Dual Diode Array and Fourier Transform Near Infrared Reflectance Spectrometer Calibrations for Composition Analysis of Single Soybean Seeds for Genetic Selection, Cross-Breeding Experiments

[*#]Ion C. Baianu, [+]Tiefeng You,
and
[**]Randall L. Nelson

[+ *]Department of Food Science and Human Nutrition,
Department of Nuclear, Plasma, and Radiological Engineering,
and
[**]National Soybean Research Center,
College of ACES, University of Illinois, Urbana, IL 61801, U.S.A.
Emails: i-baianu@uiuc.edu, tyou@uiuc.edu, , rlnelson@uiuc.edu

Abstract

- We are presenting the results of extensive evaluations aimed at developing NIR calibrations for single soybean seed composition analysis on both Dual Diode Array and Fourier Transform NIR reflectance instruments. Single soybean seed, bulk, and powder calibrations were developed on four different NIRS spectrometer models that are commercially available.

- Accurate and reliable measurements were performed on single soybean seeds for protein, oil, and moisture. Our results show that state-of-the-art DA and FT- NIR instruments can be employed to obtain high quality calibrations for single soybean seeds. Seed-to-seed variations in protein and oil contents can thus be monitored on single soybeans of selected soybean accessions.

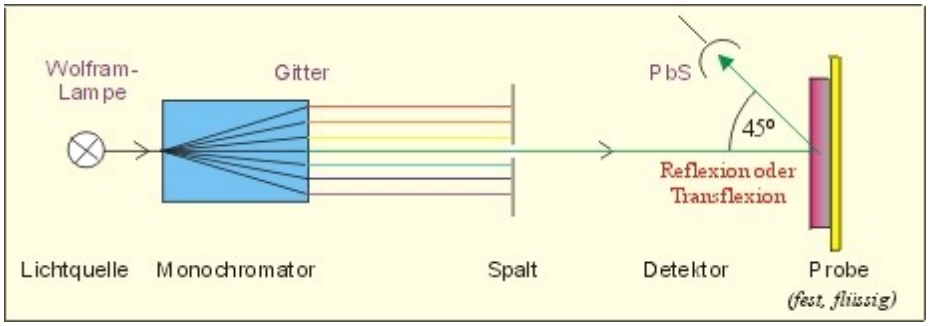
Introduction

- Rapid, accurate, reliable, and cost-effective composition analyses of single soybean seeds are important for genetic breeding and selection programs aimed at improving the efficiency of soybean breeding and soybean quality.

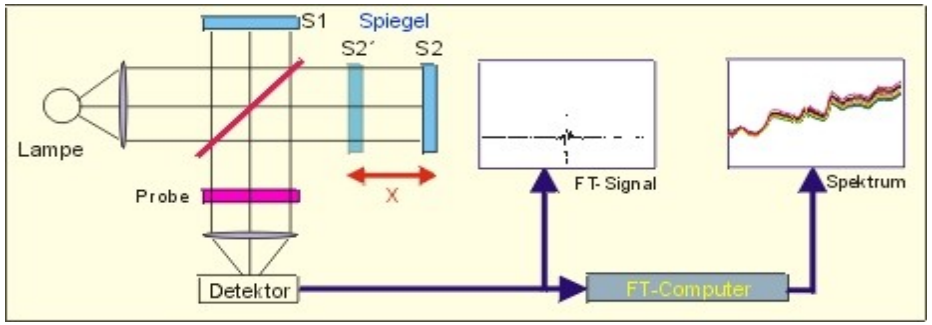
- Previously NIR spectroscopy has been applied for rapid composition analysis of bulk grain and oilseed samples. Novel NIR instrument designs have markedly improved sensitivity and reproducibility, and therefore provide the potential for single seed soybean composition analysis. In our study, five different state-of-the-art NIR instruments were evaluated and calibrated for single seed soybean composition analysis.

NIR Instrumentation Techniques

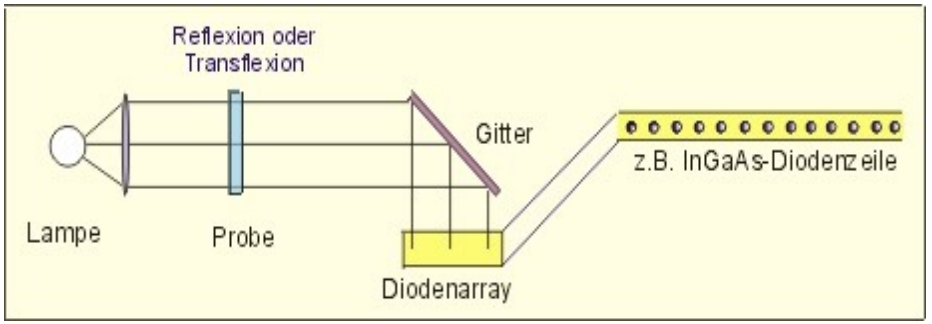
Dispersive



Fourier Transform



Dual Diode Array



http://www.pirnet.bafz.de/enirintro.htm

NIR Instruments

1. Perten's DA7000 Dual Diode-Array NIR Spectrometer
   - Detector: Silicon, InGaAs
   - Spectral Range: 400 ~ 1700 nm

1. Nicolet's Antaris FT-NIR Spectrometer
   - Detector: InGaAs
   - Spectral Range: 12000 ~ 4000 cm^{-1}

NIR Instruments

1. Bruker's Vector-22 FT-NIR Spectrometer
   - Detector: PbSe
   - Spectral Range: 10000-4000 cm^{-1}

2. Bruker's Vector-33 FT-IR/NIR Spectrometer
   - Detector: PbSe
   - Spectral Range:10000- 4000 cm^{-1}

3. Perkin Elmer Spectrum One
   NTS: InGaAs, 12,000-4,000 $cm^{-1.}$

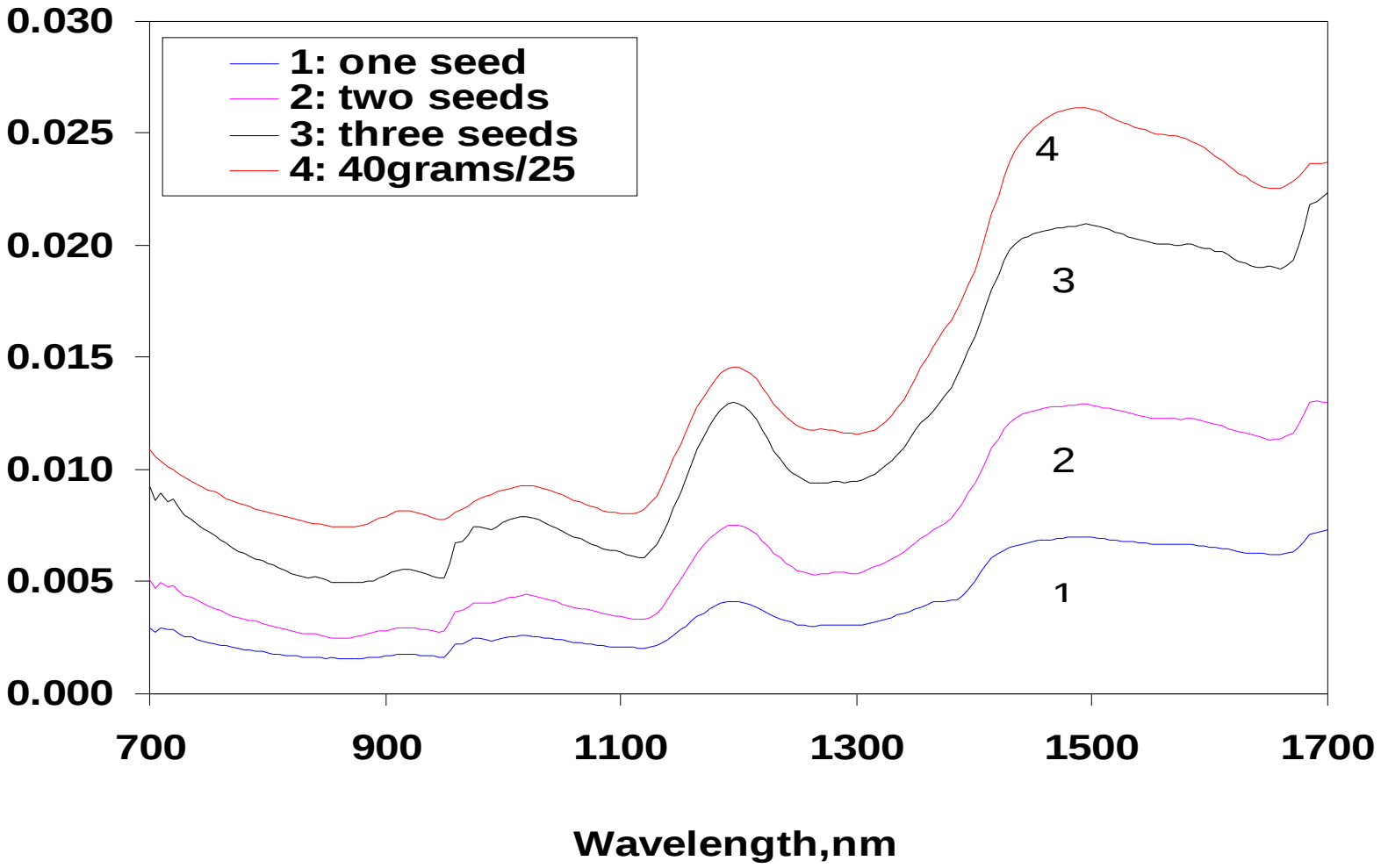
Standard Soybean Samples

- Standard soybean samples for the calibration development in this study were selected from soybean cultivars available at the National Soybean Germplasm Collection (Urbana, IL) and from developmental soybean lines cross-bred by Dr. R.L.Nelson. The selection of such standard samples was based on their protein, oil, and moisture contents, to ensure that the ranges of standard sample constituent content covered the full range of possible constituent variations of samples that are to be measured.
- The constituent ranges of the selected standard samples are: from 34% to 55% for protein, from 11% to 22% for oil, from 2% to 16% for moisture; all such composition data are here reported on a wet basis.

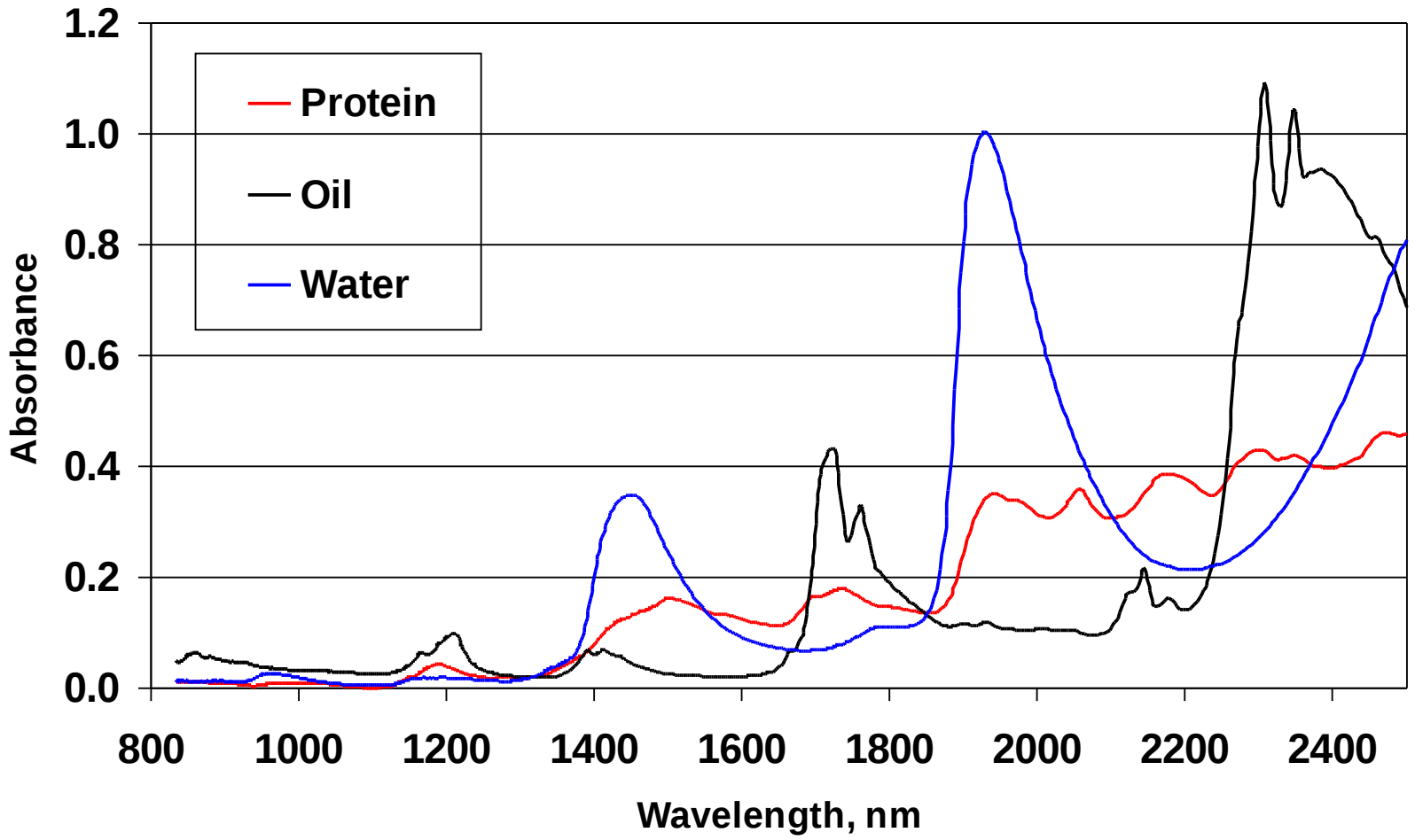
Acquisition of Single Seed Soybean NIR Spectra

- All NIR spectra of single soybean seeds in this study were collected in reflection mode employing the full spectral range of each NIR instrument
- On the Perten's DA-7000 DA-NIR instrument, an 1x1 cm Teflon was placed on the top center of the Spectralon, single soybean seeds were then placed on top of the Teflon and spectra were collected between 400 and 1700 nm
- On the Perkin Elmer's Spectrum One NTS FT-NIR instrument, single seed spectra were collected between 12000 and 4000 cm^{-1} with an Integrating Sphere accessory
- On the Nicolet's Antaris, Bruker's Vector-22, Vector-33 FT-NIR instruments, single seed spectra were collected between 10000 and 4000 cm^{-1} with an Integrating Sphere accessory.

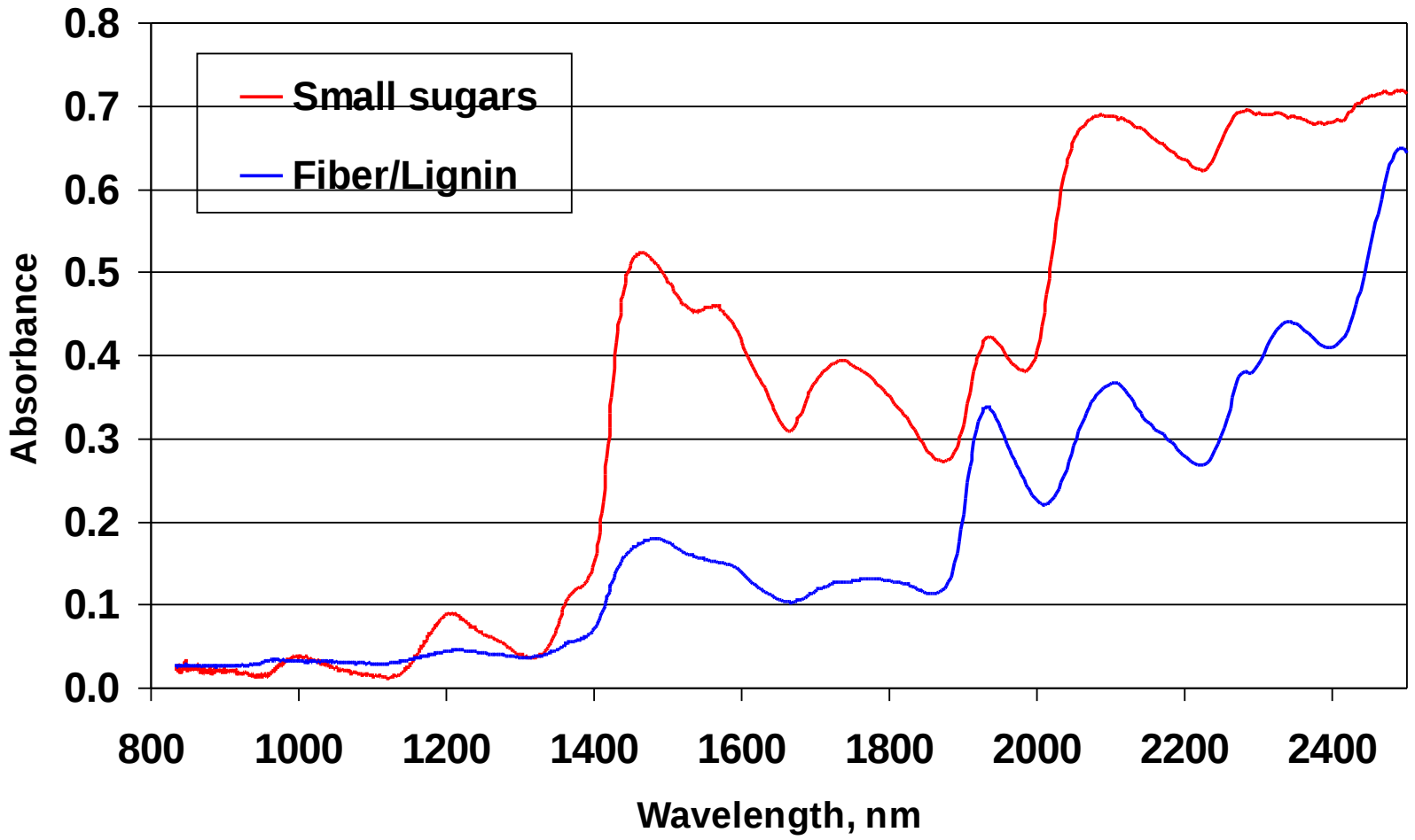
Comparison of DA-NIRS Spectra of Single Soybean Seeds with Bulk Sample



Pure Component FT-NIR Spectra of Major Soybean Constituents: Protein, Oil, Moisture



Pure Component FT-NIR Spectra of Major Soybean Constituents: Small sugars, and Fiber



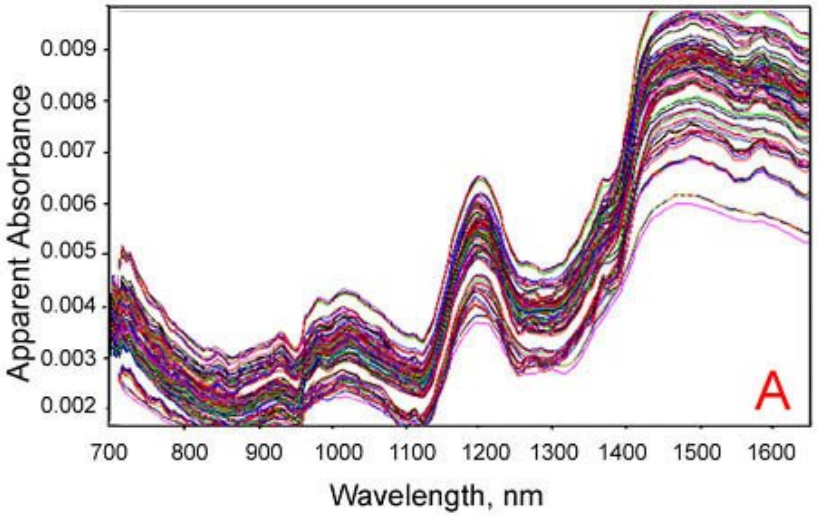
Single Seed NIR Spectra Preprocessing

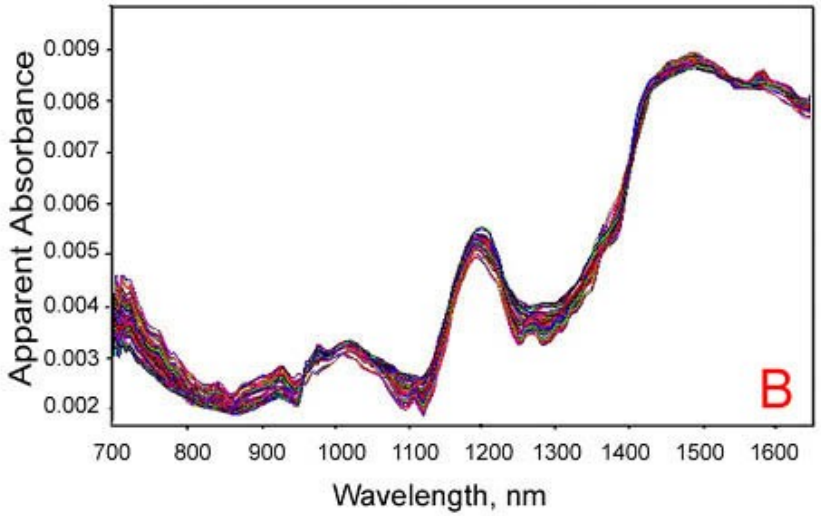
- For calibrations developed on all five NIR instruments, before proceeding to single seed calibration, raw NIR spectra were corrected for spectral variations that are not attributed by sample composition, such as light scattering effects and baseline variations.

- Random spectral variations and light scattering effects were corrected by application of baseline correction and Multiplicative Scattering Correction (MSC).

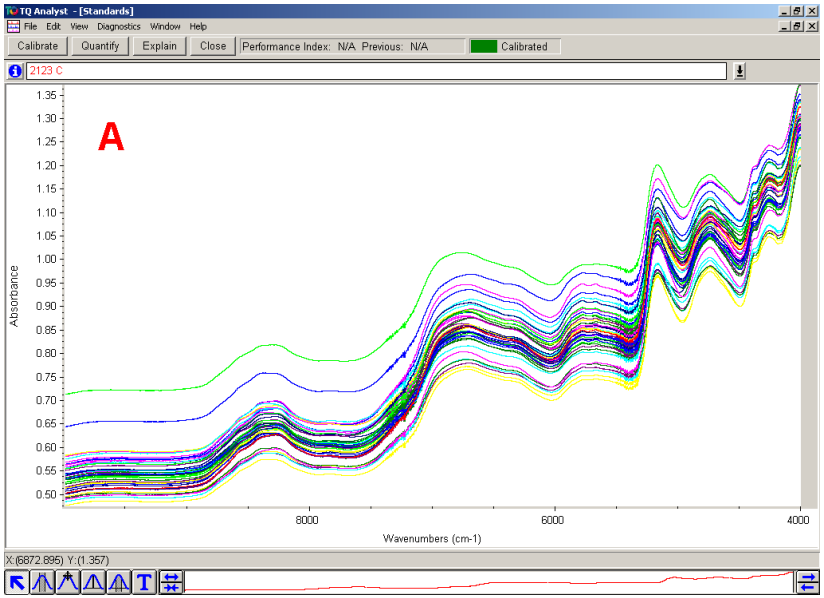
- Baseline variations were mainly corrected by taking the spectra derivatives with the Savitsky-Golay algorithms. For the Perkin Elmer Spectrum One NTS, baseline corrections were also carried out with an interactive spline function correction method.

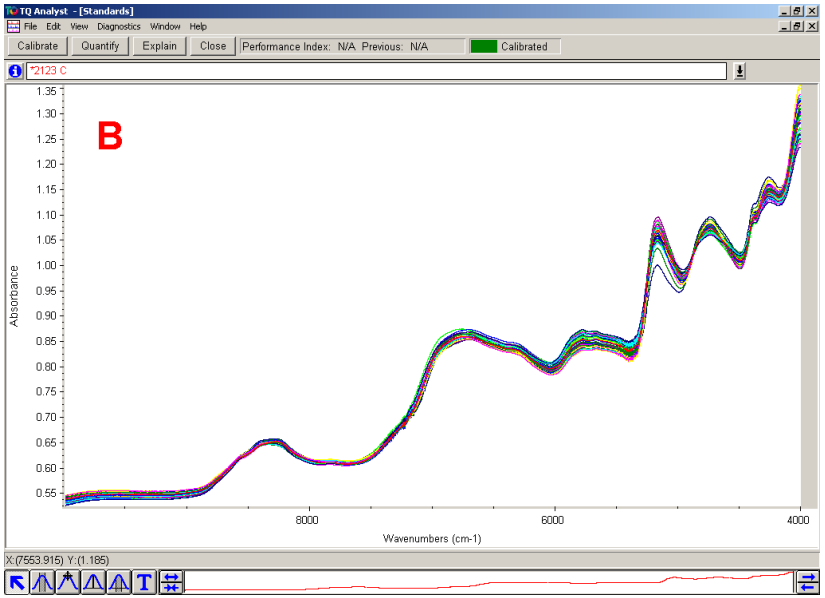
Overlay Plot of DA-NIRS Spectra of Single Soybean Seeds Obtained with the Perten DA-7000 Instrument. A: Raw, B: Multiplicative Scattering Correction(MSC)



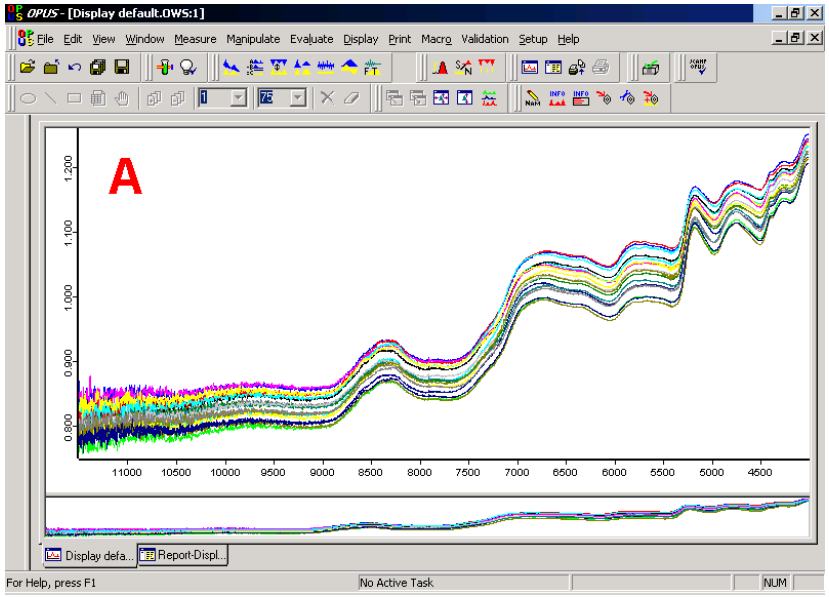


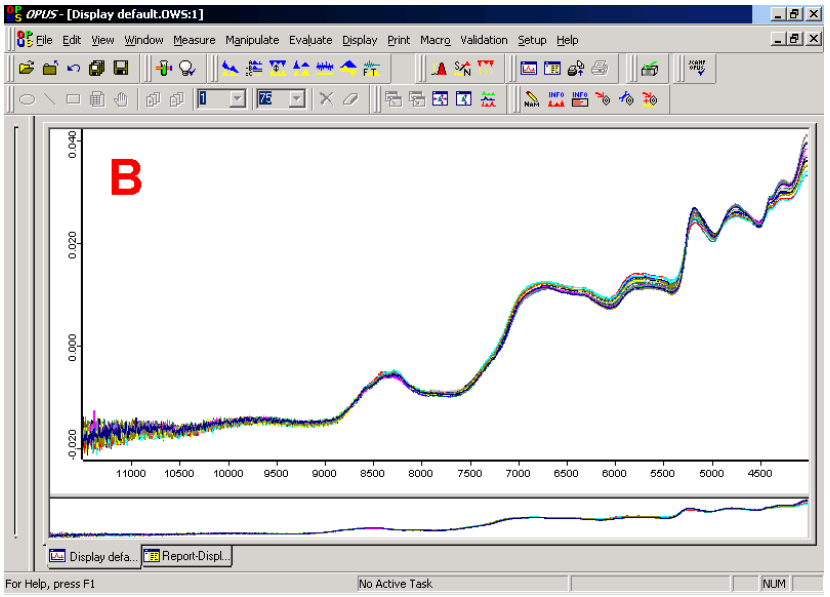
Overlay Plot of FT-NIRS Spectra of Single Soybean Seeds Obtained with the Nicolet Antaris Instrument. A: Raw, B: MSC





Overlay Plot of FT-NIRS Spectra of Single Soybean Seeds Obtained with the Bruker Vector-33 Instrument. A: Raw, B: MSC

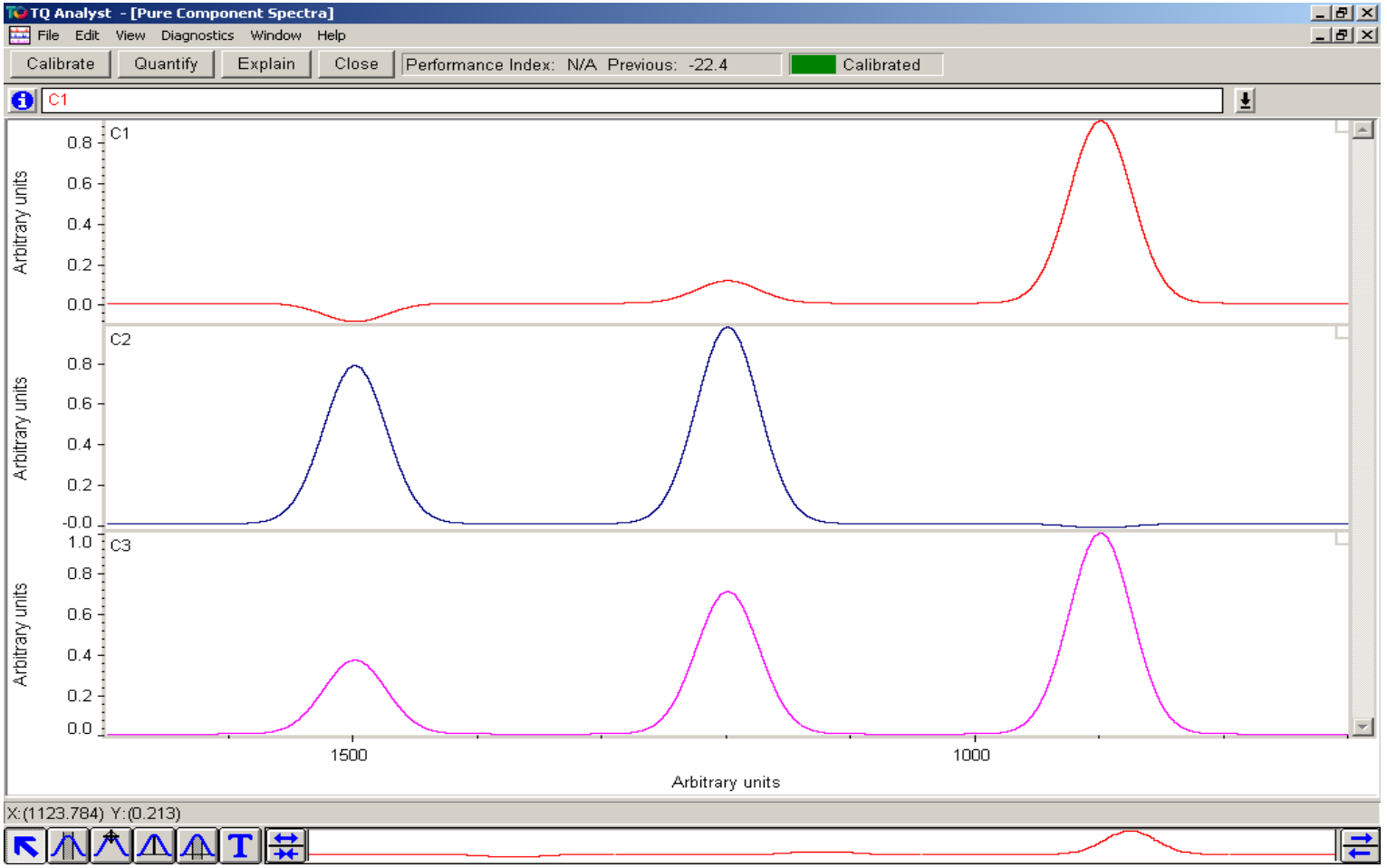




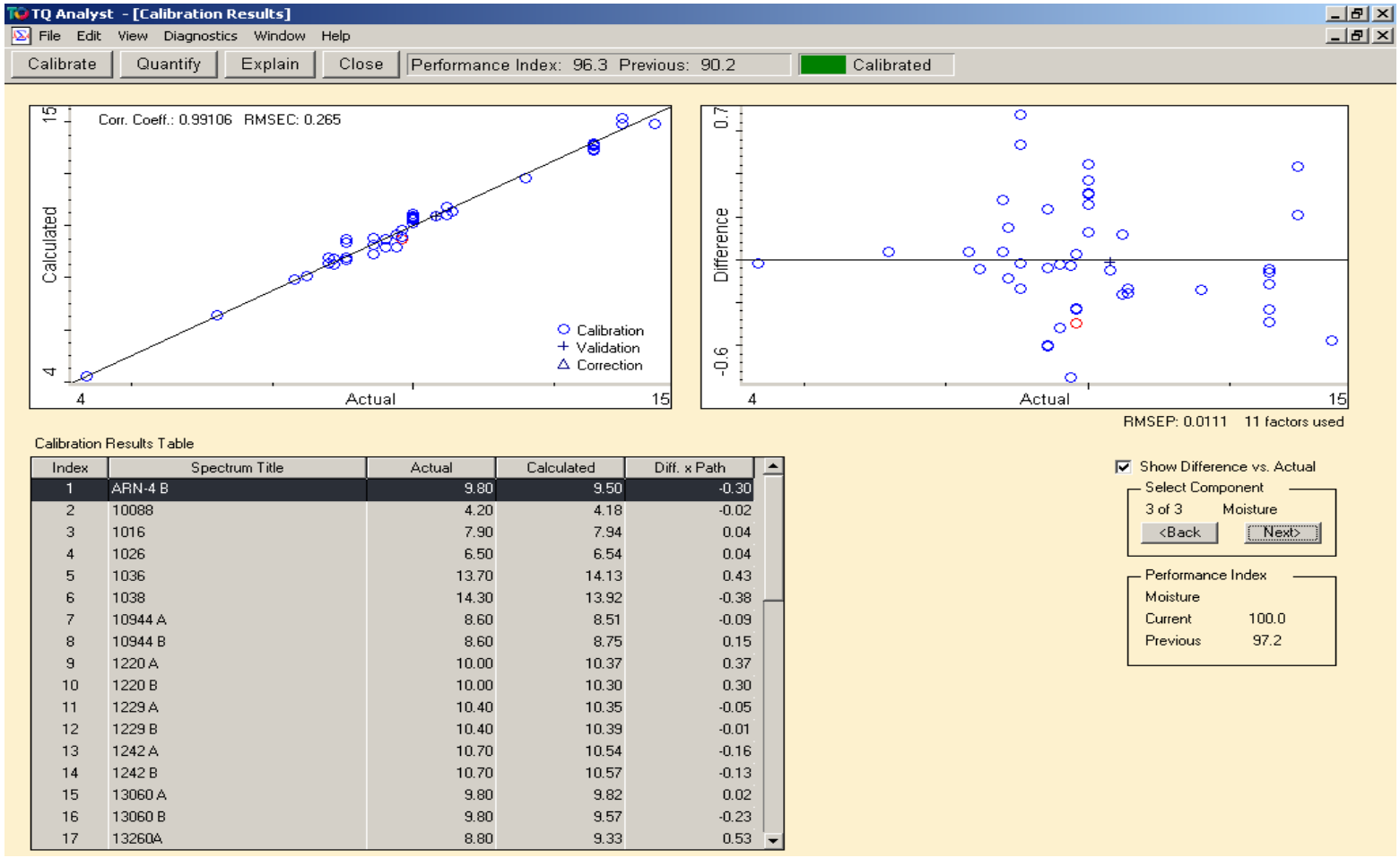
Single Seed Soybean Calibration Models

- Based on our previous experience with bulk soybean calibration development, Partial Least Squares Type 1 (PLS-1) model was applied for single seed NIR calibration on all five NIR instruments in this study.
- Three different software packages were employed for calibration development, for the reason that some NIR instrument stores the collected spectra in a format only recognizable by its own chemometrics packages.
- Calibrations for the Perten's DA-7000 instrument were developed with the PLSplus/IQ program included in the GRAMS/32 software package from the Galactic Industries Corporation (Salem, NH).
- Calibrations for the Bruker's Vector 22 and Vector 33 were developed with the OPUS software package provided by Bruker.
- Calibrations for the Nicolet's Antaris were developed with the "TQ Analysis" software provided by Nicolet.

Possible Failure of the PLS-1 Algorithm for the Special Case of Highly Correlated Samples, such as Soybeans Measured at Constant Moisture with a Strong Inverse P-O Correlation: Modeling with A Synthetic Spectra Calculation

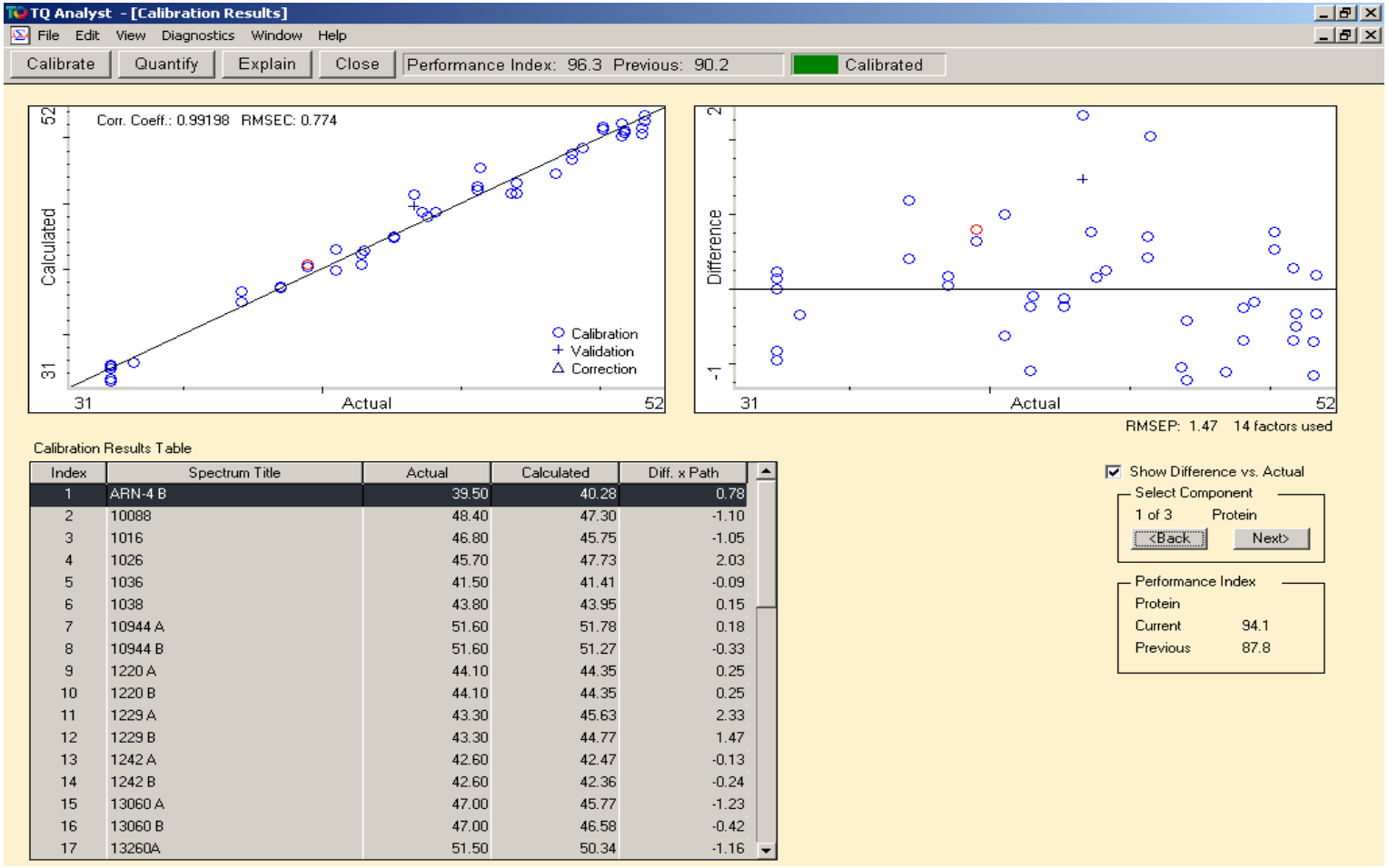


NIR Predicted vs Reference Moisture Values by Nicolet Antaris Single Seed Calibration



NIR Predicted vs Reference Protein Values by Nicolet Antaris Single Seed Calibration

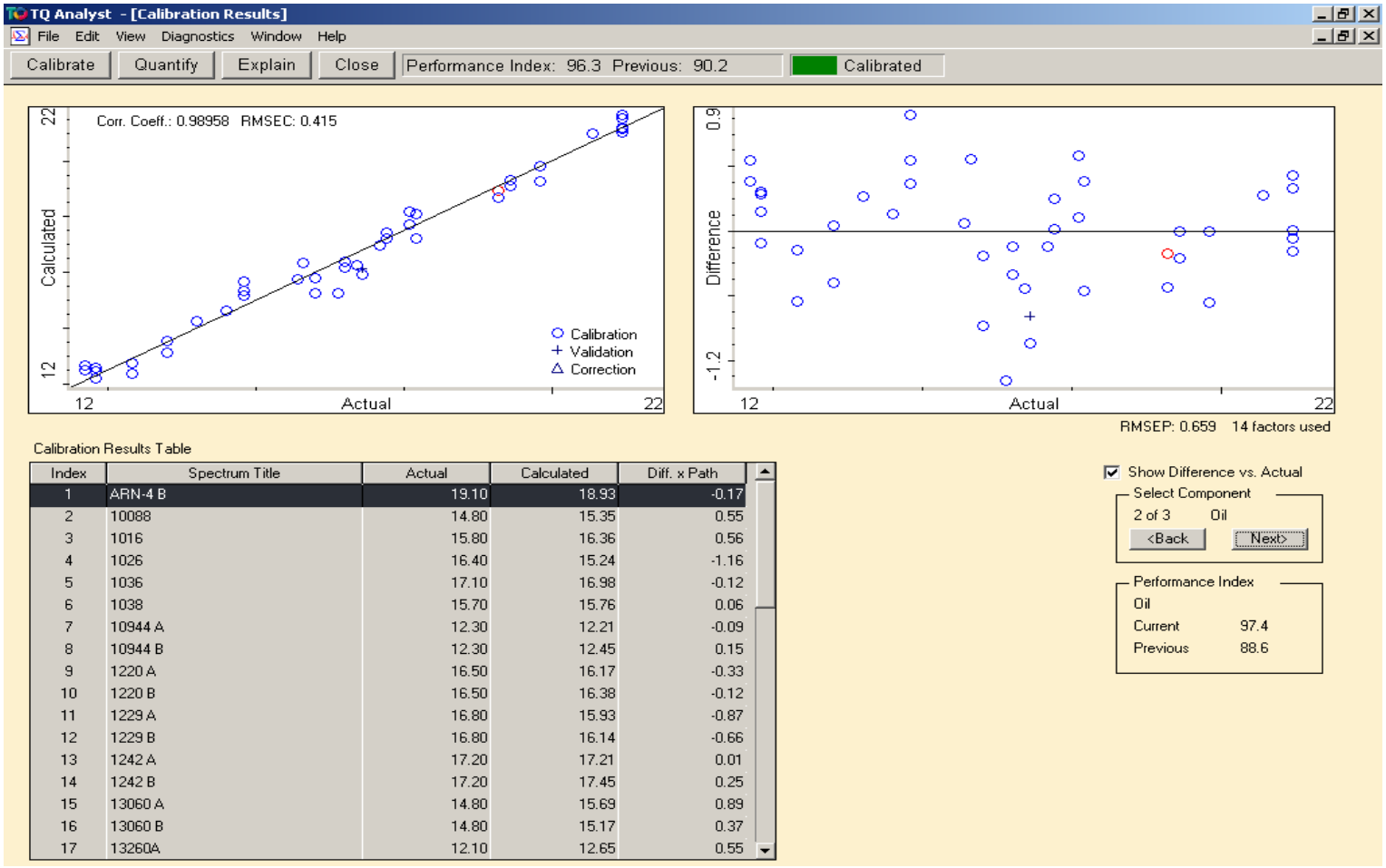
Nature Precedings : doi:10.1038/npre.2011.6195.1 : Posted 4 Aug 2011



Index	Spectrum Title	Actual	Calculated	Diff. x Path
1	ARN-4 B	39.50	40.28	0.78
2	10088	48.40	47.30	-1.10
3	1016	46.80	45.75	-1.05
4	1026	45.70	47.73	2.03
5	1036	41.50	41.41	-0.09
6	1038	43.80	43.95	0.15
7	10944 A	51.60	51.78	0.18
8	10944 B	51.60	51.27	-0.33
9	1220 A	44.10	44.35	0.25
10	1220 B	44.10	44.35	0.25
11	1229 A	43.30	45.63	2.33
12	1229 B	43.30	44.77	1.47
13	1242 A	42.60	42.47	-0.13
14	1242 B	42.60	42.36	-0.24
15	13060 A	47.00	45.77	-1.23
16	13060 B	47.00	46.58	-0.42
17	13260A	51.50	50.34	-1.16

NIR Predicted vs Reference Oil Values by Nicolet Antaris Single Seed Calibration

Nature Precedings : doi:10.1038/npre.2011.6195.1 : Posted 4 Aug 2011



Index	Spectrum Title	Actual	Calculated	Diff. x Path
1	ARN-4 B	19.10	18.93	-0.17
2	10088	14.80	15.35	0.55
3	1016	15.80	16.36	0.56
4	1026	16.40	15.24	-1.16
5	1036	17.10	16.98	-0.12
6	1038	15.70	15.76	0.06
7	10944 A	12.30	12.21	-0.09
8	10944 B	12.30	12.45	0.15
9	1220 A	16.50	16.17	-0.33
10	1220 B	16.50	16.38	-0.12
11	1229 A	16.80	15.93	-0.87
12	1229 B	16.80	16.14	-0.66
13	1242 A	17.20	17.21	0.01
14	1242 B	17.20	17.45	0.25
15	13060 A	14.80	15.69	0.89
16	13060 B	14.80	15.17	0.37
17	13260A	12.10	12.65	0.55

Correlation Coefficients (R) and Standard Errors of Cross Validation (SECV) for Single Seeds Analysis on the DA-7000 Instrument

Component	Number of Factors	R	SECV
Protein	15	98.5%	1.1
Oil	16	98.5%	0.5
Moisture	16	99.0%	0.3

Correlation Coefficients (R) and Standard Error of Cross Validation (SECV) for Single Seed Analysis on the Nicolet Antaris FT-NIR Instrument

Component	Number of Factors	R	SECV
Protein	14	99.2%	0.77
Oil	14	99.0%	0.42
Moisture	11	99.1%	0.27

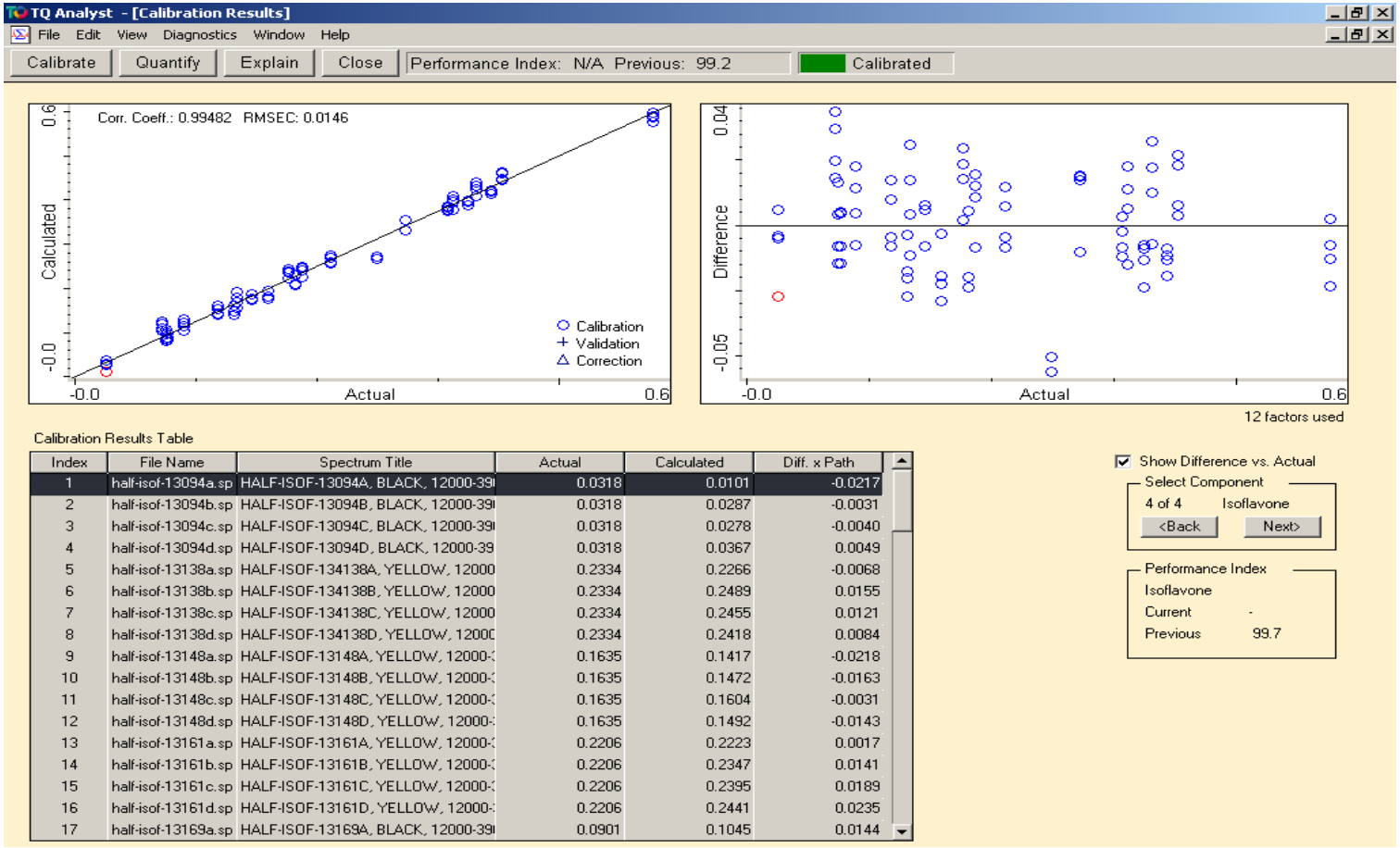
Correlation Coefficients (R) and Standard Error of Cross Validation (SECV) for Single Seed Analysis on the Bruker Vector-22 FT-NIR Instrument

Component	Number of Factors	R	SECV
Protein	6	98.1%	1.15
Oil	6	98.0%	0.56
Moisture	6	94.8%	0.38

Correlation Coefficients (R) and Standard Error of Cross Validation (SECV) for the Soybean **Half Seed** Calibration Obtained with the Bruker Vector-22 Instrument

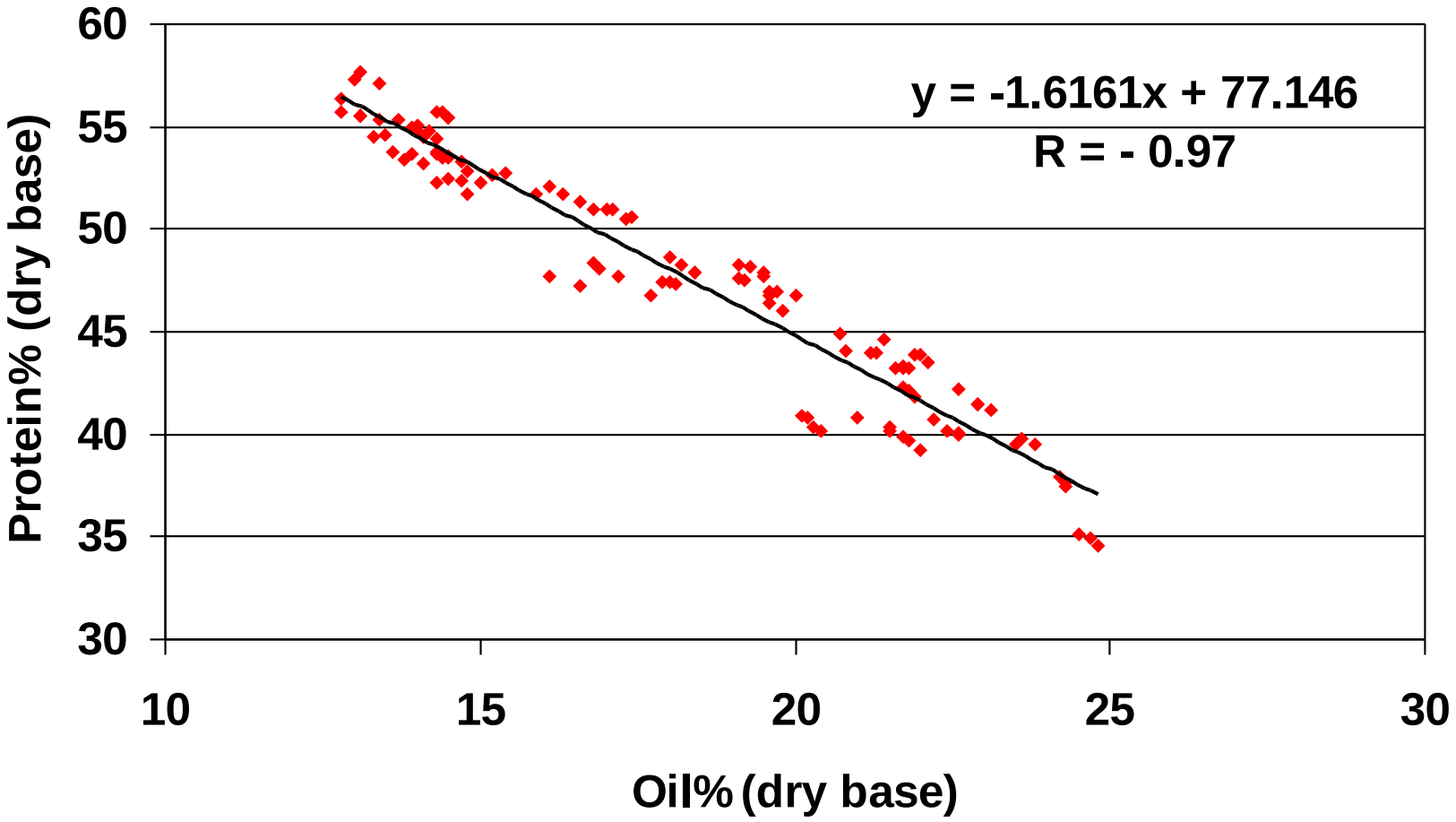
Component	Number of Factors	R	SECV
Total Isoflavones	12	99.5%	0.015
Protein	12	99.8%	0.16
Oil	12	99.8%	0.10
Moisture	12	99.7%	0.05

NIR Predicted vs Reference Values of Total Isoflavone Content (**Half Soybean Seed** Calibration Obtained with the Bruker Vector-22)



Index	File Name	Spectrum Title	Actual	Calculated	Diff. x Path
1	half-isof-13094a.sp	HALF-ISOF-13094A, BLACK, 12000-39	0.0318	0.0101	-0.0217
2	half-isof-13094b.sp	HALF-ISOF-13094B, BLACK, 12000-39	0.0318	0.0287	-0.0031
3	half-isof-13094c.sp	HALF-ISOF-13094C, BLACK, 12000-39	0.0318	0.0278	-0.0040
4	half-isof-13094d.sp	HALF-ISOF-13094D, BLACK, 12000-39	0.0318	0.0367	0.0049
5	half-isof-13138a.sp	HALF-ISOF-134138A, YELLOW, 12000	0.2334	0.2266	-0.0068
6	half-isof-13138b.sp	HALF-ISOF-134138B, YELLOW, 12000	0.2334	0.2489	0.0155
7	half-isof-13138c.sp	HALF-ISOF-134138C, YELLOW, 12000	0.2334	0.2455	0.0121
8	half-isof-13138d.sp	HALF-ISOF-134138D, YELLOW, 12000	0.2334	0.2418	0.0084
9	half-isof-13148a.sp	HALF-ISOF-13148A, YELLOW, 12000-	0.1635	0.1417	-0.0218
10	half-isof-13148b.sp	HALF-ISOF-13148B, YELLOW, 12000-	0.1635	0.1472	-0.0163
11	half-isof-13148c.sp	HALF-ISOF-13148C, YELLOW, 12000-	0.1635	0.1604	-0.0031
12	half-isof-13148d.sp	HALF-ISOF-13148D, YELLOW, 12000-	0.1635	0.1492	-0.0143
13	half-isof-13161a.sp	HALF-ISOF-13161A, YELLOW, 12000-	0.2206	0.2223	0.0017
14	half-isof-13161b.sp	HALF-ISOF-13161B, YELLOW, 12000-	0.2206	0.2347	0.0141
15	half-isof-13161c.sp	HALF-ISOF-13161C, YELLOW, 12000-	0.2206	0.2395	0.0189
16	half-isof-13161d.sp	HALF-ISOF-13161D, YELLOW, 12000-	0.2206	0.2441	0.0235
17	half-isof-13169a.sp	HALF-ISOF-13169A, BLACK, 12000-39	0.0901	0.1045	0.0144

Protein-Oil Inverse Correlation for 120 Single Seed Soybean Samples



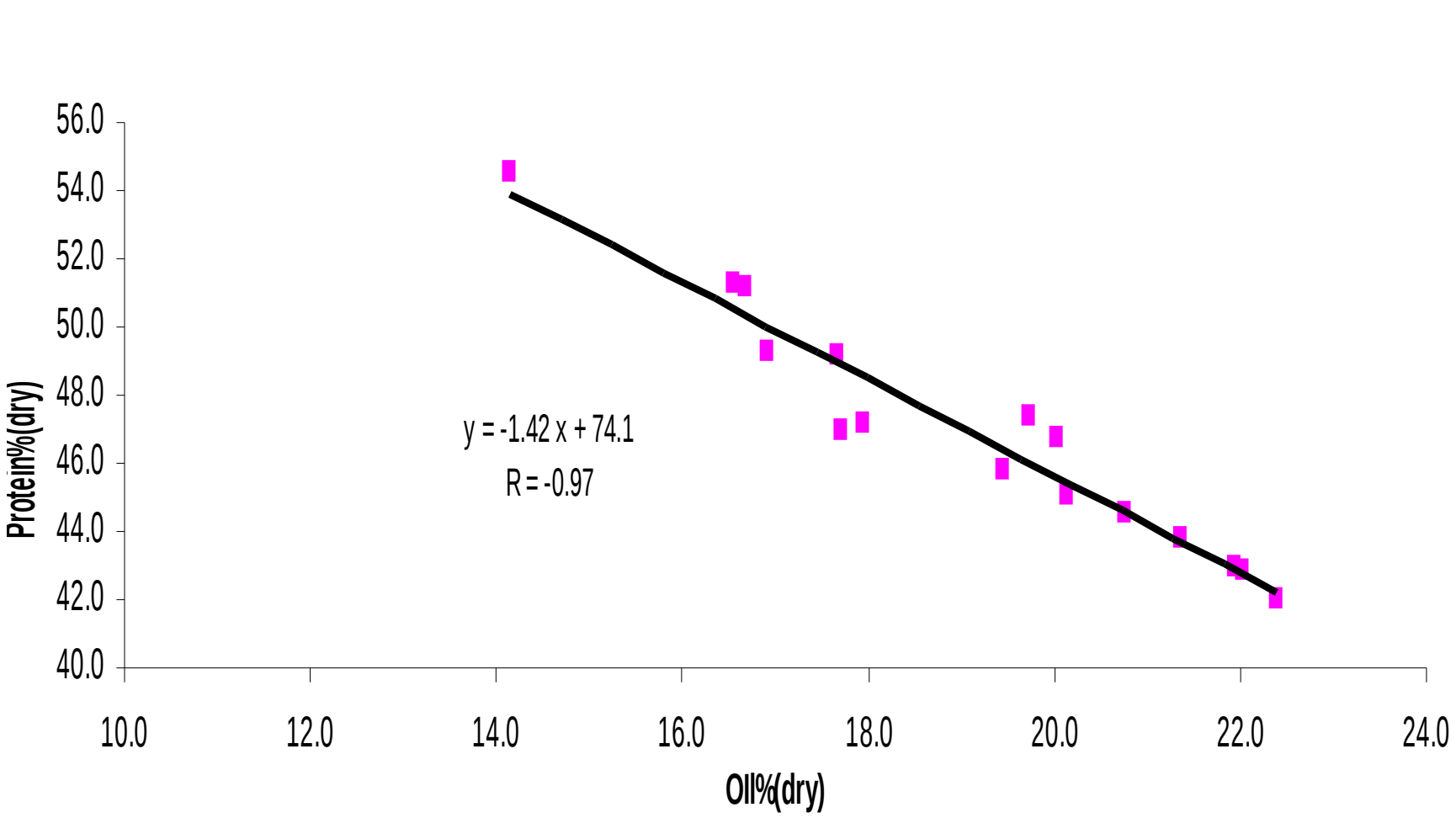


Figure 6(A). Median protein%vs. mean protein%correlation for 17 bulk soybean selection groups

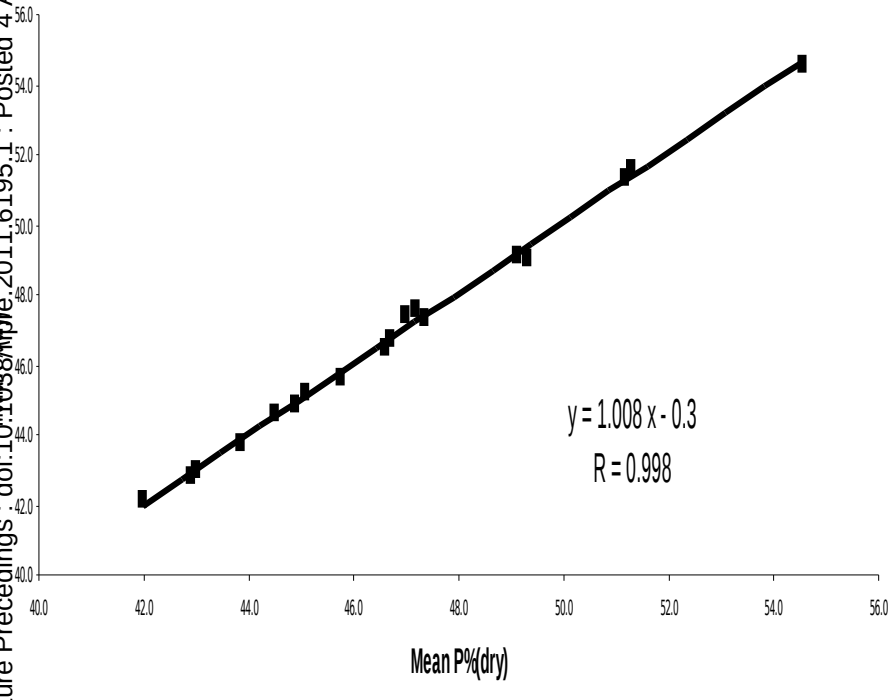
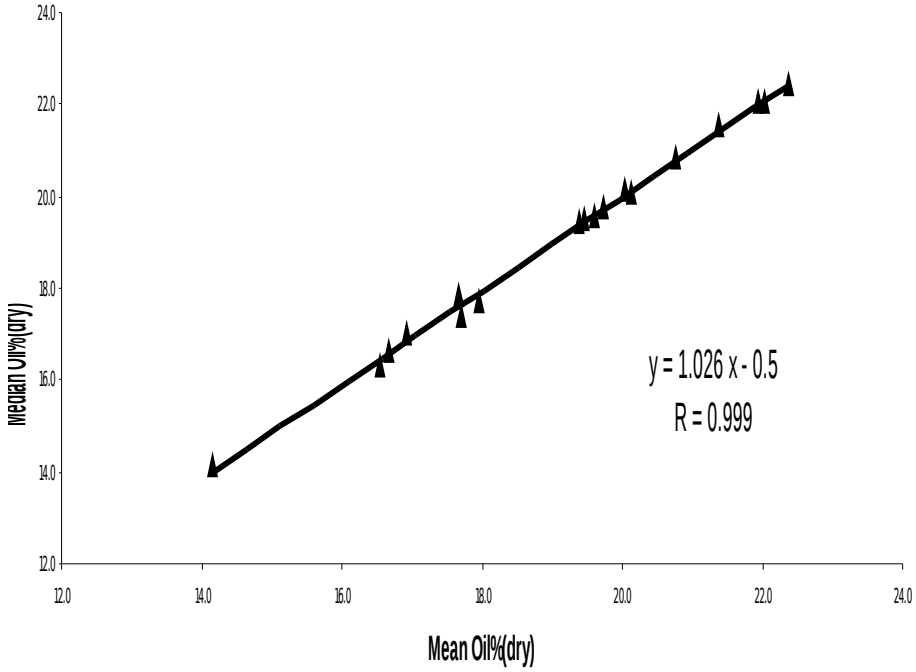
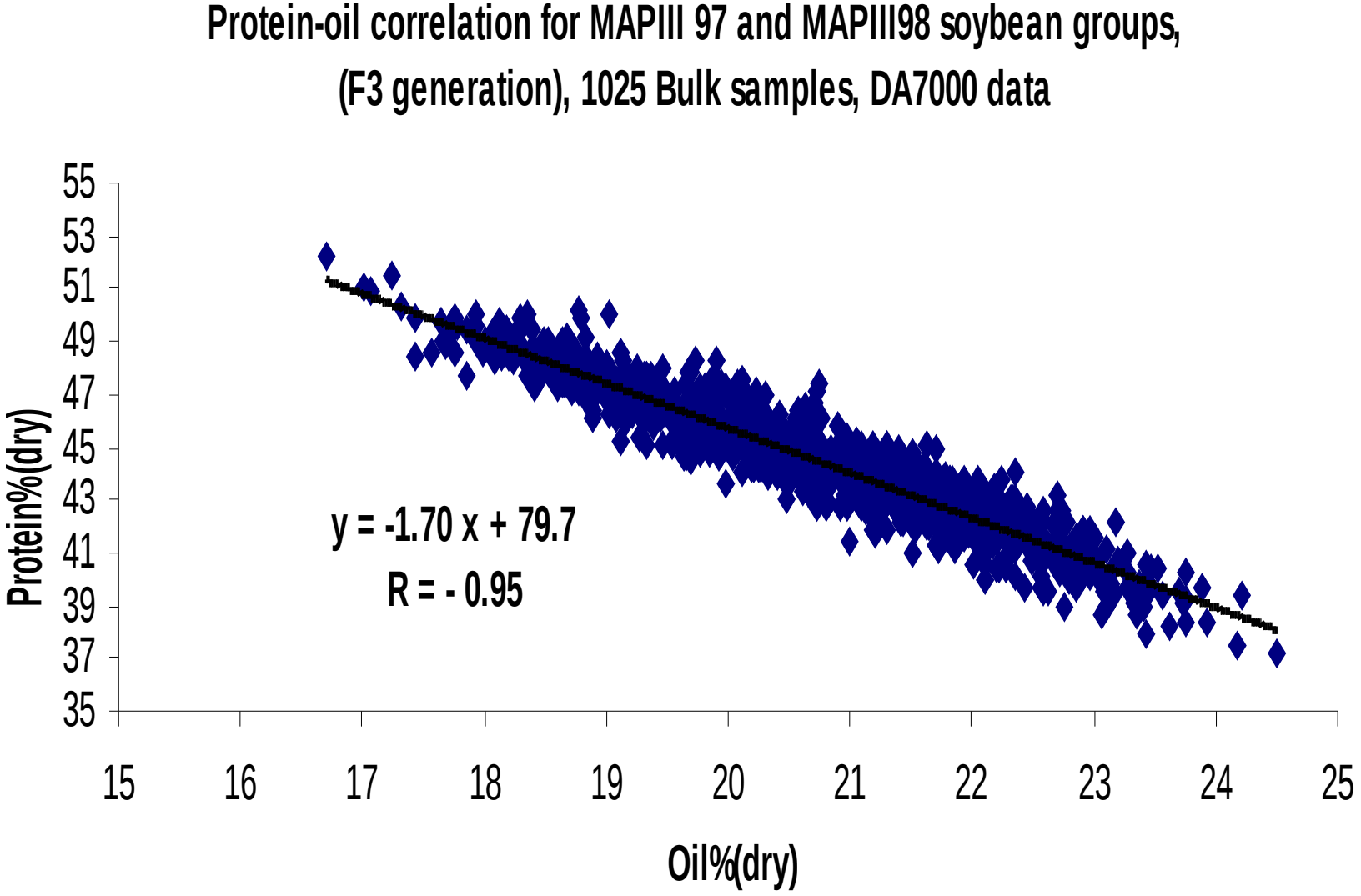
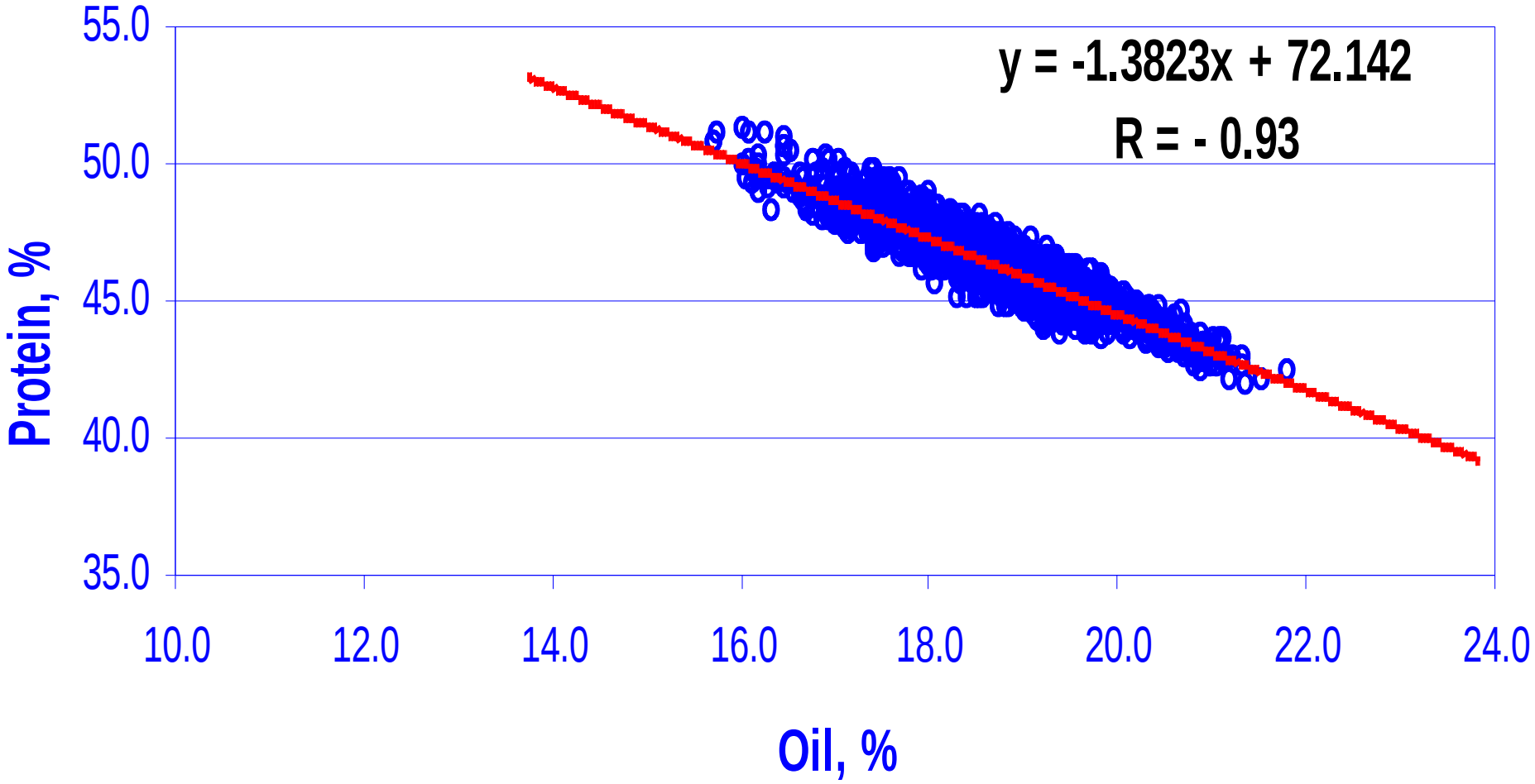


Figure 6(B). Median oil%vs. mean oil%correlation for 17 bulk soybean selection groups





Protein-Oil Inverse Correlation for 4500 Map3 (F3 Generation, Harvested in '97 & '98) Bulk Soybean Samples



Discussion

- Conventional soybean composition analysis methods (such as AOCS official methods) usually require relatively large amounts of samples and are also time-consuming, therefore impractical for single seed composition analysis
- The very high sensitivity of the novel design FT-NIR detectors/instruments/accessories makes it possible to develop a series of new applications that are potentially important for molecular genetics, proteomics, biotechnology, soybean improvements, health foods, human nutrition, disease prevention and its early detection; some of these novel FT-NIR applications are summarized on the next slide.

What Other Interesting Analyses can be done with such novel FT-NIR Techniques & Methods ?

- Amino Acid composition analyses of both bulk and single/half soybean seeds, calibrated with HR-NMR data
- Fatty acid composition
- Phytic acid (for breeding and genetic selection programs)
- Nucleic acid analyses
- High-resolution FT-NIR Chemical Imaging and Microspectroscopy of single/half seeds, developing and mature embryos
- ***Picogram*** Microanalysis by FT-NIR microscopy of ***Microarrays*** for genetic selection, proteomics, metabolic engineering, other related biotechnology and biomedical applications
- For further details and illustrations, please visit the five related posters #505-509 on these topics in the poster session.

Conclusions (1)

- NIR calibrations for single seed soybean composition analysis were successfully developed on four state-of-the-art NIR instruments that utilize the novel Dual Diode Array and Fourier Transform techniques. Seed-to-seed variations in protein and oil content can thus be *monitored for the first time on single soybeans* of selected soybean accessions
- The results presented here show that Multiplicative Scattering Correction (MSC) significantly reduces spectral variations that are not related to concentration changes, thereby improves the reliability of our calibrations.
- We have extensively evaluated NIR techniques and instrumentation for genetic selection and cross-breeding purposes. The best available NIR instrumentation is found to provide reliable secondary analyses for : protein, oil, moisture, small sugars, fiber and isoflavones of both bulk and single soybean seeds.
- A very strong, inverse correlation (-R>0.90) is found between protein and oil contents of soybeans for more than 10,000 different soybean lines and a large number of locations. This is consistent with tight energetic constraints for protein and oil synthesis in the soybean.

Conclusions (2)

- We are reporting the first High-resolution (1um) FT-NIR Chemical Imaging and Microspectroscopy of single/half soy seeds, developing and mature embryos.

- Our results also indicate that picogram microanalysis of Microarrays for molecular genetics, proteomics/ metabolic engineering and other related biotechnology/biomedical applications are now possible (Poster #509, this meeting).

Acknowledgements

- We are gratefully acknowledging the partial support by the Illinois Soybean Operation Board and the USDA

- We are also grateful for the loan of the DA7000 instrument by Perten Instruments Inc., and especially Dr. Carrasco and Mgr. Walter Munday for helpful suggestions

- Thanks are also due to Dr. Dean Roberts from Bruker and Dr. Herman He from Thermo Nicolet for helping generate the FT-NIR data analyzed in our study.

References

- Baianu, I.C. Value Added Soybean Summit, Washington, D.C. 1993

- Baianu, I.C. and T. F. Kumosinski. 1992. Physical Chemistry of Food Processes. Vol. 1.

- Geladi, P., MacDougall, D., Martens, H. 1985. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. Applied Spectroscopy. 39(3): 491-500

- Isaksson, T., Naes, T. 1988. The effect of Multiplicative Scatter Correction (MSC) and Linearity Improvement in NIR Spectroscopy. Applied Spectroscopy. 42(7): 1273-1284

- Nielsen S. S. 1994. Introduction to the Chemical Analysis of Foods: 139-159

- Shadow, W., Carrasco, A., 2000. Practical single-kernel NIR/Visible analysis for small grains. Cereal Foods World. 45(1): 16-18

- Specht, J.E., et al, 2002. *A Genomic Perspective on the Soybean Protein(+)/Oil(-)/Yield(-) Enigma.* **Paper #202 at this meeting.**

References

- Williams, P, Norris, K. 1987. Near-Infrared technology in the agricultural and food industries. AACC, Inc., St.Paul, MN,USA

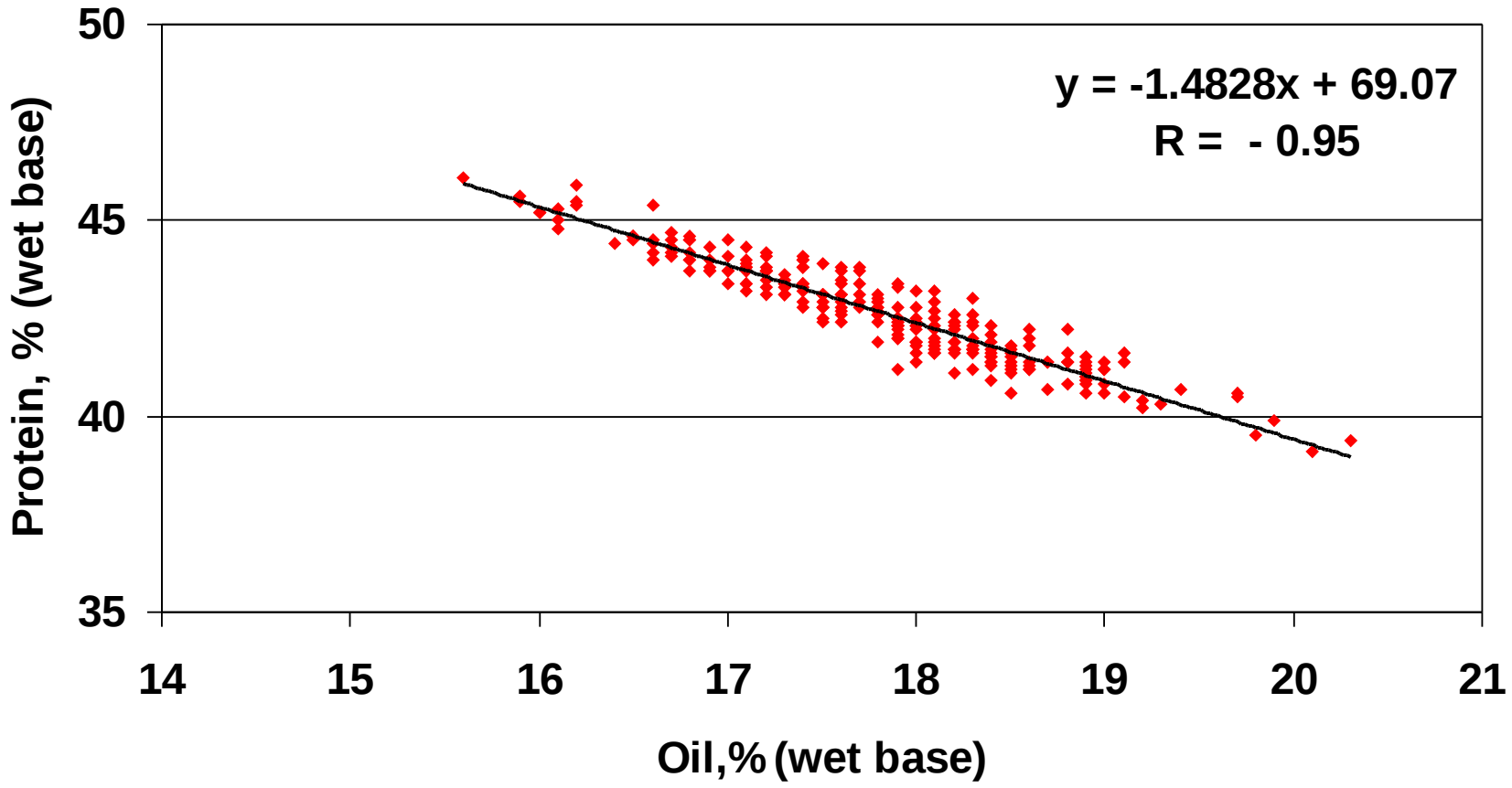
- You, T., Guo J., Baianu, I.C., 1999. Soybean composition analysis by Near Infrared Spectroscopy. Global Soy1999 (Abstract).

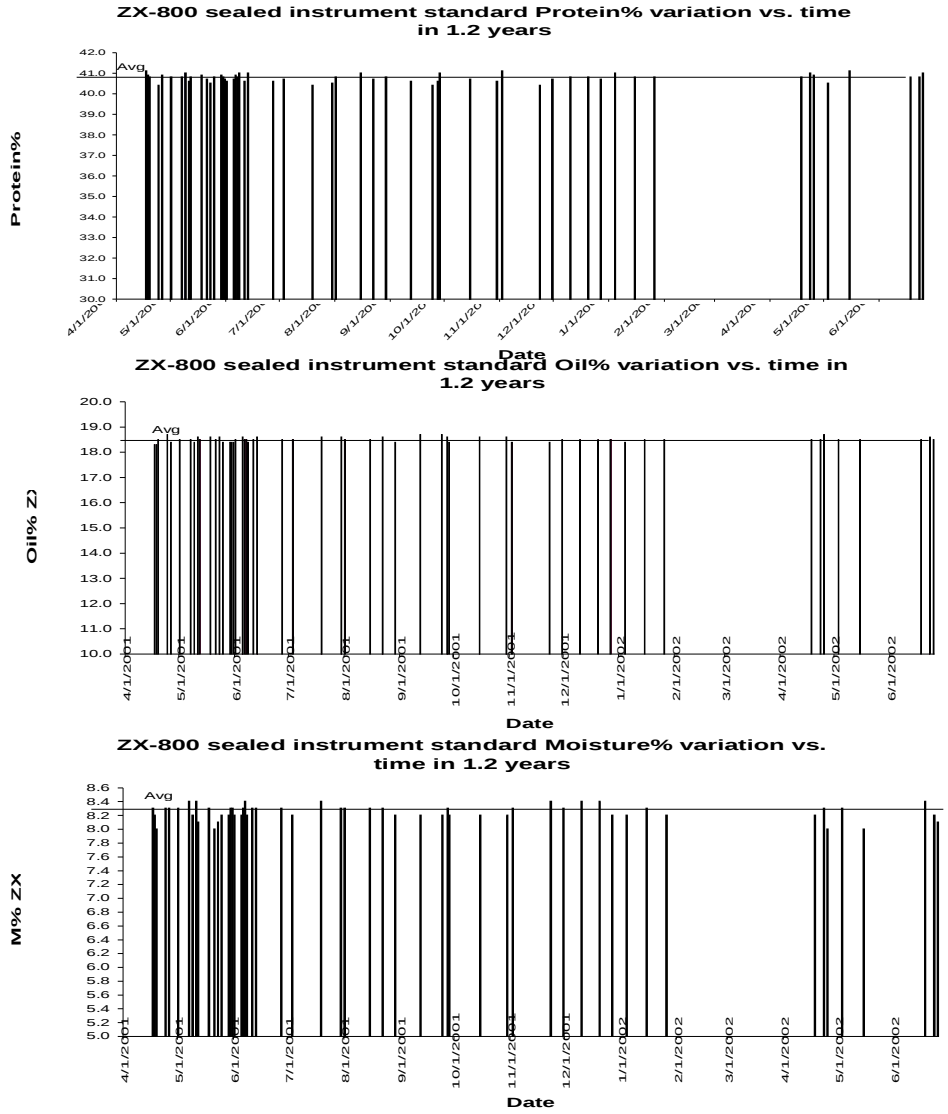
- You, T., Guo J., Baianu, I.C., Nelson, R.L., 2002. ***Determination of Isoflavone Contents for Selected Soybean Lines by Fourier Transform Near Infrared Reflectance Spectroscopy***. **Poster #P505 at this meeting.**

Mean values and variances of protein and oil for 17 Bulk Soybean groups.

Soybean groups and instruments used	Sample number	P-mean	P-var	O-mean	O-var
Germplasm98, DA7000	400	46.2	13.55	19.6	2.23
MAPIII97-mix, F3, DA7000	496	43.8	5.53	21.4	1.50
MAPIII98-mix, F3, DA7000	529	45.1	6.56	20.1	1.73
MAPIII97-Bell, F3, IM9100	707	47.3	3.84	19.7	1.25
MAPIII97-Fisher, F3, IM9100	704	46.7	3.85	20.0	1.03
MAPIII97-EG1000, F3, IM9100	703	43.0	3.80	21.9	0.95
OPMAP98, F3, IM9100	379	44.5	6.19	20.8	2.54
Germplasm99, ZX-50	961	44.9	2.61	19.4	0.96
99ProSel-F3, ZX-50	72	51.2	3.69	16.7	0.28
99ProSel-F4, ZX-50	206	49.1	4.40	17.7	0.90
OPMAP99-Hume, F3, ZX-800	495	47.2	6.83	17.9	2.76
OPMAP99-Bell, F3, ZX-800	490	47.0	6.91	17.7	3.07
99ProGen, F5, ZX-800	362	49.3	6.31	16.9	1.71
2000YLDMAP-Hume, F6, ZX-800	380	42.9	1.68	22.0	0.55
2000YLDMAP-MIV, F6, ZX-800	380	42.0	2.70	22.4	0.75
2000 ProSel and BCPro, F8, ZX-800	79	51.3	3.67	16.5	1.14
2000 ProSel, F9, ZX-800	24	54.6	0.99	14.2	0.27

Note 1: P-mean, P-var, O-mean, O-var represent protein mean, protein variance, oil mean, oil variance, respectively.





Time graph for long term stability of the ZX-800 instrument for protein, oil and moisture measurements with its sealed instrument standard sample over 1.2 years: (A) Protein, stdev = 0.2, (B) Oil, stdev = 0.1, (C) Moisture, stdev = 0.1.

Protein Correlation Plot : PLS-1 Prediction from Pure Component Spectra—'Synthetic' Data