Portable Near-Infrared Spectroscopy in Physical Anthropology and Primate Ecology Field Settings

Sydney Martin
Sydney.Martin@Colorado.EDU

Follow this and additional works at: https://scholar.colorado.edu/honr_theses

Part of the Anthropology Commons

Recommended Citation
Undergraduate Honors Theses. 1392.
https://scholar.colorado.edu/honr_theses/1392
Portable Near-Infrared Spectroscopy in Physical Anthropology and Primate Ecology Field Settings

Sydney Martin

Thesis Advisor
Matt Sponheimer, Department of Anthropology

Committee Members
Robin Bernstein, Department of Anthropology
William Bowman, Department of Ecology
# Table of Contents

Abstract 4

Chapter I: Introduction
   A. Wet Chemistry as a Traditional Method 5
   B. Near Infrared Spectroscopy and Portable NIR in the Field 5
   C. Main Purpose of this Thesis 8

Chapter II: Methods
   A. Preparation 10
   B. Analysis Using Wet Chemistry 12
   C. Master Excel Sheet 18
   D. Near Infrared Spectroscopy 18
   E. Master Excel Sheet and New Spreadsheet 20
   F. Software 21
   G. Regression Models 22

Chapter III: Results and Analysis 33

Chapter VIII: Conclusion and Recommendation for Future Work 36

Chapter: Appendices
   Appendix 1 38
   Appendix 2 39
   Appendix 3 41
   Appendix 4 43
   Appendix 5 45
   Appendix 6 47
   Appendix 7 49

Chapter IX: References 50
This thesis is the result of persistence, perseverance, and a unique support system.

Mom and Scott, thanks for reminding me of who I am when I have forgotten what I’ve been working towards. Also thank you for the contribution from Emily and David; it is one of the most sentimental gifts I have ever received.
Grace, Eri, Alison, and Dad, thank you for your unconditional love and friendship.

To the people undertaking their own research that took time out of their days, nights, and holidays to help me grow:

This research was supported and advised by Matt Sponheimer for whom I am extremely grateful. Thank you for introducing me into your world and the research you do. The experiences I’ve gained from working with you have been most valuable.

Thank you Oliver, you are instrumental, not only in this study letting me use your raw material, but also for giving me my first opportunity in lab experience.

Erin your assistance with sample preparation and analysis was instrumental not only in my research but for this entire thesis. You always made me feel a little more comfortable in a lab setting knowing I could ask you anything.

Dr. Suzanne Nelson, thank you for always supporting my dreams and pushing, as well as inspiring, me to do my best.

Carla, bless you. You give the word advisor a new meaning. None of us could have written our theses without your support and care. Leaving your office I always felt a little stress and an enormous amount of hope. You are the reason I persevered.

Leslie, thank you for helping me formulate my thoughts and encouraging the improvement of my writing skills throughout this project.

Anthropology thesis students and Shaelyn, you gave me a community where I was always inspired by your passion, reassured by your reluctance, entertained with Claudia’s humor, and kept well fed by Emily. We’ve watched each other’s struggles and victories.

Congratulations on your thesis!

To my friends:

Deltas, ZBTs, and Zach you are my sanity. You are the best siblings, loving and encouraging me when I’m frustrated and helping me start fresh. Zach you are the sunshine when I feel most low, bringing me warmth and laughter.
Abstract

Portable near infrared (NIR) spectroscopy is a fast, dependable, and cost-effective method for assessing nutritional chemistry in different plant species. Portable NIR can streamline analysis and investigation within ecological field settings, allowing researchers to evaluate significant nutritional variation for among populations and species of plants. The use of NIR in agriculture is common, but the application of portable NIR in fieldwork is still new in ecology. NIR in field settings has the potential to open new avenues of research, and can also save time and money when used in more traditional projects. To test the applicability of portable NIR spectroscopy in the field, I used nutritional data from South African savanna plant samples generated using wet chemistry, and NIR spectra of the same samples, to create regression models to predict plant nutritional properties from spectral data. I created regression models to predict crude fat, crude protein, and acid detergent fiber (ADF) of all plant organs (root, leaf, inflor, fruit, rhizome, and stem) and also specifically for leaf. My findings indicate that portable NIR can be used to predict crude protein and acid detergent fiber of South, African savanna plant samples. My findings also indicate that portable NIR is valuable for researchers in field settings, providing them with instant access to information.
Chapter I: Introduction

Wet Chemistry as a Traditional Method

Wet chemistry has been used in agricultural settings and field biology research to determine the nutritional contents of samples. Wet chemistry uses different chemical washes to extract the nutrient from a sample, making this an invasive approach. Once a sample is put through the process of wet chemistry, it is destroyed, limiting further studies using the same sample. Wet chemistry has been used to analyze the nutritional chemistry for the macronutrient content of agricultural food samples (Norris et al. 1976; Shenk et al. 1979); this method has become the standard for analyzing nutrient content.

Near Infrared Spectroscopy and Portable NIR in the Field

Near infrared spectroscopy, in comparison to wet chemistry, assesses nutritional values quickly and non-invasively. A sample is irradiated using near infrared light and spectra are reflected based on the chemical bonds, creating spectral peaks. The spectral peaks of the tested sample allow an estimation of different spectral properties determined by the specific chemical characteristics of the tested sample (Coates 2000; Rothman et al. 2009). NIR is a derived method, as the assessments and calculations are based on a previous calculation relating the value of a primary analytical technique (in this case traditional wet chemistry) to a calibration set from the spectra.

Portable NIR spectroscopy is more efficient and less costly than wet chemistry as it requires less time for analysis and less sample preparation.
Portable NIR can be used to analyze many samples in the amount of time it would take someone to prepare a few samples using the traditional wet chemistry method. NIR can predict crude protein and fiber across many different plant species and can be used broadly across many ecological environments (da Costa et al. 2005; Lawler et al. 2006; Locher et al. 2005; Rothman et al. 2009; Woolnough and Foley 2002).

Agricultural and primatological field work have used NIR spectroscopy in nutritional research to better understand how nutritional properties such as crude protein, acid detergent fiber, and moisture affect animal ecology within given ecosystems (Batten et al. 1998; Foley et al. 1998; McIlwee et al. 2001; Moore and Foley 2005). Near infrared spectroscopy has been used to analyze and predict feeding rates of folivorous marsupials (McIlwee et al. 2001). NIR has also been used to understand reproductive success in herbivores based on the amount of food intake from quality food resources (DeGabriel et al. 2008; DeGabriel et al. 2009). To understand digestibility, scientists have predicted the botanical and nutritional composition of different animal diets using the spectra of feces (Glasser et al. 2008; Parveen et al. 2008; Steen et al. 1998). NIR has been used to study the quality of diet for donkeys, deer, ostriches, and cattle (Kidane et al. 2008; Landeau et al. 2006; Lyons and Stuth 1992; Showers et al. 2006).

Primatologists use NIR rapid assessment of food resources to examine what nutritional properties drive primate food choice (Felton et al. 2009; Ortmann et al. 2006; Rothman et al. 2009). The analysis of protein and fiber has proven particularly important for assessing the nutritional quality of primate
diets (Chapman et al. 2002; Milton et al. 1979; Rothman et al. 2009). The ratio of protein to acid detergent fiber in primate foods affects folivorous primate diet choices based on the variation of seasonal food resources (Chapman et al. 2002; Ganzhorn 1992; Oates et al. 1990). Hohmann et al. (2006) reviewed the ecological, physiological, and behavioral aspects of nutrition within primate diets while Janmaat et al. (2006) observed spatial memory of wild mangabeys as the animals tried to remember where successful procurement of fruit had previously taken place. Altman (1991 and 1998) observed nutrition in the context of primate reproductive lifespan and the probable number of infants that would survive. Other studies consider food as a selective force in primates, focusing on primate ecology and the nutritional effects food choice has on primate development (Chapmann et al. 2012; Clutton-Brock 1977; Dunham et al. 2016; Milton 1980). Variation in food quantity and quality requires species of animals to modify their diets to meet nutrition requirements (Rothman 2006). Seasonal fluctuations influence the availability of certain species of foods, affecting the density and distribution of primate populations (Kay et al. 1997; Milton 1990; Rothman et al. 2006; Van Schaik et al. 1993; Worman and Chapman 2005). Other studies found that varying seasonal resources, and the nutritional quality of these resources, affect female and male primates differently (Ganzhorn 2002; Koenig et al. 1998; Rothman 2006; Watts 1998; Worman and Chapman 2006). The nutrition of the foods used by primate mothers strongly affects the future health of their juvenile primates (Altman 1991; Altman 1998; Rothman 2006).
Main Purpose of this Thesis

Understanding how nutrition affects primate biology and behavior is significant for future research in primatology and physical anthropology. Portable NIR has the ability to make this research more efficient and cost effective. There is a remarkable amount of nutrient variation among plant species depending on the ecological season. This requires the analysis of many samples (Rothman et al. 2009; Rothman et al. 2012), typically examined using traditional wet chemistry. Wet chemistry is extremely time-consuming, taking days of preparation for a set of samples. For example, preparing the plant samples I used in this study, took one to three hours per session. During each session I sorted the samples by organ and ground them. This process took one to two hours of loading the samples into bags, and less than an hour for each machine to process the samples for specific nutrients. I repeated this process, multiple times a week, over numerous sessions between 2014 to 2016. The data for this study were generated using wet chemistry and took several hundred hours. Generating the NIR spectra for an equal number of 316 samples took me three hours. Portable near infrared spectroscopy (NIR) will create monumental change for researchers in the field. Portable NIR allows researchers the ability to expand on discoveries immediately after sample analysis. Results found using NIR spectroscopy are available immediately (Coates 2000). Portable NIR makes it possible for researchers to gather, analyze, and synthesis results in the same location where they originally collected samples. NIR spectroscopy is also safer and more cost effective than traditional wet chemistry, requiring no lab, no hazardous chemicals, (Conklin-
Brittain et al. 1998; Rothman et al. 2012; Wrangham et al. 1998) or the cost for acquiring those dangerous chemicals (Van Soest et al. 1991). Infrared technology consists of exposing a sample to infrared light and observing the reflectance peaks that characterize molecular bonds showing different macro and micronutrients (Dunham et al. 2016; McKelvy et al. 1996). However, NIR will never be as accurate as wet chemistry because wet chemistry is the standard for measuring nutrients while NIR compares spectral wavelengths to referential data. I therefore ask: Is portable NIR a reliable field method for the assessment of crude protein, crude fat, and acid detergent fiber in South, African savanna plant matter? My findings suggest that there is preliminary evidence to support that it is. I argue that portable near infrared spectroscopy has considerable promise for use in field settings.
Chapter II: Methods
Preparation

Over a two-year period (2014-2016) I assisted Oliver Paine, a doctoral student at the University of Colorado Boulder, analyze plant samples he collected from the Cradle Nature Reserve, Cradle of Humankind, South Africa from both dry and wet seasons. We used wet chemistry to analyze whole plant samples collected in South Africa. These had been dehydrated and labeled in the field, stored in brown paper lunch bags, and returned to the United States, where they were kept at room temperature on shelves in a lab in the Cristol chemistry building at the University of Colorado Boulder. The bags were labeled with the plant species, the season (wet or dry), the year the sample was collected, and the bag number labeling how many of the same type of sample were collected. Under Paine’s guidance, I took each individual bag and, with scissors, divided the plant by organ: root, leaf, inflor, fruit, rhizome, and stem with assistance from Karen Hillson, Kendall Abady, and Garrett Heidrick. We put the dehydrated sections through an industrial grinder, the Retsch® ZM 200 (see image 1), one plant organ at a time. The Retsch® ZM 200 uses centrifugal technology to reduce the size of fibrous material. This process is a standard in agriculture and chemical industries.

For a diagram of the methods used, please refer to Appendix 7.
The powdered plant samples were then split into two groups, one to be tested using wet chemistry, the second to be tested with portable NIR. The first sample set was placed in plastic containers, (1x1½ inch around) and were to be used for the wet chemistry analysis. The second sample set was placed in glass jars (about 2 inches tall with a 1 inch diameter). Each glass jar and plastic container was labeled with the same information as the bag containing the original samples. The glass jars were then stored in a drawer in the lab until they were to be used for portable NIR spectroscopy analysis.
**Analysis Using Wet Chemistry**

The labeled plastic containers containing the samples were placed on racks in an Excalibur ® food dehydrator to remove water content until the samples were ready to be tested using wet chemistry. Concentrations of the sample’s ADF, crude fat, and crude protein were determined using an ANKOM® Fiber Analyzer, ANKOM\textsuperscript{XT15} Extractor, and Leco FP-528 respectively.

To analyze fiber and fat concentrations, suggested amounts of each sample were placed into ANKOM F57 Fiber Filter Bags and XT4 Fat Filter Bags. To calculate fiber, the suggested sample amount was 0.5g to 1g while the suggested amount to calculate fat was 1g to 3g. These bags were labeled with a Sharpie®. The season, the year, an abbreviation of the plant species, the type of plant organ, the number of the bag the organ was from, and the batch number, were noted on the bag. For example, if the sample contained *Paspalum urvillei* leaves from the wet season in 2015 from Bag 2 and the batch number was 17, the bag would be labeled “Wet 2015 TR P.urv. Leaf 2 B17.”

We then created a table to record the analysis information for each sample. The columns included the batch, the sample ID (listed in the example above), the filter bag weight, the sample weight, and the sample weight after the process analysis. We first used the Sartorius mechatronics CPA1245 laboratory scale to find the weight of the fiber bag without the sample and recorded the information. After recording the weight of the bag, the Sartorius CPA1245 was tared so the sample could be added. The sample was added into the bag with small metal scoops, while the bag was held open in a small black cylinder (Bag
Weight Holder) sold by ANKOM. The suggested amount of sample was added into the specific analysis bag depending on the nutrient we were analyzing. Once enough of the sample was added to the bag, the ANKOM Heat Sealer was used to seal the bag and encapsulate the sample. We recorded the weight of the contents of the bag. The weight of the bag by itself was subtracted from the weight of the bag with the added contents, to give the exact weight of the sample that had been added. The samples were stored in a Moisture Stop pouch with desiccant packs to absorb moisture until the samples were ready to be analyzed.

We used the ANKOM2000 Fiber Analyzer (See image 2) to analyze the fiber concentrations for the plant samples. 24 sample bags made up one batch and they were placed in the bag suspender supplied by ANKOM. Once the proper analytical method was chosen, the machine introduced the suitable chemicals to perform each rinse, where the chemical solution was heated and solublization was able to occur. The remaining residue after the sample goes through each detergent solution is the fiber concentration (Rothman 2009; and Van Soest et al. 1991).
To analyze fat concentrations the ANKOM$^{XT15}$ Extractor (see image 3) used Soxhlet extraction technology to extract lipids from a solid material. The machine processed a batch of 15 samples for a time period between 30 and 60 minutes at 90°C. Solvent was added to the samples and was then heated to extract for crude fat.
Image 3. The ANKOM\textsuperscript{XT15} Extractor.


After the machines finished processing the samples for fiber and fat, the sample bags were taken out to dry and placed in the Precision™ Oven\textsuperscript{5121126} (see image 7) to remove water content. Once the bags were dry, each bag and its contents' new remaining weight were again recorded. The new measurement of the bag and its contents was then subtracted from the original measurement of the bag and its contents to give the measurement of the crude fat.
To analyze protein I used the Leco FP-528 Nitrogen analyzer (see image 5). The plastic containers containing each sample were removed from the dehydrator. Instead of putting the samples into bags, which had been done for the calculation of fiber and fat, the samples were placed on foil squares referred to as sample holders. Each foil square can contain up to 250mg of a single sample. The foil square was folded up around the sample and the top of the remaining part of the foil was twisted. When the sample was ready to be placed into the Leco FP-528 it resembled a raindrop (see image 6). The sample holder was weighed before the sample was added and the sample, in the holder, was again weighed using the CPA1245 Laboratory Scale and the information was recorded. The samples were then loaded into the machine where atmospheric gases were removed. CO₂, H₂0, NOₙ, and N₂ passed through a furnace and thermoelectric cooler and finally into a ballast collection system where all evolved gasses were
collected. The gases collected here were equilibrated and a small portion was used to convert the gases while the thermal conductivity cell, measuring for nitrogen, reduced the rest of the sample. The machine then communicated the concentration of protein for each plant sample to a PC computer.

Image 5. Leco FP-528.

Image 6. Leco FP-528 Sample.
Leco FP-528. Available from: Pictures from https://encrypted-tbn2.gstatic.com/images?q=tbn:ANd9GcQXGt966f3XwMa7VwaKC646R8YqroE-I0BACfgojxsZeQwFB36https://encrypted-tbn2.gstatic.com/images?q=tbn:ANd9GcQXGt966f3XwMa7VwaKC646R8YqroE-I0BACfgojxsZeQwFB36
Master Excel Sheet

Paine recorded all of the nutrient concentration information and created a sheet in Excel with all of the data. Each row in the sheet listed the lab, the field season, the species, the plant type, the transect, the habitat, the bag number, the organ, the structure, the organ group, and then had a variety of macro and micro nutrients listed with the concentrations. For example, a row would read as follows: ‘CU, DRY 2014, Searsia lancea, Tree, DOW, Open, 1, Leaf, Leaf, Dicot Leaf, NDF 27.22’. I later used this Excel sheet as my reference point while I created my calibration models with the NIR.

Near Infrared Spectroscopy

Using portable near infrared spectroscopy (NIR), I collected spectra for 316 plant samples Paine collected in the Cradle Nature Reserve, Cradle of Humankind, South Africa during wet and dry seasons between 2013 and 2015. I combined the spectra and original wet chemistry data of crude fat, acid detergent fiber, and crude protein, running a total of 94 experiments. I collected the spectra for the samples and compared the raw material for crude protein and fiber using the ASD LabSpec® Near Infrared Spectrometer and the Indico® Pro Spectral Acquisition software. Data collection took place over a period of three different lab sessions, each one around two and a half hours, collecting data for roughly 120 plant samples each session. I had help from Erin Smith, a graduate student at the University of Colorado Boulder. We analyzed the plant organ samples by each transect: Dolomite Open Woodland (DOW), Bloed veld (BV), Kudu Hill (KH),
Pieter's Vlei (PV), Tick River (TR), Baboon Food (BF), or No Transect (NT). The jars were taken out of the drawers in these subdivisions and lined up by transect and plant species, to make the recording process more efficient. We used the ASD software and checked the accuracy of the portable NIR by running a spectral analysis on a plastic sheet provided by the company. This sheet came with information where spectral peaks should occur on the sample analysis. Coates (2002) argues that NIR results are only as good as the reference values found in the lab. Because of these findings, our team decided it was practical to take multiple scans from the same sample, and have multiple samples of the same plant species and organs. We analyzed the calibration sheet three times for the most accurate reading and compared where the spectral peaks should be to the reference data. The peaks were always correct, allowing us to take a baseline reading. Once the baseline was taken, each sample was analyzed three times. One person recorded the name of the sample on the lid of the jar and on a separate sheet of paper, the name of the .asd file name of the plant sample. Another person handed the jars off while the other stuck the metal probe against the sample. Each sample took 30 seconds to be processed. Each of the three times the sample was analyzed, the probe was moved to a different location on the sample. One spectrum per plant sample was saved, which was the average of the three analyses, and stored as an .asd file. A new baseline measure was taken each time 10 samples were analyzed to ensure the spectra were correct.
Master Excel Sheet and New Spreadsheet

I added a column containing the .asd file of the NIR spectrum to its corresponding analyzed plant sample using wet chemistry on the Excel sheet. This information was recorded in the Microsoft Excel sheet master file containing the wet chemistry data. I coded the .asd file name incorporating the NIR spectrum file in Excel and recorded the file next to its matching plant sample on the Excel sheet. I inserted four columns after the field season and before the specimen column. The first column contained the word ‘Spectrum,’ the second column contained the number of the spectrum file, the third column contained the type of file name to be used for Grams IQ™ (the program with which I created chemometric regression models using partial least squares regression) and the fourth column contained the combined file name of the spectrum from the previous columns. The command I used was =$(the column spectrum was written in as well as the number of the row &TEXT($the column the number of the file was scripted in, as well as the number of the row, “00000”)&the column the file
type was scripted in, as well as the number of the row. For example row 1050's command is:

\[ =E1050&\text{TEXT}(F1050,\text{"00000")}\&G1050 \]

Once all of the file names were created, I generated a new Excel spreadsheet. On the new Excel spreadsheet the data were sorted by plant organ, with each organ having its own tab.

**Software**

I converted the NIR samples from the .asd file to a file suitable for the Grams IQ™ Spectroscopy software. Once I had converted the spectra to .cfl and .spc files, I uploaded them onto Grams IQ™ and paired them with their constituent reference information: crude protein, crude fat, or acid detergent fiber. Grams IQ™ delivers multivariate analysis for all applied spectra allowing me to create qualitative and quantitative chemometric calibration models.

The software came with a *Chemometrics Training Manual for Grams IQ/IQ Predict v9.1* (ASD Inc.). The manual explained how to convert files and configure training sets, set-up the experiments and the procedures, use the experiment diagnostic options and understand the results. I grouped the spectra by plant organ: fruit, inflor, leaf, rhizome, root, stem, stem base, and underground storage organ (USO). I uploaded the converted .spc files in groups by plant organ to create regression models that analyze the relationship between the variables for each plant organ spectra and the corresponding data for wet chemistry.
Regression Models

Grams IQ™ explained how to create quantitative models using the training manual. To produce the simplest predictive models, I tried to obtain the fewest factors, the highest $R^2$ values, as well as the lowest Standard Error of Cross Validation (SECV). My organization process for creating the regression models consisted the three different saved Grams IQ™ files for crude protein, crude fat, and acid detergent fiber. I uploaded the spectral .asd files as well as the corresponding data from the Excel sheet. I separated experiments by individual organs and I constructed an experiment for all the organs combined. I ran each experiment for crude fat, crude protein, and acid detergent fiber, originally making regression models in the preliminary experiment only for ‘All Organs.’ The Model Setup selected the calibration type, factors, and the diagnostic type, and I applied certain types of preprocessing. I chose to only display the raw data for the spectral data type.

![Experiment Properties Plots](image8.png)

Image 8. Experiment Properties Plots.

While initiating the Model Setup, I chose ‘PLS-1’ as the calibration type. PLS-1 is Partial Least Squares Regression combining the constituent information (y-axis) with the spectral information to create a model that can predict accuracy.
This model is used when the constituents are independent of one another. I used Cross Validation as the diagnostic type, as it provides an estimate of predictive error. The program also suggested putting the samples in sequential order.

![Image 9. Experiment Properties Model Setup.](image)

For the preprocessing, I chose ‘Mean Center’ as the data preparation because it is usually used for spectral modeling. Mean center subtracts the average spectrum from the sample set allowing the differences to become more visible, resulting in better models. I did not adjust for baseline, and for the pathlength correlation option I originally chose ‘Multiplicative Scatter Correction’ (MSC). I used MSC when the spectra appeared to follow the same pattern, but the spectra were on different sections of the y-axis. MSC creates an “idealized
spectrum" using the average of the sample spectra. The slope and offset effects are removed over the different regions of the spectrum. I selected the option to use Savitzky-Golay (SG) first-order derivatives as it results in less wavelength shift while also showing the differences between samples. I preprocessed using derivatives when wavelength features were overlapping extensively. I chose an initial amount of 50 smoothing points and included the entire spectral region from 350 nanometers to 2500 nanometers. Using the factor loadings plot discussed below, I found where noise was present and removed the excess noise from models by changing the spectral regions (see Image 18 in Appendix).

![Image 10. Experiment Properties Preprocess.](image)

For each experiment I changed one variable at a time, removing outliers if necessary or focusing on different parts of the model setup. I changed wavelength range frequently and used preprocessing to prevent possible outliers so that I
could identify what changes, if any, contributed to the enhancement or deterioration of the experiment. I also reviewed different plots including: PRESS, Actual vs. Predicted, Spectral Residuals, Factor Loadings, Score vs. Score, Concentration Residuals vs. Spectral Residuals, Leverage vs. Studentized Residuals, and Standard Error Cross Validation (SECV). The PRESS and the SECV plots note the number of factors and margin of error that exists with each experiment. The Actual vs. Predicted plot show the $R^2$ value between zero and one. As the $R^2$ value approaches one, the regression is stronger, meaning the calibration model can better predict unknowns. Points that did not fall on the trend line were potential concentration outliers. They were then compared across multiple plots to prove their status as an outlier.

![Image 11. Actual vs. Predicted Plot.](image.png)
The Spectral Residual plot shows spectra that are spectrally different from one another. As the point moves farther away on the y-axis, it has a greater potential to be an outlier.


The Leverage vs. Studentized Residuals plot shows how distinct the independent variable values are resulting from division of a residual by standard deviation.

Image 13. Leverage vs. Studentized Residuals Plot
The ‘Factor Loadings’ plot shows areas along the models where noise occurred. I reviewed important spectral regions and reevaluated potential region(s) for the next experiment. This plot identified areas that had importance in the model as well as areas that did not.

The ‘Concentration Residuals vs. Spectral Residuals’ use both x- and y-axis parameters. This plot creates a mean cluster with potential outliers farther up on the y-axis and shifted farther on the x-axis.

Image 15. Concentration Residuals vs. Spectral Residuals Plot
I evaluated each plot for potential outliers (specifically focusing on y-axis outliers) and compared across all plots before any samples were defined. This allowed me to verify samples that may have been problematic.

Image 16. Results Page Setup.

Once I reviewed the problematic points for each plot, I cross-referenced them to the Master Excel Sheet and reviewed the type of organ and the species to verify whether they were valid explanations for the outliers.

I named each experiment based on the changes compared to the previous experiment. For example: ‘All Organs Changing Spectral Range: 1850-2500’ (See image 17).
Image 17. Experiment Setup.
I ran a total 94 experiments looking at crude protein, crude fat, and acid detergent fiber (ADF), looking in-depth at potential changes to make each regression stronger. The regression revealed concentrations of each plant sample, showing where the crude protein, crude fat, and acid detergent fiber contents were already known to the spectra. This allowed me to compare the applications and identify the accuracy and applicability of portable NIR.
Chapter III: Results and Analysis

Table I. Predictive power and error associated with calibration equations used to estimate characteristics of savanna African plant nutrition.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>5</td>
<td>R²=.845</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>350-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>5</td>
<td>R²=.858</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>350-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>7</td>
<td>R²=.825</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>350-750, 1350-1650, 1700-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>4</td>
<td>R²=.883</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1350-1650</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>7</td>
<td>R²=.885</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1350-1650</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>4</td>
<td>R²=.861</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1700-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>5</td>
<td>R²=.889</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1700-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>All Organs</td>
<td>6</td>
<td>R²=.882</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1700-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>Leaf</td>
<td>7</td>
<td>R²=.801</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>350-800, 1350-1650, 1900-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>Leaf</td>
<td>4</td>
<td>R²=.923</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1350-1650</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>Leaf</td>
<td>7</td>
<td>R²=.902</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1800-2500</td>
</tr>
<tr>
<td>Cradle ADF</td>
<td>Leaf</td>
<td>6</td>
<td>R²=.924</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1800-2500</td>
</tr>
<tr>
<td>Cradle CP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Cradle CP</td>
<td>All Organs</td>
<td>7</td>
<td>R²=.793</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1850-2500</td>
</tr>
<tr>
<td>Cradle CP</td>
<td>Leaf</td>
<td>6</td>
<td>R²=.868</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1850-2500</td>
</tr>
<tr>
<td>Cradle CP</td>
<td>Leaf</td>
<td>7</td>
<td>R²=.886</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1850-2500</td>
</tr>
<tr>
<td>Cradle CP</td>
<td>Leaf</td>
<td>8</td>
<td>R²=.909</td>
<td>Raw</td>
<td>PLS-1</td>
<td>25</td>
<td>Cross Validation</td>
<td>Sequential</td>
<td>1</td>
<td>Mean Center</td>
<td>No Baseline</td>
<td>MCS</td>
<td>SG1</td>
<td>50</td>
<td>1850-2500</td>
</tr>
</tbody>
</table>

Cradle ADF= acid detergent fiber; Cradle CP= crude protein; SECV= standard error of cross validation; Actual vs. Predicted= relationship between spectra and wet chemistry data; Model Setup and Preprocessing see Regression Models in the Methods section.
Chapter III: Results and Analysis

I compared the standard error of cross validation (SECV) and R² value for each experiment. Aside from crude fat, the equations for each nutritional attribute had high predictive power 80-90% of the time. The predictive power for acid detergent fiber in leaves was relatively high (R² = .923); the predictive power of acid detergent fiber for all organs and the crude protein of leaves had R² value above .8; the predictive power of crude protein for all organs was lower with R² = .793 (see Table 1) In the industry, regression coefficients above .95 are considered excellent. Crude fat had a very low predictive power around R² = .001 (See Appendix 4). The experiments with the highest predictive power and lowest SECV for crude protein and acid detergent fiber for ‘leaves’ and ‘all organs’ are bolded in Table 1. The spectral display and plots for these experiments are located in Appendix 2, Appendix 3, Appendix 5, and Appendix 6. The spectral display and plots for crude fiber “All Organs” are located under Appendix 4.

Other studies in primatology have utilized NIR to analyze nutrients of tropical plants. Rothman et al. (2009) found that NIR had a high predictive power for herbs and trees eaten by gorillas using modified partial least-squares regression. Neutral detergent fiber, acid detergent fiber, and crude protein had an r² value of .95 or greater. The NIR results I found for the savanna grasses had a lower predictive power for crude protein and acid detergent fiber, but all organs of a plant sample were incorporated in this analysis while Rothman et al. (2009) strictly used tree leaves, herbaceous leaves, inner stem core and outer herb peel.
Overall, the models made using portable near infrared spectroscopy were able to accurately predict crude protein and acid detergent fiber in savanna African plants, but predicting the crude fat proved to be more difficult. After running 94 experiments, removing outliers and changing variables within the model setup and preprocessing, making predictive models for crude fat was ineffective. Of all the data collected, some samples were excluded because they were labeled as duplicate plants while others did not have reference data from the original wet chemistry analysis. When I made models for “All Organs,” some of the plant samples had to be discarded because they were spectral outliers. These included samples such as spectral sample ‘00089’ Plantago, a plantain that is rarely sampled within this ecosystem. For the experiments originally containing “All Organs” a total of 234 plants were analyzed before the removal of the outliers. For the predictive models made for leaf plant organs, a total of 89 leaf samples were used before the removal of the outliers. Paine and I analyzed the nutritional values of plant organs spanning over 70 plant species and examined spectral data for 39 different plant species, finding organ specific nutrient properties for species spanning savanna South African ecology.

Limitations in NIR

Portable NIR will never be as accurate as wet chemistry as NIR is limited to the reliability of a sample’s composition. If one sample does not accurately represent the species it can provide inaccurate information. It is important to test multiple samples of the same species, continuously monitoring and updating the
sample information. When generating predictive equations, particular regions of the spectra provide better predictive equations. A model containing the full spectral region has more variables than a region specific model, which can cause the predictive equation to be less accurate. For this reason it is important to note where there are areas of importance on the spectra and limit the regression to those areas of importance. While analyzing samples and making regression models, researchers should use the most fitting analytical technique so the predictive equations are more accurate.

My results indicate that portable near infrared spectroscopy is applicable when generating robust equations to predict nutritional concentrations of crude protein and acid detergent fiber for ‘all plant organs’ and ‘leaves’ in savanna South African plants.
Chapter VIII: Conclusion and Recommendation for Future Work

Portable near infrared spectroscopy is an emerging technology with potential in academia and industry. With portable NIR, researchers can gather samples and get nutritional data on site. My results indicate that portable near infrared spectroscopy is useful for predicting the nutritional content of certain plant organs in South African savanna plants. The equations I generated for the South African savanna plant samples accurately predicted crude protein and acid detergent fiber, but proved less consistent when predicting crude fat, as the $R^2$ value was very low.

There have currently been no studies that have utilized NIR to access the nutritional properties of the broad suite of savanna plants analyzed here. This can be revolutionary for scientific fields studying savanna animals with diverse herbivorous diets because portable NIR broadens the ability to instantly access information, making it possible to expand on findings immediately and efficiently, negating the problem of additional permits and sample storage and transport.

Future research using portable near infrared spectroscopy ought to focus on changes in ecological systems, tracking varying nutrient changes in specific plant species over different seasons and years. Researchers should take tracking changes in nutrient further by comparing sexual dimorphism in primates based on the diets they have and the nutrient content within these specific diets over a longer period of time (e.g., over one season or multiple seasons).

Furthering this study, portable NIR allows researchers to increase sample size and scope. For instance, researchers can take spectra of everything the
baboons in the Cradle of Humankind, South Africa eat, instead of relying on a few samples that can be collected and shipped back to the lab. It would also be of interest to use these regression models and compare them with savanna plant samples collected from Kenya, East Africa to understand the applicability of these regression models across different environments in Africa. Portable NIR has the ability to transform the effectiveness of field research by providing access to nearly instantaneous information once predictive models have been produced.
Chapter VI: Appendices

Appendix 1. Image 18. Change in Spectral Region. Region 1850-2500nm highlighted.
For explanation of table refer to Table 1 description in Results.
Plots display of acid detergent fiber for "All Organs" changing spectra 1350-1650nm.
Spectral display and experiment setup of acid detergent fiber for plant organ „Leaf“.

Appendix 3. Cradle ADP Leaves
Plots display of acid detergent fiber for plant organs. Leaf changing spectra 1350-1650nm removing outlier spectrum 00080.
Appendix 4. Cradle CF All Organs

Spectral display of crude fat for "All Organs".
Plots display crude fat for "All Organs," not changing the spectral region.
For explanation of table refer to Table 1 description in Results.

Appendix 5, Cradle CP All Organs
Plots display of crude protein for "All Organs" changing spectra 1850-2500nm, removing outliers 52 and 251.
Special display and experiment setup of crude protein for plant organ “leaf”.

For explanation of table refer to Table 1 description in Results

Appendix 6. Cradle CP Leaves
Plot display of crude protein for plant organ "Leaf" changing spectra 1850-2500nm.
Appendix 7.

Diagram of Methods section.
Chapter IX: References


Coates, D. B. (2002). Is near infrared spectroscopy only as good as the laboratory reference


Ecol. 6: 493-499.


Precision™ Oven 5121126 Precision™ Oven 5121126. Available from: Pictures from http://www.americaninstrument.com/imagesfromdatabase/3698C_OVEN.jpg


