Analysis of Seed Protein and Oil from Soybean Northern and Southern Region Uniform Tests

R. W. Yaklich,* B. Vinyard, M. Camp, and S. Douglass

ABSTRACT

Protein and oil concentration are seed constituents that make soybean [Glycine max (L.) Merr.] an important crop. Reports have been published that the oil concentration and recently the protein concentration of U.S. soybean has been decreasing with respect to other exporting countries. The objective of this study was to utilize 51 yr of two data sets, The Uniform Soybean Tests, Northern Region and Southern Region to determine long-term trends in these constituents. Oil and protein composition data of lines from Regional Summaries of Maturity Groups OO through VIII were used for the analysis. Mean oil concentration was higher in the Southern Region compared with the Northern Region (209 and 206 g kg⁻¹, respectively), as was protein concentration (411 and 407 g kg⁻¹, respectively). Maturity group means for oil and protein concentration ranged from 198 to 212 g kg⁻¹ and 404 to 414 g kg⁻¹, respectively. Oil concentration increased in both data sets to 1973 and abruptly decreased in 1974 (21 g kg⁻¹) to a low in 1982 (184 g kg⁻¹) to higher than average values in 1983 (217 g kg⁻¹) and by linear regression analysis, decreasing in the Northern Region since then. Analysis of lines evaluated from 1973 to 1974 and from 1982 to 1983 indicated that genetics was not the cause of the change in oil concentrations. Temporal protein concentrations were not as distinct as those of oil but indicated groups of years when protein concentration was above or below the mean. The data indicate that oil concentrations of soybean has been erratic and decreasing since 1974 because of unknown factors.

A number of reports claim that the oil and protein concentration of soybean produced in North America has been changing compared with other countries that produce soybean for export (Breene et al., 1988; Hurburgh and Brumm, 1990; Hurburgh et al., 1987, 1990). In a cooperative study with three soybean crushing companies, Breene et al. (1988) noted a trend toward lower protein concentrations in processed soybean grown in northern versus southern locations. They found that in the years 1983 to 1986 that the protein concentration of processed soybean in Minnesota was lower than soybean processed in Indiana and Illinois. They also analyzed 11 yr of data from the Uniform Soybean Tests (UTs), Northern Region (NR) and found a significant effect of the year of production on oil concentration but not on protein concentration.

Hurburgh et al. (1987) studied soybean samples from grain elevators in four states (Iowa, Illinois, Minnesota, Ohio) for protein and oil composition over 2 yr. They found oil concentration to be higher in the hotter, dryer year with the highest soybean oil samples coming from the area with the most severe drought. These environmental factors also affected the total protein and oil concentration of the seed because the decrease in protein was greater than the increase in oil in the hot dryer regions. In a follow up study, Hurburgh and Brumm (1990) observed that differences in protein and oil concentration of soybean differed consistently at some elevator locations from year to year. They found that the standard deviation of protein and oil concentration in soybean deliveries to grain elevators was 1.0 and 0.5 percentage points, respectively. In the 3 yr of the study, 15% of the samples were above average for both protein and oil concentration and about 15% were below average for both constituents. The remaining 70% were divided about evenly between above average for either protein or oil and below average for the other component.

Hurburgh et al. (1990) cited data for the years 1972 to 1988 that claimed that exported U.S. No. 2 soybean contained about 1.0 to 1.5% less oil than Brazilian soybean and that protein concentration had also fallen behind that of Brazil. In a larger 3-yr study (Hurburgh et al., 1990), soybean samples were received from growers in 573 U.S. counties that produce more than one million bushels of soybean and represent about 80% of total U.S. production. A consistent pattern over 3 y of the study revealed that soybean grown in the Corn Belt states were generally lower in protein and higher in oil concentration than the East Coast and southern soybean growing regions. Also, the protein and oil concentration was more variable between years in the western Corn Belt, which the authors attributed to a greater variability of weather in these states.

Green and Williams (1988) used the data from the UTs to determine protein and oil patterns in soybean evaluated at various locations for the years 1957 to 1986. They used the average values for each line evaluated at a location to calculate means for yield and protein and oil concentration to summarize the data for state, regional, and national levels. They found that the overall U.S. mean protein concentration remained nearly constant in this period. However, soybean grown in the Southern Region (SR) had increased protein levels, while Corn Belt soybean seed protein levels fell about 1.0% relative to the national average.

The soybean UTs, aside from their functional utility for providing a mechanism for the evaluation and release of soybean lines, are also a chronology of the agronomic improvement of soybean in the USA and...
Canada. Previous studies have utilized a few locations and years to answer regional questions (Breeene et al., 1988) or utilized yearly means of the lines evaluated at all of the locations (Green and Williams, 1988) as a method to look at trends in yield and protein and oil concentration. The objective of this study was to utilize 51 yr of two data sets, The Uniform Soybean Tests, Northern Region and Southern Region to determine long-term trends in protein and oil concentration.

**MATERIALS AND METHODS**

The NR and SR UTs critically evaluate the best of the experimental soybean lines developed by federal and state research personnel in the USA and Canada. Lines are adapted to Maturity Groups (MGs) O through VIII. The soybean lines evaluated in the UTs were selected from Preliminary Tests that evaluated all submitted lines at a limited number of locations for one year. Promising new lines from the Preliminary tests were evaluated in the UTs and grown at more locations. Lines evaluated in the UTs are grown in replicated field plots and evaluated for yield, maturity, lodging, height, seed size, seed composition, shattering, iron chlorosis, emergence, and disease susceptibility. These evaluations were conducted by cooperators in the states of Alabama, Arkansas, Delaware, Florida, Georgia, Iowa, Illinois, Indiana, Kansas, Kentucky, Louisiana, Maryland, Michigan, Minnesota, Missouri, Mississippi, Nebraska, North Carolina, New Jersey, North Dakota, Ohio, Oklahoma, Pennsylvania, South Dakota, South Carolina, Tennessee, Texas, Virginia, Wisconsin, and the Canadian provinces of Manitoba, Ontario, and Quebec. Results from these tests are assembled by the cooperators and sent to a central location, edited by one of the cooperators and published. Minor changes in the name of the published results have occurred since their inception.


Protein and oil concentration were determined from a composite sample from all replications of the line at a location and listed on a moisture-free basis. The method of analysis for these constituents changed as faster and less expensive methods became available. In 1974, the method of analysis for protein and oil concentration changed from the Kjeldahl method and nuclear magnetic resonance, respectively, to the use of infrared reflectance. In 1993, the method of analysis for these constituents changed to near infrared transmittance. The chemical analyses were performed at the National Center for Agricultural Utilization Research (changes in the name have occurred). USDA-ARS, Peoria, IL. Protein and oil were reported on a percent-dry-weight basis and converted to SI units. Total protein and oil concentration were the sum of the individual entry values for protein and oil. Protein to oil ratio was derived by dividing the individual entry protein concentration by its oil concentration.

A one-way ANOVA with MG mean comparisons was conducted for protein, oil, total protein and oil concentration, and protein to oil ratio by means of the pdiff option of the LS means statement in SAS Proc Mixed (Littell et al., 1996). MG means were compared on the basis of among line and year variability. For each MG, linear and quadratic regressions were fit across years with data that were a mean of all lines grown each year. The overall regression models were fit using the NR and SR MG means separately and combined for each year. Regression models were also calculated for the years 1948 to 1973, 1974 to 1982, and 1983 to 1998 because graphs of the data indicated data trends for these groupings of years. Regressions were fit by SAS Proc Reg (SAS, 1996). Pearson linear correlation coefficients were calculated for all pairs of response variables for all MGs separately and combined.

**RESULTS AND DISCUSSION**

Mean oil concentration by MG ranged from 198 to 212 g kg\(^{-1}\) (Table 1) and ranged from 198 to 211 g kg\(^{-1}\) in the NR MGs and 207 to 212 g kg\(^{-1}\) in the SR MGs. Maturity Group IVS had the greatest mean oil concentration. The mean oil concentration in the NR and SR were 206 and 209 g kg\(^{-1}\), respectively. The mean oil concentration for the entire UTs was 207 g kg\(^{-1}\). The

<table>
<thead>
<tr>
<th>Maturity group</th>
<th>Oil concentration</th>
<th>Protein concentration</th>
<th>Total protein and oil</th>
<th>Protein to oil ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>O</td>
<td>199g±</td>
<td>404g</td>
<td>601g</td>
<td>2.04a</td>
</tr>
<tr>
<td>II</td>
<td>205g±</td>
<td>409g</td>
<td>614g</td>
<td>2.06a</td>
</tr>
<tr>
<td>III</td>
<td>210b±</td>
<td>406g</td>
<td>616g</td>
<td>1.92±</td>
</tr>
<tr>
<td>IV</td>
<td>210b±</td>
<td>410b</td>
<td>620b</td>
<td>1.95±</td>
</tr>
<tr>
<td>V</td>
<td>208c±</td>
<td>406d</td>
<td>615c</td>
<td>1.95±</td>
</tr>
<tr>
<td>VI</td>
<td>207d±</td>
<td>414a</td>
<td>620b</td>
<td>2.01b</td>
</tr>
<tr>
<td>VII</td>
<td>209e±</td>
<td>414a</td>
<td>623a</td>
<td>1.99c</td>
</tr>
<tr>
<td>VIII</td>
<td>209e±</td>
<td>414a</td>
<td>623a</td>
<td>1.99c</td>
</tr>
<tr>
<td>North region</td>
<td>206b±</td>
<td>407b</td>
<td>613b</td>
<td>1.98a</td>
</tr>
<tr>
<td>South region</td>
<td>209a±</td>
<td>409b</td>
<td>616b</td>
<td>1.97a</td>
</tr>
<tr>
<td>Uniform tests</td>
<td>207±</td>
<td>409</td>
<td>616</td>
<td>1.97±</td>
</tr>
</tbody>
</table>

† Means in the same column followed by the same letter are not significantly different under Fisher’s protected LSD (P = 0.05).
MG means demonstrate that there was a relationship between MG and oil concentration of soybean evaluated in the USA and Canada. The soybeans evaluated in MGs II-IVS contain the highest average oil concentration and represent the majority of the soybean acreage grown in the USA. These values were significantly different from the remaining MGs, but on average contained only 2 g kg$^{-1}$ more oil than most of the remaining SR MGs. These values confirm other reports (Hurburgh et al., 1990; Piper and Boote, 1999) that indicated that soybean from the northern production areas are higher in oil concentration and are bred to produce more oil. However, an exception is SR MG IVS that had the highest mean oil concentration. The remaining SR MGs (V–VIII) are similar to and in some cases higher in oil concentration than some of the NR MGs and also have a higher overall mean oil concentration than the NR MGs. The lower oil concentrations in MGs OO-I have been attributed to lower temperatures and a shorter growing season (Piper and Boote, 1999).

The mean yearly oil concentrations (Fig. 1) for the UTs were on an upward trend between 1948 and 1973. Between 1974 and 1982, mean oil concentration decreased to its lowest value. In 1983, the mean yearly oil concentration increased to an above average concentration and appears to be decreasing to 1998. In the NR, mean yearly oil concentration decreased below 200 g kg$^{-1}$ in 1950, 1979, 1981, 1982, and 1992, and in the SR in 1979, 1980, 1981, 1982, and 1997. In the NR, mean yearly oil concentration was greater than 220 g kg$^{-1}$ in 1973 and 1983 and in the SR in 1969 and 1973.

Yearly deviations from the overall mean oil concentration for the NR and SR (Fig. 2) indicates that prior to 1974, the yearly oil concentrations were generally greater than or equal to the mean. All of the yearly means in the SR were equal to or greater than the SR mean and only 7 yr when the yearly oil concentrations in the NR were below the NR mean. The 26 yr from 1948 to 1973 constitutes 26 yr of data from the NR and SR tests and represent 52 crop years of data. The oil concentration was equal to or greater than the mean oil concentration in 87% of the crop years to 1974. From 1974 to 1998 (25 yr) there were 14 crop years (28%) when the oil concentration was equal to or greater than the overall mean. These results indicate that something dramatic happened after 1973 that affected the oil concentration of the lines evaluated in the UTs.

Linear regression of oil concentration onto years showed that the mean oil concentration of the NR and SR has been significantly decreasing with $r^2$ values of 0.0018 and 0.1332, respectively. However, the $r^2$ values indicate there is large variability from the regression line. Quadratic regression did not improve the estimate of oil trend with time. Likewise, linear and quadratic regression of oil concentration onto year by MG gave small, near zero $r^2$ values. Analyzing the yearly means for oil concentration as three different time periods gave a better estimate of the temporal trend in oil concentration during the 51 yr. From 1948 through 1973, linear regression (Table 2) indicated that yearly oil concentrations were increasing in the NR, SR, and UTs. Linear
regression by MG showed that all of the MGs except MGs IVS and VII had significant probabilities ($r^2$ values) for this time period. From 1974 to 1982, linear regression (Table 2) showed that yearly oil concentrations were decreasing in the NR, SR, and UTs. Analysis by MG showed that MGs OO, II, and VI did not have significant probabilities during this time period. Linear regression for the time period 1983 to 1998 showed that oil concentration was decreasing in the NR and UTs. In the NR, MGs OO and O did not have significant probabilities, but $r^2$ values of MGs II through IV ranged from 0.6073 to 0.7482. In the SR, MGs IVS and VI had significant probabilities and $r^2$ values of 0.6208 and 0.2362, respectively. These data indicate that oil concentration increased in most of the MGs until 1973 and decreased in most of the MGs from 1974 to 1982. From 1983 to 1998 oil concentration decreased in six of the MGs with the major decrease in MGs II to IVS, an area that represents a large portion of soybean production acreage.

The question then becomes whether the 21 g kg$^{-1}$ decrease in oil concentration between 1973 and 1974 is genetic or environmental. During these years the SR evaluated 12 superior lines in each MG for a total of 60 superior lines each year. Of these 60 superior lines, 33 (55%) were evaluated in both 1973 and 1974. The average oil concentration of these lines was 227 and 204 g kg$^{-1}$ in 1973 and 1974, respectively. In the NR no fixed number of superior lines was evaluated each year. There were 50 and 59 superior lines evaluated in 1973 and 1974, respectively. Of these, 31 superior lines were evaluated each year representing 62 of the 109 superior lines (57%) evaluated in 1973 and 1974, respectively. The average oil concentration of these lines was 226 and 203 g kg$^{-1}$ in 1973 and 1974, respectively. All of the superior lines evaluated in 1973 were higher in oil concentration.

Table 2. Slope ($b$), probability values ($P$), goodness of fit ($r^2$), and $y$ intercept for linear regression of oil concentration onto years for the Soybean Uniform Tests Northern Region and the Soybean Uniform Tests Southern Region, 1948–1998.

<table>
<thead>
<tr>
<th>Region</th>
<th>Years</th>
<th>Regression</th>
<th>$b$</th>
<th>$P$</th>
<th>$r^2$</th>
<th>$y$ Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>North</td>
<td>1948–1973</td>
<td>Linear</td>
<td>0.4193</td>
<td>0.002</td>
<td>0.3223</td>
<td>183.966</td>
</tr>
<tr>
<td>South</td>
<td>1948–1973</td>
<td>Linear</td>
<td>0.2980</td>
<td>0.003</td>
<td>0.3116</td>
<td>196.530</td>
</tr>
<tr>
<td>Overall</td>
<td>1948–1973</td>
<td>Linear</td>
<td>0.3586</td>
<td>0.001</td>
<td>0.2397</td>
<td>190.248</td>
</tr>
<tr>
<td>North</td>
<td>1974–1982</td>
<td>Linear</td>
<td>-2.6537</td>
<td>0.014</td>
<td>0.6003</td>
<td>408.783</td>
</tr>
<tr>
<td>South</td>
<td>1974–1982</td>
<td>Linear</td>
<td>-2.3585</td>
<td>0.015</td>
<td>0.5906</td>
<td>384.679</td>
</tr>
<tr>
<td>Overall</td>
<td>1974–1982</td>
<td>Linear</td>
<td>-2.5048</td>
<td>0.001</td>
<td>0.5923</td>
<td>396.731</td>
</tr>
<tr>
<td>North</td>
<td>1983–1998</td>
<td>Linear</td>
<td>-0.9651</td>
<td>0.004</td>
<td>0.4516</td>
<td>293.963</td>
</tr>
<tr>
<td>South</td>
<td>1983–1998</td>
<td>Linear</td>
<td>-0.1276</td>
<td>0.508</td>
<td>0.0318</td>
<td>216.563</td>
</tr>
<tr>
<td>Overall</td>
<td>1983–1998</td>
<td>Linear</td>
<td>-0.5464</td>
<td>0.005</td>
<td>0.2266</td>
<td>255.263</td>
</tr>
</tbody>
</table>
concentration than in 1974 indicating that genetics was probably not the cause of the decrease in oil concentration. We used the same method to compare oil concentrations between 1982 and 1983. Oil concentration of 33 superior lines evaluated both years in the SR had an average oil concentration of 184 and 204 g kg⁻¹ in 1982 and 1983, respectively. In the NR, 62 superior lines evaluated each year had an average oil concentration of 184 and 224 g kg⁻¹ in 1982 and 1983, respectively. All of the superior lines had lower oil concentrations in 1982 than in 1983.

We believe these data indicate that the change in oil concentration between 1973 and 1974 was not due to genetics because of the large change in oil concentration from one year to the next and the observation that values from superior lines evaluated both years were consistent with the overall values for all of the lines that were evaluated. These results also confirm the observations by others (Breene et al., 1988; Hurburgh and Brumm, 1990; Hurburgh et al., 1987, 1990) that the oil concentration of U.S. soybean was decreasing. The large change in yearly mean oil concentration in 1974, the continued decrease in oil concentration to 1982, and the decrease in oil concentration of six MGs since then, lead us to believe the apparent decrease in oil in soybean produced in the UTs was not from any fundamental difference in germplasm or methods of production. The data presented indicate that the production environment and most probably unstable weather patterns has been the cause of decreased oil in the lines evaluated in the UTs since 1973.

The mean protein concentration by MG ranged from 404 to 414 g kg⁻¹ (Table 1) and ranged from 404 to 410 g kg⁻¹ in the NR and from 406 to 414 g kg⁻¹ in the SR. Maturity Group VIII had the greatest mean protein concentration. The mean protein concentration in the NR and the SR was 407 and 411 g kg⁻¹, respectively. The mean protein concentration for the entire UTs was 409 g kg⁻¹.


The deviations from the mean for the yearly protein concentrations for the NR and SR (Fig. 4) were mixed compared with the oil (Fig. 2). Thirty-two (62%) of the crop years to 1973 were below the mean. This consisted of 20 and 12 crop years from the SR and NR, respectively. From 1974 to 1998 there were 27 crop years (40%) above the mean, 8 and 12 crop years from the SR and NR, respectively. There were sequential years, 1963 to 1968 (11 crop years) and 1986 to 1993 (14 crop years), when protein and oil concentrations were below the mean. Likewise, there were sequential years, 1969 to 1973 (10 crop years) and 1994 to 1998 (eight of 10 crop years), when protein and oil concentrations were above the mean.

The mean protein to oil ratio by MG ranged from 1.92 to 2.06 (Table 1) and ranged from 1.92 to 2.06 in the NR and 1.92 to 2.01 in the SR. Maturity Group O had the highest mean protein to oil ratio. The mean protein to oil ratio in the NR and SR was 1.98 and 1.97, respectively. The mean protein to oil ratio for the UTs was 1.97.

The mean yearly protein to oil ratios for the UTs (Fig. 7) demonstrated that the protein to oil ratio was lowest (1.82) in 1973 and above 2.10 in 1980, 1981, and 1982. The mean yearly protein to oil ratios in the NR were at or below 1.82 in 1973 and 1983. In the NR, mean yearly protein to oil ratios were above 2.10 in 1981 and 1982 and in the SR in 1980, 1981, 1982, and 1998.

The deviations from the mean to protein oil ratios for the NR and SR (Fig. 8) indicates that before 1974, the yearly protein to oil ratios were generally less than or equal to the mean. Two of the crop years and five of the crop years in the SR and NR, respectively, were equal to or greater than their respective means. This indicates that the protein to oil ratio was less than or equal to the mean in 86% of the crop years and indicates that oil concentration relative to protein concentration was greater in soybean produced before 1974. From 1974 to 1998 there were 37 crop years when the protein
Fig. 3. Mean yearly protein concentration for the strains evaluated in the Soybean Uniform Tests between 1948 and 1998.

Fig. 4. Yearly Deviations from the overall mean protein concentrations of the Soybean Uniform Tests, Northern Region and Southern Region, 1948 to 1998.
Fig. 5. Mean yearly total protein and oil concentration for the strains evaluated in the Soybean Uniform Tests between 1948 and 1998.

Fig. 6. Yearly Deviations from the overall mean total protein and oil concentrations of the Soybean Uniform Tests, Northern Region and Southern Region, 1948 to 1998.
Fig. 7. Yearly mean protein to oil ratios for the strains evaluated in the Soybean Uniform Tests between 1948 and 1998.

Fig. 8. Yearly Deviations from the overall mean protein to oil ratios of the Soybean Uniform Tests, Northern Region and Southern Region, 1948 to 1998.
to oil ratio was greater than or equal to the mean. This indicates that 74% of the crop years in the NR and SR were greater than the average mean in this time period and indicates that the soybean contained more protein relative to oil.

The correlations by MG (Table 3) showed that protein and oil were negatively correlated with a range of correlation coefficients of −0.176 to −0.616. Maturity Group VII had the highest negative correlation and the correlation coefficients ranged from −0.176 to −0.445 and −0.443 to −0.616 in the NR and SR, respectively. The overall correlation for protein and oil in the NR and SR was −0.325 and −0.522, respectively. The overall correlation between protein and oil in the UTs was −0.385.

Another issue addressed in Hurburgh et al. (1990) was the concern that Brazilian soybean have become higher in protein concentration than those produced in the USA. On the surface, this issue may have some merit because the MG means (Table 1) demonstrated that the SR MGs were generally higher in protein concentration and generally contained more total protein and oil than the NR MGs. Brazilian germplasm was developed from lines introduced from the USA and 11 ancestors collectively represent 89% of the Brazilian gene pool (Hiromoto and Vello, 1986). These ancestors are SR germplasm and would indicate Brazilian soybean contain more protein and oil based on the SR means (Table 1). However, negative correlations between protein and oil were stronger in the SR MGs. This indicates that it would be difficult to have a soybean line with both high protein and oil. Hurburgh and Brumm (1990) plotted protein and oil values from a three-year-survey and found that 15% of the seed lots contained both above average protein and oil concentrations. We plotted the 24 839 individual line–location protein and oil concentrations in the NR tests and found that 20% were above average and 17% were below average in oil and protein concentration. The remaining 63% were divided about evenly between above average for either protein or oil and below average for the other component. These data and those presented by Hurburgh and Brumm (1990) indicate that one of every five seed lots produced were above average in protein and oil concentration and lead us to believe that it would be difficult for Brazilian soybean production, on average, to be both high in protein and oil concentration. Negative correlations between protein and oil concentration further illustrate the degree of difficulty in producing a soybean line with more protein that would not have an impact on oil concentration (Burton, 1987; Yaklich 2001). The yearly deviation of protein to oil ratios (Fig. 8) also indicates that soybean, since 1973, contains more protein relative to oil.

Researchers (Andresen et al., 2001, and references therein) have noted a time period in the climate record between the mid 1950s and 1970s called the benign climate period. During this period relatively favorable weather led to consistently high agricultural productivity across the Corn Belt region of the central USA. This period was cooler and wetter than normal and was both preceded and followed by weather of greater variability. During the benign climate the average July–August temperature was 2°C cooler in the 1960s than in the 1930s and there were increasing amounts of July and August precipitation. This contributed to consistent and increasing yield during this twenty year time period. The twenty-year periods prior to and following the benign climate had favorable years but were marked by years with extremes in summer and winter weather. The graph of the mean yearly oil concentration (Fig. 1) indicates that during the benign climate of the 1950s to the 1970s soybean seed oil concentration was increasing. Thus, the benign climate may have been favorable for increased soybean seed oil concentration. The fluctuations in soybean seed oil concentration after 1973 may be due to a return to more normal weather conditions that are characterized by years with extremes in precipitation and temperature.

**ACKNOWLEDGMENTS**

REFERENCES


