

**Determination of Percent
Oleic and Linoleic Acid
In Sunflower Seeds
Using AOTF-NIR Spectroscopy**

Executive Summary

During the past few years there has been an increased interest to breed sunflower seeds for high oleic acid. Once the inbred lines, often referred to as parent lines, have been chosen and the first generation cross is made the objective of the breeder is then to identify in the progeny those seeds with high oleic content. Once determined he can then make the second generation back cross and once again work to identify those seeds with high oleic content. A current method for determining oleic acid is to perform GC analysis on an aliquot of the embryo of the shelled seed. That part of the embryo not destroyed for GC analysis can still be planted and will, in most cases, germinate. However, without the protection of the shell the embryo will soon undergo chemical changes, which render it useless. A non-destructive method by which oleic acid could be determined quickly in large seed populations and still allow for them to be preserved by virtue of their natural shell would be most useful.

The study reported herein shows that it is possible to measure the percent oleic acid as well as other constituents in sunflower seeds without first shelling the seed. The instrument used was the automated Brimrose Luminar 3076 Seed Meister. Seeds are first placed on a small moving conveyor belt, which moves the seed through the NIR beam. A spectrum is collected to which a PLS1 model is applied and prediction for percent oleic acid is made. Seeds are automatically sorted down corresponding chutes based upon the predicted value.



Brimrose Corporation of America
19 Loveton Circle
Sparks, MD 21152-9201 USA
Phone: +1 410 472-7070
Fax: +1 410 472-7960
E-Mail: process@brimrose.com
Web: <http://www.brimrose.com>

BRIMROSE



I. 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.

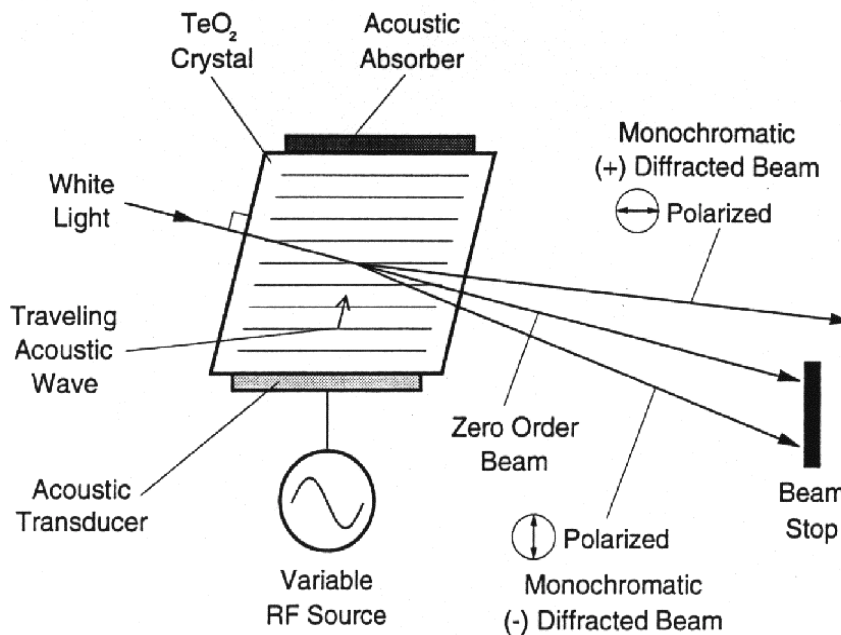


Figure 1. Schematic of the AOTF



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.



II. Methodology

Brimrose was supplied with 200 sunflower seeds, individually packaged in small key envelopes, each of which was labeled with an identifying number. The 200 seeds made up the calibration set on which the Seed Meister was to be calibrated. Also, supplied were 4 additional packets, each containing 20 seeds, to be used for validation. The spectrometer used for scanning and collecting spectral data on the seeds was the Brimrose Luminar 3076 Seed Meister. The wavelength range was set between 900nm and 1400nm with a resolution of 2nm. Before collecting spectra of the seeds, the small conveyor belt of the analyzer on which the seeds are placed was calibrated to a speed, which enabled approximately 40 scans to be collected on each seed as it passed through the NIR beam. The total number of scans collected on each seed are then averaged by the software program into a spectrum. Seeds were manually placed on the conveyor belt. With the exception of 2 seeds, the entire population was scanned twice. For the first repetition the seeds were oriented such that the tip entered the NIR light first and the crown was oriented first for the second repetition. In both repetitions, the side of the seed facing up was random. The seeds were packaged and then returned to a lab for GC analysis. A period of 2 weeks past before the GC analysis was obtained and made available to Brimrose.

After receiving the data a regression model was built which was then enabled in the Brimrose application, Seedm.exe. Before scanning, the seeds of the 4 validation sets were first separated into individual key envelopes and labeled. The 4 populations were then scanned through the Seed Meister 3 times. As each seed was scanned the model was applied to the spectrum and the value predicted was then automatically written to a log file that had previously been enabled in the Seedm.exe application. After the scanning process was complete the seeds were returned to the lab for GC analysis to which NIR measurement would be compared.



III. Results

Spectra were collected in transmission mode and processed to absorbance. No other special processing or treatment of the spectra was performed. The absorbance spectra of the calibration set can be seen below in Fig. 2.

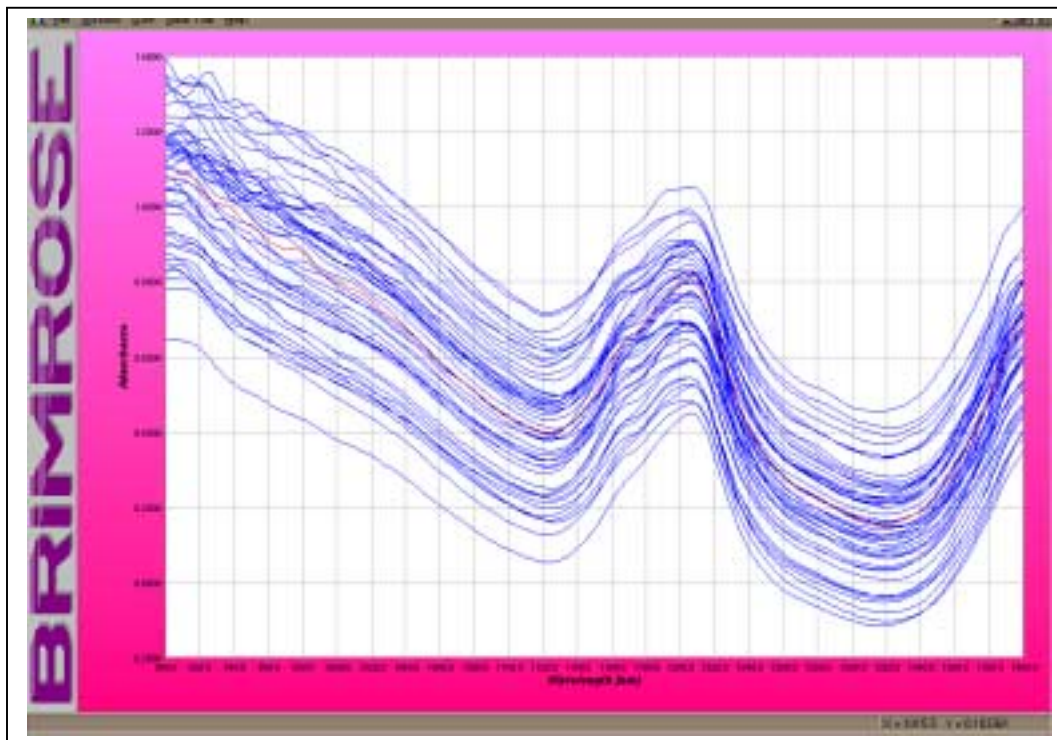


Figure 2. Absorbance spectra of sunflower seeds

IV. Regression Analysis

The spectral data along with the GC lab analysis for oleic, linoleic, palmitic and stearic acid was imported into the chemometric software package Unscrambler, available commercially from CAMO. Calibration models were built using the Partial Least Squares (PLS) method. The initial regression on oleic acid, which is the primary constituent of interest, requiring only



7 Principle Components, looks very good. Without removing any potential outliers the correlation for the calibration and the validation is 0.94 and 0.93 respectively, (see Figure 3). Seed samples SD080 and SD123 both showed to have high influence on the model and could be considered outliers. Removing the 2 samples slightly improved the model requiring only 6 Principle Components and reducing the SEP slightly from 10.1 to 9.9, (see Figure 4).

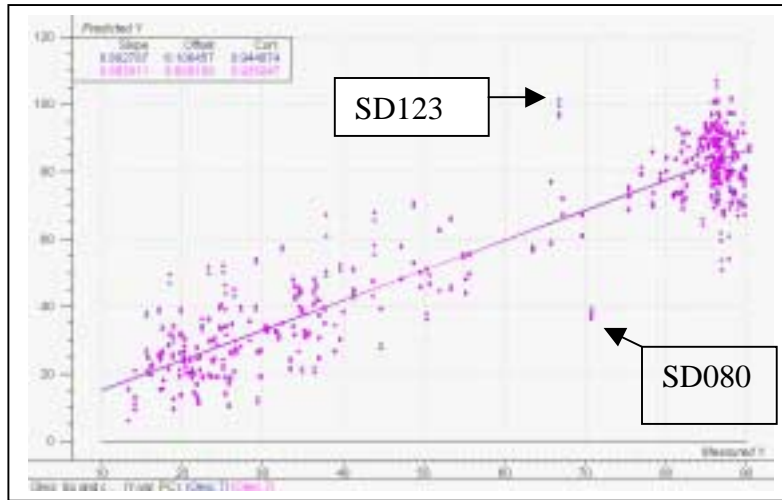


Figure 3. Regression with zero outliers removed, SEP equal to 10.1.

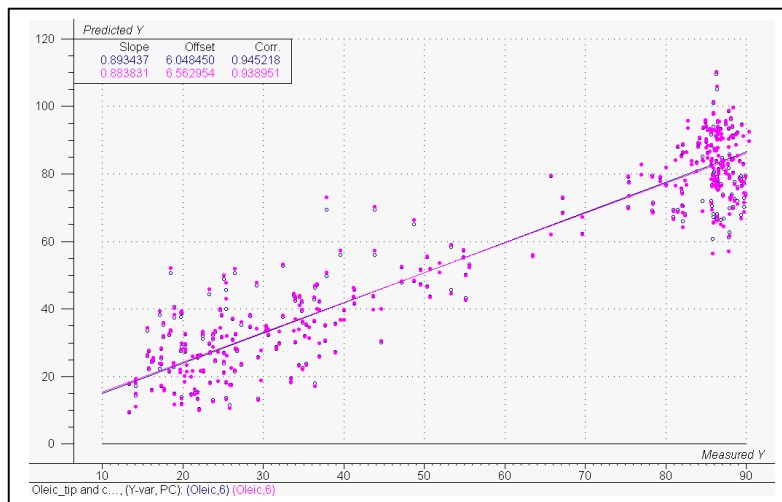


Figure 4. Regression with outliers removed, SEP equal to 9.9.



The X-loading weights indicate those regions of the spectra that contribute more strongly to the model. According to reference data, found in “The Atlas of Near Infrared Spectra” by Sadler, oleic acid creates absorbing peaks between approximately 920nm and 930nm, a shoulder approximately between 1150nm and 1160nm and a large peak slightly after 1200nm. As can be seen in Figure 5, the X-loadings for the first Principle Component, which explains 93 percent of the variation, are high in precisely those regions indicated above.

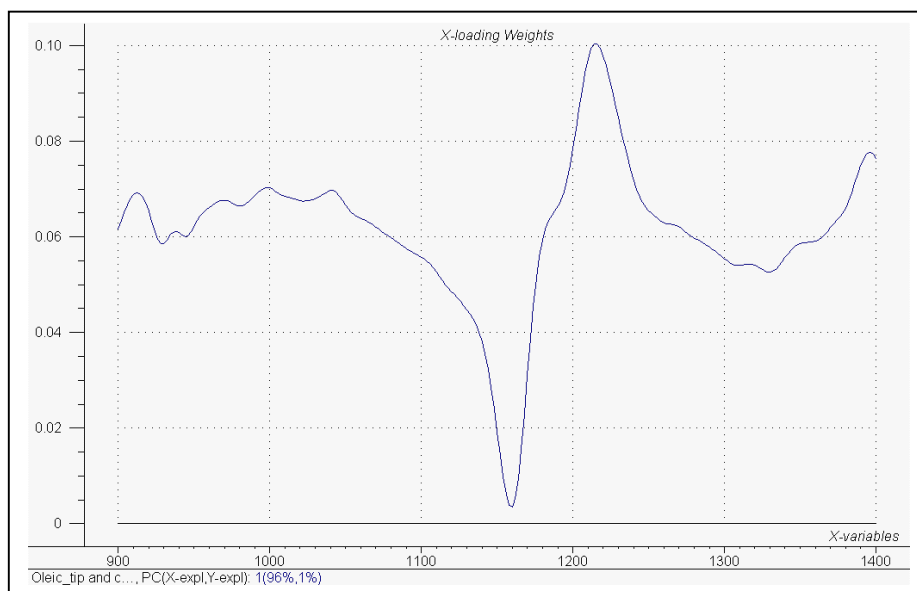


Figure 5. X-loading for the first Principle Component for oleic acid, explaining 93% of spectral variation.

The regression coefficients which summarize the relationship between all predictors (spectral data) and a given response, in this case oleic acid are also high in the corresponding spectral regions. Figures 6 and 7 below shows the regression coefficients for the PCs 3 and 4 that cumulatively describe 74 percent of the “Y” response values, showing peaks quite clearly in the areas of 1160nm (CH₂ bonds) and 1220nm (CH bonds). Less pronounced but still significant is the slight peak in the 920 region. The model for Linoleic was similar to that for Oleic, (see Figure 8) and as expected resulted in an inverse scenario regarding the X-loading and the regression coefficients, (see Figures 9 and 10).



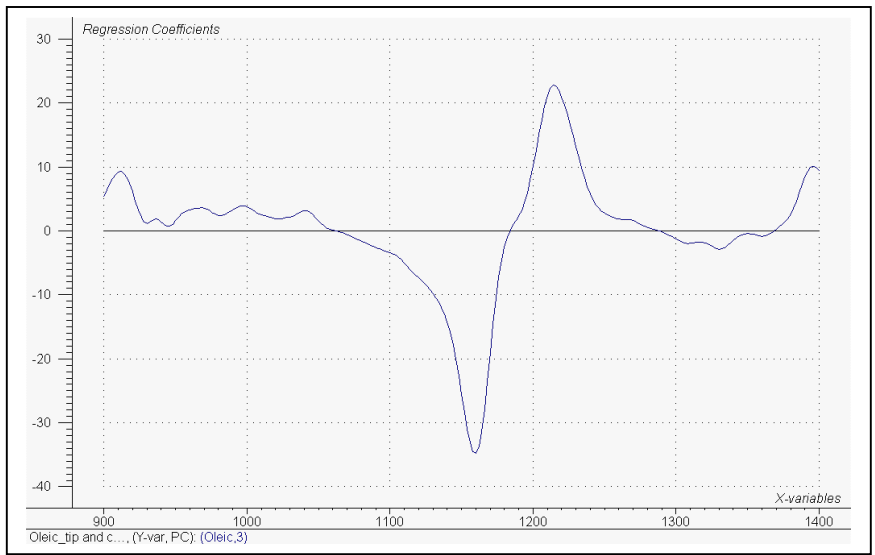


Figure 6. Regression Coefficients for oleic acid for the 3rd PC explaining 21% of the Y variation.

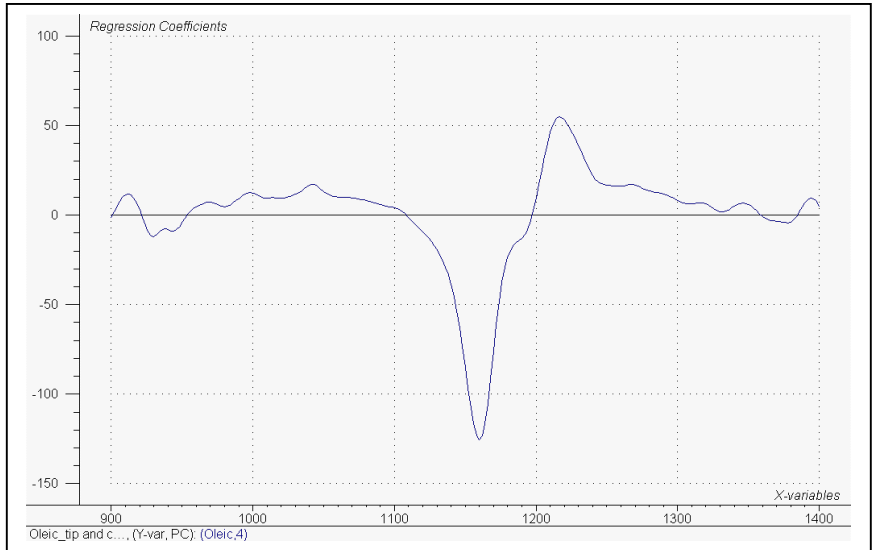


Figure 7. Regression Coefficients for oleic acid for 4th PC explaining 53% of the Y variation.



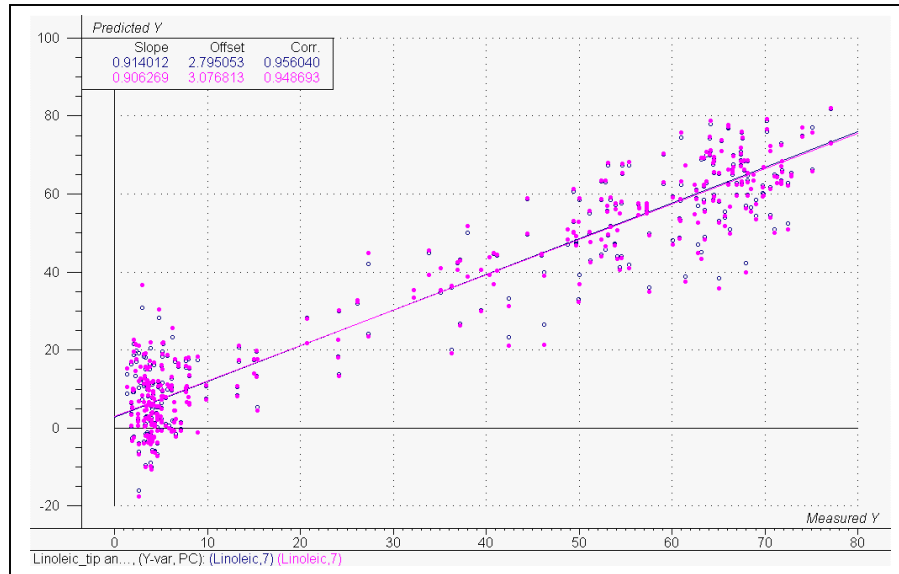


Figure 8. Regression on Linoleic with 2 outliers removed. SEP equal to 8.8.

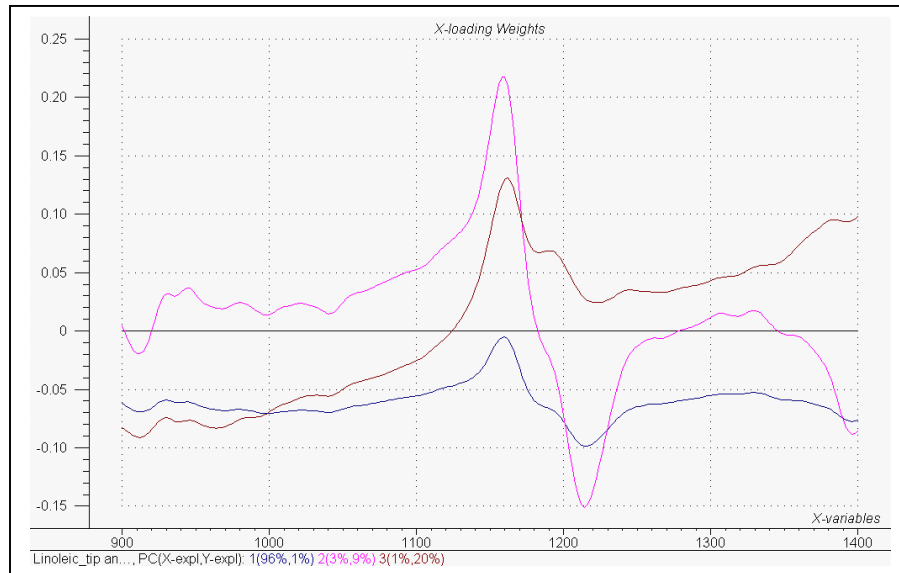


Figure 9. X-loading for PCs 1,2 and 3 for regression on Linoleic.



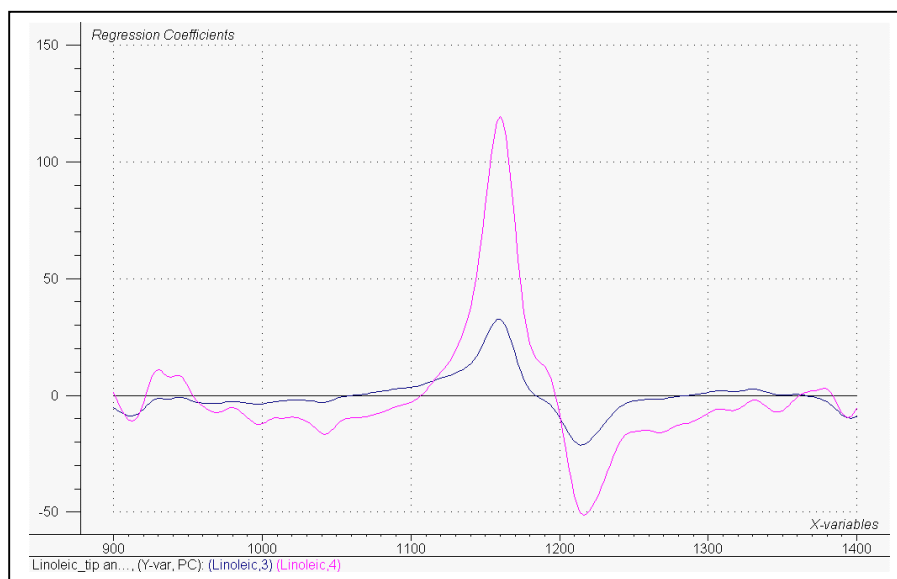


Figure 10. Regression coefficients for PCs 3 and 4 explaining 73% of the Y variation.

V. Predictions

Each of the 4 packets of seeds was scanned 3 times with no attention given to the orientation of the seed on the conveyor belt. The prediction results of the 3 repetitions were then averaged. In an effort to reduce the amount of GC analysis only 9 of the possible 20 seeds were selected for analysis. The selection was based on the NIR predictions, i.e. 3 seeds from both the extreme high and low range and 3 seeds from the middle range. The results are given in Tables 1 through 4. Corresponding regression results and statistics are given in Figures 11 through 14 respectively. Since the results on linoleic were similar only the results on oleic acid will be given and discussed.



Population 104-2		
	Oleic Acid	
Sample	GC Results	NIR Ave. of 3 Rep
104-2_01	23.65	11.17
104-2_03	19.96	18.17
104-2_05	85.61	78.52
104-2_06	89.12	85.25
104-2_07	72.25	63.29
104-2_09	18.57	16.04
104-2_14	86.76	78.91
104-2_16	64.30	55.07
104-2_19	88.65	77.64
Total Average	60.99	53.78

Table 1. GC analysis and the average of 3 repetitions of NIR for selected samples of 104-2.

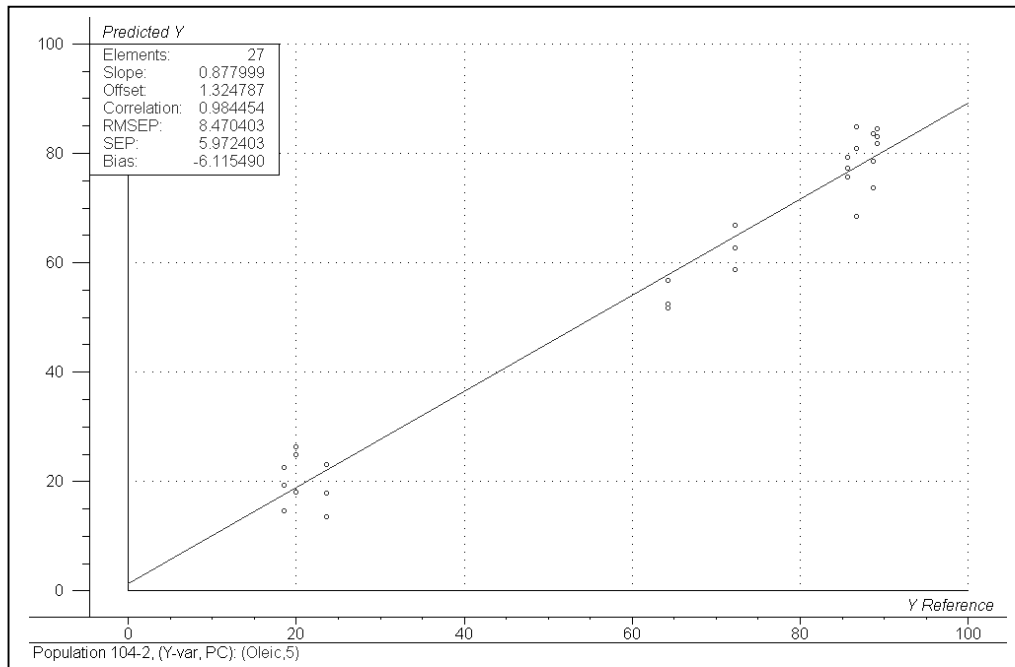


Figure 11. Line fit and statistics for selected samples of population 104-2



Population 121-4		
	Oleic Acid	
Sample	GC Results	NIR Ave. of 3 Rep
121-4_02	82.42	78.05
121-4_04	89.55	82.38
121-4_05	89.28	84.32
121-4_06	88.06	90.90
121-4_10	21.16	17.92
121-4_14	88.90	90.49
121-4_15	80.08	65.93
121-4_17	25.36	28.78
121-4_19	17.40	17.27
Total Average	64.69	61.78

Table 2. GC analysis and the average of 3 repetitions of NIR for selected samples of 121-4.

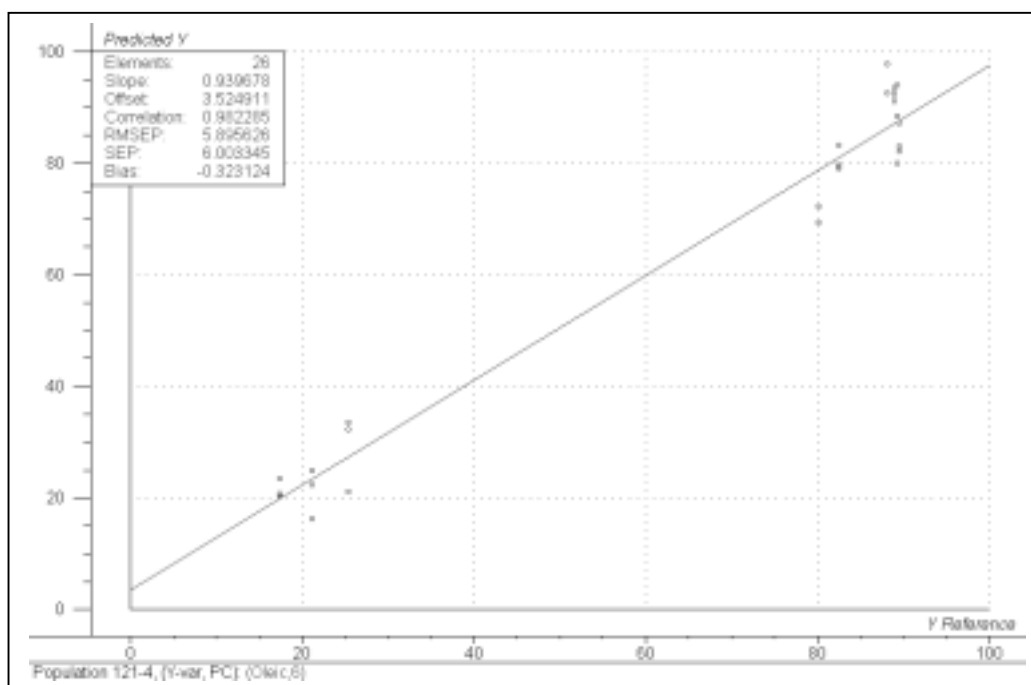


Figure 12. Line fit and statistics for selected samples of population 121-4



Population 174-4		
	Oleic Acid	
Sample	GC Results	NIR Ave. of 3 Rep
174-4_01	84.14	90.46
174-4_02	20.63	24.26
174-4_04	85.03	87.94
174-4_07	18.62	29.07
174-4_09	83.77	81.69
174-4_10	17.72	27.87
174-4_16	86.21	84.11
174-4_17	85.90	86.68
174-4_19	17.21	25.15
Total Average	55.47	59.69

Table 3. GC analysis and the average of 3 repetitions of NIR for selected samples of 174-4

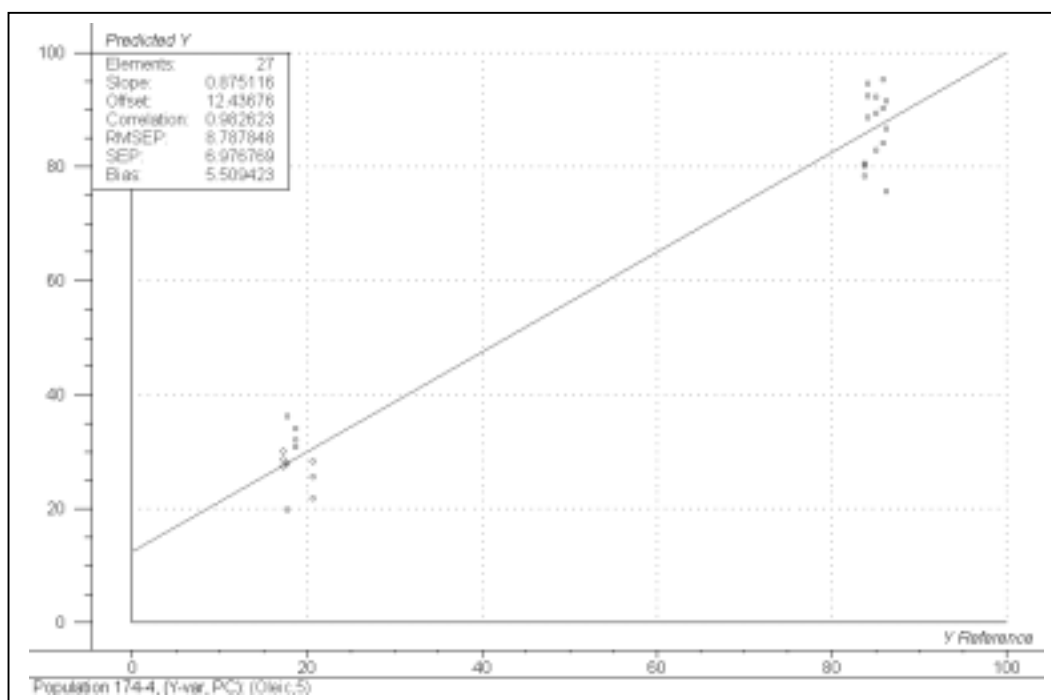


Figure 13. Line fit and statistics for selected samples of population 174-4



Population 187-2		
	Oleic Acid	
Sample	GC Results	NIR Ave. of 3 Rep
187-2_01	67.31	52.55
187-2_03	81.35	74.36
187-2_05	78.35	73.53
187-2_08	84.61	79.53
187-2_09	23.95	13.36
187-2_10	82.45	76.38
187-2_11	21.29	17.91
187-2_12	18.57	11.99
187-2_15	85.77	77.24
Total Average	60.41	52.98

Table 4. GC analysis and the average of 3 repetitions of NIR for selected samples of 187-2

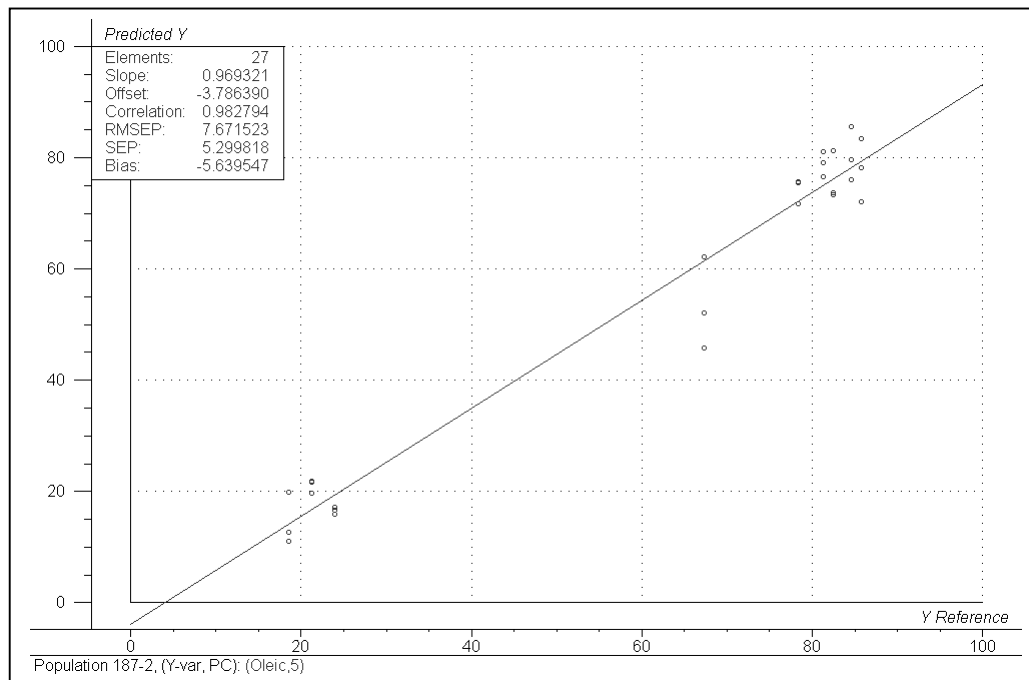


Figure 14. Line fit and statistics for selected samples of population 174-4



In addition to the 4 population identified exclusively for validation, a 5th validation population was configured. From the total sample population that now included the 4 populations addressed above, samples were randomly selected from within the entire range of 13.33 – 90.39 oleic acid. The samples were excluded from those samples used for building the calibration model. The model built was then used to predict those excluded samples. The results can be seen below in Figure 15.

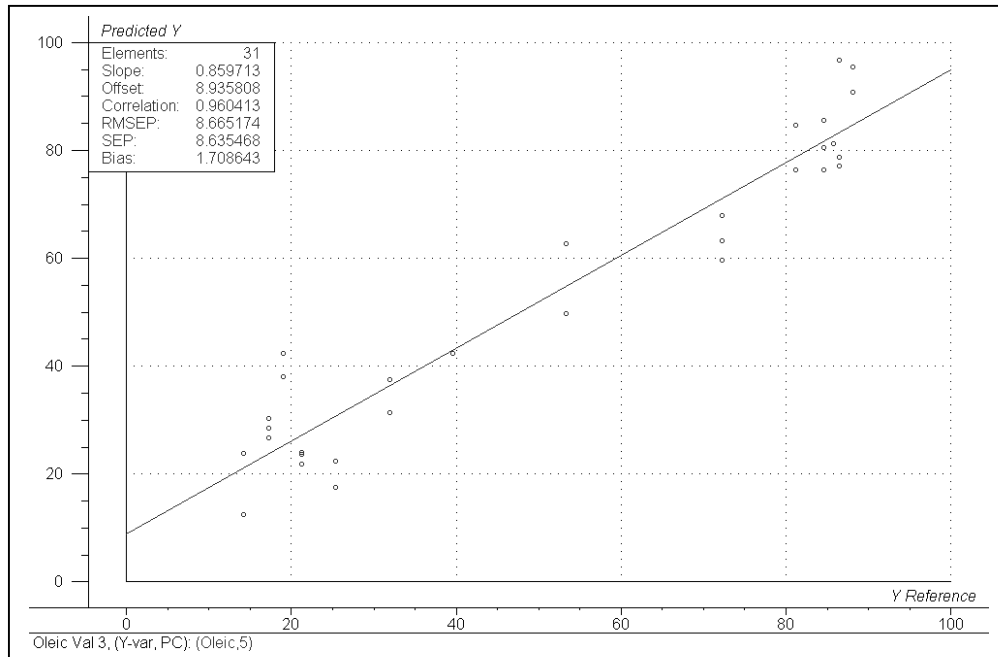


Figure 15. Fifth validation set randomly selected from total sample population.



VI. Discussion and Conclusion

Very good correlation of 0.94 and 0.93 for the calibration and validation respectively was achieved in the PLS1 cross validation regression analysis of the calibration population consisting of 198 seeds. The SEP with only 2 samples removed was 9.9 and is given in actual units of oleic acid with the range being from 13.33 to 90.39 percent. The high SEP is largely due to the shell of the seed causing a degree of noise in the spectral data. However, the gain is much greater than the loss because we are able to sort the seeds quickly and obtain a rather accurate measurement without damaging the seed, which can now be stored under normal conditions until it is to be planted. Under conventional methods the seed is carefully cut open and a portion of the embryo cut off for GC analysis. The seed must soon thereafter be planted in the greenhouse, which is less favorable than planting in the field.

Regarding the four individual validation populations, based on the high correlation between predicted and measured it should be safe to assume that had GC analysis been performed on all samples similar results would have been obtain. The correlation for predicted versus measured for populations 104-2, 121-4, 174-4 and 187-2 is 0.96, 0.98, 0.98 and 0.96 respectively while the SEP is 5.97, 6.00, 6.98, and 5.29 respectively. The results give us a rather high degree of confidence that with subsequent populations with the same similarities as those on which the calibration was built, the same accuracy in prediction and sorting can be achieved.

The 5th validation set that was randomly selected is very indicative of what can be expected with prediction made on subsequent populations of sunflower seeds using a model built from the existing sample population of 233 samples. The correlation was still excellent at 0.96 and the SEP was only slightly higher than that obtained on any of the 4 validation sets. We point out to the reader that at 8.64 the SEP calculated for the 5th validation set is lower than the 9.9 calculated for the regression model built using the total sample population.



Even though the regression analysis was quite good as well as prediction results on the validation sets, even better results can be expected. The argument for this statement is evident after evaluating the frequency distribution of the total sample population, (see Figure 16 below). The calibration set is extremely right skewed with very little representation of samples between 40% and 75% oleic acid. This condition means that the prediction made on any seed with an oleic acid content between 40% and 75% is going to be an interpolation. With continued efforts to add samples to the calibration set, filling in the lightly represented area, a very robust model will definitely be achieved. The resulting model will accurately predict percent oleic acid in subsequent seed populations.

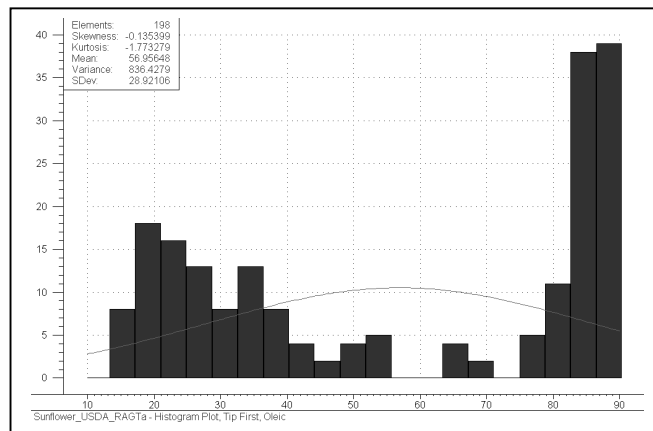


Figure 16. Frequency distribution of sample population.

With this condition met the Seed Meister will definitely prove to be an invaluable tool for the breeder of sunflower seeds. With the capability to non-destructively sort out, with a very high degree of accuracy and repeatability, those seeds with the value added traits the breeder can save both time and money.

