

**Determination of Percent
Protein and Oil in Soybean Seed
Using AOTF-NIR Spectroscopy**

Executive Summary

Sixty soybean seeds, fifteen from each of 4 different populations, were scanned using the Brimrose Luminar 3076 Seed Meister. Each of the 4 populations varied in percent protein and total oil. Partial Least Squares regression models were then calculated to determine the degree of correlation between the 2 constituents of interest, i.e., protein and total oil and the spectra. Very good correlation was calculated for both constituents and predictions on a subset of 14 samples gave very good results.

The Brimrose Luminar 3076 Seed Meister is an excellent tool for breeders to sort the best of seeds from large segregating F_2 and F_3 populations. The measurement is non-destructive and sorting is fast at 30 or more seeds per minute. The breeder can reduce the number of seeds planted in trial plots by as much as 80 percent saving both time and money.



1 Introduction

The principle of the solid-state, non moving parts, Acousto-Optic Tunable Filter (AOTF) is based upon the acoustic diffraction of light in an an-isotropic medium. The device consists of a piezo-electric transducer bonded to a birefringent crystal. When the transducer is excited by an applied RF signal, acoustic waves are generated in the crystal. The propagating acoustic wave produces a periodic modulation of the refractive index. This provides a moving phase grating that under proper conditions will diffract portions of an incident light beam. For a fixed acoustic frequency, a narrow band of optical frequencies satisfies the phase matching conditions and is cumulatively diffracted. As the RF frequency is changed, the center of the optical band-pass is changed accordingly so that the phase matching condition is maintained.

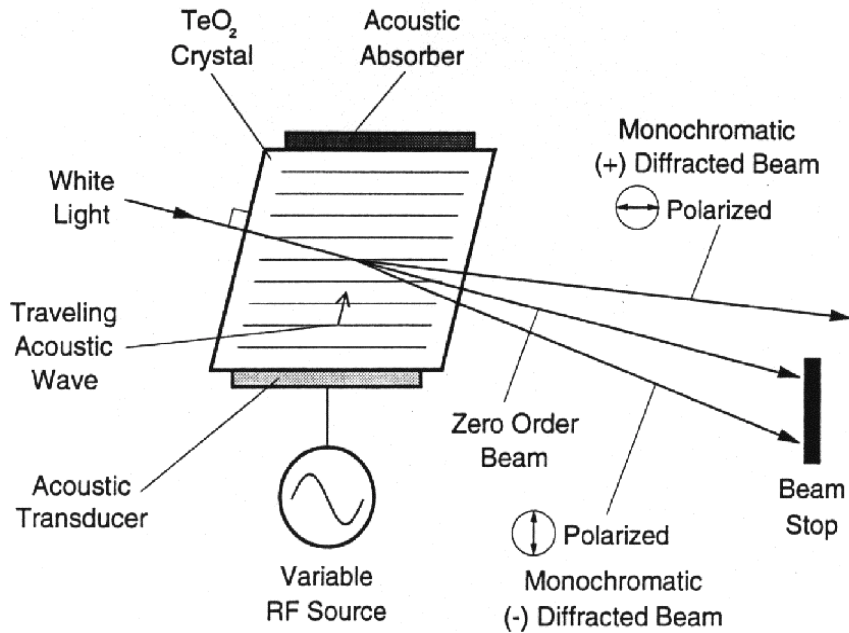


Figure 1. Schematic of the AOTF



The near infrared region of the spectrum extends from 800nm to 2500nm. The absorption bands that are most prominent in this region are due to overtones and combinations of the fundamental vibrations active in the mid-infrared region. The energy transitions are between the ground state and the second or third excited vibrational states. Because higher energy transitions are successively less likely to occur, each overtone is successively weaker in intensity. Since the energy required to reach the second or third excited state is approximately twice or three times that needed for a first order transition, the absorption bands occur at about one-half and one-third the wavelength of the fundamental. In addition to the simple overtones, combination bands also occur. These usually involve a stretch plus one or more bending or rocking modes. Many different combinations are possible and therefore the NIR region is complex, with many bands partially overlapping each other.

Near Infrared Spectroscopy is currently being used as a quantitative tool which relies on chemometrics to develop calibrations relating a reference analysis of the constituent to that of the NIR optical spectrum. The mathematical treatment of NIR data includes Multi- Linear Regression (MLR), Principle Component Analysis (PCA), Principle Component Regression (PCR), Partial Least Squares (PLS) and discriminant analysis. All of these algorithms can be used singularly or in combination to yield the resultant goal of quantitative prediction and qualitative description of the constituents of interest.



2. Methodology

Brimrose was supplied with 4 populations of soybean seeds, individually packaged with a protein and oil percentage written on each package. The spectrometer used for collecting spectra was the Brimrose Luminar 3076 Seed Meister. The wavelength range was set between 900nm and 1400nm with a resolution of 2nm. Fifteen seeds from each population, for a total of 60, were scanned. Spectra were collected in transmission mode and processed to absorbance. No other special processing or treatment of the data was performed. The absorbance spectra of the calibration set can be seen below in Fig. 2.

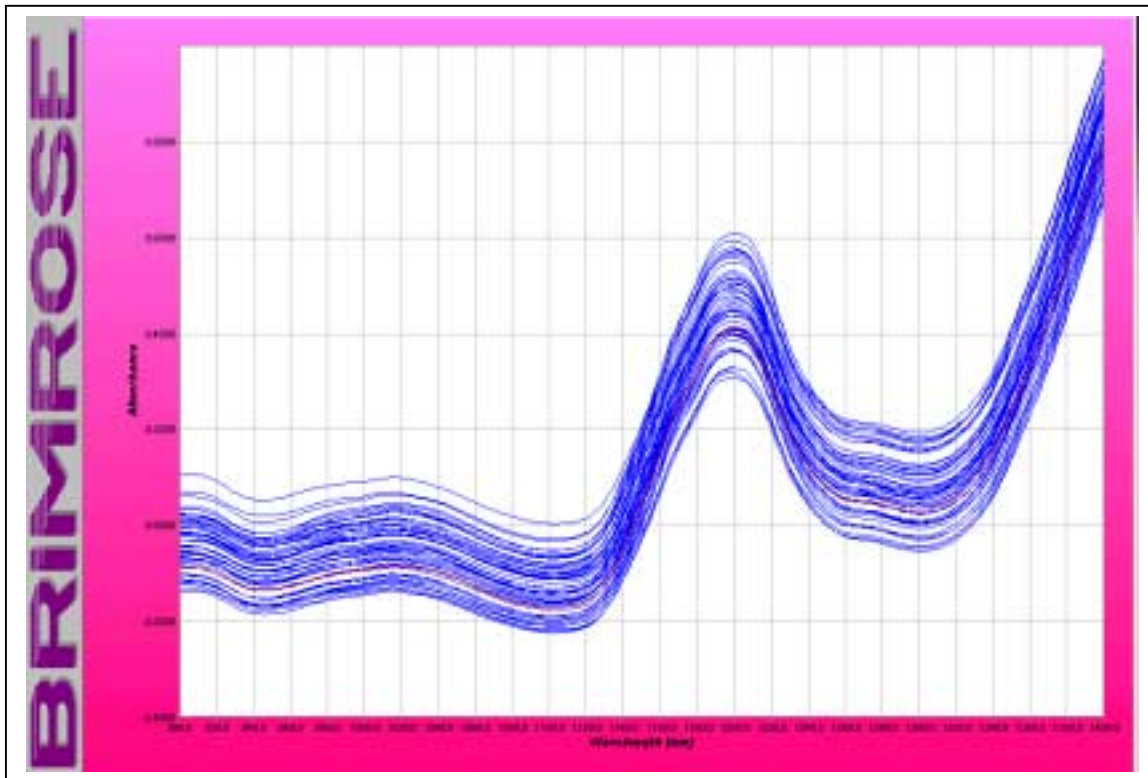


Figure 2. Absorbance spectra of sunflower seeds



What is not easily seen by analyzing the absorbance spectra are minute differences due to the changing concentrations of the various constituents in the seeds. These differences are more easily seen in a second derivative of the spectra. Figure 3 seen below is an expanded view of the wavelength region between 1100nm and 1250nm which is a strong absorbing region for different C-H stretch and C=C bonds all of which are found in oil, lipids and protein.

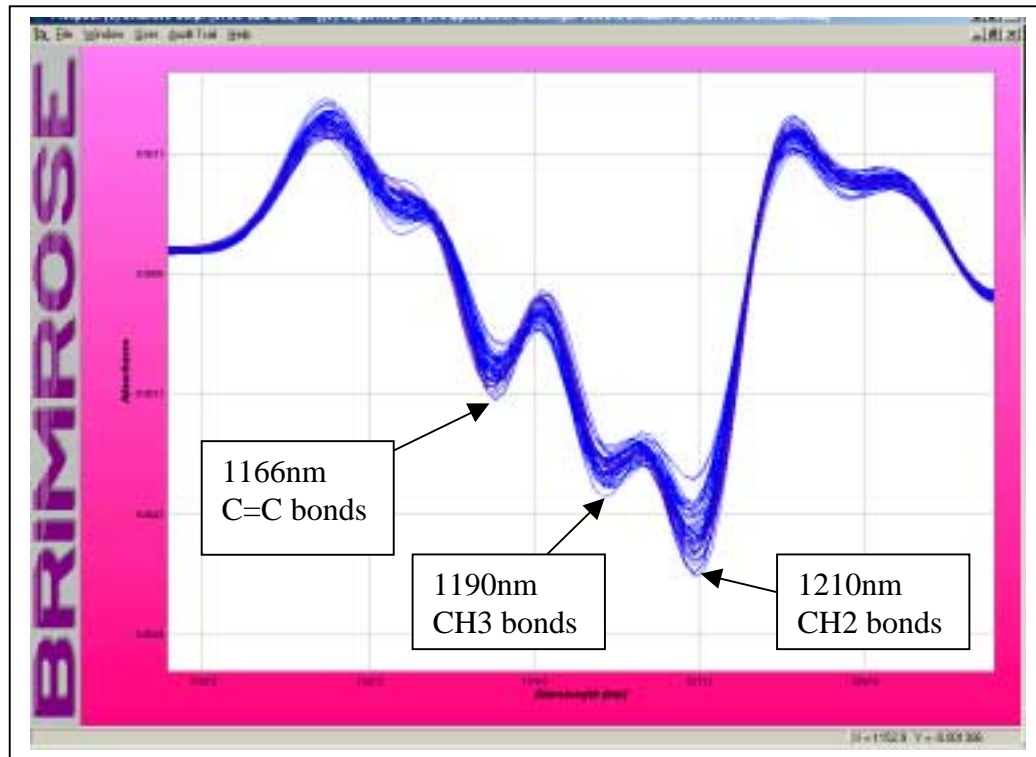


Figure 3. 2nd derivative of absorbance spectra showing specific areas of absorbance.



3. Regression Analysis

3a Regression results on percent protein

The software used for performing regression analysis was Unscrambler, available commercially from CAMO. A Partial Least Squares with a full cross validation regression analysis was calculated. The correlation with 2 samples removed as outliers, is 0.97 and 0.96 for the calibration and validation respectively. The Standard Error of Prediction (SEP) which is calculated as the standard deviation of the difference between predicted and measured, is 1.00 and is given in units of protein.

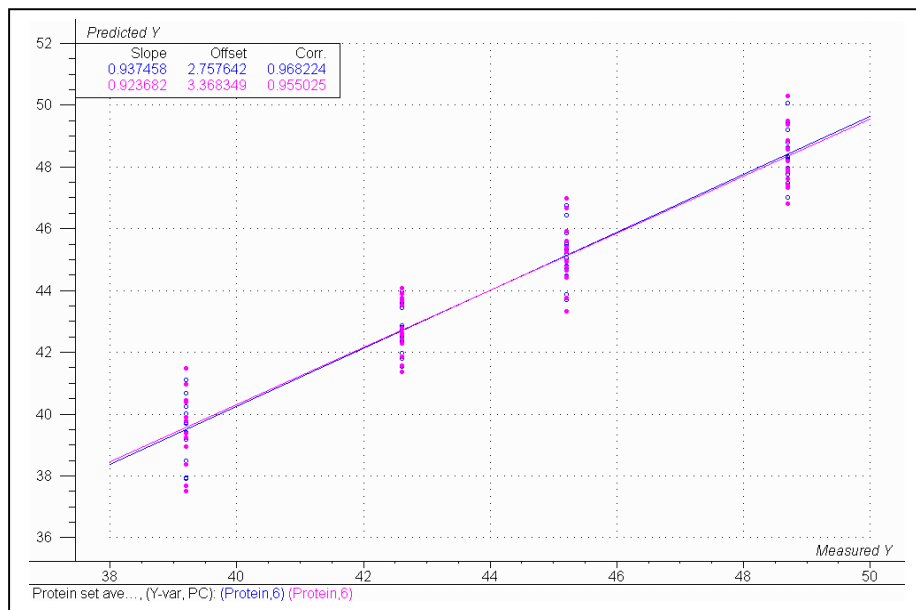


Figure 4. Regression for protein on 60 soybean samples. SEP equals 1.0

3b Regression results on percent oil

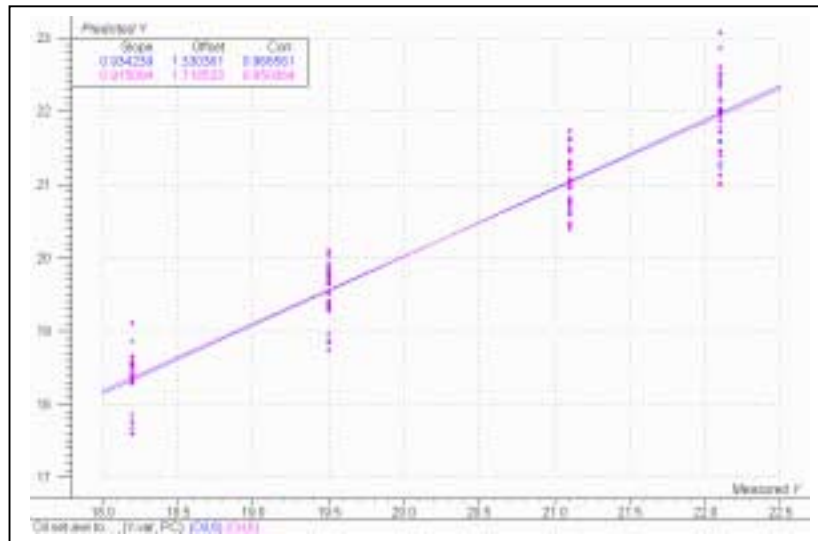


Figure 5. Regression for oil on 60 soybean samples. SEP equals 0.45

The regression on oil gave similar results as the protein. With the same 2 outliers removed from the data set the correlation is 0.97 and 0.95 for the calibration and validation respectively. The SEP is 0.46 in units of percent oil.



4. Predictions

Since there was not a population of seeds identified for the purpose of validating the model a simple technique of randomly excluding a sample population of seeds from the calibration set was used. In all 14 spectra were excluded from the total of 60. A regression model was then calculated on the 46 remaining samples and used to predict the excluded 14. The results are shown in Figures 6 and 7 for protein and oil respectively.

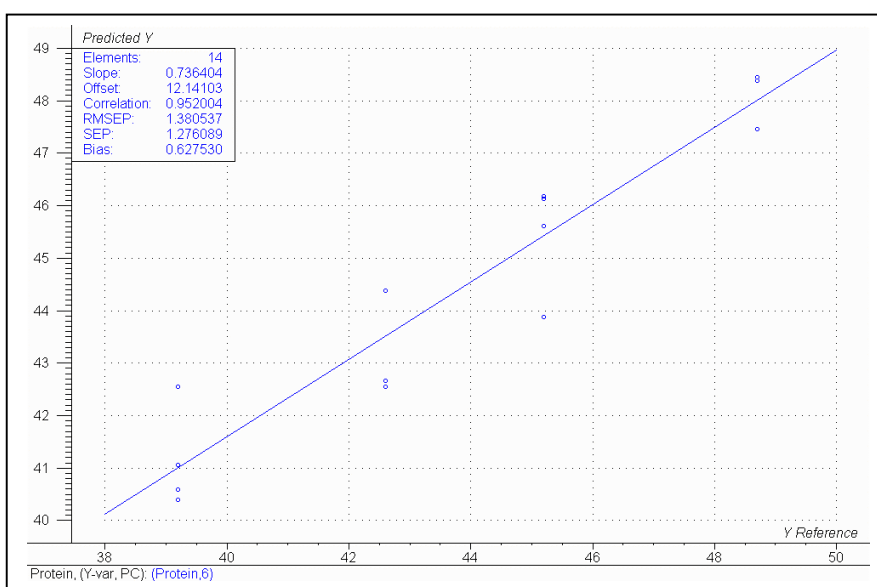


Figure 6. Predicted versus measured for protein on sub set of 14 samples.

The correlation between predicted versus measured for protein is 0.95 and the standard error of prediction (SEP) given in percent units of protein is 1.28.



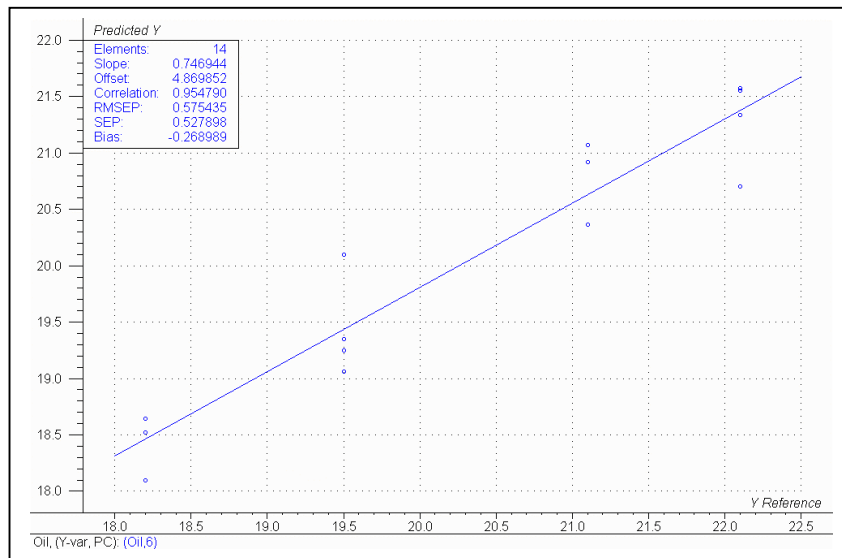


Figure 7. Predicted versus measured for oil on sub set of 14 samples.

The correlation between predicted versus measured for oil is also 0.95 and the standard error of prediction given in percent units of oil is 0.53.

5. Discussion and Conclusion

The result of this feasibility was excellent given the method by which it was performed. The ideal method for creating a calibration set is to identify a population of seeds that would represent the entire seeds of interest. By this we mean different varieties, seasonal changes, seeds grown in different regions etc. Spectra would then be collected on each individual seed and lab analysis conducted to determine the percent protein and total oil for each seed. In some cases averages can be done on 2 possibly 3 seeds. This technique creates over time a robust model that will predict accurately subsequent populations of seeds. The study reported herein included only 4 populations of seeds on which the percent protein and total oil was determined by NIR prediction on a bulk sample from which they came.

With 60 seeds in the calibration set we were able to get excellent correlation of 0.97 and 0.95 for the calibration and validation for both protein and total oil respectively. The SEP for protein with a range of 39.2 to 48.7 was very good at 1.00 and for oil with a range of 18.2 to 22.0 the SEP was 0.46. The correlation between predicted versus measured for both protein and total oil for the randomly selected validation set of 14 samples was 0.95. This is consistent with that calculated for the validation set of the full cross validation on the 60 samples, (see Figures 4 and 5). The SEP on the sub set of 14 for protein was only slightly higher at 1.28 and likewise for total oil at 0.53. However, this is to be expected since the calibration set was reduced by 14 samples resulting in a less robust model.

Based on the result of this study, it is our conclusion that soybean seed can be accurately sorted for high protein and oil using the Brimrose Seed Meister. If for example high protein is the primary interest then the seed meister can be configured such that only those seeds predicted within a certain range would be shunted down a given chute. All other seeds are shunted to other chutes depending upon the configuration. There are 6 chutes on the seed meister, 5 that collect accurately scanned and predicted seeds, and therefore many different sorting configurations can be arranged. The breeder has total control over this parameter. As



many as 5 different calibration models can be applied to the spectra and all predictions are written to a text file.

The Seed Meister is a solid-state instrument built for continued use especially during months when breeders are sorting harvested seeds from trial plots. It is very low maintenance requiring only periodic lamp changes, which can be done by the user. Adjustments to the calibration are not necessary since the lamp is pre-aligned at the factory.

