) and spectrograph (Kaiser Optical Systems, Holospec *f*/1.8i) were used for all measurements. Both the spectrograph and Raman probe use holographic gratings and

filters that result in very high throughput and high spatial resolution with excellent rejection of scattered laser light. The excitation radiation was delivered to the sample through a 50 μm optical fiber in the probe and focused onto the sample with an attached Olympus microscope objective lens that has a focal length of 6 mm. The scattered radiation was collected by the same microscope objective lens via the probe in at 180-degree geometry and delivered to the spectrograph through a 100 μm optical fiber. No silica background was observed. The Raman spectra were obtained with a spectral resolution of $\sim 1\text{cm}^{-1}$ over a spectral range of 50-2000 cm^{-1} .

Similarly to the previous set of Raman spectra, this spectra is presented in both raw and corrected form. The spectra were treated in the same fashion as the previous set of Raman spectra; they were reduced to an area of interest, baseline corrected and then normalized to unit area. Just as before, the spectra were baseline corrected by fitting a second order polynomial to each spectrum and a subtracting that polynomial resulting in a flat baseline.

This set of Raman spectra only contains 99 of the 100 samples contained in this data set. That is a result of sample 8E, a sample of toasted sesame oil, being too fluorescent to be analyzed via Raman. This sample is still retained for the other spectroscopic techniques.

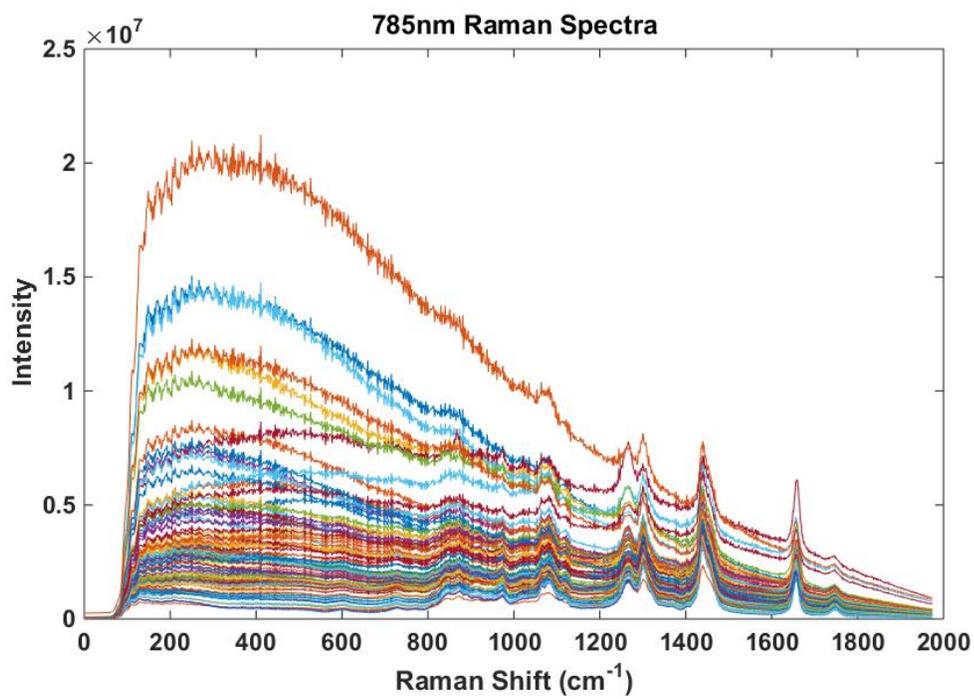


Figure 15: LLNL Raman spectra of 99 samples

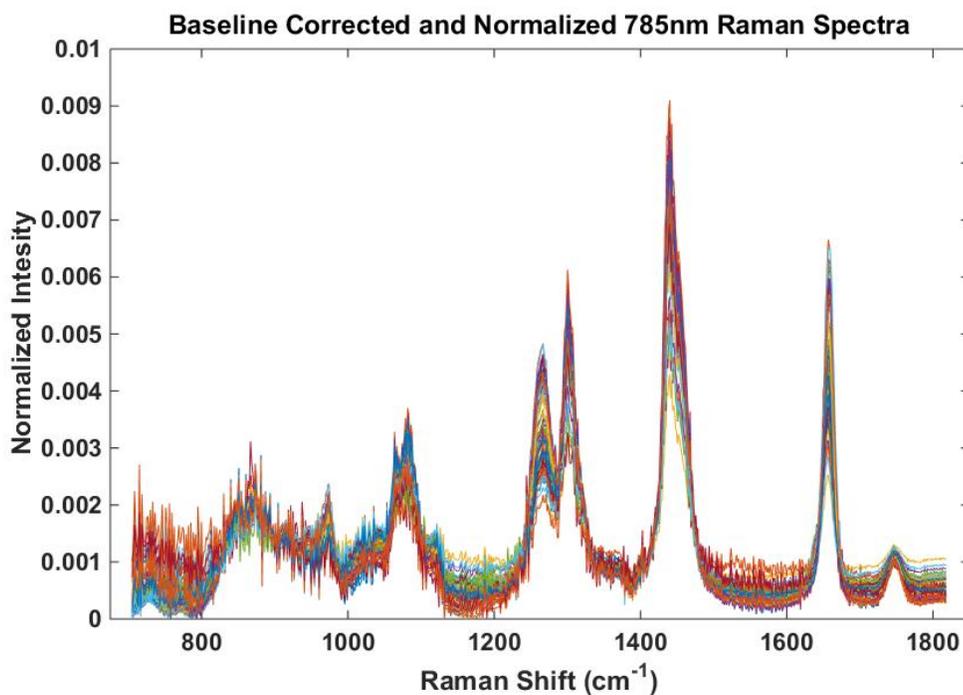


Figure 16: LLNL Raman spectra after baseline correction, normalization and reduction to area of interest.

3.3.2 Mid Infrared Spectroscopy

Mid IR data was collected on a Bruker Vertex 70 FTIR, equipped with a liquid nitrogen cooled MCT detector. The instrument was set to a 1 mm aperture using a Pike Technologies flow cell with a 50 μm path length. The spectra were collected for 256 scans with a 2 cm^{-1} spectral resolution from $697.15 - 4002.62\text{ cm}^{-1}$. Figure 17 below displays the spectra as collected.

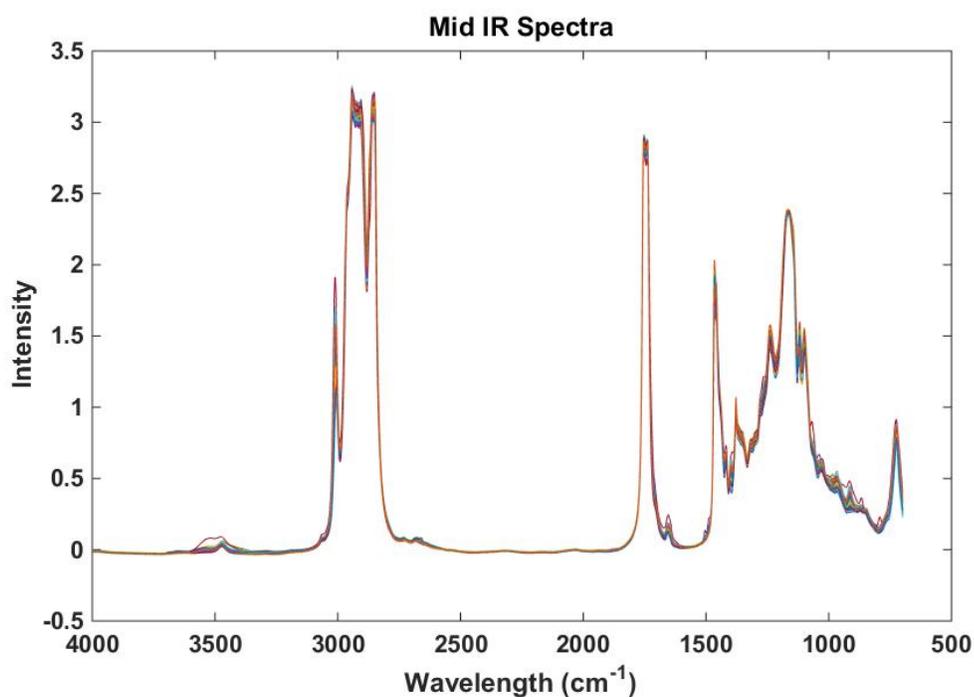


Figure 17: Mid IR spectra of 100 edible oil samples.

3.3.3 Near Infrared Spectroscopy

Both the 24 mm and 8 mm path length NIR data sets were collected at a 1mm aperture size using a room temperature indium gallium arsenide (InGaAs) detector. The data was collected for 64 scans per sample with a 2 cm⁻¹ spectral resolution. The 24mm path length samples were collected from 3797.2 – 15002.8 cm⁻¹ and were collected in triplicate for each sample. The 8mm path length samples were measured from 3895.5 – 15002.8 cm⁻¹ with only one measurement per sample. Below are figures of the 24 mm and 8 mm path length spectra. The 8mm pathlength spectra exhibited a small degree of baseline shift and were therefore baseline corrected as was previously described (via a polynomial fit), the spectra displayed below is that baseline corrected spectra.

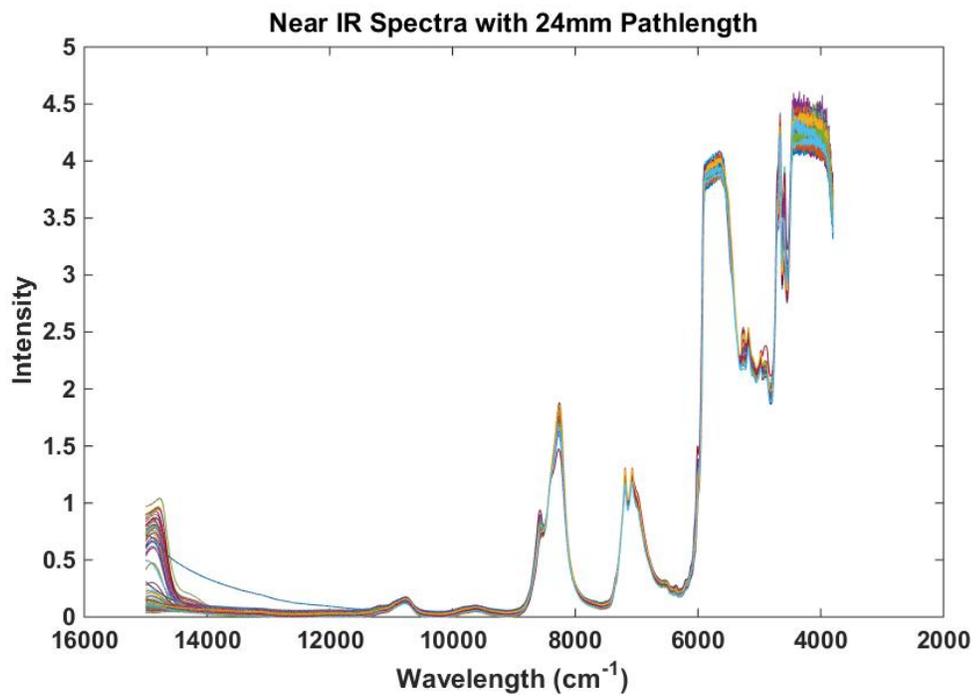


Figure 18: 24 mm path length NIR spectra

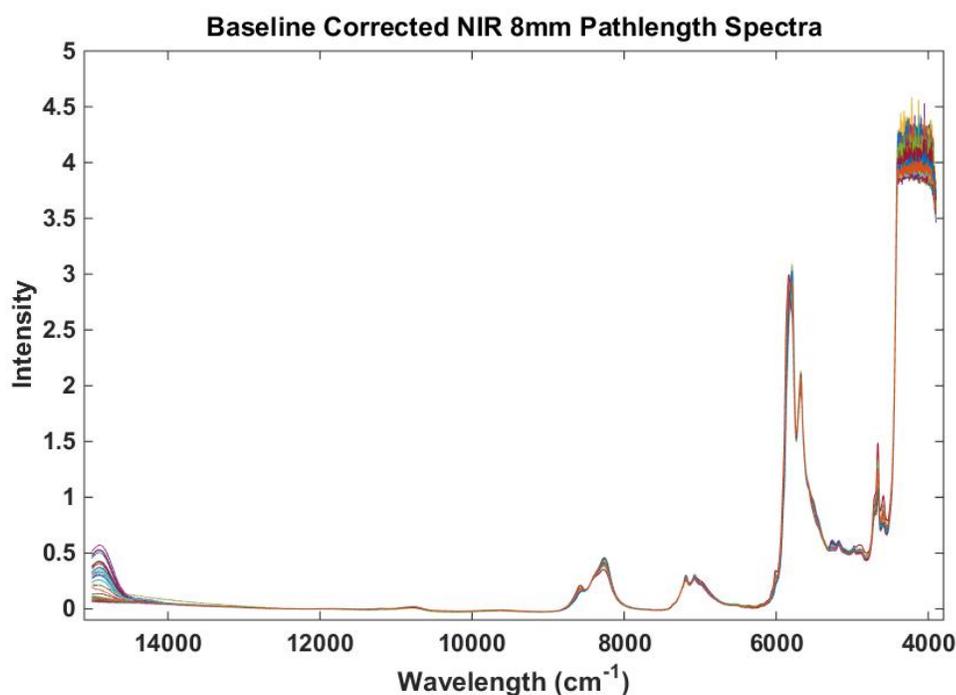


Figure 19: 8 mm path length NIR spectra (with baseline correction)

3.4 Peroxide Value Determinations

Peroxide values were determined based on the American Oil Chemists Society method Cd-8b (137). This method was accomplished via a Mettler Toledo Easy Pro automatic titrator for the data set collected at the University of Delaware and a Mettler Toledo T50 automatic titrator for the data collected at Lawrence Livermore National Lab. Regardless of the titrator used, the procedure is as follows: weigh out approximately 2g of samples, add 20mL of a 60/40 acetic acid and chloroform mixture, add 1ml saturated KI solution, stir and then let stand for 1 minute in the dark, after 1 minute add 50 mL of H₂O and begin the titration using sodium thiosulfate.

3.5 Peroxide Value and Classification Data Tables

Below in Table 9 are the peroxide value and classification identifier for all the samples in Data Set 1. As shown in Table 9, the peroxide values for the samples range from 0.46 milliequivalent/kg to 80.46 milliequivalent/kg. Any value above 10 milliequivalent/kg constitutes a rancid or spoiled oil and a value of 100 milliequivalent/kg can cause food poisoning (*151*). The data set also consists of 15 different types of oils with the largest percentage being oils which are split into classes 1, 2, and 3 for Extra Virgin, Virgin, and Pure Olive oil, respectively, as laid out in Table 10.

Similarly, Table 10 below contains the peroxide values and classification identifiers for the samples in Data Set 2. Table 10 shows that Data Set 2 displays a much larger range of peroxide values, ranging from 1.52 milliequivalent/kg to 168.40 milliequivalent/kg. This is due to the fact that the Data Set 1 consists primarily of oils purchased and measured within the same calendar year, whereas Data Set 2 has oils that are over 4 years old. Data Set 2 contains 19 different types of edible oils, again with a large percentage of the samples coming from 1 of the 3 Olive Oil categories. Finally Table 11 identifies each class of oil.

Table 9: Sample IDs ,Peroxide Values, and Class IDs for University of Delaware Data Set (Data Set 1)

| Samp. ID | Peroxide Value | Class | Samp. ID | Peroxide Value | Class | Samp. ID | Peroxide Value | Class |
|----------|----------------|-------|----------|----------------|-------|----------|----------------|-------|
| 1 | 11.2133 | 7 | 19 | 20.1733 | 2 | 37 | 12.6567 | 1 |
| 2 | 11.4567 | 6 | 20 | 32.0067 | 1 | 38 | 14.8033 | 2 |
| 3 | 4.0233 | 6 | 21 | 38.5433 | 3 | 39 | 12.5933 | 8 |
| 4 | 12.1200 | 1 | 22 | 31.2167 | 1 | 40 | 57.2733 | 8 |
| 5 | 1.0567 | 18 | 23 | 11.4900 | 8 | 41 | 26.5167 | 13 |
| 6 | 4.0833 | 1 | 24 | 6.0633 | 10 | 42 | 80.4567 | 7 |
| 7 | 2.4100 | 5 | 25 | 12.3367 | 9 | 43 | 21.3100 | 13 |
| 8 | 1.6767 | 1 | 26 | 7.8067 | 6 | 44 | 25.7933 | 1 |
| 9 | 5.2100 | 1 | 27 | 12.5933 | 8 | 45 | 5.0700 | 13 |
| 10 | 0.8300 | 13 | 28 | 17.3667 | 10 | 46 | 32.1700 | 19 |
| 11 | 39.2133 | 3 | 29 | 22.6833 | 10 | 47 | 13.4200 | 3 |
| 12 | 10.6450 | 3 | 30 | 11.8833 | 1 | 48 | 10.6367 | 8 |
| 13 | 8.3300 | 2 | 31 | 10.7367 | 15 | 49 | 2.3025 | 17 |
| 14 | 10.2967 | 1 | 32 | 11.7967 | 13 | 50 | 6.1167 | 3 |
| 15 | 2.7500 | 16 | 33 | 2.0367 | 7 | 51 | 12.8033 | 1 |
| 16 | 2.4867 | 7 | 34 | 2.0233 | 13 | 52 | 0.5433 | 13 |
| 17 | 20.8467 | 3 | 35 | 4.7475 | 2 | 53 | 0.4600 | 7 |
| 18 | 10.4367 | 17 | 36 | 11.9333 | 1 | | | |

Table 10: Sample IDs, Peroxide Value, and Class IDs for Lawrence Livermore National Lab Data Set (Data Set 2)

| Samp. ID | Peroxide Value | Class | Samp. ID | Peroxide Value | Class | Samp. ID | Peroxide Value | Class |
|----------|----------------|-------|----------|----------------|-------|----------|----------------|-------|
| 1A | 4.2823 | 8 | 5D | 8.5020 | 7 | 9G | 30.0680 | 6 |
| 2A | 12.1284 | 1 | 6D | 25.3629 | 7 | 10G | 12.5910 | 1 |
| 3A | 27.0728 | 12 | 7D | 14.8180 | 1 | 1H | 16.5277 | 8 |
| 4A | 9.2389 | 1 | 8D | 22.3575 | 3 | 2H | 11.5350 | 6 |
| 5A | 4.5023 | 1 | 9D | 13.0338 | 1 | 3H | 1.5240 | 19 |
| 6A | 8.5960 | 3 | 10D | 12.3427 | 13 | 4H | 25.1979 | 12 |
| 7A | 88.0056 | 16 | 1E | 14.8593 | 13 | 5H | 1.9079 | 7 |
| 8A | 15.5798 | 2 | 2E | 11.2809 | 18 | 6H | 1.5326 | 8 |
| 9A | 168.3691 | 1 | 3E | 5.9051 | 9 | 7H | 5.1241 | 7 |
| 10A | 5.4091 | 1 | 4E | 3.7311 | 1 | 8H | 6.6563 | 15 |
| 1B | 13.7396 | 2 | 5E | 13.1598 | 2 | 9H | 11.4082 | 3 |
| 2B | 11.0739 | 1 | 6E | 44.3814 | 1 | 10H | 26.9384 | 6 |
| 3B | 13.7396 | 1 | 7E | 20.6042 | 1 | 1I | 18.3439 | 7 |
| 4B | 7.7146 | 1 | 8E | 11.9963 | 15 | 2I | 31.0285 | 17 |
| 5B | 1.5487 | 15 | 9E | 47.0155 | 7 | 3I | 15.8701 | 9 |
| 6B | 19.6149 | 3 | 10E | 26.2818 | 17 | 4I | 6.6844 | 13 |
| 7B | 19.5255 | 13 | 1F | 10.8816 | 1 | 5I | 2.9771 | 17 |
| 8B | 37.0341 | 3 | 2F | 39.3852 | 2 | 6I | 6.2129 | 13 |
| 9B | 103.9495 | 13 | 3F | 11.9177 | 5 | 7I | 45.9692 | 12 |
| 10B | 17.5244 | 3 | 4F | 9.0931 | 1 | 8I | 34.2026 | 20 |
| 1C | 14.1921 | 1 | 5F | 5.8004 | 1 | 9I | 55.0535 | 8 |
| 2C | 4.2071 | 1 | 6F | 22.8678 | 2 | 10I | 49.1640 | 1 |
| 3C | 7.0759 | 3 | 7F | 32.8143 | 10 | 1J | 19.3281 | 6 |
| 4C | 1.7687 | 1 | 8F | 16.2763 | 13 | 2J | 11.3021 | 8 |
| 5C | 9.6059 | 1 | 9F | 22.3206 | 14 | 3J | 11.4322 | 20 |
| 6C | 4.1524 | 1 | 10F | 24.9607 | 2 | 4J | 16.3954 | 10 |
| 7C | 10.0417 | 1 | 1G | 3.8927 | 13 | 5J | 3.9732 | 12 |
| 8C | 32.2463 | 13 | 2G | 2.2956 | 6 | 6J | 22.5182 | 12 |
| 9C | 17.9176 | 1 | 3G | 5.8209 | 7 | 7J | 46.9173 | 17 |
| 10C | 8.1630 | 1 | 4G | 15.0724 | 3 | 8J | 18.1342 | 3 |
| 1D | 36.5767 | 8 | 5G | 4.5513 | 5 | 9J | 20.9850 | 20 |
| 2D | 17.3062 | 7 | 6G | 17.3479 | 8 | 10J | 10.3424 | 8 |
| 3D | 18.2620 | 7 | 7G | 6.5984 | 15 | | | |
| 4D | 27.6878 | 1 | 8G | 18.5038 | 11 | | | |

Table 11: Class Identifiers

| Class Number | Oil Type | Class Number | Oil Type |
|--------------|------------------------|--------------|--|
| 1 | Extra Virgin Olive Oil | 11 | Flaxseed Oil |
| 2 | Extra Light Olive Oil | 12 | Almond Oil |
| 3 | Pure Olive Oil | 13 | Canola Oil |
| 4 | Coconut Oil | 14 | Avocado, Flaxseed, and Olive Oil Blend |
| 5 | Avocado Oil | 15 | Sesame Oil |
| 6 | Peanut Oil | 16 | Canola and Vegetable Oil Blend |
| 7 | Corn Oil | 17 | Vegetable Oil |
| 8 | Grapeseed Oil | 18 | Canola, Sunflower, and Soybean Oil Blend |
| 9 | Safflower Oil | 19 | Sunflower Oil |
| 10 | Hazelnut Oil | 20 | Walnut Oil |

3.6 Data Collection Conclusions

Two edible oil spectroscopy data sets were collected along with their corresponding peroxide value and classification. The results of using both data sets for the prediction of both peroxide value and class will be discussed in the next chapter. Additionally the use of Data Set 2 to compare the prediction efficiency of 3 different spectroscopic approaches will be discussed in the next chapter.

Chapter 4

EDIBLE OILS DATA ANALYSIS

4.1 Introduction

As previously discussed the freshness and identity of edible oils is of significant concern in the oil industry. The main cause for loss of freshness is related to the oxidation of fats within the oil. This oxidation process starts with the creation of organic peroxides within the oil. These organic peroxides then break down into lower molecular weight compounds such as alcohols, aldehydes and ketones leading to a process known as auto-oxidative rancidity (*152, 153*). The level of these organic peroxides, or peroxide value (PV), in an oil is traditionally determined via the accepted American Oil Chemists Society method Cd-8b (*137*), a titration based method.

This titration-based method is not only time consuming, but also involves several organic solvents with varying levels of toxicity. In the interest of moving away from this method several different analytical approaches for both determining PV as well as identity of the edible oil have been demonstrated in the literature (*145-150*). Analytical techniques, such as chromatography, have demonstrated the ability to accurately analyze edible oil samples (*154-167*). However, due to the complex nature of edible oil samples, chromatography techniques are exceedingly difficult and time consuming to interpret. This reduces the practicality of using any chromatographic technique for routine quality assurance or rapid determination of the freshness of an

edible oil sample. This has been the catalyst for the development of various spectroscopy methods.

The majority of these spectroscopic approaches focus on the use of either NIR or MIR (168-174). Raman, though less investigated, has recently shown promise in both determining both PV and identity of edible oils samples (175-178). Though a search thorough literature will produce examples of using simply peak heights and ratio to determine both PV and identity of edible oils (179), the complex nature of the sample matrix typically necessitates that some kind of chemometric approach be applied.

The objectives of this chapter are to demonstrate the viability of using Raman spectroscopy for both determining the peroxide value and identity of edible oil samples when paired with chemometrics. A second goal is to perform an in-depth comparison of the most popular types spectroscopy used in edible oil research, NIR, MIR, and Raman, and their ability to predict both identity and peroxide value on the same set of edible oil samples.

4.2 Materials and Methods

4.2.1 Data Sets

As detailed in the previous chapter, this study consists of two edible oil data sets. Data set 1 was a collection of edible oils purchased in the Newark, DE area at various grocery stores. This data set consists of 53 unique oil samples of varying types and brands. Each sample was measured in at least triplicate using a 785 nm Ocean

Optics Raman system. Each sample was then titrated in triplicate using a Mettler Toledo auto titrator to determine the peroxide value.

Data Set 2 also consists of edible oils purchase in the Newark, DE area. These samples were then transported to Lawrence Livermore National Lab for data collection. Spectra of these samples were collected in the NIR and MIR using a Vertex 70 spectrometer and Raman was collected using the setup described in the previous chapter based around a Holospec *f*/ 1.8i spectrograph. The peroxide value of each was also determined using a Mettler Toledo M50 autotitrator. For more details about the specifics of either data set, refer to the previous chapter.

4.2.2 Software

All chemometric analysis was done in MatLab 2014b (The Mathworks, Natick, MA). The chemometric analysis was done using a combination of author written software and portions of the PLS toolbox version 8.2 (Eigenvector Research, Inc., Manson, WA). The specifics of the chemometric and preprocessing steps will be detailed within the results and discussion.

4.3 Results and Discussion

4.3.1 Data Set 1

4.3.1.1 Classification

Partial Least Squares Discriminant Analysis (PLSDA) was used to create a model for the predication of edible oil class. For the purpose of this study, Data set 1 was reduced to just those classes that contained 5 or more samples. This resulted in a

reduced data set that consisted of classes 1, 2, 3, 7, 8, and 13 or Extra Virgin Olive Oil, Virgin Olive Oil, Pure Olive Oil, Corn Oil, Grapeseed Oil and Canola Oil respectively. The data was pretreated as was described in the previous chapter. For the purpose of these models, all three of the olive oils were considered as the same class. This is due to the fact that despite technical definitions concerning the differences between different types of oils there is no guarantee that various manufactures and supplies adhered to the same interpretation of these definitions. Therefore this combining of classes was to eliminate any unnecessary complexity the modeling that was outside the research scope, i.e. our goal was not to determine which brands of Extra Virgin Olive Oil are actually extra virgin.

Once the data set was reduced and preprocessed it was then further broken into both a calibration set consisting of 78 spectra and a prediction set consisting of 69 spectra. This calibration set was then used to create a 6 factor PLSDA model that was then applied to the test set. As can be seen in Figure 20 below this resulted in a perfect prediction for the calibration set and 2 sets of samples, or 5 spectra, misclassified in the test set. This equates to a 92.75% accuracy for the model overall. If broken down by class, this model represents 100% correct identification of olive oils with the presence of two false positives, 100% correct identification of corn oil with the presence of three false positives, 57.14% correct identification of Grapeseed oil with three samples being false negatives, and finally 87.5% correct identification of Canola oil with 2 samples being false negatives. The lack of perfect classification of both Grapeseed and Canola oils could likely be mitigated by the presence of additional samples in both classes.

This lack of samples within a class was a consistent problem within this data set. The lack of samples was due to the desire to acquire unique samples, i.e. different brands of oils. There are simply many more unique brands and sub classes of some types of oils than others. This is most clearly illustrated by the fact that a full third of the data set is some variety of olive oil and there is little to no duplication of specific brand and type combinations.

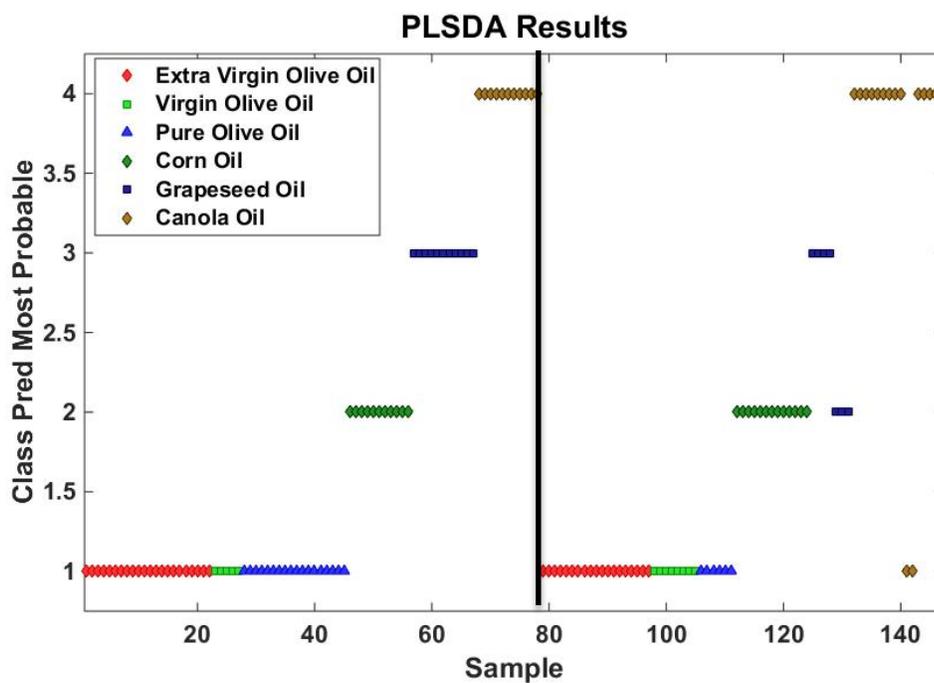


Figure 20: PLSDA results for Data Set 1, the black line represents the difference between the calibration set and the test set

4.3.1.2 Predicting Peroxide Values

Table 12 below shows the results of models built to predict the peroxide values of samples within Data set 1. Independent models were built with just Extra Virgin Olive Oil, all types of Olive Oil and all types of oil present in the data set. For the model built exclusively for Extra Virgin Olive Oil the data set was split into 30 spectra in the calibration set and 11 spectra in the test set. For all the olive oils the data set was broken into a 60 spectra calibration set and a 20 spectra test set. Finally for the model built to encompass all the oils was constructed using 165 spectra for the calibration set and 56 spectra for the test set.

As can be seen in Table 12 the test set prediction error is 5.57 for the data set comprised of extra virgin olive oil. This error can be compared to a mean peroxide value of 14.30 milliequivalent/kg for that portion of the data. The prediction error of 7.32 for all olive oil types can be compared to the mean value of 15.84 milliequivalent/kg for that portion of the data set. Finally, the prediction error of 10.08 for all errors can be compared to the mean value of 14.37 milliequivalent/kg for the entire data set. While these prediction error are comparable to what has been reported in literature (174) there is certainly room for improvement.

There is a clear degradation of the model as additional oils are introduced, shown by the increasing prediction error. This degradation is a result of both the addition complexity of attempting to model multiple types of oil at once as well as the increased spread of peroxide values as addition oils are introduced. The Extra Virgin Olive Oil portion of the data set has a minimum peroxide value of 1.68 and a maximum of 32.00, the Olive oil portion a minimum of 1.68 and maximum of 39.21

and the whole data set a minimum of 0.46 and a maximum of 80.47. This increase in spread of peroxide values coupled with the addition of various other types of oils leads to increased complexity in the model, which is demonstrated in the increase of latent variable used in to model the entire data set.

Table 12: PLS Peroxide Value Prediction

| Data Set | LV | RMSEC | RMSECV | RMSEP |
|------------------------|----|-------|--------|-------|
| Extra Virgin Olive Oil | 4 | 3.67 | 6.13 | 5.57 |
| All Olive Oils | 4 | 4.65 | 7.94 | 7.32 |
| All Oils | 5 | 11.39 | 12.74 | 10.08 |

4.3.2 Data Set 2

4.3.2.1 Classification

Classification was attempted on this data set but proved to be unsuccessful. We were unable to construct models that had better than 50% classification accuracy on the calibration set. There are multiple possible causes for this, the most likely is a simple lack of spectra and samples to define a class. With the exception of the 24mm path length NIR, each sample had a single spectrum taken for any given technique. Pair this with the fact that other than Extra Virgin Olive Oil, with 27 samples, no oil type had more than 9 samples. So this lack of data paired with variability in the samples with in a class (i.e. different brands, supplies, ages, etc.) made an already difficult chemometrics problem impossible. Of course these problems could be

remedied with the collection of additional data but unfortunately that is outside the scope of work for this dissertation.

4.3.2.2 Predicting Peroxide Values

Similarly to Data Set 1, PLS was used to build models to predict peroxide values however for this data set there were four different spectra for each sample (with the exception of 1). This data set was collected with the goal of comparing the suitability of NIR, at two path lengths, MIR and Raman to accurately predict peroxide value. For all of the sets of spectra, the samples were split into a calibration set consisting of 60 spectra for Raman, MIR, and NIR with 8 mm path length and 180 spectra for NIR with 24 mm path length, and a test set consisting of the other 40 spectra for Raman, MIR, and NIR with 8mm path length and 120 spectra for NIR with 24mm path length. For more detailed information about the data collection of each spectral set, please see the previous chapter.

Unlike the previous data set various outliers and further variable selection had to be done to this data set before prediction models were created. For the all the spectroscopic techniques those samples with larger than 60 millequivalent/kg were eliminated for their high leverage within the model. Additional samples were removed based on a PCA done on each spectroscopic technique after each data set was properly preprocessed. This resulted in 32 spectra from the NIR 24 mm data set, 11 samples from the NIR 8mm set, 8 from the Raman set and 12 from the MIR set. Also for the NIR 24mm data set, the 5200-6000 cm^{-1} was eliminated due to the saturation of the detector.

The NIR spectra were collected at two different path lengths to examine the difference in the information captured by each path length. Though the NIR spectra were collected from 3895.5 – 15002.8 cm^{-1} and 3797.2 – 15002.8 cm^{-1} for the 8mm path length and 24 mm path length respectively as discussed in the previous chapter, the spectra were reduced to 4477 to 11300 cm^{-1} for the purposes of this analysis. As can be seen in the Figure below with the 8 mm path length the peaks at 5792 cm^{-1} and 5679 cm^{-1} can be clearly resolved, whereas, in the 24 mm those two peaks combine and saturate the detector. However, the 8 mm path length spectra sacrifice the detail in the 4800 to 5200 cm^{-1} and 6200 to 6600 cm^{-1} regions that are clearly evident in the 24 mm path length spectra.

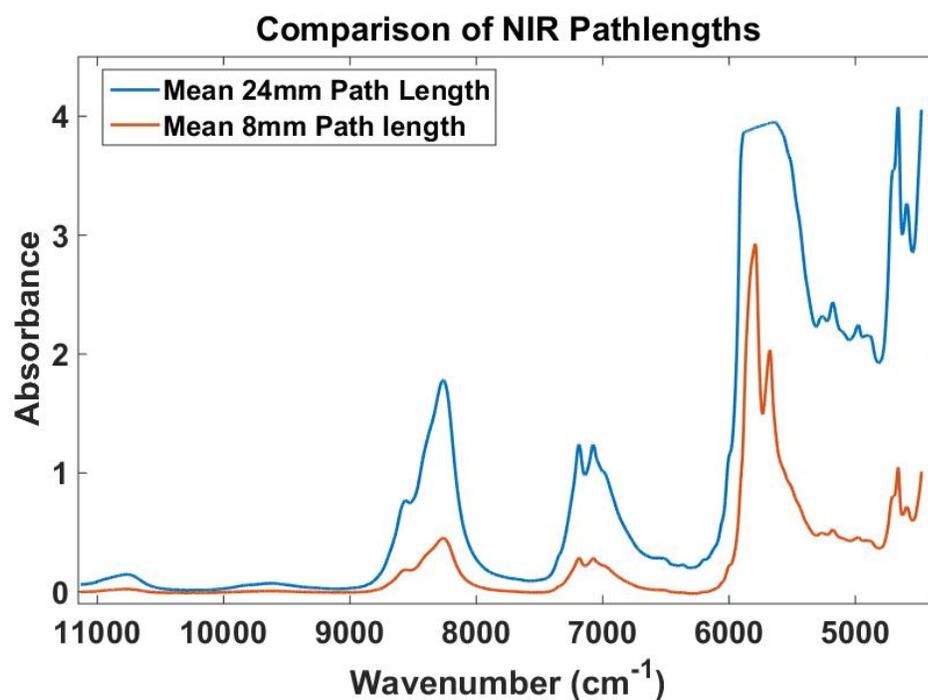


Figure 21: Comparison between 24mm and 8 mm path length NIR spectra

As can be seen in Table 13, this difference in NIR path lengths does not contribute to a significant difference in the prediction error. The 8 mm path length data set has an RMSEP of 3.93 millequivalent/kg whereas the 24 mm data set has an RMSEP of 3.60 both compared to an average of 19.61 millequivalent/kg. This lack of any significant difference in the prediction error indicates that the region from 5600 to 6000 cm^{-1} , which represent the first overtone of the C – H stretching (179) from various functional groups within the edible oils, does not play a crucial role in the prediction of peroxide value. This instead implies that the smaller peaks are more crucial for prediction.

The MIR and Raman data sets show comparable RMSEPs to that of the NIR 24mm path length data set. However both MIR and Raman data sets show RMSECV that are roughly double that of their RMSEC whereas the NIR 24mm path length data set's RMSEC and RMSECV are equivalent. This indicates that there is some instability within those two models and that, given a different data split, the prediction could get significantly worse.

Table 13 PLS Peroxide Value Prediction (milliequivalents/kg)

| Spectroscopy | # Factors | RMSEC | RMSECV | RMSEP |
|----------------------|-----------|-------|--------|-------|
| NIR 8mm path length | 7 | 3.70 | 5.37 | 3.93 |
| NIR 24mm path length | 6 | 3.60 | 3.89 | 3.65 |
| MIR | 7 | 5.87 | 10.26 | 6.61 |
| Raman | 7 | 6.83 | 11.70 | 8.72 |

4.4 Conclusions

4.4.1 Data Set 1

With this data set the ability to build models both to predict the type of edible oil as well its peroxide value was demonstrated. The classification models exhibited nearly perfect classification errors, with only 5 spectra within the 69 spectra test set being incorrectly classified via PLSDA. This did necessitate the reduction of the data set to just those classes with 5 or more spectra, however this is reasonable to ensure a class is adequately described within the calibration set. In the future with additional data the model could be expanded to include different types of oil.

The models to predict peroxide value showed clear ability to accurately determine peroxide values with RMSEPs of 5.57, 7.32, and 10.08 milliequivalent/kg for Extra Virgin Olive Oil, All types of Olive Oil, and All Oils, respectively. The quality of the prediction dropped as expected when model was expanded to include additional types of oils. This is due to not only the additional complexity of attempting to model across numerous types of oils, but also because the range of peroxide values the model must accurately cover increased as additional highly oxidized oils are introduced to the model. In the future additional oils, other than Olive oil, with higher peroxide values, could be introduced to the data set to allow for the creation of a model that would represent all oils more accurately

4.4.2 Data Set 2

Unfortunately, this data set proved ineffective for building models for classification however, it was extremely effective at building models for predicting peroxide values. This data set indicates that NIR may be the best spectroscopic

technique for the prediction of peroxide values, with an RMSEP of 3.93 milliequivalent/kg for the NIR with an 8 mm path length and 3.60 milliequivalent/kg for the 24 mm path length NIR as compared to 6.61, and 8.72 milliequivalent/kg for MIR, and Raman respectively. This data set also highlights that there is little gained from different path lengths in NIR. However, in the future additional studies could be performed with varying path length for both the NIR and MIR to more thoroughly determine path length's effect.

Chapter 5

CONCLUSIONS

5.1 ARSE Conclusions

This dissertation has demonstrated a novel chemometric algorithm, Adaptive Regression via Subspace Elimination, or ARSE. This algorithm has shown promise in its ability to mitigate the contribution of uncalibrated spectral interferences within an unknown sample. This mitigation is accomplished by determining those variables within the calibration set that are uncontaminated by contributions from the uncalibrated spectral interference in the new unknown sample and then eliminating those variables. By this approach, the original calibration data can still be used, saving time in the lab as well as money.

ARSE is not without its drawbacks, namely variable selection. The crucial component of ARSE is determining which variables within the calibration set to retain. One must ensure that the variables are not only uncontaminated but that they also retain predictive power with regards to the analyte of interest. This first entails maximizing the number of variables to choose from which is the reason wavelets are being utilized. Once the number of variables is maximized, the proper subset of variables must be retained. ARSE has shown great promise in determining which variables to retain, however future work should be done to further optimize this process.

5.2 Edible Oil Conclusions

This dissertation has also demonstrated the successful development of both classification and calibration models for edible oils. This modeling of edible oils is novel in two respects, the sampling and the data collection. To the authors' knowledge, this is the first time edible oil models have been built using such a wide range of oil types, procured in from such a wide range of sources. By obtaining the oils from various grocery stores across several different brands our models were able to incorporate much of the sample to sample variation that is missing in other studies. This also represent the first time, to the authors knowledge, that various types of spectroscopy have been applied to the same data set for the purpose of building models for the prediction of peroxide values.

5.3 Future Directions

5.3.1 ARSE

There is a multitude of ways to expand the applications of ARSE. As mentioned above, the first step would be additional work to further optimize the selection process for the retention of variables. Once accomplished, ARSE has the ability to expand into the realm of classification model, with minor alteration. Ideally, ARSE could be adapted to correctly classify a samples based on its major component within a mixture. For example, if a calibration set consists samples of pure Olive oil and Corn oil while a new sample is Olive Oil mixed with some amount of Sunflower oil. In this scenario the model would likely fail due to the uncontaminated interferent. With ARSE, the sample could still be correctly identified as Olive oil, as well as possible helping identify the interferent based on the variables removed.

5.3.2 Edible Oils

There are many directions to pursue regarding the edible oil research discussed within this dissertation. The comparison of different spectroscopic techniques, discussed in Chapter 4, could be expanded to include any number of other spectral techniques. Additionally, an interesting study could be done comparing a wide range of path lengths both within the NIR and perhaps the MIR.

Additional work could also be done with regards to building models to predict peroxide values. Though not included in this dissertation, small scale studies of artificially aging edible oils have been done. These studies could be expanded to include multiple different oils over a longer time scale. This would allow for the creation of extensive data sets with a wide range of both oil types as well as peroxide values.

BIBLIOGRAPHY

1. Avramov, I. M.; Antanasijevic, J.; Trisovic, N.; Antanasijevic, D.; Lovic, J.; Mijin, D.; Petrovic, S. A Chemometrical analysis of voltammetric data for simultaneous determination of phenobarbital sodium and paracetamol obtained at a gold electrode. *Int. J. Electrochem. Sci.* **2016**, *11*, 5935-5951.
2. Cramer, C. J.; Johnson, J. L.; Kamel, A. M. Prediction of Mass Spectral Response Factors from Predicted Chemometric Data for Druglike Molecules. *J. Am. Soc. Mass Spectrom.* **2016**, Ahead of Print.
3. Gao, F.; Xu, Z.; Wang, W.; Lu, G.; Vander Heyden, Y.; Zhou, T.; Fan, G. A comprehensive strategy using chromatographic profiles combined with chemometric methods: Application to quality control of *Polygonum cuspidatum* Sieb. et Zucc. *J. Chromatogr. A* **2016**, *1466*, 67-75.
4. Naguib, I. A.; Abdelaleem, E. A.; Zaazaa, H. E.; Hussein, E. A. Partial least-squares and linear support vector regression chemometric methods for simultaneous determination of amoxicillin trihydrate and dicloxacillin sodium in the presence of their common impurity. *J. AOAC Int.* **2016**, *99*, 972-979.
5. Patel, M. N.; Kothari, C. S. Multivariate approaches for simultaneous determination of avanafil and dapoxetine by uv chemometrics and HPLC-QbD in binary mixtures and pharmaceutical product. *J. AOAC Int.* **2016**, *99*, 649-663.
6. Tang, J.; Li, X.; Feng, Y.; Liang, B. Simultaneous Determination of Amiloride and Hydrochlorothiazide in a Compound Tablet by Diffuse Reflectance Spectroscopy and Chemometrics. *J. Appl. Spectrosc.* **2016**, *83*, 710-716.
7. Tawakkol, S. M.; El-Zeiny, M. B.; Hemdan, A. Full spectrum and selected spectrum based chemometric methods for the simultaneous determination of Cinnarizine and Dimenhydrinate in laboratory prepared mixtures and pharmaceutical dosage form. *Spectrochim. Acta, Part A* **2017**, *173*, 892-896.

8. Tewari, J.; Strong, R.; Boulas, P. At-line determination of pharmaceuticals small molecule's blending end point using chemometric modeling combined with Fourier transform near infrared spectroscopy. *Spectrochim. Acta, Part A* **2017**, *173*, 886-891.
9. Wang, L.; Tian, X.; Wei, W.; Chen, G.; Wu, Z. Fingerprint analysis and quality consistency evaluation of flavonoid compounds for fermented Guava leaf by combining HPLC-TOF-ESI-MS and chemometric methods. *J. Sep. Sci.* **2016**, *39*, 3906-3916.
10. Zhang, M.; Zhao, C.; Liang, X.; Ying, Y.; Han, B.; Yang, B.; Jiang, C. Fingerprint analysis of *Desmodium triquetrum* L. based on ultra performance liquid chromatography with photodiode array detector combined with chemometrics methods. *J. Chromatogr. Sci.* **2016**, *54*, 706-712.
11. Barac, N.; Skrivanj, S.; Bukumiric, Z.; Zivojinovic, D.; Manojlovic, D.; Barac, M.; Petrovic, R.; Corac, A. Distribution and mobility of heavy elements in floodplain agricultural soils along the Ibar River (Southern Serbia and Northern Kosovo). Chemometric investigation of pollutant sources and ecological risk assessment. *Environ. Sci. Pollut. Res.* **2016**, *23*, 9000-9011.
12. Capuano, E.; van der Veer, G.; Boerrigter-Eenling, R.; Elgersma, A.; Rademaker, J.; Sterian, A.; van Ruth, S. M. Verification of fresh grass feeding, pasture grazing and organic farming by cows farm milk fatty acid profile. *Food Chem.* **2014**, *164*, 234-241.
13. Erich, S.; Schill, S.; Annweiler, E.; Waiblinger, H.; Kuballa, T.; Lachenmeier, D. W.; Monakhova, Y. B. Combined chemometric analysis of ¹H NMR, ¹³C NMR and stable isotope data to differentiate organic and conventional milk. *Food Chem.* **2015**, *188*, 1-7.
14. Laursen, K. H.; Schjoerring, J. K.; Kelly, S. D.; Husted, S. Authentication of organically grown plants - advantages and limitations of atomic spectroscopy for multi-element and stable isotope analysis. *TrAC, Trends Anal. Chem.* **2014**, *59*, 73-82.
15. Low, K. H.; Idris, N. S. U.; Md. Zain, S.; Kamaruddin, A. F.; Md. Salleh, K. Evaluation of Elemental Distributions in Wild-Caught and Farmed *Pangasius* sp. Using Pattern Recognition Techniques. *Int. J. Food Prop.* **2016**, *19*, 1489-1503.
16. Luan, L.; Wang, Y.; Li, X.; Hu, W.; Li, K.; Li, J.; Yang, K.; Shu, R.; Zhao, L.; Lao, C. Application of multiple classifier fusion in the discriminant analysis of

- near infrared spectroscopy for agricultural products. *J. Near Infrared Spectrosc.* **2016**, *24*, 363-372.
17. Peng, J.; Liu, F.; Zhou, F.; Song, K.; Zhang, C.; Ye, L.; He, Y. Challenging applications for multi-element analysis by laser-induced breakdown spectroscopy in agriculture: A review. *TrAC, Trends Anal. Chem.* **2016**, *85*, 260-272.
 18. Potorti, A. G.; Di Bella, G.; Lo Turco, V.; Rando, R.; Dugo, G. Non-toxic and potentially toxic elements in Italian donkey milk by ICP-MS and multivariate analysis. *J. Food Compos. Anal.* **2013**, *31*, 161-172.
 19. Ravikanth, L.; Jayas, D. S.; White, N. D. G.; Fields, P. G.; Sun, D. Extraction of Spectral Information from Hyperspectral Data and Application of Hyperspectral Imaging for Food and Agricultural Products. *Food Bioprocess Technol.* **2016**, Ahead of Print.
 20. Jaworski, A.; Wikiel, H.; Holverson, P.; Nelson, A.; Wikiel, K. Detection and Diagnosis of Various Contaminants in an Electroplating Bath by a Voltammetric Sensor: a Case Study. *Electroanalysis* **2015**, *27*, 144-155.
 21. Jaworski, A.; Wikiel, H.; Wikiel, K.; Holverson, P.; Nelson, A. Case study of early detection of iron contamination in copper damascene plating process by in-situ electrochemical sensor. *ECS Trans.* **2015**, *64*, 91-108.
 22. Mukherjee, P.; Lim, S. J.; Wrobel, T. P.; Bhargava, R.; Smith, A. M. Measuring and Predicting the Internal Structure of Semiconductor Nanocrystals through Raman Spectroscopy. *J. Am. Chem. Soc.* **2016**, *138*, 10887-10896.
 23. Nallon, E. C.; Schnee, V. P.; Bright, C. J.; Polcha, M. P.; Li, Q. Discrimination Enhancement with Transient Feature Analysis of a Graphene Chemical Sensor. *Anal. Chem. (Washington, DC, U. S.)* **2016**, *88*, 1401-1406.
 24. Wang, H.; Yao, M. Fault detection of batch processes based on multivariate functional kernel principal component analysis. *Chemom. Intell. Lab. Syst.* **2015**, *149*, 78-89.
 25. da Costa, G. B.; Fernandes, D. D. S.; Gomes, A. A.; de Almeida, V. E.; Veras, G. Using near infrared spectroscopy to classify soybean oil according to expiration date. *Food Chem.* **2016**, *196*, 539-543.
 26. Jovic, O.; Smolic, T.; Primožic, I.; Hrenar, T. Spectroscopic and Chemometric Analysis of Binary and Ternary Edible Oil Mixtures: Qualitative and Quantitative Study. *Anal. Chem. (Washington, DC, U. S.)* **2016**, *88*, 4516-4524.

27. Jovic, O.; Smrecki, N.; Popovic, Z. Interval ridge regression (iRR) as a fast and robust method for quantitative prediction and variable selection applied to edible oil adulteration. *Talanta* **2016**, *150*, 37-45.
28. Li, X.; Kong, W.; Shi, W.; Shen, Q. A combination of chemometrics methods and GC-MS for the classification of edible vegetable oils. *Chemom. Intell. Lab. Syst.* **2016**, *155*, 145-150.
29. Mahboubifar, M.; Yousefinejad, S.; Alizadeh, M.; Hemmateenejad, B. Prediction of the acid value, peroxide value and the percentage of some fatty acids in edible oils during long heating time by chemometrics analysis of FTIR-ATR spectra. *J. Iran. Chem. Soc.* **2016**, *13*, 2291-2299.
30. Rohman, A. Infrared Spectroscopy for Quantitative Analysis and Oil Parameters of Olive Oil and Virgin Coconut Oil: A Review. *Int. J. Food Prop.* **2016**, Ahead of Print.
31. Rueda, A.; Samaniego-Sanchez, C.; Olalla, M.; Gimenez, R.; Cabrera-Vique, C.; Seiquer, I.; Lara, L. Combination of analytical and chemometric methods as a useful tool for the characterization of extra virgin argan oil and other edible virgin oils. Role of polyphenols and tocopherols. *J. AOAC Int.* **2016**, *99*, 489-494.
32. Spatari, C.; De Luca, M.; Ioele, G.; Ragno, G. A critical evaluation of the analytical techniques in the photodegradation monitoring of edible oils. *LWT--Food Sci. Technol.* **2016**, Ahead of Print.
33. Tu, A.; Du, Z.; Qu, S. Rapid profiling of triacylglycerols for identifying authenticity of edible oils using supercritical fluid chromatography-quadruple time-of-flight mass spectrometry combined with chemometric tools. *Anal. Methods* **2016**, *8*, 4226-4238.
34. Upadhyay, R.; Sehwal, S.; Niwas Mishra, H. Chemometric approach to develop frying stable sunflower oil blends stabilized with oleoresin rosemary and ascorbyl palmitate. *Food Chem.* **2017**, *218*, 496-504.
35. Zhang, W.; Li, N.; Feng, Y.; Su, S.; Li, T.; Liang, B. A unique quantitative method of acid value of edible oils and studying the impact of heating on edible oils by UV-Vis spectrometry. *Food Chem.* **2015**, *185*, 326-332.
36. Ammari, F.; Jouan-Rimbaud-Bouveresse, D.; Boughanmi, N.; Rutledge, D. N. Study of the heat stability of sunflower oil enriched in natural antioxidants by different analytical techniques and front-face fluorescence spectroscopy combined with Independent Components Analysis. *Talanta* **2012**, *99*, 323-329.

37. Armenta, S.; Garrigues, S.; De la Guardia, M. Determination of edible oil parameters by near infrared spectrometry. *Anal. Chim. Acta* **2007**, *596*, 330-337.
38. Cordella, C. B. Y.; Tekye, T.; Rutledge, D. N.; Leardi, R. A multiway chemometric and kinetic study for evaluating the thermal stability of edible oils by ¹H NMR analysis: Comparison of methods. *Talanta* **2012**, *88*, 358-368.
39. Goncalves, R. P.; Marco, P. H.; Valderrama, P. Thermal edible oil evaluation by UV-Vis spectroscopy and chemometrics. *Food Chem.* **2014**, *163*, 83-86.
40. Khan, M. N.; Sarwar, A.; Wahab, M. F. Chemometric assessment of thermal oxidation of some edible oils. *J. Therm. Anal. Calorim.* **2010**, *102*, 369-374.
41. Le Dreau, Y.; Dupuy, N.; Artaud, J.; Ollivier, D.; Kister, J. Infrared study of aging of edible oils by oxidative spectroscopic index and MCR-ALS chemometric method. *Talanta* **2009**, *77*, 1748-1756.
42. Meng, X.; Ye, Q.; Nie, X.; Pan, Q.; Jiang, L. Assessment of Interval PLS (iPLS) Calibration for the Determination of Peroxide Value in Edible Oils. *J. Am. Oil Chem. Soc.* **2015**, *92*, 1405-1412.
43. Nunes, C. A. Vibrational spectroscopy and chemometrics to assess authenticity, adulteration and intrinsic quality parameters of edible oils and fats. *Food Res. Int.* **2014**, *60*, 255-261.
44. Petersen, K. D.; Kleeberg, K. K.; Jahreis, G.; Busch-Stockfisch, M.; Fritsche, J. Comparison of analytical and sensory lipid oxidation parameters in conventional and high-oleic rapeseed oil. *Eur. J. Lipid Sci. Technol.* **2012**, *114*, 1193-1203.
45. Quinones-Islas, N.; Meza-Marquez, O. G.; Osorio-Revilla, G.; Gallardo-Velazquez, T. Detection of adulterants in avocado oil by Mid-FTIR spectroscopy and multivariate analysis. *Food Res. Int.* **2013**, *51*, 148-154.
46. Sikorska, E.; Romaniuk, A.; Khmelinskii, I.; Sikorski, M.; Koziol, J. Characterization of edible oils using synchronous scanning fluorescence spectroscopy. *Pol. J. Food Nutr. Sci.* **2003**, *12*, 108-112.
47. Wojcicki, K.; Khmelinskii, I.; Sikorski, M.; Sikorska, E. Near and mid infrared spectroscopy and multivariate data analysis in studies of oxidation of edible oils. *Food Chem.* **2015**, *187*, 416-423.
48. Wold, S. Chemometrics; what do we mean with it, and what do we want from it? *Chemometrics Intellig. Lab. Syst.* **1995**, *30*, 109-115.

49. Andries, E.; Martin, S. Sparse methods in spectroscopy: an introduction, overview, and perspective. *Appl. Spectrosc.* **2013**, *67*, 579-593.
50. Esteban, M.; Arino, C.; Diaz-Cruz, J. M. Chemometrics in electroanalytical chemistry. *Crit. Rev. Anal. Chem.* **2006**, *36*, 295-313.
51. Geladi, P. Some recent trends in the calibration literature. *Chemom. Intell. Lab. Syst.* **2002**, *60*, 211-224.
52. Geladi, P.; Dabakk, E. An overview of chemometrics applications in near infrared spectrometry. *J. Near Infrared Spectrosc.* **1995**, *3*, 119-132.
53. Kalivas, J. Multivariate Calibration, an Overview. *Anal. Lett.* **2005**, *38*, 2259-2279.
54. Kalivas, J. H. Overview of two-norm (L2) and one-norm (L1) Tikhonov regularization variants for full wavelength or sparse spectral multivariate calibration models or maintenance. *J. Chemom.* **2012**, *26*, 218-230.
55. Leardi, R. In *In Chemometrics in data analysis*. Section Title: Biochemical Methods; 2003; , pp 299-320.
56. Leardi, R. Chemometrics: from classical to genetic algorithms. *Grasas Aceites (Sevilla, Spain)* **2002**, *53*, 115-127.
57. Pearson, K. The Law of Ancestral Heredity. *Biometrika* **1903**, *2*, 211-228.
58. Yule, G. U. On the Theory of Correlation. *Journal of the Royal Statistical Society* **1897**, *60*, 812-854.
59. Frank, I. E.; Friedman, J. H. A Statistical View of Some Chemometrics Regression Tools. *Technometrics* **1993**, *35*, 109-135.
60. Helland, I. S. Partial Least Squares Regression and Statistical Models. *Scandinavian Journal of Statistics* **1990**, *17*, 97-114.
61. Jolliffe, I. T. A Note on the Use of Principal Components in Regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **1982**, *31*, 300-303.
62. Ottaway, J.; Kalivas, J. H.; Andries, E. Spectral multivariate calibration with wavelength selection using variants of Tikhonov regularization. *Appl. Spectrosc.* **2010**, *64*, 1388-1395.

63. Gerlach, R. W.; Kowalski, B. R.; Wold, H. O. A. Partial least-squares path modelling with latent variables. *Anal. Chim. Acta* **1979**, *112*, 417-421.
64. Baroni, M.; Costantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; Clementi, S. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct. -Act. Relat.* **1993**, *12*, 9-20.
65. Dyrby, M.; Engelsen, S. B.; Norgaard, L.; Bruhn, M.; Lundsberg-Nielsen, L. Chemometric quantitation of the active substance (containing CN) in a pharmaceutical tablet using near-infrared (N-IR) transmittance and N-IR FT-Raman spectra. *Appl. Spectrosc.* **2002**, *56*, 579-585.
66. Janik, L. J.; Skjemstad, J. O. Characterization and analysis of soils using mid-infrared partial least-squares. II. Correlations with some laboratory data. *Aust. J. Soil Res.* **1995**, *33*, 637-650.
67. Maeda, H.; Ozaki, Y.; Tanaka, M.; Hayashi, N.; Kojima, T. Near infrared spectroscopy and chemometrics studies of temperature-dependent spectral variations of water: relationship between spectral changes and hydrogen bonds. *J. Near Infrared Spectrosc.* **1995**, *3*, 191-201.
68. Ni, Y.; Wang, L.; Kokot, S. Simultaneous determination of nitrobenzene and nitro-substituted phenols by differential pulse voltammetry and chemometrics. *Anal. Chim. Acta* **2001**, *431*, 101-113.
69. Pistonesi, M. F.; Di Nezio, M. S.; Centurion, M. E.; Palomeque, M. E.; Lista, A. G.; Fernandez Band, B. S. Determination of phenol, resorcinol and hydroquinone in air samples by synchronous fluorescence using partial least-squares (PLS). *Talanta* **2006**, *69*, 1265-1268.
70. Thissen, U.; Pepers, M.; Ustun, B.; Melssen, W. J.; Buydens, L. M. C. Comparing support vector machines to PLS for spectral regression applications. *Chemom. Intell. Lab. Syst.* **2004**, *73*, 169-179.
71. Trygg, J.; Holmes, E.; Lundstedt, T. Chemometrics in Metabonomics. *J. Proteome Res.* **2007**, *6*, 469-479.
72. Wise, B. M.; Gallagher, N. B. The process chemometrics approach to process monitoring and fault detection. *J. Process Control* **1996**, *6*, 329-348.
73. Wold, S.; Trygg, J.; Berglund, A.; Antti, H. Some recent developments in PLS modeling. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 131-150.

74. Smekal, A. Zur Quantentheorie der Dispersion. *Naturwissenschaften* **1923**, *11*, 873-875.
75. Gardiner, D. J. *Practical Raman Spectroscopy*; Springer-Verlag: 1989; .
76. Placzek, G. In *Hdb. der Radiologie*; Rayleigh Streuung und Raman Effekt; 1934; pp 209.
77. Long, D. *The Raman Effect: A unified Treatment of the theory of Raman Scattering by Molecules*; Wiley and Sons: 2002; .
78. Wikipedia. Raman Spectroscopy.
https://en.wikipedia.org/wiki/Raman_spectroscopy (Accessed 10/15, 2016).
79. Herschel, W. Experiments on the Refrangibility of the Invisible Rays of the Sun. By William Herschel, LL. *Philosophical Transactions of the Royal Society of London* **1800**, *90*, 284-292.
80. Bornemann, L.; Welp, G.; Brodowski, S.; Rodionov, A.; Amelung, W. Rapid assessment of black carbon in soil organic matter using mid-infrared spectroscopy. *Org. Geochem.* **2008**, *39*, 1537-1544.
81. Calderon, F. J.; Reeves, J. B., III; Collins, H. P.; Paul, E. A. Chemical differences in soil organic matter fractions determined by diffuse-reflectance mid-infrared spectroscopy. *Soil Sci. Soc. Am. J.* **2011**, *75*, 568-579.
82. Cozzolino, D.; Holdstock, M.; Damberg, R. G.; Cynkar, W. U.; Smith, P. A. Mid infrared spectroscopy and multivariate analysis: A tool to discriminate between organic and non-organic wines grown in Australia. *Food Chem.* **2009**, *116*, 761-765.
83. Ludwig, B.; Nitschke, R.; Terhoeven-Urselmans, T.; Michel, K.; Flessa, H. Use of mid-infrared spectroscopy in the diffuse-reflectance mode for the prediction of the composition of organic matter in soil and litter. *J. Plant Nutr. Soil Sci.* **2008**, *171*, 384-391.
84. Mauer, L. J.; Chernyshova, A. A.; Hiatt, A.; Deering, A.; Davis, R. Melamine Detection in Infant Formula Powder Using Near- and Mid-Infrared Spectroscopy. *J. Agric. Food Chem.* **2009**, *57*, 3974-3980.
85. Koperska, M. A.; Pawcenis, D.; Bagniuk, J.; Zaitz, M. M.; Missori, M.; Lojewski, T.; Lojewska, J. Degradation markers of fibroin in silk through infrared spectroscopy. *Polym. Degrad. Stab.* **2014**, *105*, 185-196.

86. Nagy, E. E.; Adamsons, K.; Moomaw, K. Disfiguring organic residues on industrially produced sheet metal coupons simulating copper and brass works of art by Donald Judd: Attenuated total reflectance Fourier transform infrared spectroscopy analysis and treatment recommendations. *Stud. Conserv.* **2013**, *58*, 245-255.
87. Newman, R. Some applications of infrared spectroscopy in the examination of painting materials. *J. Am. Inst. Conserv.* **1979**, *19*, 42-62.
88. Ayvaz, H.; Plans, M.; Towers, B. N.; Auer, A.; Rodriguez-Saona, L. E. The use of infrared spectrometers to predict quality parameters of cornmeal (corn grits) and differentiate between organic and conventional practices. *J. Cereal Sci.* **2015**, *62*, 22-30.
89. Ayvaz, H.; Santos, A. M.; Moysenko, J.; Kleinhenz, M.; Rodriguez-Saona, L. E. Application of a Portable Infrared Instrument for Simultaneous Analysis of Sugars, Asparagine and Glutamine Levels in Raw Potato Tubers. *Plant Foods Hum. Nutr. (N. Y., NY, U. S.)* **2015**, *70*, 215-220.
90. Kamal, M.; Karoui, R. Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review. *Trends Food Sci. Technol.* **2015**, *46*, 27-48.
91. McParland, S.; Lewis, E.; Kennedy, E.; Moore, S. G.; McCarthy, B.; O'Donovan, M.; Butler, S. T.; Pryce, J. E.; Berry, D. P. Mid-infrared spectrometry of milk as a predictor of energy intake and efficiency in lactating dairy cows. *J. Dairy Sci.* **2014**, *97*, 5863-5871.
92. Mellado-Mojica, E.; Seeram, N. P.; Lopez, M. G. Maple syrup vs. different natural sweeteners: Sensory evaluation, physicochemical properties, carbohydrate profiles, and botanical classification by mid infrared spectroscopy-chemometrics. *Abstracts of Papers, 247th ACS National Meeting & Exposition, Dallas, TX, United States, March 16-20, 2014* **2014**, AGFD-21.
93. Misra, N. N.; Sullivan, C.; Cullen, P. J. Process Analytical Technology (PAT) and Multivariate Methods for Downstream Processes. *Curr. Biochem. Eng.* **2015**, *2*, 4-16.
94. Wisniewska, P.; Boque, R.; Borrás, E.; Busto, O.; Wardencki, W.; Namiesnik, J.; Dymerski, T. Authentication of whisky due to its botanical origin and way of production by instrumental analysis and multivariate classification methods. *Spectrochim. Acta, Part A* **2017**, *173*, 849-853.

95. Reich, G. Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications. *Adv. Drug Delivery Rev.* **2005**, *57*, 1109-1143.
96. Beebe, K. R.; Kowalski, B. R. An introduction to multivariate calibration and analysis. *Anal. Chem.* **1987**, *59*, 1007A-1010A, 1012A, 1014A-1017A.
97. Thomas, E. V.; Haaland, D. M. Comparison of multivariate calibration methods for quantitative spectral analysis. *Anal. Chem.* **1990**, *62*, 1091-1099.
98. Blanco, M.; Coello, J.; Iturriaga, H.; Maspocho, S.; Rovira, E. Wavelength calibration transfer between diode array UV-visible spectrophotometers. *Appl. Spectrosc.* **1995**, *49*, 593-597.
99. Bouveresse, E.; Massart, D. L.; Dardenne, P. Modified Algorithm for Standardization of Near-Infrared Spectrometric Instruments. *Anal. Chem.* **1995**, *67*, 1381-1389.
100. Candolfi, A.; Massart, D. L. Model Updating for the Identification of NIR Spectra from a Pharmaceutical Excipient. *Applied Spectroscopy* **2000**, *54*, 48-53.
101. Capron, X.; Walczak, B.; de Noord, O. E.; Massart, D. L. Selection and weighting of samples in multivariate regression model updating. *Chemom. Intell. Lab. Syst.* **2005**, *76*, 205-214.
102. Chen, Z.; Morris, J.; Martin, E. Extracting Chemical Information from Spectral Data with Multiplicative Light Scattering Effects by Optical Path-Length Estimation and Correction. *Anal. Chem.* **2006**, *78*, 7674-7681.
103. Greensill, C. V.; Wolfs, P. J.; Spiegelman, C. H.; Walsh, K. B. Calibration transfer between PDA-based NIR spectrometers in the NIR assessment of melon soluble solids content. *Appl. Spectrosc.* **2001**, *55*, 647-653.
104. Anderson, C. E.; Kalivas, J. H. Fundamentals of calibration transfer through procrustes analysis. *Appl. Spectrosc.* **1999**, *53*, 1268-1276.
105. Igne, B.; Roger, J.; Roussel, S.; Bellon-Maurel, V.; Hurburgh, C. R. Improving the transfer of near infrared prediction models by orthogonal methods. *Chemom. Intell. Lab. Syst.* **2009**, *99*, 57-65.
106. Isaksson, T.; Kowalski, B. Piece-wise multiplicative scatter correction applied to near-infrared diffuse transmittance data from meat products. *Appl. Spectrosc.* **1993**, *47*, 702-709.

107. Feudale, R. N.; Woody, N. A.; Tan, H.; Myles, A. J.; Brown, S. D.; Ferre, J. Transfer of multivariate calibration models: a review. *Chemom. Intell. Lab. Syst.* **2002**, *64*, 181-192.
108. Adhietty, I. S.; McGuire, J. A.; Wangmaneerat, B.; Niemczyk, T. M.; Haaland, D. M. Achieving transferable multivariate spectral calibration models: demonstration with infrared spectra of thin-film dielectrics on silicon. *Anal. Chem.* **1991**, *63*, 2329-2338.
109. Andrews, D. T.; Wentzell, P. D. Applications of maximum likelihood principal component analysis: incomplete data sets and calibration transfer. *Anal. Chim. Acta* **1997**, *350*, 341-352.
110. Barring, H. K.; Boelens, H. F. M.; De Noord, O. E.; Smilde, A. K. Optimizing meta-parameters in continuous piecewise direct standardization. *Appl. Spectrosc.* **2001**, *55*, 458-466.
111. Bouveresse, E.; Massart, D. L.; Dardenne, P. Calibration transfer across near-infrared spectrometric instruments using Shenk's algorithm: effects of different standardization samples. *Anal. Chim. Acta* **1994**, *297*, 405-416.
112. Bouveresse, E.; Massart, D. L. Improvement of the piecewise direct standardization procedure for the transfer of NIR spectra for multivariate calibration. *Chemom. Intell. Lab. Syst.* **1996**, *32*, 201-213.
113. Chu, X.; Yuan, H.; Lu, W. Model transfer in multivariate calibration. *Guangpuxue Yu Guangpu Fenxi* **2001**, *21*, 881-885.
114. de Noord, O. E. Multivariate calibration standardization. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 85-97.
115. Duponchel, L.; Ruckebusch, C.; Huvenne, J. P.; Legrand, P. Standardisation of near infrared spectrometers using artificial neural networks. *J. Near Infrared Spectrosc.* **1999**, *7*, 155-166.
116. Fearn, T. Standardisation and calibration transfer for near infrared instruments: a review. *J. Near Infrared Spectrosc.* **2001**, *9*, 229-244.
117. Geladi, P.; Barring, H.; Dabakk, E.; Trygg, J.; Antti, H.; Wold, S.; Karlberg, B. Calibration transfer for predicting lake-water pH from near infrared spectra of lake sediments. *J. Near Infrared Spectrosc.* **1999**, *7*, 251-264.

118. Vogt, F.; Mizaikoff, B. Secured PCR (sPCR) for detection and correction of PCR calibration model failures induced by uncalibrated spectral features. *J. Chemom.* **2003**, *17*, 225-236.
119. Goswami, J.; Chan, A. *Fundamentals of Wavelets: Theory, Algorithms, and Applications*; Wiley: New York, 1999; .
120. Jiang, J.; Berry, R. J.; Siesler, H. W.; Ozaki, Y. Wavelength interval selection in multicomponent spectral analysis by moving window partial least-squares regression with applications to mid-infrared and near-infrared spectroscopic data. *Anal. Chem.* **2002**, *74*, 3555-3565.
121. Simopoulos, A. P. Omega-3 fatty acids in health and disease and in growth and development. *Am. J. Clin. Nutr.* **1991**, *54*, 438-463.
122. Baeten, V.; Aparicio, R. Edible oils and fats authentication by Fourier transform Raman spectrometry. *Biotechnol. , Agron. , Soc. Environ.* **2000**, *4*, 196-203.
123. Elzey, B.; Pollard, D.; Fakayode, S. O. Determination of adulterated neem and flaxseed oil compositions by FTIR spectroscopy and multivariate regression analysis. *Food Control* **2016**, *68*, 303-309.
124. Farhad, S. F. U.; Abedin, K. M.; Islam, M. R.; Talukder, A. I.; Haider, A. F. M. Y. Determination of ratio of unsaturated to total fatty acids in edible oils by laser Raman spectroscopy. *J. Appl. Sci.* **2009**, *9*, 1538-1543.
125. Gurdeniz, G.; Ozen, B. Detection of adulteration of extra-virgin olive oil by chemometric analysis of mid-infrared spectral data. *Food Chem.* **2009**, *116*, 519-525.
126. Jovic, O. Durbin-Watson partial least-squares regression applied to MIR data on adulteration with edible oils of different origins. *Food Chem.* **2016**, *213*, 791-798.
127. Karunathilaka, S. R.; Kia, A. F.; Srigley, C.; Chung, J. K.; Mossoba, M. M. Nontargeted, Rapid Screening of Extra Virgin Olive Oil Products for Authenticity Using Near-Infrared Spectroscopy in Combination with Conformity Index and Multivariate Statistical Analyses. *J. Food Sci.* **2016**, *81*, C2390-C2397.
128. Kim, M.; Lee, S.; Chang, K.; Chung, H.; Jung, Y. M. Use of temperature dependent Raman spectra to improve accuracy for analysis of complex oil-based samples: Lube base oils and adulterated olive oils. *Anal. Chim. Acta* **2012**, *748*, 58-66.

129. Maggio, R. M.; Cerretani, L.; Chiavaro, E.; Kaufman, T. S.; Bendini, A. A novel chemometric strategy for the estimation of extra virgin olive oil adulteration with edible oils. *Food Control* **2010**, *21*, 890-895.
130. Marigheto, N. A.; Kemsley, E. K.; Defernez, M.; Wilson, R. H. A comparison of mid-infrared and Raman spectroscopies for the authentication of edible oils. *J. Am. Oil Chem. Soc.* **1998**, *75*, 987-992.
131. Oussama, A.; Elabadi, F.; Devos, O. Analysis of Argan Oil Adulteration Using Infrared Spectroscopy. *Spectrosc. Lett.* **2012**, *45*, 458-463.
132. Rohman, A.; Man, Y. B. C. Application of Fourier Transform Infrared Spectroscopy for Authentication of Functional Food Oils. *Appl. Spectrosc. Rev.* **2012**, *47*, 1-13.
133. Zhang, L.; Li, P.; Sun, X.; Wang, X.; Xu, B.; Wang, X.; Ma, F.; Zhang, Q.; Ding, X. Classification and adulteration detection of vegetable oils based on fatty acid profiles. *J. Agric. Food Chem.* **2014**, *62*, 8745-8751.
134. Zhang, Q.; Liu, C.; Sun, Z.; Hu, X.; Shen, Q.; Wu, J. Authentication of edible vegetable oils adulterated with used frying oil by Fourier Transform Infrared Spectroscopy. *Food Chem.* **2012**, *132*, 1607-1613.
135. Zhang, X.; Zou, M.; Qi, X.; Liu, F.; Zhang, C.; Yin, F. Quantitative detection of adulterated olive oil by Raman spectroscopy and chemometrics. *J. Raman Spectrosc.* **2011**, *42*, 1784-1788.
136. Zhao, H.; Wang, Y.; Xu, X.; Ren, H.; Li, L.; Xiang, L.; Zhong, W. Detection of adulterated vegetable oils containing waste cooking oils based on the contents and ratios of cholesterol, β -sitosterol, and campesterol by gas chromatography/mass spectrometry. *J. AOAC Int.* **2015**, *98*, 1645-1654.
137. American Oil Chemists' Society (AOCS) In *Official method Cd 8b-90-Peroxide value acetic acid-isoctane method* ; 2003; .
138. Javidnia, K.; Parish, M.; Karimi, S.; Hemmateenejad, B. Discrimination of edible oils and fats by combination of multivariate pattern recognition and FT-IR spectroscopy: A comparative study between different modeling methods. *Spectrochim. Acta, Part A* **2013**, *104*, 175-181.
139. Ren, H.; Lin, W.; Shi, W.; Qi, S. Classification of Edible Oils by Infrared Spectroscopy with Optimized K-Means Clustering by a Hybrid Particle Swarm Algorithm. *Anal. Lett.* **2013**, *46*, 2727-2738.

140. Ren, H.; Lin, W.; Shi, W.; Shen, Q.; Wang, S. Characterization of Peanut Oil by Infrared Spectroscopy with an Improved Gaussian Mixture Model. *Anal. Lett.* **2014**, *47*, 2548-2559.
141. Song, Z.; Zhang, H.; Zheng, X.; Ren, X.; He, D. Research progress on detection of edible vegetable oil based on the near infrared spectroscopy. *Wuhan Gongye Xueyuan Xuebao* **2013**, *32*, 1-5.
142. Zheng, W.; Fu, X.; Ying, Y. Spectroscopy-based food classification with extreme learning machine. *Chemom. Intell. Lab. Syst.* **2014**, *139*, 42-47.
143. Zhou, X.; Dai, L.; Li, S. Fast discrimination of edible vegetable oil based on Raman spectroscopy. *Guangpuxue Yu Guangpu Fenxi* **2012**, *32*, 1829-1833.
144. Zhou, Y.; Liu, T.; Li, J. Rapid identification between edible oil and swill-cooked dirty oil by using a semi-supervised support vector machine based on graph and near-infrared spectroscopy. *Chemom. Intell. Lab. Syst.* **2015**, *143*, 1-6.
145. Al Majidi, M. I. H.; Bader, A. T. Physicochemical characteristics of some imported edible vegetable oils in Iraq. *Res. J. Pharm. , Biol. Chem. Sci.* **2015**, *6*, 488-494.
146. Li, L.; Wei, M. Quality investigation and analysis of edible peanut oil made by traditional way. *Zhongguo Weisheng Jianyan Zazhi* **2015**, *25*, 3575-3576.
147. Liang, P.; Chen, C.; Zhao, S.; Ge, F.; Liu, D.; Liu, B.; Fan, Q.; Han, B.; Xiong, X. Application of Fourier transform infrared spectroscopy for the oxidation and peroxide value evaluation in virgin walnut oil. *J. Spectrosc. (N. Y. , NY, U. S.)* **2013**, 138728, 6.
148. Vunguturi, S.; Sarwar, A. Evaluation of peroxide value in various fresh and reused edible oils by iodometric titrations. *Eur. J. Biomed. Pharm. Sci.* **2016**, *3*, 636-638.
149. Xu, L.; Fei, T.; Li, Q.; Yu, X.; Liu, L. Qualitative analysis of edible oil oxidation by FTIR spectroscopy using a mesh "cell". *Anal. Methods* **2015**, *7*, 4328-4333.
150. Xu, L.; Yu, X.; Liu, L.; Li, M.; Zhang, R. A rapid method for evaluating the edible oil oxidative stability during ambient storage by FTIR spectroscopy using a mesh cell. *Anal. Methods* **2016**, *8*, 5117-5122.
151. Gotoh, N.; Wada, S. The importance of peroxide value in assessing food quality and food safety. *J. Am. Oil Chem. Soc.* **2006**, *83*, 473-474.

152. Jimenez-Sanchidrian, C.; Ruiz, J. R. Use of Raman spectroscopy for analyzing edible vegetable oils. *Appl. Spectrosc. Rev.* **2016**, *51*, 397-410.
153. van de Voort, F. R.; Ismail, A. A.; Sedman, J.; Dubois, J.; Nicodemo, T. The determination of peroxide value by Fourier transform infrared spectroscopy. *J. Am. Oil Chem. Soc.* **1994**, *71*, 921-926.
154. Alberdi-Cedeno, J.; Ibargoitia, M. L.; Cristillo, G.; Sopelana, P.; Guillen, M. D. A new methodology capable of characterizing most volatile and less volatile minor edible oils components in a single chromatographic run without solvents or reagents. Detection of new components. *Food Chem.* **2016**, Ahead of Print.
155. Andrikopoulos, N. K. Chromatographic and spectroscopic methods in the analysis of triacylglycerol species and regiospecific isomers of oils and fats. *Crit. Rev. Food Sci. Nutr.* **2002**, *42*, 473-505.
156. Barranco, A.; Alonso-Salces, R. M.; Bakkali, A.; Berrueta, L. A.; Gallo, B.; Vicente, F.; Sarobe, M. Solid-phase clean-up in the liquid chromatographic determination of polycyclic aromatic hydrocarbons in edible oils. *J. Chromatogr. A* **2003**, *988*, 33-40.
157. Biedermann, M.; Grob, K.; Mariani, C. Transesterification and on-line liquid chromatography-gas chromatography (LC-GC) for determining the sum of free and esterified sterols in edible oils and fats. *Fett Wiss. Technol.* **1993**, *95*, 127-133.
158. De Koning, S.; Janssen, H.; Van Deursen, M.; Brinkman, U. A. T. Automated on-line comprehensive two-dimensional LC \bar{A} — GC and LC \bar{A} — GC-ToF MS: instrument design and application to edible oil and fat analysis. *J. Sep. Sci.* **2004**, *27*, 397-409.
159. Gliszczynska-Swiglo, A.; Sikorska, E. Simple reversed-phase liquid chromatography method for determination of tocopherols in edible plant oils. *J. Chromatogr. A* **2004**, *1048*, 195-198.
160. Hajimahmoodi, M.; Vander Heyden, Y.; Sadeghi, N.; Jannat, B.; Oveisi, M. R.; Shahbazian, S. Gas-chromatographic fatty-acid fingerprints and partial least squares modeling as a basis for the simultaneous determination of edible oil mixtures. *Talanta* **2005**, *66*, 1108-1116.
161. Jira, W.; Ziegenhals, K.; Speer, K. Gas chromatography-mass spectrometry (GC-MS) method for the determination of 16 European priority polycyclic aromatic

- hydrocarbons in smoked meat products and edible oils. *Food Addit. Contam. , Part A* **2008**, *25*, 704-713.
162. Lee, D.; Noh, B.; Bae, S.; Kim, K. Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics. *Anal. Chim. Acta* **1998**, *358*, 163-175.
163. Masukawa, Y.; Shiro, H.; Nakamura, S.; Kondo, N.; Jin, N.; Suzuki, N.; Ooi, N.; Kudo, N. A new analytical method for the quantification of glycidol fatty acid esters in edible oils. *J. Oleo Sci.* **2010**, *59*, 81-88.
164. Moreda, W.; Perez-Camino, M. C.; Cert, A. Gas and liquid chromatography of hydrocarbons in edible vegetable oils. *J. Chromatogr. A* **2001**, *936*, 159-171.
165. Purcaro, G.; Morrison, P.; Moret, S.; Conte, L. S.; Marriott, P. J. Determination of polycyclic aromatic hydrocarbons in vegetable oils using solid-phase microextraction-comprehensive two-dimensional gas chromatography coupled with time-of-flight mass spectrometry. *J. Chromatogr. A* **2007**, *1161*, 284-291.
166. Veyrand, B.; Brosseaud, A.; Sarcher, L.; Varlet, V.; Monteau, F.; Marchand, P.; Andre, F.; Le Bizec, B. Innovative method for determination of 19 polycyclic aromatic hydrocarbons in food and oil samples using gas chromatography coupled to tandem mass spectrometry based on an isotope dilution approach. *J. Chromatogr. A* **2007**, *1149*, 333-344.
167. Yang, M.; Lin, H.; Choong, Y. A rapid gas chromatographic method for direct determination of BHA, BHT, and TBHQ in edible oils and fats. *Food Res. Int.* **2002**, *35*, 627-633.
168. Hong, J.; Yamaoka-Koseki, S.; Yasumoto, K. Determination of palmitic acid, oleic acid and linoleic acid by near-infrared transfectance spectroscopy in edible oils. *Food Sci. Technol. , Int. (Tsukuba, Jpn.)* **1996**, *2*, 146-149.
169. Li, H.; Van de Voort, F. R.; Sedman, J.; Ismail, A. A. Rapid determination of cis and trans content, iodine value, and saponification number of edible oils by Fourier transform near-infrared spectroscopy. *J. Am. Oil Chem. Soc.* **1999**, *76*, 491-497.
170. Li, H.; Van de Voort, F. R.; Ismail, A. A.; Sedman, J.; Cox, R. Trans determination of edible oils by fourier transform near-infrared spectroscopy. *J. Am. Oil Chem. Soc.* **2000**, *77*, 1061-1067.

171. Mossoba, M. M.; Tyburczy, C.; Delmonte, P.; Fardin-Kia, A. R.; Rader, J. I.; Azizian, H.; Kramer, J. K. G. In *In Application of gas chromatography and infrared spectroscopy for the determination of the total trans fatty acid, saturated fatty acid, monounsaturated fatty acid, and polyunsaturated fatty acid contents in edible fats and oils*. Section Title: Food and Feed Chemistry; 2014; , pp 89-121.
172. Scotter, C. N. G.; Wilson, R. In *In Infrared spectroscopy*. Section Title: Food and Feed Chemistry; 1998; , pp 76-96.
173. Yang, H.; Irudayaraj, J.; Paradkar, M. M. Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food Chem.* **2005**, *93*, 25-32.
174. Armenta, S.; Garrigues, S.; De la Guardia, M. Determination of edible oil parameters by near infrared spectrometry. *Anal. Chim. Acta* **2007**, *596*, 330-337.
175. Ayora-Canada, M. J.; Dominguez-Vidal, A.; Lendl, B. In *In Monitoring oxidation of lipids in edible oils and complex food systems by vibrational spectroscopy*. Section Title: Food and Feed Chemistry; 2010; Vol. 1, pp 277-296.
176. Carmona, M. A.; Lafont, F.; Jimenez-Sanchidrian, C.; Ruiz, J. R. Raman spectroscopy study of edible oils and determination of the oxidative stability at frying temperatures. *Eur. J. Lipid Sci. Technol.* **2014**, *116*, 1451-1456.
177. Muik, B.; Lendl, B.; Molina-Diaz, A.; Ayora-Canada, M. J. Direct monitoring of lipid oxidation in edible oils by Fourier transform Raman spectroscopy. *Chem. Phys. Lipids* **2005**, *134*, 173-182.
178. Muik, B.; Lendl, B.; Molina-Diaz, A.; Valcarcel, M.; Ayora-Canada, M. J. Two-dimensional correlation spectroscopy and multivariate curve resolution for the study of lipid oxidation in edible oils monitored by FTIR and FT-Raman spectroscopy. *Anal Chim Acta* **2007**, *593*, 54-67.
179. Wang, P.; Sun, J.; Zhang, T.; Liu, W. Vibrational spectroscopic approaches for the quality evaluation and authentication of virgin olive oil. *Appl. Spectrosc. Rev.* **2016**, *51*, 763-790.

Appendix A

PERMISSION FOR MATERIAL REPRINT

Chapter 4 was reprinted with permission from Adaptive Regression via Subspace Elimination, Joshua Ottaway, Joseph P. Smith, and Karl S. Books; 40 Years of Chemometrics – From Bruce Kowalski to the Future. January 1, 2015, 241-256.
Copyright 2015 American Chemical