Novel NIR Spectroscopy Correlation Approach to Amino Acid Analysis of Soybean Proteins for Composition Improvements

Valentin I. Prisecaru and I.C. Baianu

FSHN Department – College of ACES
AFC-NMR & NIR Microspectroscopy Facility

University of Illinois at Urbana-Champaign
Urbana, Illinois USA
# Table of Contents

- Proposed Project Summary
- Proposed Project Description

1. Introduction
   - Overall Goals
   - Research Objectives
   - Previous Studies: Referenced Summary of Previous Work
   - Preliminary Data

1. Rationale and Significance

2. Research Methods
   - Proposed Studies and Techniques
   - Application of Results, Future
   - Limitations and Advantages of Methodologies

1. References

2. Appendices
Proposed Project Summary
Overall goal: To generate and improve NIR calibrations for the estimation of amino acid composition in soybeans and soybean proteins. This proposed project is focused on improving the reliability, transferability and speed of NIR spectroscopic methods of analysis for amino acids in whole soybeans and soy proteins for food and feed precise matrix formulations.

Potential for: Improvements in agricultural efficiency, human and animal food functionality, and human health.

Finding soybean varieties that vary in specific amino acids, such as essential, or conditionally essential, amino acids, is critical for finding new uses of specific soybean lines with improved traits for food and feed applications.

**PROJECT SUMMARY**

<table>
<thead>
<tr>
<th>ID</th>
<th>Spectrum Title</th>
<th>Index</th>
<th>Phe</th>
<th>Arg</th>
<th>Leu</th>
<th>Ala</th>
<th>His</th>
<th>Performance Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>88459</td>
<td>Calibration</td>
<td>9</td>
<td>1.210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15521</td>
<td>Calibration</td>
<td>10</td>
<td>1.240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12067</td>
<td></td>
<td>11</td>
<td>1.300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hs934118</td>
<td></td>
<td>12</td>
<td>1.400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8061</td>
<td></td>
<td>13</td>
<td>1.450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12131</td>
<td></td>
<td>14</td>
<td>1.480</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11195</td>
<td></td>
<td>15</td>
<td>1.490</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1197</td>
<td></td>
<td>16</td>
<td>1.500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>561363</td>
<td></td>
<td>17</td>
<td>1.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>290149</td>
<td></td>
<td>18</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>423912</td>
<td></td>
<td>19</td>
<td>1.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>423914b</td>
<td></td>
<td>20</td>
<td>1.200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>417414a</td>
<td></td>
<td>21</td>
<td>1.200</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>473277</td>
<td></td>
<td>22</td>
<td>1.050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15520</td>
<td></td>
<td>23</td>
<td>1.300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
I plan to develop calibration plots of three selected amino acid groups that include essential amino acids for identified soybean accessions.

Conventional “wet chemistry” analytical methods are time-consuming and costly. As a result, soybean breeders and researchers have an imperative need to utilize faster and less expensive methods. NIR Spectroscopy is a rapid and inexpensive method for composition analysis for academia and industry. Recent advancements in instrumentation design, such as the application of the Diode Array (DA) technique and the Fourier Transform (FT) IR and NIR techniques, have significantly improved overall instrument performance and advancement in the field of grain analysis.
In the following research proposal, the state-of-the-art FT-NIR instruments will be evaluated and utilized, data analysis and calibration methodology will be substantially improved, in order to allow rapid NIR analysis of the large number of soybean samples required for improving the genetics of soybean seed composition for agricultural cost savings and human health food applications.
Proposed Project Description
1. Introduction

- Overall Goals
- Research Objectives
- Background
- Referenced Summary of Relevant Work
Overall Goals

- To design improved approaches to NIR calibrations for amino acid composition analysis of soybean proteins, soybeans and soybean food and feed matrix formulations

- To develop NIR calibration transfer methodology between different laboratories for various food and feed applications
Research Objectives
Research Objectives

Objectives that need to be addressed in order to realize NIR’s utility in practical food applications and lab analysis:

1. Generate an NIRS calibration for all of the amino acid residues in soybean proteins that are determined by a primary method.

2. Obtain an NIRS calibration for selected groups of three amino acid triplets, such as Arg-Lys-Glx, Asx-Ala(or Val)-Pro, or (Met+Cys)-Arg-Val. (Can such determinations result in significant savings of time and money in research and food industry labs?)

1. Obtain appropriate and accurate values for use in precision formulation matrices.
Research Objectives

4. To produce NIR amino acid calibrations for a small group of three amino acids.

5. To investigate and develop methodology for NIR spectra preprocessing and data analysis to improve both the accuracy and reliability of NIR measurements of soybean seed composition.

6. To develop and optimize NIR Spectroscopy calibrations for determination of the above-mentioned amino acids in soybean seeds.
Research Objectives

7. To obtain reproducibly calibration plots of the above-mentioned four individual amino acids in soybean lines, using PLS-1 and PLS-2 regression algorithms.

8. To perform multivariate analysis to better resolve the individual amino acid calibrations.

9. To carry out the analysis and comparison of soy proteins of similar amino acid composition in powder vs. gel vs. liquid suspension, in order to investigate the matrix effect on the NIR calibrations.

10. To compare reduced and unreduced soybean protein calibrations, to improve the cysteine vs. cystine NIR calibrations.
Research Objectives

11. To compare our results with the results of other laboratories’ data, in order to investigate the transferability of NIR calibrations among different laboratories.

1. To compare results with $^{13}$C (Waltz) NMR and GC-MS (WAHU-HA) primary data for amino acid composition. (possibly remove)
Objectives, cont’d

13. (read the 60 page review, then write something about multivariate analysis that is better) Evaluate multivariate analysis methodology for NIR calibrations in order to determine soybean protein and soybean amino acid residue contents (go to wiki, or google, etc., and learn some)

14. Investigate the potential of NIR spectroscopy for developing calibrations for amino acids and amino acid mixtures [amino acid triple matrix method] [specify which amino acids and which mixtures – specify aa’s, but for the mixtures, refer to the table in the next page]

15. Generate NIR calibrations of the soybean protein amino acid residues specified in the table on the next page, based on high-resolution nuclear magnetic resonance analysis ((make calibration plots))
Background
Background

The soybean:

- more than 3 Billion bushels produced in the US each year (USDA, 2007)
- Major source of plant protein and oil (and a high-level plant source of Methionine and Tryptophan)
- Protein content from different soybean cultivars vary greatly
  - Some have over 50% protein (dry wt.)
  - Some accessions show significantly higher Methionine (~19%*) and Cysteine levels

*Kuiken, 1948*
Background

- NIR spectroscopy has been widely applied to the analysis of major components in grains and oilseeds since the 1960s. However, both the accuracy and reliability of previous applications have been limited by the instrumentation and data analysis techniques.

- Until 2004 the transferability of calibration data in a systematic and verifiable manner have not been possible because of the lack of suitable instrumentation and methodology.
Soybean Uses

Main Soybean Growing Countries:
- United States, Brazil, Argentina, China and India

Some final products from soybean processing:
- Foods, Nutraceuticals, soy isoflavones
- TVP
- Animal feed
- Adhesives, Fibers, Lining
- Foams
- Fertilizers
Usage in Industry

• Developmental Labs and Grain Labs in industry have been reluctant to use NIR because of the low quality of instruments available (until recently) and to an extent, the lack of proper calibrations.

• It has already been used in the area of new grain development, genetic selection and cross-breeding.

• Because of its high sensitivity, NIR is useful as a rapid and inexpensive screening tool, despite not having very high resolution, if a robust and accurate calibration can be generated.
Current Status

- The Food Industry and Nutritional Sciences have a great need for rapid techniques that are economical, accurate, reproducible and nondestructive.
- Protein Quality is an important processing and nutritional attribute
Current Status, cont’d

• Accuracy and reliability of previous applications have been limited by the instrumentation and data analysis techniques.

• In recent years NIRS instrumentation has evolved.

• Diode Array (DA-NIR spectrometers) and Fourier Transform (FT-NIR spectrometers) techniques have significantly improved speed, sensitivity, resolution and reproducibility in comparison with previous conventional instruments.
Soybean protein and amino acids – best method

• The methods that exist for judging protein quality are mostly destructive, and possess severe limitations, like changing the structure of amino acids before they’re quantitated.

• SS NMR is an established method used to identify the aa residues, but has resolution limitations. However, NMR can be done in liquids or gels that improves the resolution.

• Drawback: NMR takes much longer than NIR

• NIR is a powerful secondary technique
UIUC NIR Soybean Database

• Our high-resolution NIR calibrations and methodologies were employed to carry out a large number of protein and oil composition analyses of soybean seeds (~50,000) for breeding and selection purposes, over a period of three years.

• A wide variety of soybean developmental lines and more than 2,000 exotic soybean germplasm accessions were thus characterized accurately and reproducibly (Source: UIUC Soybean NIR Database).

• Such results demonstrate the usefulness of this novel NIR approach for soybean selection and breeding purposes. They also validate our NIR calibrations undertaken in parallel with the higher resolution (but slower and more expensive) NMR measurements.
Practical Implications of our High-resolution NIR Analyses of Soybean Seeds

- High-resolution FT-NIR Reflection Spectroscopy is capable of rapid, reproducible and accurate analyses of food grains and foods when careful calibrations and appropriate data corrections are carried out.
- This can result in huge savings both in research and food industry labs.

• (combine this slide and the next one)
Applications

- R & D
  - Food Formulations and Protein Quality / Amino Acid Composition Analysis
- Food Developments
- Food Safety and Microbiology Applications
- Health Foods
- Nutraceuticals
- Nutrition Research
  - Agricultural Feeds and Pet Foods
Biomedical Applications

High-resolution NIR Chemical Imaging may also enable rapid and sensitive analyses with micro-arrays for Nucleic Acids, multiple Molecular Bioassays, Automated Proteomics, Biotechnology, Biomedical & Pharmaceutical Applications, such as those aimed at early Detection of Cancer and Prevention.
Referenced Summary of Previous Work

- VIP Publications
- ICB et al. Publs
- Orf et al. Publications, etc.
VIP References


• Assessing vitamin C and E inadequacies in U.S. adults and suggesting ways to reduce inadequacies. ADA Conference, St. Louis, MO 2001.

• The effects of various macronutrient ingredients in energy bars on blood glucose levels. ADA Conference, St. Louis, MO 2001.

• The effects of various macronutrient ingredients in snack bars on energy levels, satiety and hunger. FFH Conference, Urbana, IL, 2002. Abstract.

• The starting point for a healthy diet, ADA 2003 Oct. Abstract.


• Poster presentation: The effects of various protein and carbohydrate ingredients in snack bars on blood glucose levels. ADA Conference, St. Louis, MO 2001.

• Poster presentation: Assessing vitamin C and E deficiencies in U.S. adults and suggesting ways to reduce deficiencies. IDA meeting, Chicago, 2002.

• Poster presentation: The effects of various protein and carbohydrate ingredients in laboratory-made energy bars on blood glucose levels. IDA meeting, Chicago, 2002.

• Poster presentation: The effects of various protein and carbohydrate ingredients in snack bars on energy levels, satiety and hunger. FFH Conference, Urbana, IL, 2002.

• Poster presentation: The starting point for a healthy diet. ADA 2003 Oct.

ICB References
Other References

- Orf et al. refs

- Other refs
Previous Studies

Brazil, 2003 to 2006

• Collected soybean samples in several different states in Brazil
• Analyzed amino acids in soybeans using:
  • HPLC
  • Derivatized HPLC
• Individual amino acid compositions of the soybean samples showed significantly different amino acid mean levels (p<0.01 and p<0.02) – Met, Lys, and Thr
• Tentative Conclusion: it is possible to calculate the a.a. content of a sample for several amino acids by comparison with a primary method; however, primary data was not available for several essential amino acids.
Rationale and Significance
Rationale and Significance

Both major advancements in instrumentation, and improved data analysis/novel calibration methodologies are necessary to improve the accuracy and reliability of NIR for measuring low-level components such as individual amino acids.
• High-protein, high-yield cultivars increase the soybean crop value. Conventional, or “wet chemistry”, methods are time-consuming, expensive and impractical for repetitive measurements required for genetic selection and breeding experiments to increase both protein content and the agronomic yield values of soybean cultivars

• Faster and less expensive methods for protein, oil, moisture and amino acid analysis of soybeans are needed
Rationale and Significance

Novel NIR instrumentation techniques – combined with improved data analysis and calibration methodologies – are essential for selecting soybean cultivars with both high quality protein composition and high agronomic yield.

Such improved NIR analysis can also result in enormous cost and time savings for the amino acid composition analysis that is required for example by soybean breeders in genetic selection experiments.
Preliminary Data
Flow-Chart of the Steps in a Novel Approach for Amino Acid NIRS Analysis of Soybeans and Proteins

Selection of Standard Samples with the widest Range of Equally-Spaced Values (ES-ROV)

Primary Methods of Analysis for:
(AA, P, M, and O (oil))

Correlation Test with Protein %D

Selection Criteria:
High ROV > 20%
and Low Correlation: < 90%

Precision Formulation Matrix for Calculating Concentration Steps

NIRS Measurements

Next Group of Amino Acids

PLS-1

PLS-2 (n<6: 3 AA+P+M+O)

Selection of an Amino Acid Triplet Group for Calibration

3 AA Composition Results
Protein Calibration for Bulk Soybean Analysis on the Spectrum One NTS FT-NIR Instrument

65 calibration standards, 20 grams for each standard, 8.9mm NIR beam size

Source: Soybean NIR Database, UIUC
Comparison of NIR Dry Soy Protein Data (Primary data: Sigma Method – Lowry modified)

Val’s DP ZX50 vs. DP Peoria

$n = 45$ data points
ROV = 20%

$y = 0.839x + 9.726$
$R^2 = 0.92$
$R = 0.96$
Correlations of Amino Acids with Crude Protein in Soybeans

<table>
<thead>
<tr>
<th>A.A.</th>
<th>R²</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASX</td>
<td>0.634</td>
<td>0.796</td>
</tr>
<tr>
<td>THR</td>
<td>0.005</td>
<td>0.071</td>
</tr>
<tr>
<td>SER</td>
<td>0.578</td>
<td>0.760</td>
</tr>
<tr>
<td>GLX</td>
<td>0.690</td>
<td>0.831</td>
</tr>
<tr>
<td>PRO</td>
<td>0.480</td>
<td>0.693</td>
</tr>
<tr>
<td>GLY</td>
<td>0.628</td>
<td>0.792</td>
</tr>
<tr>
<td>ALA</td>
<td>0.619</td>
<td>0.787</td>
</tr>
<tr>
<td>VAL</td>
<td>0.640</td>
<td>0.800</td>
</tr>
<tr>
<td>ILE</td>
<td>0.631</td>
<td>0.794</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A.A.</th>
<th>R²</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>LEU</td>
<td>0.644</td>
<td>0.802</td>
</tr>
<tr>
<td>TYR</td>
<td>0.536</td>
<td>0.732</td>
</tr>
<tr>
<td>PHE</td>
<td>0.474</td>
<td>0.689</td>
</tr>
<tr>
<td>HIS</td>
<td>0.499</td>
<td>0.706</td>
</tr>
<tr>
<td>LYS</td>
<td>0.652</td>
<td>0.807</td>
</tr>
<tr>
<td>ARG</td>
<td>0.687</td>
<td>0.829</td>
</tr>
<tr>
<td>MET</td>
<td>0.548</td>
<td>0.740</td>
</tr>
<tr>
<td>CYS</td>
<td>0.498</td>
<td>0.706</td>
</tr>
<tr>
<td>M+C</td>
<td>0.538</td>
<td>0.733</td>
</tr>
</tbody>
</table>
Amino Acids Highly-Correlated with the Dry Protein Content:

- Histidine: $R = 0.93$
- Arginine: $R = 0.90$
- Glx: $R = 0.88$
- Valine: $R = 0.87$
- Leucine: $R = 0.85 \rightarrow \text{borderline}$

and the imino acid Proline: $R = 0.87$
Histidine

- Basic
- Polar
- Positively charged
- $pK_1 = 1.8$
- $pK_2 = 9.2$
- $pK_R = 6.0$ (NH$^+$)
Histidine: **His vs. % Dry Protein**

![Graph showing the relationship between CP and HIS with a linear regression line.](natureprecedings://doi:10.1038/npre.2011.6231.3)

- Equation: $y = 0.029x - 0.158$
- $R^2 = 0.87$
- $R = 0.93$
- $n = 65$
Histidine: **His vs. % Dry Protein**

**Graph:**
- Title: DP new% (UIUC) vs HIS
- Equation: $y = 0.0289x - 0.2118$
- $R^2 = 0.499$
- Data points plotted with a linear trend line.
- Sample size: $n=49$
Arginine

- Polar
- Positively charged
- $pK_1 = 2.2$
- $pK_2 = 9.0$
- $pK_R = 12.5$

http://www.biology.arizona.edu/biochemistry/problem_sets/aa/Arginine.html
Amino Acid Contents of Soybeans Determined by GCMS and Correlation with Total Dry Protein

\[ y = 0.142x - 3.047 \]

\[ R^2 = 0.81 \]

\[ R = +0.90 \]
Amino Acid Content of Soybeans Determined by GCMS: Correlation with Total Dry Protein

![Graph showing correlation between DP new% (UIUC) and ARG](image)

**Equation:**

\[ y = 0.1561x - 3.9777 \]

**R^2:** 0.6868
Glutamine

- Neutral
- Polar (uncharged)
- $pK_1 = 2.2$
- $pK_2 = 9.1$
Glutamine plus Glutamic Acid, “GLX”, as Total Dry Weight % vs. Dry Soybean Protein %

\[ y = 0.227x - 2.085 \]

\[ R^2 = 0.78 \]

\[ R = +0.88 \]
Glutamine plus Glutamic Acid, “GLX”, as Total Dry Weight % vs. Dry Soybean Protein %

DP new% (ZX-50) vs GLX% (HPLC)

\[ y = 0.2361x - 2.9332 \]
\[ R^2 = 0.6902 \]
Glutamic acid

- Acidic
- Polar (charged)
- $pK_1 = 2.2$
- $pK_2 = 9.7$
- $pK_R = 4.3$
Valine

DP new% (UIUC) vs VAL

n=49

y = 0.0384x + 0.3006

R² = 0.6402
Valine

- Aliphatic
- Non-polar
- $pK_1 = 2.3$
- $pK_2 = 9.6$
PRO vs. Protein %

The graph shows a linear relationship between Protein by N and PRO with the following equation:

\[ y = 0.047x + 0.142 \]

The correlation coefficient is:

\[ R = 0.87 \]

The coefficient of determination is:

\[ R^2 = 0.75 \]
Imino Acid – Proline

- Cyclic
- Non-polar
- Similar to aliphatic group amino acids
  - $pK_1 = 2.0$
  - $pK_2 = 11.0$
Leucine

DP new% (UIUC) vs LEU

n=49

y = 0.0756x - 0.1558

R² = 0.6439
Leucine

- Aliphatic
- Non-polar
- Usually buried in folded proteins
- $pK_1 = 2.4$
- $pK_2 = 9.6$
Amino Acids Highly-Correlated with the Dry Protein Content:

- Histidine: $R = 0.93$
- Arginine: $R = 0.90$
- Glx: $R = 0.88$
- Valine: $R = 0.87$
- Leucine: $R = 0.85$

and the Imino Acid Proline: $R = 0.87$
Amino Acid Primary Data (GC-MS)
## Amino Acid Primary Data

(sample of 401 data points out of a total of 3,618)

<table>
<thead>
<tr>
<th>#</th>
<th>ASP</th>
<th>THR</th>
<th>SER</th>
<th>GLU</th>
<th>PRO</th>
<th>GLY</th>
<th>ALA</th>
<th>VAL</th>
<th>ILE</th>
<th>LEU</th>
<th>TYR</th>
<th>PHE</th>
<th>HIS</th>
<th>LYS</th>
<th>ARG</th>
<th>MET</th>
<th>CYS</th>
<th>M+S</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.8</td>
<td>1.6</td>
<td>2.0</td>
<td>7.3</td>
<td>2.1</td>
<td>1.7</td>
<td>1.7</td>
<td>2.0</td>
<td>1.8</td>
<td>3.2</td>
<td>1.4</td>
<td>2.1</td>
<td>1.1</td>
<td>2.6</td>
<td>2.8</td>
<td>0.7</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>2</td>
<td>4.8</td>
<td>1.6</td>
<td>2.0</td>
<td>7.3</td>
<td>2.0</td>
<td>1.7</td>
<td>1.7</td>
<td>2.0</td>
<td>1.8</td>
<td>3.2</td>
<td>1.2</td>
<td>2.0</td>
<td>1.0</td>
<td>2.5</td>
<td>2.7</td>
<td>0.7</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>3</td>
<td>5.0</td>
<td>1.6</td>
<td>2.1</td>
<td>7.8</td>
<td>2.1</td>
<td>1.8</td>
<td>1.8</td>
<td>2.0</td>
<td>1.9</td>
<td>3.2</td>
<td>1.3</td>
<td>2.1</td>
<td>1.1</td>
<td>2.7</td>
<td>3.1</td>
<td>0.7</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>4</td>
<td>4.8</td>
<td>1.6</td>
<td>2.1</td>
<td>7.5</td>
<td>2.0</td>
<td>1.7</td>
<td>1.7</td>
<td>2.0</td>
<td>1.9</td>
<td>3.2</td>
<td>1.3</td>
<td>2.1</td>
<td>1.1</td>
<td>2.6</td>
<td>3.0</td>
<td>0.7</td>
<td>0.7</td>
<td>1.4</td>
</tr>
<tr>
<td>5</td>
<td>4.9</td>
<td>1.6</td>
<td>2.1</td>
<td>7.4</td>
<td>2.1</td>
<td>1.8</td>
<td>1.8</td>
<td>2.0</td>
<td>1.9</td>
<td>3.2</td>
<td>1.2</td>
<td>2.1</td>
<td>1.0</td>
<td>2.6</td>
<td>2.8</td>
<td>0.7</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>6</td>
<td>4.9</td>
<td>1.6</td>
<td>2.1</td>
<td>7.6</td>
<td>2.1</td>
<td>1.8</td>
<td>1.8</td>
<td>2.0</td>
<td>1.9</td>
<td>3.2</td>
<td>1.2</td>
<td>2.2</td>
<td>1.1</td>
<td>2.6</td>
<td>3.0</td>
<td>0.7</td>
<td>0.7</td>
<td>1.4</td>
</tr>
<tr>
<td>7</td>
<td>5.3</td>
<td>1.8</td>
<td>2.4</td>
<td>8.4</td>
<td>2.4</td>
<td>1.9</td>
<td>1.9</td>
<td>2.1</td>
<td>2.0</td>
<td>3.4</td>
<td>1.5</td>
<td>2.2</td>
<td>1.2</td>
<td>2.8</td>
<td>3.5</td>
<td>0.6</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>5.0</td>
<td>1.6</td>
<td>2.1</td>
<td>7.8</td>
<td>2.2</td>
<td>1.8</td>
<td>1.8</td>
<td>2.1</td>
<td>1.9</td>
<td>3.3</td>
<td>1.4</td>
<td>2.1</td>
<td>1.1</td>
<td>2.7</td>
<td>3.1</td>
<td>0.7</td>
<td>0.8</td>
<td>1.5</td>
</tr>
<tr>
<td>9</td>
<td>5.1</td>
<td>1.7</td>
<td>2.3</td>
<td>7.9</td>
<td>2.2</td>
<td>1.9</td>
<td>1.9</td>
<td>2.1</td>
<td>2.0</td>
<td>3.3</td>
<td>1.3</td>
<td>2.2</td>
<td>1.1</td>
<td>2.7</td>
<td>3.1</td>
<td>0.7</td>
<td>0.9</td>
<td>1.7</td>
</tr>
<tr>
<td>10</td>
<td>5.7</td>
<td>1.4</td>
<td>2.7</td>
<td>9.1</td>
<td>2.3</td>
<td>2.1</td>
<td>2.0</td>
<td>2.2</td>
<td>2.2</td>
<td>3.6</td>
<td>1.6</td>
<td>2.4</td>
<td>1.2</td>
<td>3.0</td>
<td>3.7</td>
<td>0.8</td>
<td>1.0</td>
<td>1.9</td>
</tr>
<tr>
<td>11</td>
<td>6.2</td>
<td>1.6</td>
<td>2.9</td>
<td>9.7</td>
<td>2.5</td>
<td>2.2</td>
<td>2.6</td>
<td>2.3</td>
<td>2.2</td>
<td>3.9</td>
<td>1.6</td>
<td>1.8</td>
<td>1.3</td>
<td>3.1</td>
<td>4.1</td>
<td>0.9</td>
<td>1.0</td>
<td>1.9</td>
</tr>
<tr>
<td>12</td>
<td>6.1</td>
<td>1.7</td>
<td>2.8</td>
<td>9.5</td>
<td>2.5</td>
<td>2.2</td>
<td>2.1</td>
<td>2.4</td>
<td>2.3</td>
<td>3.8</td>
<td>1.7</td>
<td>2.5</td>
<td>1.4</td>
<td>3.2</td>
<td>4.2</td>
<td>0.9</td>
<td>1.1</td>
<td>1.9</td>
</tr>
<tr>
<td>13</td>
<td>5.9</td>
<td>1.5</td>
<td>2.8</td>
<td>9.3</td>
<td>2.4</td>
<td>2.1</td>
<td>2.0</td>
<td>2.3</td>
<td>2.2</td>
<td>3.8</td>
<td>1.6</td>
<td>2.5</td>
<td>1.2</td>
<td>3.1</td>
<td>3.8</td>
<td>0.9</td>
<td>1.1</td>
<td>1.9</td>
</tr>
<tr>
<td>14</td>
<td>5.2</td>
<td>1.5</td>
<td>2.4</td>
<td>8.3</td>
<td>2.2</td>
<td>1.9</td>
<td>1.9</td>
<td>2.1</td>
<td>2.0</td>
<td>3.4</td>
<td>1.4</td>
<td>2.1</td>
<td>1.2</td>
<td>2.8</td>
<td>4.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>6.0</td>
<td>1.5</td>
<td>2.8</td>
<td>9.2</td>
<td>2.3</td>
<td>2.1</td>
<td>2.1</td>
<td>2.3</td>
<td>2.2</td>
<td>3.8</td>
<td>1.5</td>
<td>2.4</td>
<td>1.3</td>
<td>3.0</td>
<td>4.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>5.7</td>
<td>1.5</td>
<td>2.7</td>
<td>8.8</td>
<td>2.4</td>
<td>2.0</td>
<td>2.0</td>
<td>2.3</td>
<td>2.0</td>
<td>3.6</td>
<td>1.6</td>
<td>2.3</td>
<td>1.2</td>
<td>2.9</td>
<td>4.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>6.0</td>
<td>1.5</td>
<td>2.8</td>
<td>8.7</td>
<td>2.3</td>
<td>2.0</td>
<td>2.0</td>
<td>2.2</td>
<td>2.1</td>
<td>3.6</td>
<td>1.6</td>
<td>2.4</td>
<td>1.2</td>
<td>2.9</td>
<td>3.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>5.9</td>
<td>1.5</td>
<td>2.7</td>
<td>9.3</td>
<td>2.4</td>
<td>2.1</td>
<td>2.1</td>
<td>2.3</td>
<td>2.2</td>
<td>3.7</td>
<td>1.6</td>
<td>2.4</td>
<td>1.3</td>
<td>3.1</td>
<td>3.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>5.7</td>
<td>1.8</td>
<td>2.5</td>
<td>8.9</td>
<td>2.4</td>
<td>2.0</td>
<td>2.0</td>
<td>2.2</td>
<td>2.1</td>
<td>3.6</td>
<td>1.6</td>
<td>2.4</td>
<td>1.2</td>
<td>3.0</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>5.8</td>
<td>1.4</td>
<td>2.8</td>
<td>9.1</td>
<td>2.6</td>
<td>2.0</td>
<td>2.0</td>
<td>2.2</td>
<td>2.2</td>
<td>3.7</td>
<td>1.6</td>
<td>2.4</td>
<td>1.4</td>
<td>3.0</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>5.9</td>
<td>1.8</td>
<td>2.6</td>
<td>9.3</td>
<td>2.5</td>
<td>2.1</td>
<td>2.0</td>
<td>2.3</td>
<td>2.2</td>
<td>3.8</td>
<td>1.6</td>
<td>2.2</td>
<td>1.2</td>
<td>3.0</td>
<td>3.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Amino Acid ROVs

<table>
<thead>
<tr>
<th>AA</th>
<th>ASX</th>
<th>THR</th>
<th>SER</th>
<th>GLX</th>
<th>PRO</th>
<th>GLY</th>
<th>ALA</th>
<th>VAL</th>
<th>ILE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>6.20</td>
<td>1.82</td>
<td>2.90</td>
<td>9.67</td>
<td>2.60</td>
<td>2.18</td>
<td>2.57</td>
<td>2.37</td>
<td>2.27</td>
</tr>
<tr>
<td>Min</td>
<td>4.73</td>
<td>1.35</td>
<td>2.04</td>
<td>7.23</td>
<td>1.95</td>
<td>1.72</td>
<td>1.72</td>
<td>1.90</td>
<td>1.78</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.24</td>
<td>0.26</td>
<td>0.30</td>
<td>0.25</td>
<td>0.25</td>
<td>0.21</td>
<td>0.33</td>
<td>0.20</td>
<td>0.22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AA</th>
<th>LEU</th>
<th>TYR</th>
<th>PHE</th>
<th>HIS</th>
<th>LYS</th>
<th>ARG</th>
<th>MET</th>
<th>CYS</th>
<th>met+cys</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max</td>
<td>3.89</td>
<td>1.66</td>
<td>2.53</td>
<td>1.36</td>
<td>3.20</td>
<td>4.67</td>
<td>0.88</td>
<td>1.08</td>
<td>1.95</td>
</tr>
<tr>
<td>Min</td>
<td>3.06</td>
<td>1.11</td>
<td>1.54</td>
<td>1.00</td>
<td>2.52</td>
<td>2.73</td>
<td>0.65</td>
<td>0.72</td>
<td>1.38</td>
</tr>
<tr>
<td>Ratio</td>
<td>0.21</td>
<td>0.33</td>
<td>0.39</td>
<td>0.26</td>
<td>0.21</td>
<td>0.41</td>
<td>0.27</td>
<td>0.34</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Legend:
- **Red** (poor; low)
- **Green** (acceptable)
- **Blue** (good; high)
Comparison Between Amino Acid Contents of Soybean Seeds Determined by Primary Data: $^{13}$C Liquid State HR NMR and IEC

<table>
<thead>
<tr>
<th></th>
<th>Ala</th>
<th>Val</th>
<th>Leu</th>
<th>Ile</th>
<th>Gly</th>
<th>Asn</th>
<th>Asp</th>
<th>Asx</th>
<th>Gln</th>
<th>Glu</th>
<th>Glx</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NMR</strong></td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>4.5</td>
<td>4</td>
<td>7</td>
<td>5</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td><strong>IEC</strong></td>
<td>4</td>
<td>5.2</td>
<td>7.3</td>
<td>4.7</td>
<td>2.8</td>
<td>ND</td>
<td>ND</td>
<td>12.1</td>
<td>ND</td>
<td>ND</td>
<td>21.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Ser</th>
<th>Thr</th>
<th>Arg</th>
<th>Lys</th>
<th>Trp</th>
<th>Tyr</th>
<th>His</th>
<th>Phe</th>
<th>Cys</th>
<th>Met</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NMR</strong></td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td><strong>IEC</strong></td>
<td>4.6</td>
<td>3.6</td>
<td>9.5</td>
<td>7.8</td>
<td>ND</td>
<td>3.5</td>
<td>2.6</td>
<td>5.5</td>
<td>ND</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Source: (Baianu, You, Costescu, and Prisecaru, 2005, page.....)
## SOY PROTEIN and MOISTURE DATA, with Soybean Accession Identifiers

<table>
<thead>
<tr>
<th>ID</th>
<th>coat color</th>
<th>cultivar</th>
<th>seed source</th>
<th>Prot</th>
<th>Moist</th>
<th>Oil</th>
<th>DP</th>
<th>DP calc</th>
<th>DO calc</th>
</tr>
</thead>
<tbody>
<tr>
<td>PI536636</td>
<td>Y</td>
<td>Ripley</td>
<td>93U-5051</td>
<td>35.57</td>
<td>11.10</td>
<td>20.50</td>
<td>40.01</td>
<td>40.01</td>
<td>23.06</td>
</tr>
<tr>
<td>PI548518</td>
<td>Y</td>
<td>Cutler 71</td>
<td>92U-901</td>
<td>36.91</td>
<td>10.00</td>
<td>20.76</td>
<td>41.00</td>
<td>41.01</td>
<td>23.07</td>
</tr>
<tr>
<td>PI548477</td>
<td>Lgn</td>
<td>Ogden</td>
<td>99S-4040</td>
<td>41.75</td>
<td>10.41</td>
<td>17.06</td>
<td>46.60</td>
<td>46.60</td>
<td>19.04</td>
</tr>
<tr>
<td>PI548379</td>
<td>Y</td>
<td>Mandarin</td>
<td>00U-109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PI548603</td>
<td>Y</td>
<td>Perry</td>
<td>98U-1459</td>
<td>41.44</td>
<td>10.30</td>
<td>10.02</td>
<td>46.20</td>
<td>46.20</td>
<td>11.17</td>
</tr>
<tr>
<td>PI533655</td>
<td>Y</td>
<td>Burlison</td>
<td>95U-2238</td>
<td>41.10</td>
<td>10.16</td>
<td>18.60</td>
<td>45.75</td>
<td>45.75</td>
<td>20.70</td>
</tr>
<tr>
<td>PI548311</td>
<td>Y</td>
<td>Capitol</td>
<td>97U-2231</td>
<td>36.71</td>
<td>9.12</td>
<td>20.60</td>
<td>40.40</td>
<td>40.39</td>
<td>22.67</td>
</tr>
<tr>
<td>PI548659</td>
<td>Y</td>
<td>Braxton</td>
<td>94S-6</td>
<td>38.03</td>
<td>10.84</td>
<td>20.24</td>
<td>42.65</td>
<td>42.65</td>
<td>22.70</td>
</tr>
<tr>
<td>PI567551</td>
<td>ltGn,1/10</td>
<td>Huang li</td>
<td>94U-785</td>
<td>39.52</td>
<td>10.44</td>
<td>19.34</td>
<td>44.13</td>
<td>44.13</td>
<td>21.59</td>
</tr>
<tr>
<td>PI458057</td>
<td>Dk Gn</td>
<td></td>
<td>96U-2173</td>
<td>40.17</td>
<td>9.04</td>
<td>18.80</td>
<td></td>
<td>44.16</td>
<td>20.67</td>
</tr>
<tr>
<td>PI567704</td>
<td>Y, 1/2 brn</td>
<td>Fu yang (23)</td>
<td>93U-2482</td>
<td>40.67</td>
<td>10.25</td>
<td>18.30</td>
<td>45.31</td>
<td>45.31</td>
<td>20.39</td>
</tr>
</tbody>
</table>

**Val’s data**

Example:

10 Data Points

**TOTAL:**

3,816 data points

Improving NIR Calibrations

1. Checking the Correlation between amino acid and total protein content (dry)\%

2. Select the amino acids for the NIR calibration following two criteria:
   
   I. Correlation coefficient between a.a. and dry protein less than $\sim 90\%$ ($R^2 < 0.81$)
   
   II. The widest available range of values in the calibration standard set ($ROV > \sim 25\%$)
   
   III. Primary measurement error for such amino acid composition measurements less than 5\%.
3. Select a set of suitable calibration standards following the triangle matrix of amino acid triplet concentrations:

a. Ser-Tyr-Ala (0.30; 0.33; 0.33) → best selection
b. Ileu-Tyr-Phe (0.22; 0.34; 0.39)
c. Phe-Lys-Ala ROV’s: (0.39; 0.21; 0.33)
d. Ser-Phe-Lys (0.30; 0.39; 0.21)
e. Asx-Ile- Tyr (0.24; 0.25; 0.33)
f. Asx-Leu- Tyr (0.24; 0.25; 0.33)
g. (Met+Cys)-Tyr-Phe ? Met +Cys few samples !

Old: Arg-Lys-Glx, Old: Asx-Ala(or Val)-Pro

Improving NIR Calibrations for Soybean Amino Acids
Phenylalanine

![Graph showing the relationship between DP new\% (UIUC) and PHE with the equation y = 0.0409x + 0.2812 and R² = 0.4743.](image)
Phenylalanine

- Aromatic
- Non-polar
- Very hydrophobic
- $pK_1 = 1.8$
- $pK_2 = 9.1$
Cysteine

DP new% (UIUC) vs CYS
n=13

y = 0.0287x - 0.4768
R² = 0.4982
Cysteine

- Sulfur
- Polar (uncharged)
- $pK_1 = 2.0$
- $pK_2 = 10.3$
Cystine – Disulfide bonds

- Sulfur
- Polar (uncharged)
- $pK_1 = 2.0$
- $pK_2 = 10.3$
Methionine

DP new% (UIUC) vs MET
n=13

\[ y = 0.0204x - 0.2286 \]
\[ R^2 = 0.5483 \]
Methionine

- Sulfur
- Non-polar (hydrophobic)
- \( pK_1 = 2.3 \)
- \( pK_2 = 9.2 \)
Tyrosine

- Aromatic
- Non-polar
- Not as hydrophobic as Phe
- $pK_1 = 2.2$
- $pK_2 = 9.1$
Tyrosine vs. %DP

DP new% (UIUC) vs TYR
n=49

y = 0.0382x - 0.3778
R² = 0.5361
Amino Acid Primary Data (GC-MS), cont’d

**DP new% (UIUC) vs. SER (Peoria)**

- Sample size: n=49
- Equation: \( y = 0.0796x - 1.3815 \)
- Coefficient of determination: \( R^2 = 0.5777 \)

**DP new% (ZX-50) vs GLX% (HPLC)**

- Sample size: n=49
- Equation: \( y = 0.2361x - 2.9332 \)
- Coefficient of determination: \( R^2 = 0.6902 \)
Amino Acid Primary Data (GC-MS), cont’d

DP new% (UIUC) vs PRO
n=49

\[ y = 0.0458x + 0.0563 \]
\[ R^2 = 0.4804 \]

DP new% (UIUC) vs GLY
n=49

\[ y = 0.0425x - 0.1133 \]
\[ R^2 = 0.6275 \]
Amino Acid Primary Data (GC-MS), cont’d

DP new% (UIUC) vs ALA,
(n=48)

\[ y = 0.0376x + 0.1043 \]
\[ R^2 = 0.619 \]

DP new% (UIUC) vs VAL
(n=49)

\[ y = 0.0384x + 0.3006 \]
\[ R^2 = 0.6402 \]
Amino Acid Primary Data (GC-MS), cont’d

DP new% (UIUC) vs ILE
n=49

y = 0.0447x - 0.1292
R² = 0.6309

DP new% (UIUC) vs LEU
n=49

y = 0.0756x - 0.1558
R² = 0.6439
Amino Acid Primary Data (GC-MS), cont’d

**DP new\% (UIUC) vs TYR**

n=49

\[ y = 0.0382x - 0.3778 \]
\[ R^2 = 0.5361 \]

**DP new\% (UIUC) vs PHE**

n=47 (2 outliers removed)

\[ y = 0.0409x + 0.2812 \]
\[ R^2 = 0.4743 \]
Amino Acid Primary Data (GC-MS), cont’d

**DP new% (UIUC) vs HIS**

- $y = 0.0289x - 0.2118$
- $R^2 = 0.499$

**DP new% (UIUC) vs LYS**

- $y = 0.0608x - 0.1074$
- $R^2 = 0.6518$
Amino Acid Primary Data (GC-MS), cont’d

DP new% (UIUC) vs ARG

\[ y = 0.1561x - 3.9777 \]
\[ R^2 = 0.6868 \]

n=49

DP new% (UIUC) vs MET

\[ y = 0.0204x - 0.2286 \]
\[ R^2 = 0.5483 \]

n=13
Amino Acid Primary Data (GC-MS), cont’d

DP new\% (UIUC) vs CYS
n=13

\[
y = 0.0287x - 0.4768
\]
\[R^2 = 0.4982\]

DP new\% (UIUC) vs MET+CYS
n=13

\[
y = 0.0491x - 0.7054
\]
\[R^2 = 0.5379\]
Amino Acid Summary

• Total protein measurement is by wet chemistry analysis, for which a ~96% correlation with our lab’s NIR analysis is confirmed.

• Some of the amino acid correlations depend on the type of amino acid residue.

• There is a potential for selecting the composition of amino acids seen in soybeans.
Amino Acids, p.2

• A few amino acid residues vary much more across some soybean varieties than others

• The combined NIR, NMR and GC-MS data for amino acids in soybean seeds shows the possibility of generating reliable calibrations for selected triplet groups of amino acids using FT-NIR Spectroscopy.

• NIRS data for AA can be validated and may thus become a very useful tool for cross-breeding and genetic selection purposes.
3. Proposed Research Methods

I. Techniques, Data analysis and Expected Results
   - Primary and secondary techniques; Chemical and Hyperspectral NIR Imaging; Fluorescence Correlation Spectroscopy and Microspectroscopy)
   - Data Analysis: Principles of NIRS, Data Corrections, Regression Algorithms and Multivariate analysis.

II. Limitations and Advantages of proposed methods

II. Tentative schedule to conduct major steps
I. Techniques, Data analysis and Expected Results
I. Techniques, Data analysis and Expected Results

A. TECHNIQUES:

• NIR Spectroscopy,
• GC-MS
• HPLC and IEC
• Chemical Imaging,
• Fluorescence Correlation Spectroscopy

B. Data Analysis

C. Expected Results
NIR Analysis of Amino Acids & Proline in Soybeans

- Wet chemistry analysis:
  - Ion Exchange Chromatography (IEC) / UV / vis Abs. / Fluorescence
  - Derivatization HPLC
  - GC-MS (USDA – Peoria)

- NMR as Primary Method
Soybean Proteins Content by C-13 NMR and Sigma Methods

Principles of NIR Spectroscopy

- IR absorption spectra occur because the atom-to-atom bonds within molecules can vibrate and rotate thus generating series of different energy levels among which rapid transitions can occur.

- According to Quantum Mechanics, the vibro-rotational energy levels of a molecule can be approximated by the following equations:

\[
E_{\text{NIR}} = E_{\text{rot}} + E_{\text{vib}} + E_{\text{anh}} = j(j + 1)Bhc + [1 - x(n + 1/2)]h\nu
\]

- where: 
  - \(j\): rotation quantum number = 0,1,2, …
  - \(n\): vibration number = 0,1,2,…;
  - \(E\): Energy eigenvalues , and
  - \(x\): anharmonicity constant (\(\sim 0.01\)).
Current Near-Infrared Instruments: Techniques

Current NIR instruments utilize EM radiation with wavelengths from ~750 nm to 2500 nm. Their operation is based on the fact that molecular bonds stretch and/or bend, thus causing absorption bands at certain characteristic IR and NIR wavelengths that are proportional to the amount of the absorbing components present in the sample, e.g., amide 1 and 2 bands.
Measuring the Absorbance

• The absorbance of a sample is often difficult to measure directly.

• In practice, the absorption is often calculated indirectly by measuring reflectance \( A = \log \frac{1}{R} \), or transmittance \( A = \log \frac{1}{T} \), that can be readily measured even for thick samples.

• This assumes the samples do not possess a composite structure, such as thick, multiple layers with different compositions.
The calculated absorbance is usually referred to as the “apparent absorbance,” and it can be significantly affected by Specular Reflection and Light Scattering, even for thin samples.

Therefore, in order to obtain reliable NIR quantitation, Spectral Pre-Processing and Corrections are always required.
Data Correction Problems

- We found that the NIR methods currently employed in industry for:
  - spectral pre-processing
  - correction of light scattering
  - specular reflection effects

are in need of substantial improvements in order to produce high accuracy, robust and stable calibrations for rapid composition analyses of seeds.
Spectral variations between soybean samples can be caused by:

- chemical composition differences
  - (i.e., what you want to measure)
- Spurious effects*
  - specular reflection
  - scattering effects
  - internal reflection

*These do not monitor chemical composition -- and are therefore measurement artifacts that are undesirable and distort the data.
Fourier Transform NIR

- Recent NIR Spectroscopy techniques utilizing Fourier Transform (FT) fulfill all these conditions, but require pre-calibration by AOCS-approved wet chemistry techniques, using well-defined and stable sample standard sets of 50 to 100 different samples.
FT-IR Spectrometer *Spectrum One* and FT-NIR Spectrometer *SpectrumOne-NTS*

- Introduced in 2001 by Perkin Elmer Co. (Shelton, CT, USA) for High Sensitivity, high-resolution and long-term stability
- *SpectrumOne* and *SpectrumOne-NTS* have a similar look but are configured for different spectral ranges (e.g., IR and NIR, respectively).

(Perkin Elmer Co., USA)
Comparison of Soybean Spectra Collected with either Perten DA7000 or the PE Spectrum One NTS (with Extd. InGaAs/NIRA) NIR Instruments

![Graph comparing soybean spectra collected with either Perten DA7000 or the PE Spectrum One NTS (with Extd. InGaAs/NIRA) NIR Instruments.](image_url)
Lambert-Beer’s Law

Absorption is a universal spectroscopic phenomenon that has immediate chemometric applications, because it is directly related to the constituent concentration as described by:

Lambert-Beer’s Law, which states that …

\[ A = \varepsilon \times C \times L \]

where:

\[ A = \text{True Absorbance} \]
\[ \varepsilon = \text{Extinction coefficient of analyte} \]
\[ C = \text{Concentration} \]
\[ L = \text{Pathlength of light} \]
Lambert-Beer’s Law, Con’t

- The absorbance of a sample is difficult to measure directly.
- In practice, the absorption is often calculated indirectly from the measurement of the reflectance \( A = \log \frac{1}{R} \), or transmittance \( A = \log \frac{1}{T} \), that can be readily measured even for thick samples, provided these do not possess a composite structure, such as thick, multiple layers with different compositions.
- The calculated absorbance is usually referred to as the ‘apparent absorbance,’ and it can be significantly affected by Specular reflection and light scattering even for thin (e.g., 5 mm) samples.
- Therefore, in order to obtain reliable NIR quantitation, spectral pre-processing and corrections are always required.
Light Scattering Corrections for Soybean NIR Spectra

- Spectral variations between soybean samples can be caused by chemical composition differences, specular reflection, as well as light scattering effects (that do not monitor chemical composition—and are therefore a measurement artifact)

- The effects of light scattering and/or specular reflection on the NIR spectra of soybean need be investigated, and eliminated if at all possible
One finds that the methods that are currently employed by the NIR industry for spectra pre-processing and corrections of light scattering and/or specular reflection effects are in need of substantial improvements in order to produce calibrations that are:

- Highly accurate
- Robust
- Lead to stable calibrations for rapid composition analyses of seeds
SpectrumOne NTS Spectra of Bulk Soybean Samples, before (A) and after (B) Multiple Scattering Correction (MSC)

Source: UIUC Soybean NIR Database, 2007
Detrimental Effects of Light Scattering on the Accuracy of NIR Analysis for Whole Soybeans

(measured with the FT-NIR, SpectrumOne NTS Spectrometer)

<table>
<thead>
<tr>
<th>Component</th>
<th>R</th>
<th>SECV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No MSC</td>
<td>MSC</td>
</tr>
<tr>
<td>Protein</td>
<td>99.5</td>
<td>99.9</td>
</tr>
<tr>
<td>Oil</td>
<td>99.3</td>
<td>99.9</td>
</tr>
<tr>
<td>Moisture</td>
<td>99.8</td>
<td>99.9</td>
</tr>
</tbody>
</table>

**R**: Correlation coefficient  
**SECV**: Standard Error of Cross Validation  
(Tested with 65 bulk whole soybean standards)

Source: T. You, 2006
FT-NIR Spectra of Five Major Soybean Components

Collected on the Perkin-Elmer SpectrumOne NTS FT-NIR Spectrometer

[Source: T. You, 2006]

[Figure: FT-NIR Spectra of Soybean Components with Absorbance on the y-axis and Wavelength on the x-axis. The spectra are color-coded as follows: Protein (pink), Oil (yellow), Water (light blue), Sugars (blue), Fiber (red), and Soybean (teal).]
SpectrumOne NTS FT-NIR Spectra of Soy Protein Isolates (SPI) in H₂O, before (A) and after (B) Multiple Scattering Correction (MSC) (Source: I.C. Baianu et al., 2009)
Detrimental Effects of Light Scattering on the Accuracy of NIR Analysis for Hydrated Soy Proteins (SPI; measured with a Perkin-Elmer FT-NIR SpectrumOne NTS Spectrometer)

<table>
<thead>
<tr>
<th>Component</th>
<th>Number of Factors</th>
<th>R</th>
<th>SECV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>R: Correlation coefficient</td>
<td>SECV: Standard Error of Cross Validation</td>
</tr>
<tr>
<td>SPI</td>
<td>NoMSC</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>MSC</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td>H2O</td>
<td>NoMSC</td>
<td>0.998</td>
<td>0.999</td>
</tr>
<tr>
<td></td>
<td>MSC</td>
<td>0.998</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Source: T. You, 2006
The Standard Error of Cross Validation (SECV) is defined as:

\[ SECV = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \]

where \( n \) is the total number of samples, \( y_i \) denotes the standard value of the component concentration, and \( \hat{y}_i \) denotes the predicted component concentration.
The **standard error of prediction** is the standard deviation of the sample mean estimate of a population mean.

 Usually estimated by the sample estimate of the population standard deviation (sample standard deviation) divided by the square root of the sample size:

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

where

- $s$ is the sample standard deviation (i.e. the sample based estimate of the standard deviation of the population), and
- $n$ is the size (number of observations) of the sample.
Illustration of the Interactive, Spline Baseline Correction (BC)* of the FT-NIR Spectrum of a Whole Soybean Seed, with Coat

• Note: Only the PE Spectrum program supports this interactive, spline-function, Fx, baseline correction (BC), shown above in purple color.
Illustration of the Interactive, Spline Baseline Correction (BC) of the FT-NIR Spectrum of A Whole Soybean Seed, with Coat

*Note: Only the PE Spectrum program supports this interactive, spline-function, Fx, baseline correction (BC), shown above in purple color.
True FT-NIR Absorption Spectrum of Soybeans with Coat
Calibration

- Generate or Select a suitable set of Standard Samples of Known composition
- Obtain Raw FT-NIR data
- Correct Data for Multiple Scattering
- Use Lambert-Beer Law in conjunction with iterated data regression by PLS-1 or 2; check up on PLS-1 precision and correct computation;
- Examine the Calibration’s Linear Correlations and Validation/ Composition Predictions
Computer Simulation Study of the PLS-1 Calibration Algorithm

N=21, Calibration Standards

Source: T. You, 2006
Computer Simulation of PLS-1 Algorithm with 3 Components (%)
Loading Vectors and SECV’s of Components C1 to C3 for the Simulated PLS-1 Calibration Algorithm, without any Noise... except from negligible PC computation errors!

<table>
<thead>
<tr>
<th>Components</th>
<th>Number of Factors</th>
<th>R</th>
<th>SEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>2</td>
<td>100%</td>
<td>$9 \times 10^{-6}$</td>
</tr>
<tr>
<td>C2</td>
<td>2</td>
<td>100%</td>
<td>$1 \times 10^{-6}$</td>
</tr>
<tr>
<td>C3</td>
<td>2</td>
<td>100%</td>
<td>$2 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

R: Correlation coefficient
SEP: Standard Error of Prediction
*‘Ideal’ conditions, that is, without Noise!
**Predicted vs. Reference** Concentrations of Component C1 for our PLS-1 Simulation Study

![Graph showing predicted vs. reference values](image)

- Corr. Coeff.: 1.00000
- RMSEC: (0.339e-5)

R = 1.0

SECV = 0.000003

Source: T. You, 2006
II. Limitations and Advantages of Proposed Procedures
Major Limitations

- Amino Acids affected by acid hydrolysis of the protein:
  - Absence of Trp data. Trp is just an example of what hydrolysis does to amino acids. Corrections to hydrolysis is one of the limitations caused by hydrolysis.
  - No data extrapolation to Zero
  - Correlation of amino acids with Dry Protein
  - Extremely limited range of ROV’s makes it impossible to do the calculations for those amino acids
Matrix Effects

- A major obstacle that exists in the comparison of NIRS calibration data for different types of samples with the same chemical/biochemical composition but in different form or phase is the so-called “matrix effect.”

- This “matrix effect” depends on the state in which the molecules are in: Solid, Liquid phases and also on different: Texture/Morphology, Particle Size Distribution, Molecular Alignments.

- Other related causes: internal gaps, different interfaces, internal sample changes in refractive index, and so on.
NMR Advantages

- Trp data included
- \textit{in situ} data acquisition
Derivatized HPLC Advantages

- Derivatization prior to acid hydrolysis followed by HPLC eliminates the limitations caused by the protein treatment with acid.
III. Tentative Schedule for major steps
NIRS Calibrations for Selected AA Groups

• It is very important that there are enough knowns in the protein spectra to solve for all of the amino acid residues selected

• This research project will be using four groups of amino acid triplets such as:
  • Arg—Ala—Val
  • Cys (and/or Met)—Lys—Ser
Planned Work

- Individual AA calibrations in amino acid mixtures (may also serve pharmaceutical purposes)
- FT-NIR spectra of AA’s in solution, incl. Trp?
- FT-IR spectra of solid powder from 0 to 100% powder, using >100 samples for each trial
- Compare my Proline data to the NMR data
- Supplements: protein powders calibration
- Certain pharmaceutical drugs
- Revise the primary analysis chemistry tables with more accurate data
Planned work also includes:

- Crystalline amino acids powders vs. amino acids in gels
- Egg white vs. egg albumin data
- The making of concentrated solutions, followed by partial drying resulting in an amorphous gel rather than crystalline powders.
Timeline and Steps

1. Select the best primary method for analyzing protein and amino acids
   • GCMS, NMR
   • Both are useable and comparable
   • NMR is superior
     • No acid hydrolysis
     • Readily available

1. Obtain dry protein and moisture contents
2. Obtain PLS-1 for protein content
Model Systems Rationale

Why do we need a model system? Why can’t we skip this step and just run our soybean samples?

- We need knowledge of the bands assigned from the model systems (e.g., Tyrosine ring, Amide II band)
- Serves to show the scale of the level of errors we can expect to get
- Others have tried and failed
4. Check and validate NIR calibration for:
   • Amino acids using model systems of known and simple composition (e.g., mixtures of 3 amino acids), and
   • Proteins of known composition
     • i.e., “Model Systems”
     • By November 12, 2009
Timeline and Steps, cont’d

• In principle, all amino acid combinations should be performed, but we will do 4 triplets
• TIME FACTOR: Running the samples
  • 4 x 100 samples x 2 (duplicates)
  • = 800 samples
  • 5 minutes/sample…. 4,000 minutes
  • Or one month
• TIME FACTOR: Making the samples
  • 1 week per sample, parallel proc.
  • Total estimated time: several months
Experimental Setup

• For the selection of the individual amino acids to be analyzed, I will be using the following matrix, using pure amino acid powders:

• #1: Pro-Glu-Ala
• #2: Arg-pSer-Ile
• #3: Lys-Cys-Phe
Computer Simulation of PLS-1 Algorithm with 3 Components (%)

Adapted from T. You, 2006
Planned Work, cont’d

Similar model calibrations will be carried out for mixtures of other amino acids, triple selected, such as: Cys + Met – Arg–Ala or Met – Asn – Lys
Planned Work, cont’d

• Future work may include a “triangle” triplet of

  • Pro– Glx--Val
  • Met+Cys –Arg--Ala
  • Soy protein isolates (SPI), egg albumin and lysozyme
Moisture Calibration for Bulk Soybean Analysis on the Spectrum One NTS FT-NIR Instrument

65 calibration standards, 20 grams for each standard, 9mm NIR beam size

Source: Soybean NIR Database, UIUC
Bulk Soybean Calibration with 65 Standards for Protein, Oil, and Moisture Analysis
Developed with Data from the SpectrumOne NTS FT-NIR Spectrometer

<table>
<thead>
<tr>
<th>Component</th>
<th>Number of Factors</th>
<th>R</th>
<th>SECV</th>
<th>SEP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>13</td>
<td>99.9%</td>
<td>0.26</td>
<td>0.33</td>
</tr>
<tr>
<td>Oil</td>
<td>15</td>
<td>99.9%</td>
<td>0.13</td>
<td>0.23</td>
</tr>
<tr>
<td>Moisture</td>
<td>15</td>
<td>99.9%</td>
<td>0.17</td>
<td>0.30</td>
</tr>
</tbody>
</table>

**R**: Correlation Coefficient  
**SECV**: Standard Error of Cross Validation  
**SEP**: Standard Error of Prediction  

Source: Soybean NIR Database, UIUC
Detrimental Effects of Light Scattering on the Accuracy of NIR Analysis for Whole Soybeans
(measured with the FT-NIR, SpectrumOne NTS Spectrometer)

<table>
<thead>
<tr>
<th>Component</th>
<th>R No MSC</th>
<th>R MSC</th>
<th>SECV No MSC</th>
<th>SECV MSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>99.5</td>
<td>99.9</td>
<td>0.63</td>
<td>0.26</td>
</tr>
<tr>
<td>Oil</td>
<td>99.3</td>
<td>99.9</td>
<td>0.29</td>
<td>0.13</td>
</tr>
<tr>
<td>Moisture</td>
<td>99.8</td>
<td>99.9</td>
<td>0.26</td>
<td>0.17</td>
</tr>
</tbody>
</table>

R: Correlation coefficient
SECV: Standard Error of Cross Validation
(Tested with 65 bulk whole soybean standards)

Source: Soybean NIR Database, UIUC
A new and Improved Set of 124 Bulk Soybean Standard Samples:
Protein-Oil Inverse Correlation for the Year 2003 Calibration Standard

![Graph showing the relationship between Protein and Oil](image)

\[ y = -0.584x + 46.376 \]

\[ R = -0.99 \]

* 124 Standard samples were selected with a wide range of Protein and Oil concentrations that were uniformly distributed in 0.5% concentration steps for Protein, and in 0.2% steps for Oil. Source: Soybean NIR Database, UIUC
Chemical and Hyperspectral NIR Imaging
Chemical and Hyperspectral Imaging of Amino Acid Residues and Proteins in Soybeans
Golden Cove (Digital Color Infrared)

Gold Beach, Oregon

http://www.pbase.com/image/38188240
Images Using the NIR Autolmage

FT-NIR Microspectrometer:

- Introduced in 2003 by PerkinElmer Co. (Shelton, CT, USA) for high-resolution studies.

- Employed for our NIR Microspectroscopy and Chemical Imaging investigations of Soybean seeds.

Microscope Coupled to the FT-NIR Spectrometer
Image of the Perkin Elmer AutoImage Microspectrometer “Innards”
FT-IR Chemical Image (Left) and Visible Light Micrograph (Right) of a Black Coat Soybean with Part of the Coat Removed
FT-NIR Chemical Image of Oil Distribution in a Mature Soybean Embryo Section
Acknowledgements

• Prof. Ion Baianu – my PhD Adviser
• Fmr. Research colleagues: Dr. Tiefeng You, Dr. Jun Guo
• Prof. Randall Nelson (Crop Sciences)
• Renée – my wife
• Ali and Hannah – my daughters
• Mom and Dad

• Renessen Biotech Corp.
• AOCS – Urbana, Collaboration
• USDA Lab – fatty/amino acid analysis data
• Northern USDA Res.Ctr. for wet chemistry: GC-MS and HPLC aa data
• C-FAR
• ISPOB (Illinois Soybean Operating Board)
Fluorescence Correlation Spectroscopy and Microspectroscopy
4. References
REFERENCES


• Tiwari, P. N. and Gambhir, P. N.1995. Seed Oil Determination without Weighing and Drying the Seeds by Combined Free Induction Decay and Spin-Echo Nuclear Magnetic Resonance Signals, JAOCS, 72(9).
• Whistler, R. L. and BeMiller, J. N. 1997, Carbohydrates Chemistry for Food Scientists, Eagen Press, St. Paul, MN.


5. Appendices

A. Supporting Data:
   1) List of data tables relevant to proposed work (hard copy)
   2) Numerical Data Tables relevant to proposed work (CD)

A. List of My Publications (hard copy)

B. Files of my published work supporting this proposal (CD)
The End
Protein Calibration for Bulk Soybean Analysis on the DA-7000 Diode Array NIR Instrument

\[ R = 0.999 \]
\[ SECV = 0.13 \]

Source: Soybean NIR Database, UIUC
Moisture Calibration for Bulk Soybean Analysis on the DA-7000 Diode Array NIR Instrument

\[ R = 0.997 \]
\[ \text{SECV} = 0.04 \]

Source: Soybean NIR Database, UIUC
Calibration Results for Protein and Oil Analysis with the Perten DA-7000, Dual Diode-Array DA-NIR Instrument

<table>
<thead>
<tr>
<th>Components</th>
<th>Protein</th>
<th>Oil</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bulk Sample</td>
<td>Single Seeds</td>
</tr>
<tr>
<td>SECV</td>
<td>0.1</td>
<td>1.1</td>
</tr>
<tr>
<td>R</td>
<td>99.9%</td>
<td>98.5%</td>
</tr>
</tbody>
</table>

Source: Soybean NIR Database, UIUC