# Developing Genetic Coefficients for Crop Simulation Models with Data from Crop Performance Trials

T. Mavromatis,* K. J. Boote, J. W. Jones, A. Irmak, D. Shinde, and G. Hoogenboom

## ABSTRACT

**Successful uses of crop models in technology transfer and decision support tools require that coefficients describing new cultivars be available as soon as the cultivars are marketed. The objectives of this study were (i) to develop an approach to estimate cultivar coefficients for the CROPGRO–Soybean model from typical information provided by crop performance tests, (ii) to evaluate the suitability of yield trial data for deriving genetic coefficients and site-specific soil traits for use in crop models, and (iii) to explore the extent to which our approach allowed the crop model to reproduce observed genotype × environment (GE) interactions, cultivar ranking, and year-to-year yield variability. Crop performance tests typically record harvest maturity date, seed yield, seed size, height, and lodging. A stepwise procedure using data on 11 cultivars grown at five sites in Georgia over 4 to 10 yr efficiently decreased the root mean square error (RMSE) between observed and predicted data. For 'Stonewall', a maturity group VII cultivar, the RMSE of 769 kg ha $^{-1}$ between the actual and modeled seed yield, estimated initially by means of the existing general maturity group coefficients, was reduced to 404 kg ha $^{-1}$. For the same cultivar, the initial RMSE of 5.3 and 9.3 d between the actual and simulated anthesis and harvest maturity dates, respectively, estimated by means of the existing general maturity group coefficients, were reduced to 2.9 and 5.8 d. In addition to deriving useful information on site characteristics and cultivar traits, our approach has enabled CROPGRO to satisfactorily mimic the genotypic yield ranking and much of observed genotype × environment interactions. Across all environments, the difference in genotype ranking based on yield between measured and predicted values was one or less for 61% of the environments.**

DYNAMIC CROP MODELS have potential to quantify the contribution of environmental factors, such as temperature, daylength, and water supply, on complex GE interactions observed in yield trial data. CROPGRO–Soybean v. 3.5 is a process-oriented model (Boote et al., 1998; Hoogenboom et al., 1994) that has been used to study soybean [*Glycine max* (L.) Mer.] response to management (Egli and Bruening, 1992), environmental conditions (Curry et al., 1995), and genetic yield potential (Boote and Tollenaar, 1994). It also has been utilized to study causes of spatial yield variability (Paz et al., 1998; Allen et al., 1996). CROPGRO simulates carbon, water, and nitrogen balances for the soybean plant and soil. Model algorithms express the relationships between plant processes, including phenological development, photosynthesis, respiration, plant water uptake, biomass growth and partitioning, and environmental variables such as daily temperature, photoperiod, and soil water availability. This model also incorporates knowledge of cultivar-specific traits to predict daily growth and development as the plant responds to weather, soil characteristics, and management practices (Boote et al., 1998). These cultivar-specific factors have come to be called "genetic coefficients" (Table 1).

Estimation of genetic coefficients and soil parameters provides an important first step in crop model use for making farm management decisions (Heiniger et al., 1997), estimating large area yields (Hodges et al., 1987), predicting the performance of new cultivars (Liu et al., 1989), and testing model improvements (Boote et al., 1997). Conventionally, the estimation of crop model coefficients is a fitting process done systematically but manually by labor-intensive, somewhat subjective, trial and error procedures. The modeler manipulates those parameters until simulated output mimics field observations of phenology, growth, and yield. This approach was adopted by a number of previous investigators (Colson et al., 1995). A few studies have used optimization procedures to derive soybean cultivar coefficients for predicting flowering date (Grimm et al., 1993), to model the occurrence of reproductive stages after flowering (Grimm et al., 1994), and to improve crop models (Piper et al., 1998), by minimizing the error sum of squares between observations and predictions.

Successful use of crop models in technology transfer and as decision support tools requires that coefficients describing new cultivars be available as soon as the cultivars are marketed. Constraints imposed by time and large numbers of cultivars eliminate the possibility for scientists to conduct experiments and measure detailed growth dynamics to fully derive cultivar coefficients. While growth analysis has been successful for a limited number of cultivars, new approaches must be developed. Presently, crop performance tests are the only readily available data on the host of new cultivars released, whether done in-house by private companies or by state or other public testing agencies. Soybean performance tests typically record harvest maturity date, final seed yield, seed size, canopy height, and lodging. Sometimes flowering date is measured as well. Furthermore, yield trial data often cross a broad range of environments and they have been useful for calibrating models over large geographic areas to improve model performance (Piper et al., 1998).

The objectives of this study were (i) to develop a practical methodology to estimate cultivar coefficients

T. Mavromatis, J.W. Jones, and A. Irmak, Dep. of Agricultural and Biological Engineering, Univ. of Florida, Gainesville, FL 32611; K.J. Boote, Dep. of Agronomy, Univ. of Florida, Gainesville, FL 32611; D. Shinde, Tropical Research and Education Center, Univ. of Florida, Homestead, FL 33031; G. Hoogenboom, Dep. of Biological and Agricultural Engineering, Univ. of Georgia, 30223. Florida Agric. Exp. Stn., J. Series No R-07163. Received 14 Oct. 1999. *Corresponding author (theo@agen.ufl.edu).

Published in Crop Sci. 41:40–51 (2001).

**Abbreviations:** CSDL, critical short daylength; DOY, day of year; GE, genotype × environment; RMSE, root mean square error; SLPF, soil fertility factor.

for the soybean crop model from typical information provided by crop performance tests, (ii) to evaluate the suitability of yield trial data for deriving genetic coefficients as well as site-specific soil traits for use in crop models, and (iii) to test the extent to which our approach allowed the CROPGRO–Soybean model to reproduce observed GE interactions, cultivar ranking, and year-to-year yield variability.

## MATERIALS AND METHODS

### Yield Trial Data

Yield trial data were obtained from the "Field crop performance tests: Soybean, peanut, cotton, tobacco, sorghum, and summer annual forages" research reports from the Georgia Experiment Stations from 1987 to 1996 (Raymer et al., 1997, 1996, 1995, 1994, 1993, 1992, 1991, 1990, 1989, 1988). These crop performance tests are referred to as *yield trials*. The yield trial data include observations of harvest maturity date (day of year, DOY), seed yield (kg ha$^{-1}$), seed size (mg), and sometimes flowering date (DOY). For our study, observed yield data were decreased 13% to compare with simulated dry weight yields. These trials were conducted with planting sets of 20 to 50 cultivars over multiple years and sites. A subset of 11 cultivars grown at five locations in Georgia [Tifton (31.5° N, 83.5° W), Plains (32° N, 84.3° W), Midville (32.9° N, 82.2° W), Griffin (33.2° N, 84.7° W), and Calhoun (34.3° N, 85.1 W)] ranging in altitude from 150 to 267m over 4 to 10 yr was chosen for the development and testing of our procedure. Cultivar selection was based on each given cultivar being present in a sufficient number of years over a sufficient number of locations to provide a minimum of 20 to 25 site-year combinations per cultivar. This procedure eliminated cultivars that were only in a few tests or at a few locations. Indeed, we deleted two northern Georgia sites because they had only a partial cultivar set. The selected cultivars were Perrin, Stonewall, Hagood, Cook, Thomas, Colquitt, Brim, Young, Hutcheson, Bryan, and Deltapine 105 (Table 2). Only the rainfed plantings were used since insufficient irrigation information was available for the irrigated trials. A total of 393 cultivar × site × year combinations were available, divided between early and late plantings. Table 3 describes the mean yield and weather at each site.

### CROPGRO Model Inputs

Daily weather data of total solar radiation, daily maximum and minimum air temperature, and total precipitation are required by CROPGRO. Weather data for each site were obtained from the Georgia Automated Environmental Monitoring Network (Hoogenboom, 1996; Hoogenboom and Gresham, 1997).

CROPGRO inputs also include sowing date, plant population, row spacing, and initial soil water content. The initial soil water at planting date was set to field capacity for all years and locations. Therefore, model errors may have occurred in some cases where the initial soil water was too high for some locations in some years. Crops at all locations were sown in rows 0.76 m apart at a density of 34 plants m$^{-2}$ (as reported in the research reports). The effects of tillage, pests, and diseases were not directly considered in our simulations.

A number of cultivar-specific parameters (genetic coefficients) are used by the crop model to predict soybean daily growth and development responses to weather, soil characteristics, and management actions (Boote et al., 1998). The genetic coefficients (Table 1) describe (i) the cultivar sensitivity to daylength, (ii) durations of life cycle phases, (iii) vegetative growth traits (e.g., light saturated leaf photosynthesis rate), and (iv) reproductive growth traits (e.g., potential seed size). The life cycle phase coefficients (e.g., emergence to flowering,

**Table 1. Definition of parameters from the CROPGRO-Soybean model that were determined by the optimization procedure.**

| Parameter | Definition | Units |
|---|---|---|
| CSDL | Critical short day length below which reproductive development progresses with no daylength effect. | h |
| PPSEN | Slope of the relative response of development to photoperiod with time. | h$^{-1}$ |
| FLSH | Time between first flower and first pod. | PTD† |
| FLSD | Time between first flower and first seed. | PTD |
| SDPM | Time between first seed and physiological maturity. | PTD |
| R1PPO | Increase in daylength sensitivity after anthesis. | h |
| LFMAX | Light saturated leaf photosynthesis rate at 30°C, 350vpm $CO_2$ and high light. | mg $CO_2$/m$^2$ s$^{-1}$ |
| THRESH | The maximum ratio of (seed/(seed + shell)) at maturity. | |
| SFDUR | Seed filling duration for pod cohort at standard growth conditions. | PTD |
| PODUR | Time required for cultivar to reach full pod load under optimal conditions | PTD |
| SLPF | Soil fertility factor, 0-1. | |
| DUL | Drained upper limit, volumetric soil water content. | cm$^3$/cm$^3$ |
| LL | Drained Lower limit, volumetric soil water content. | cm$^3$/cm$^3$ |
| SAT | Saturation point, volumetric soil water content. | cm$^3$/cm$^3$ |

† PTD = photothermal days.

**Table 2. Mean yield, anthesis, and harvest maturity dates for 11 soybean cultivars in Georgia. Statistics include standard deviation (STD) of the yield, and maximum (MAX) and minimum (MIN) yield.**

| Cultivar | Maturity group | Anthesis | Harvest maturity | Yield | STD | MAX | MIN |
|---|---|---|---|---|---|---|---|
| | | dap† | | kg ha$^{-1}$ | | | |
| Perrin | VII | 60.0 | 147.2 | 2202 | 563 | 3555 | 766 |
| Stonewall | VII | 53.5 | 139.8 | 2330 | 700 | 3432 | 784 |
| Hagood | VII | 62.5 | 144.2 | 2365 | 675 | 3602 | 743 |
| Cook | VII | 58.1 | 144.2 | 2471 | 680 | 3807 | 1006 |
| Thomas | VII | 61.3 | 144.6 | 2215 | 646 | 3198 | 608 |
| Colquitt | VII | 54.6 | 141.8 | 2280 | 590 | 3596 | 1029 |
| Brim | VI | 56.5 | 138.0 | 2615 | 784 | 3935 | 608 |
| Young | VI | 61.1 | 140.3 | 2462 | 793 | 3503 | 550 |
| Hutcheson | V | 52.5 | 132.0 | 2513 | 884 | 4023 | 380 |
| Bryan | V | 57.3 | 143.3 | 2485 | 776 | 3719 | 561 |
| DP 105 | V | 53.7 | 133.2 | 2227 | 854 | 3637 | 437 |

† dap = days after planting.

**Table 3. Mean soybean yield, mean air temperature, and total precipitation during May–October from 1987 to 1996 in five locations in Georgia. Statistics include standard deviation (STD) of the yield, and maximum (MAX) and minimum (MIN) yield.**

| Location | Yield | STD | MAX | MIN | PREC | TEMP |
|---|---|---|---|---|---|---|
| | kg ha$^{-1}$ | | | | mm | C° |
| Tifton | 2063 | 588 | 3391 | 696 | 555 | 24.3 |
| Plains | 2018 | 809 | 3807 | 550 | 612 | 23.9 |
| Calhoun | 2620 | 699 | 4023 | 988 | 654 | 21.4 |
| Midville | 2489 | 957 | 3935 | 439 | 615 | 23.6 |
| Griffin | 2529 | 451 | 3596 | 1503 | 586 | 22.2 |

flowering to beginning seed, and beginning seed to physiological maturity) relate to life cycle timing and are measured in "photothermal days." The latter is a unit that combines the standard concept of degree-days with a measure of daylength. Most cultivar coefficients are generally similar for cultivars within a maturity group (Boote et al., 1997). This provides a starting point, as approximate values are known for all maturity groups (Grimm et al., 1993, 1994; Boote et al., 1997). Still, individual cultivars frequently vary from maturity group norms. This, along with site-specific and year-specific environmental variation, results in variation in cultivar performance over locations and years.

For each yield trial location, the most common soil was identified from soil surveys, as determined by Perkins et al. (1978, 1979, 1983, 1985, and 1986). Soil types and families at those locations were Dothan loamy sand (fine, siliceous thermic plinthic paleudult) at Midville; Tifton sandy loam (fine-loamy, siliceous thermic plinthic paleoudult) at Tifton; Greenville sandy clay loam (clayey, kaolinitic thermic typic rhodic paleoudult) at Plains; Rome clay loam (fine-loamy, mixed thermic typic paleoudult) at Calhoun; and Cecil sandy clay loam (clayey, kaolinitic thermic typic paleoudult) at Griffin. The soil characteristics for each profile were used to calculate the soil physical and chemical parameters required to run the CROPGRO model (Ritchie et al., 1989; Tsuji et al., 1994). However, soil properties vary within a soil series, making it difficult to estimate soil properties for a particular field by soil survey information. While those soil characteristics critically affect the crop model water balance, they are commonly not adequately described in soil survey reports. In our study, we attempted to account for uncertainties in soil characteristics in specific fields, by modifying a soil fertility factor (SLPF) and soil water holding characteristics (drained upper limit, drained lower limit, and saturated water-holding limit). In CROPGRO, SLPF is an input variable (constant for a given field site) (Table 1) that affects biomass growth rate by modifying daily canopy photosynthesis. SLPF is attributed to soil fertility differences or soil-based pests, such as nematodes. SLPF was initially set to 0.8 for the purpose of having a similar basis for comparison.

## Estimating Cultivar Coefficients and Soil Parameters

The fitting of cultivar coefficients and soil parameters was a systematic, stepwise procedure in which (i) candidate coefficients or parameters were selected; (ii) the values of the coefficients or parameters were changed by running CROPGRO in an optimization shell until the error sum of squares (simulated minus observed) was minimized; and (iii) the set of coefficients or parameters that produced the lowest RMSE was adopted. The success of this procedure was shown by a reduction in RMSE from one step to the next. Two optimization algorithms were used: (i) one- and two-dimensional linear grid search, and (ii) simulated annealing (Goffe et al., 1994). Simulated annealing is a powerful optimization technique that explores a function's entire surface and tries to optimize the function while moving both uphill and downhill. Thus, it is largely independent of the starting values, often a critical input in other algorithms.

The existing cultivar coefficients from maturity groups (MG) V through VII, provided with CROPGRO-Soybean V 3.5 (Boote et al., 1998) were first used to test the hypothesis that the selected cultivars did not deviate from the maturity group norms and that no bias existed between observed and predicted anthesis, harvest maturity, and grain yield.

Simulated annealing was used to solve for the critical short daylength (CSDL) and photoperiod sensitivity (PPSEN) (Table 1) for fitting the observed flowering date (R1) for each cultivar across all sites, years, and planting dates. For each cultivar, CSDL and PPSEN were allowed to change within the range of values for ±2 maturity groups since the assignments of the cultivars to specific groups should not necessarily limit the range of the search. The general maturity group values were initially retained for the rest of the cultivar coefficients.

For each cultivar over all locations, years, and planting dates, we optimized [FLSD + SDPM] and R1PPO to fit the observed maturity date (Table 1). In CROPGRO, R1PPO had been assumed constant for cultivars within the same maturity group. R1PPO decreases CSDL by a given number of hours, making plant development more sensitive to photoperiod after anthesis. Piper et al. (1996) provided evidence that varying CSDL after R1 often resulted in a better fit of observed phenology for MG III and later maturity group cultivars.

A two-way linear grid search was used to minimize the error sum of squares between simulated and observed maturity. For one direction of the search, a pseudo variable X (ranging from −1 to +1) shifted FLSD and SDPM together within a range supported by literature (Piper et al., 1996). On the second direction of the search, R1PPO for each maturity group was varied in the interval (R1PPO + 0.2, R1PPO − 0.5) h (Piper et al., 1996; Grimm et al., 1994). Once we had fit both FLSD and SDPM, FLSH was set proportionally to FLSD as follows:

$$\text{FLSH} = (\text{FLSH/FLSD})_{\text{MG}}\ \text{FLSD}_{\text{opt}}$$

where $\text{FLSD}_{\text{opt}}$ is the optimized value of FLSD and the ratio of $(\text{FLSH/FLSD})_{\text{MG}}$ is the ratio of the general maturity group values for the specific coefficients. This merely rescaled FLSH (time from beginning flower to beginning pod) when the FLSD (time from beginning flower to beginning seed) was changed by the optimization procedure.

For each individual location, we optimized the soil SLPF and the soil water holding limits (DUL–LL) (Table 1) that best predicted mean yield over all cultivars and planting dates at each site across multiple years. A two-dimensional linear grid search was used to find the combination of SLPF and (DUL–LL) that minimized the sum of squares of the errors between simulated and observed grain yield. For one direction of the search, a pseudo variable shifted DUL and SAT together for the specific soils. For the other search dimension, SLPF was allowed to vary from 0.7 to 1.05. This range is well within the range supported by literature (Jones et al., 1989). That study reported SLPF values as low as 0.5 for locations in India and up to 1.0 for highly fertile soils in Iowa and Illinois.

With coefficients obtained at this point of our procedure, the CROPGRO model should be able to predict site-mean yields and the part of cultivar variation in yield associated with differences in life cycle. Any remaining differences in yield among cultivars would be attributed to traits other than anthesis and maturity date. Next, we adjusted coefficients to account for yield differences among individual cultivars across

**Table 4. The maximum shifts for the pseudo-variable (X2) that shifts FLSH, FLSD, SDPM, SFDUR, and PODUR together, to decrease (X2 = −1) or increase (X2 = +1) yield.**

| VARIABLE† | X2 = −1 | X2 = +1 |
|---|---|---|
| FLSH | +2.8 | −3.1 |
| FLSD | +3.8 | −4.1 |
| SDPM | −4.2 | +4.2 |
| SFDUR | −3.0 | +3.0 |
| PODUR | +5.0 | −3.0 |

† FLSH = Time between first flower and first pod (PTD‡); FLSD = Time between first flower and first seed (PTD); SDPM = Time between first seed and physiological maturity (PTD); SFDUR = Seed filling duration for pod cohort at standard growth conditions (PTD); PODUR = Time required for cultivar to reach full pod load under optimal conditions (PTD).

‡ PTD = Photothermal days.

sites, using a two-way linear grid search. In one search direction, we created a "productivity" pseudo variable X1 that shifts LFMAX and THRESH (Table 1) together. Boote and Tollenaar (1994) reported LFMAX in a range from 0.82 to 1.39 mg $CO_2$ $m^{-2}$ $s^{-1}$ for soybean, with an average value of 1.05. In our study LFMAX was allowed to vary from 0.93 to 1.2. A maximum change in THRESH of ±2.5% was used (range from S. Welch, 1999, personal communication). For the second search dimension, we developed a second pseudo variable X2 (−1 to +1) that jointly shifted cultivar "life cycle" coefficients (FLSH, FLSD, SDPM, SFDUR, and PODUR) that affected grain yield by starting pod and seed growth sooner or later within the previously fixed time from anthesis to maturity (Table 1). These traits act to change harvest index but not total biomass. The shifts were all made together in a fashion that caused minimal changes in maturity date since we did not want to disturb the cultivar life cycle that we had already estimated. For example, to increase yield without changing maturity date, the time to beginning pod (FLSH) and beginning seed (FLSD) was shortened and pod adding duration (PODUR) was shortened, but time from beginning seed to physiological maturity (SDPM) and SFDUR were lengthened equivalently to hold constant maturity. To decrease yield, the opposite changes were made. The maximum shifts in each coefficient (for X2 = −1 and X2 = +1) are presented in Table 4. Negative and positive shifts were designed to cause simulated yield to decrease and increase, respectively, while the maximum changes in "net" maturity were held to less than one day on average.

### Stability Analysis and Cultivar Ranking

A statistical technique known as stability analysis is widely used to detect GE interactions (Specht et al., 1986). The whole dataset was not orthogonal. To evaluate the ability of our fitting approach to reproduce field-observed GE interactions with CROPGRO, we selected an othogonal subset of eight cultivars (Colquitt, Thomas, and DP 105 were excluded in this analysis) grown at four locations over a 5-yr period and we reestimated the cultivar coefficients. A total of 14 year-location combinations or environments (Table 5) were included in the stability analysis. Overall, 112 early planting cases (eight cultivars × 14 environments) were used.

Cultivar and environment main effects for measured and simulated yields were tested for significance by analysis of variance. Kolmogorov-Smirnov and equal variance tests were carried out to test normality and equal variance assumptions that the ANOVA requires. When either of those tests failed (alpha = 0.05) we used the nonparametric Kruskal-Wallis test to detect significant differences for all the pairwise comparisons of the median responses of grain yield among the different

**Table 5. The 14 year-location combinations (environments) used for the stability analysis of the orthogonal subset.**

| Environment | Year | Location |
|---|---|---|
| 1 | 1991 | Griffin |
| 2 | 1991 | Midville |
| 3 | 1991 | Tifton |
| 4 | 1992 | Griffin |
| 5 | 1992 | Midville |
| 6 | 1992 | Tifton |
| 7 | 1994 | Griffin |
| 8 | 1994 | Plains |
| 9 | 1995 | Griffin |
| 10 | 1995 | Midville |
| 11 | 1995 | Plains |
| 12 | 1996 | Griffin |
| 13 | 1996 | Midville |
| 14 | 1996 | Plains |

treatment groups. The power, or sensitivity of this test, was also calculated to determine the probability that the test will detect a difference among groups if a difference existed.

The observed yield of each cultivar was then regressed against mean observed yield at each of our 14 environments used as an "index" of test site productivity. This index is a "site" mean computed from the yield of all cultivars in each environment. The regression procedure was similarly repeated for predicted yield of each cultivar versus the predicted mean yield at each site (over all cultivars). The slopes of the regression lines of either the actual (or simulated) yields for each cultivar vs. site mean yields were then compared for significant differences.

Finally, it is important for producers to know which cultivars best respond to local environmental conditions and management practices, if genetic differences indeed exist. Traditionally, cultivar selection has been based on grain yield. To investigate the crop model ability to reproduce the observed yield-based cultivar ranking we estimated and compared the differences in ranking of cultivars for actual and modeled yield.

## RESULTS AND DISCUSSION

### Optimizing Cultivar Coefficients Related to Anthesis and Maturity

CROPGRO-Soybean with the generic MG VII coefficients simulated late flowering and maturity dates for Stonewall (Table 6), compared with observed values (Table 2). The average differences between actual and simulated dates were 4.6 and 7.1 d for flowering and maturity respectively. The crop model underestimated the observed Stonewall yield by about 15% despite the longer simulated life cycle. These results were representative of other cultivars and indicated the existence of model bias using the general MG coefficients. Our simulations also suggested that coefficients for individual cultivars differed from the maturity group norms.

Optimized CSDL values for most MG VII cultivars were within the typical MG range and were generally smaller than the values for MG V and VI as expected (Table 7). Likewise, later MG generally had larger optimized PPSEN values as expected (Grimm et al., 1993). For group V cultivars, the search algorithm found critical daylength values (CSDL) typical of MG VI genotypes, but the PPSEN values were smaller, thus allowing earlier anthesis. The differences in anthesis between MG V and VI cultivars were relatively small and a single MG shift in estimated coefficient values should not be

**Table 6. Simulated anthesis, harvest maturity, and grain yield for soybean cultivar Stonewall using the existing general maturity group values. Statistics include $r^2$ of the regression line (rsq.); root mean square error of the predicted minus observed (RMSE); intercept and slope of the regression line; index of agreement ($d$), and sample size $n$.**

| | Sim. | Intercept | Slope | $r^2$ | RMSE | $d$ | $n$ |
|---|---|---|---|---|---|---|---|
| Anthesis (dap)† | 58.1 | 12.0 | 0.861 | 0.893 | 5.3 | 0.896 | 27 |
| Harvest maturity (dap) | 146.9 | 45.7 | 0.723 | 0.89 | 9.3 | 0.898 | 45 |
| Grain yield (kg ha$^{-1}$) | 1971 | 212 | 0.755 | 0.392 | 769 | 0.752 | 49 |

† dap = days after planting.

of concern. The simulated anthesis dates for most of the cultivars were within 1 d on average from the observed dates (Table 2). The index of agreement (d) (Willmot, 1982) was higher for group VII cultivars than for VI and V (Table 7), possibly because more data ($n > 20$ versus $n < 18$) were used to fit MG VII cultivars than VI and V.

The intercept of the regression line between observed and simulated flowering for Stonewall (Fig. 1) was lower than that obtained with the general maturity group values. In addition, the slope was closer to 1. Fitting the parameters affecting flowering date also improved the model predictions of maturity date. The changes in life cycle resulted in a significant increase in the modeled mean grain yield that now overestimated the measured value by only 2.4% after fitting observed anthesis dates. The RMSE of soybean yield also decreased from 769 kg ha$^{-1}$ when the general MG VII coefficients were used to 634 kg ha$^{-1}$ after fitting CSDL and PPSEN.

The linear grid search for group VII cultivar coefficients resulted in higher values for R1PPO than the typical general MG values (Table 8), although Piper et al. (1996) found cultivar variation in this range. As expected, R1PPO increased as MG increased. FLSD and SDPM, on the other hand, were somewhat smaller than typical generic MG values. The simulated harvest maturity date for all of the cultivars was found within 1 d on average from the actual dates (Table 2). The intercepts of linear regression analysis between predictions and observations were closer to 0.0 and slopes were closer to 1.0 compared with those obtained with the general maturity group coefficients. The index of agreement for maturity dates, as for anthesis dates, was higher for MG VII cultivars than for VI and V probably because more data were available for MG VII. On the other hand, RMSE of maturity was less than 5 d for all of group V cultivars and higher than 5 d for three out of six MG VII cultivars.

The regression analysis between observed and simulated maturity dates for Stonewall (Fig. 2) showed the crop model tendency to underestimate the life cycle for early plantings, and overestimate the life cycle for late plantings, respectively. A reduction in RMSE for maturity dates (from 6.5–5.8 d) was found after fitting of the coefficients affecting maturity.

## Optimized Site Characteristics

The optimized SLPF values ranged from 0.76 to 0.94 (Table 9), well within the range supported by literature (Jones et al., 1989). Increases from +0.028 to +0.06 cm$^3$ cm$^{-3}$ in DUL and SAT for each soil layer occurred at most locations except Plains. Except at Calhoun, these increases in DUL-LL were combined with slightly reduced SLPF compared to the value of 0.8 we had used initially. The summary statistics (Table 9) demonstrate the ability of our approach to capture the site mean yield for each location. The RMSE were low, ranging from 16.1% (for Plains) to 20.2% (for Midville) of the actual site mean yields. The optimization approach we used did not explain a high percentage of the yield variation at Griffin (high intercept, low slope of ob-

**Table 7. Estimates for each soybean cultivar of CSDL and PPSEN. The simulated anthesis date (Sim.) for each cultivar is also included. Statistics for anthesis date include $r^2$ of the regression line; root mean square error of the predicted minus observed (RMSE); index of agreement ($d$); and sample size $n$.**

| Cultivar | $n$ | CSDL† | PPSEN | Sim. anthesis | $r^2$ | RMSE | $d$ |
|---|---|---|---|---|---|---|---|
| | | h | h$^{-1}$ | dap‡ | | d | |
| | | | | Group VII§ | | | |
| Perrin | 33 | 12.32 | 0.332 | 59.8 | 0.844 | 3.3 | 0.957 |
| Stonewall | 27 | 12.59 | 0.337 | 53.4 | 0.899 | 2.9 | 0.963 |
| Hagood | 21 | 12.30 | 0.344 | 62.2 | 0.872 | 3.1 | 0.966 |
| Cook | 20 | 12.33 | 0.321 | 58.0 | 0.928 | 2.2 | 0.980 |
| Thomas | 23 | 12.31 | 0.330 | 61.3 | 0.852 | 3.4 | 0.960 |
| Colquitt | 21 | 12.37 | 0.319 | 55.2 | 0.898 | 2.6 | 0.970 |
| | | | | Group VI§ | | | |
| Brim | 13 | 12.55 | 0.322 | 56.2 | 0.879 | 2.6 | 0.943 |
| Young | 18 | 12.49 | 0.335 | 59.3 | 0.777 | 3.0 | 0.906 |
| | | | | Group V§ | | | |
| Hutcheson | 18 | 12.52 | 0.306 | 53.7 | 0.772 | 2.7 | 0.918 |
| Bryan | 18 | 12.30 | 0.310 | 58.5 | 0.846 | 2.9 | 0.949 |
| DP 105 | 14 | 12.52 | 0.306 | 52.9 | 0.792 | 2.2 | 0.930 |

† CSDL = Critical short day length below which reproductive development progresses with no daylength effect (h); PPSEN = Slope of the relative response of development to photoperiod with time (h$^{-1}$).

‡ dap = days after planting.

§ Typical ranges of CSDL and PPSEN were: (a) (12.20, 12.45) (0.316, 0.325) for MG VII cultivars respectively, (b) 12.46, 12.70) (0.307, 0.315) for MG VI cultivars respectively and (c) (12.71, 12.96) (0.299, 0.307) for MG V cultivars respectively.

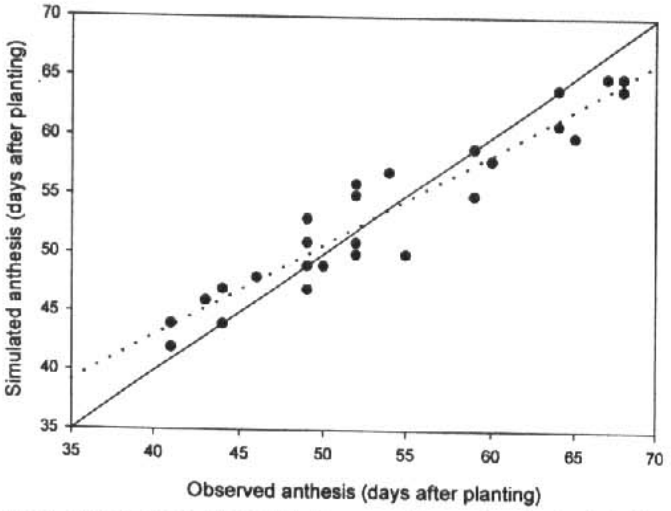


**Fig. 1. Comparison of simulated versus observed anthesis date for the soybean cultivar Stonewall. The 1:1 line (solid) and the regression line (dot) between the data are also shown.**

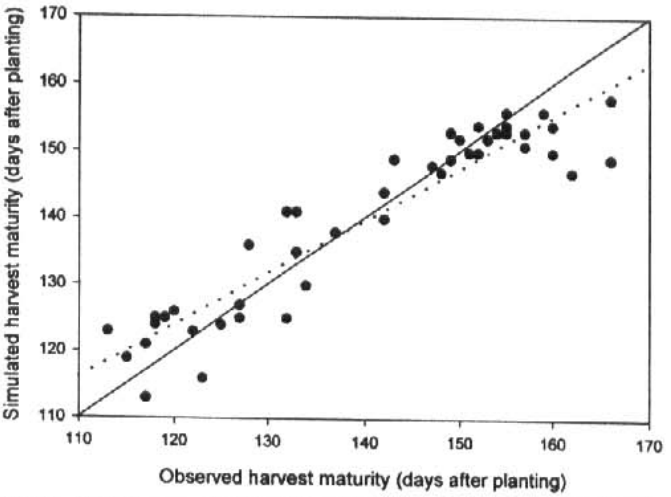


**Fig. 2. Comparison of simulated versus observed harvest maturity for the soybean cultivar Stonewall. The 1:1 line (solid) and the regression line (dot) between the data are also shown.**

served vs. predicted, and low $r^2$ (Table 3). Nevertheless, the coefficient of variation (std/mean) of yield was initially low at Griffin compared with that from other sites (17.8% versus 26.7–40.1% for the other sites) (Table 3). The historic year-to-year yield variation for early and late plantings was also well reproduced by the crop model for each location ($r^2$ was estimated from 0.55 at Tifton to 0.87 at Plains) except for Griffin (where $r^2$ was 0.24) (Fig. 3a and b, Table 9). In view of the relatively stable observed yields, but wide simulated yield range, we suspect problems with rainfall records at the Griffin site. For each site, CROPGRO tended to overestimate low yield treatments and underestimate very productive years.

The changes in soil water availability, however, slightly changed the simulated dates of anthesis and maturity because of effects of water stress on development. The RMSE between actual and simulated flowering was increased for Stonewall from 2.9 to 3.4 d. The intercept also increased (from 12.6–15.5 d) and the slope deviated more from the 1:1 slope (from 0.763–0.701). On the other hand, regression analysis gave better fits for harvest maturity. The simulated yield was increased but still underestimated the observed value by 1.4%. More importantly, the RMSE between actual and simulated yields for Stonewall was reduced from 602 kg ha$^{-1}$ to 414 kg ha$^{-1}$ after solving for site characteristics.

## Yield Characteristics Attributed to Cultivar

The linear grid search to optimize yield for each cultivar produced lower values for SDPM compared with those of general maturity groups, but the values were appropriately longer for later MGs (Table 10). THRESH did not deviate much from the maturity group norms (set to 78.0 for groups V–VII). However, it attained its upper and lower limits (80.5 and 75.5) for Brim and Thomas,

**Table 8. Estimates for each soybean cultivar of R1PPO, FLSD and SDPM. The simulated harvest maturity (Sim.) for each cultivar is also included. Statistics include $r^2$ of the regression line; root mean square error of the predicted minus observed (RMSE); index of agreement ($d$); and sample size $n$.**

| Cultivar | $n$ | R1PPO† | FLSD | SDPM | Sim. maturity | $r^2$ | RMSE | $d$ |
|---|---|---|---|---|---|---|---|---|
| | | h | PTD‡ | | dap§ | | d | |
| | | | | Group VII¶ | | | | |
| Perrin | 39 | 0.794 | 14.4 | 32.15 | 147.5 | 0.934 | 4.6 | 0.977 |
| Stonewall | 45 | 0.934 | 13.3 | 29.70 | 139.6 | 0.882 | 5.8 | 0.961 |
| Hagood | 34 | 0.759 | 13.0 | 29.00 | 144.7 | 0.943 | 4.7 | 0.978 |
| Cook | 25 | 0.864 | 13.2 | 29.35 | 144.3 | 0.932 | 4.9 | 0.973 |
| Thomas | 30 | 0.794 | 13.3 | 29.70 | 145.5 | 0.901 | 6.2 | 0.960 |
| Colquitt | 31 | 0.899 | 13.2 | 29.35 | 142.4 | 0.887 | 6.1 | 0.960 |
| | | | | Group VI¶ | | | | |
| Brim | 33 | 0.749 | 13.2 | 28.85 | 138.3 | 0.890 | 4.4 | 0.964 |
| Young | 35 | 0.259 | 14.8 | 32.70 | 139.7 | 0.748 | 5.2 | 0.909 |
| | | | | Group V¶ | | | | |
| Hutcheson | 30 | 0.389 | 12.9 | 29.05 | 132.7 | 0.661 | 4.7 | 0.866 |
| Bryan | 37 | 0.774 | 12.7 | 28.35 | 143.2 | 0.909 | 4.1 | 0.971 |
| DP 105 | 19 | 0.424 | 13.6 | 30.45 | 133.5 | 0.673 | 3.9 | 0.891 |

† R1PRO = Increase in daylength sensitivity after anthesis (h); FLSD = Time between first flower and first seed (PTD); SDPM = Time between first seed and physiological maturity (PTD).
‡ PTD = photothermal days.
§ dap = days after planting.
¶ The typical general MG values for R1PPO were set to 0.504, 0.459 and 0.414 for MG VII, VI, and V varieties respectively; Typical values of FLSD and SDPM were: (a) 16.0 and 36.00 (PTD) for MG VII cultivars respectively, (b) 16.0 and 35.50 (PTD) for MG VI cultivars respectively and (c) 15.5 and 35.00 (PTD) for MG V cultivars respectively.

**Table 9. Estimates for each location of relative soil fertility factor (SLPF) and incremental shift in water-holding capacity (change in DUL and SAT). The simulated soybean yield (Sim.) is included. Statistics include $r^2$ of the regression line (rsq.); root mean square error of the predicted minus observed yield (RMSE); intercept and slope of the regression line; index of agreement ($d$), and sample size $n$.**

| Location | $n$ | SLPF | DUL/SAT | Sim. | Int. | Slope | $r^2$ | RMSE | $d$ |
|---|---|---|---|---|---|---|---|---|---|
| | | | $cm^3/cm^3$ | kg ha$^{-1}$ | | | | kg ha$^{-1}$ | |
| Tifton | 69 | 0.78 | +0.028 | 2045 | 609 | 0.696 | 0.552 | 410 | 0.859 |
| Plains | 58 | 0.93 | −0.014 | 1960 | −97 | 1.020 | 0.870 | 324 | 0.962 |
| Calhoun | 75 | 0.94 | +0.032 | 2556 | 190 | 0.904 | 0.638 | 485 | 0.888 |
| Midville | 73 | 0.76 | +0.028 | 2470 | 689 | 0.716 | 0.724 | 504 | 0.913 |
| Griffin | 118 | 0.77 | +0.060 | 2485 | 1879 | 0.240 | 0.079 | 505 | 0.553 |

respectively. LFMAX also reached the specified minimum and maximum values (0.93 and 1.2) for the same cultivars since those coefficients were shifted together. Variation in LFMAX was generally within the expected genetic range (Boote and Tollenaar, 1994). We do not imply any true linkage between LFMAX and THRESH, and we have more confidence in the cultivar shifts in LFMAX. Furthermore, a high LFMAX could also be attributed to traits that maintain high leaf photosynthesis: including higher leaf N, slower N mobilization (stay-green), better disease resistance, and nematode resistance.

High yielding cultivars should be characterized by high values for X1 (LFMAX and THRESH) or high values for X2 (FLSH, FLSD, SDPM, SFDUR, and PODUR) or moderately high values for both. Brim and Bryan fit the first case. Cook fit the third case with moderate X1 and X2 values. Young had high X2 and moderate X1. Furthermore, traits for optimizing yield sometimes went in opposite directions, e.g., Thomas had low LFMAX and THRESH, but it had longer seed fill duration. By contrast, Brim, Young, Hutcheson, and DP105 had relatively high LFMAX and THRESH but lower X2 and shorter seed fill duration The pseudo variable X2 attained its lower limit only in the case of Young.

The simulated yield for most of the cultivars (Table 10) was within 2% of the average measured data (Table

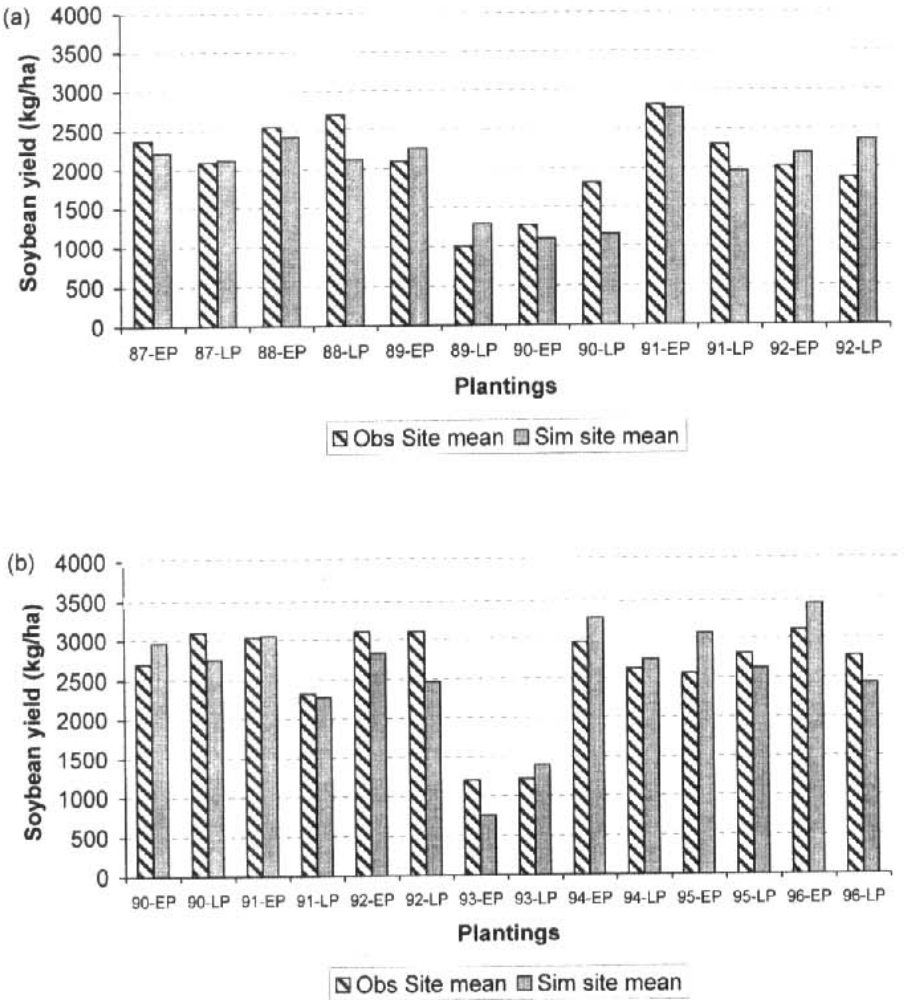


**Fig. 3. Comparison of simulated and observed seed yield at Tifton (a) and Calhoun (b) for early (EP) and late plantings (LP), averaged over all soybean cultivars at each site per planting date.**

**Table 10. Estimates for each soybean cultivar of LFMAX, THRESH, X2, FLSH, FLSD, SDPM, SFDUR and PODUR. The simulated yield (Sim.) is also included. Statistics include $r^2$ of the regression line (rsq.); root mean square error of the predicted minus observed (RMSE); intercept (Int.) and slope of the regression line; index of agreement ($d$), and sample size $n$.**

| Cultivar | $n$ | LFMAX† | THRESH | X2 | FLSH | FLSD | SDPM | SFDUR | PODUR | Sim | Int | Slope | $r^2$ | RMSE | $d$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | mg. $CO_2/m^2$-s | | | PTD‡ | | | | | — kg/ha — | | | | kg/ha | |
| **Group VII** | | | | | | | | | | | | | | | |
| Perrin | 42 | 1.01 | 77.0 | −0.16 | 9.4 | 15.0 | 31.48 | 22.5 | 10.8 | 2169 | 407 | 0.800 | 0.586 | 396 | 0.872 |
| Stonewall | 49 | 0.99 | 76.6 | 0.32 | 7.3 | 12.0 | 31.04 | 24.0 | 9.0 | 2289 | 199 | 0.897 | 0.716 | 404 | 0.917 |
| Hagood | 40 | 1.07 | 78.1 | 0.28 | 7.2 | 11.9 | 30.18 | 23.8 | 9.2 | 2345 | 596 | 0.740 | 0.640 | 414 | 0.892 |
| Cook | 28 | 1.07 | 78.1 | 0.72 | 6.0 | 10.2 | 32.37 | 25.2 | 7.8 | 2448 | 137 | 0.935 | 0.757 | 363 | 0.931 |
| Thomas | 33 | 0.93 | 75.5 | 0.88 | 5.6 | 9.7 | 33.40 | 25.6 | 7.4 | 2229 | 279 | 0.880 | 0.663 | 413 | 0.899 |
| Colquitt | 34 | 1.11 | 78.8 | −0.12 | 8.5 | 13.7 | 28.85 | 22.6 | 10.6 | 2230 | 81 | 0.942 | 0.682 | 384 | 0.902 |
| **Group VI** | | | | | | | | | | | | | | | |
| Brim | 34 | 1.20 | 80.5 | −0.36 | 8.4 | 14.6 | 27.34 | 21.9 | 11.8 | 2534 | 474 | 0.788 | 0.665 | 476 | 0.898 |
| Young | 38 | 1.08 | 78.3 | 1.00 | 11.1 | 18.6 | 28.50 | 20.0 | 15.0 | 2494 | 514 | 0.805 | 0.739 | 411 | 0.925 |
| **Group V** | | | | | | | | | | | | | | | |
| Hutcheson | 32 | 1.08 | 78.3 | −0.24 | 7.4 | 13.8 | 28.04 | 22.3 | 11.2 | 2536 | 784 | 0.697 | 0.684 | 497 | 0.901 |
| Bryan | 41 | 1.17 | 79.9 | 0.04 | 6.4 | 12.5 | 28.52 | 23.1 | 9.9 | 2468 | 390 | 0.836 | 0.720 | 424 | 0.920 |
| DP 105 | 22 | 1.11 | 78.8 | −0.96 | 9.7 | 17.2 | 26.42 | 20.1 | 14.8 | 2266 | 1094 | 0.526 | 0.579 | 559 | 0.840 |

† LFMAX = Light saturated leaf photosynthesis rate at 30°C, 350 vpm $CO_2$ and high light (mg $CL_2/m^2$-s); THRESH = The maximum ratio of (seed/(seed + shell)) at maturity; X2 = a pseudo variable that jointly shifts cultivar "life cycle" coefficients; FLSH = Time between first flower and first pod (PTD‡); FLSD = Time between first flower and first seed (PTD); SDPM = Time between first seed and physiological maturity (PTD); SFDUR = Seed filling duration for pod cohort at standard growth conditions (PTD); PODUR = Time required for cultivar to reach full pod load under optimal conditions (PTD).

‡ PTD = Photothermal days.

2). The RMSE of yield expressed as coefficient of variation (RMSE × 100/mean) was low, ranging from 14.7 (for Cook) to 25.1% (for DP105) of the actual yield means. Linear regression analysis resulted in better fits as evidenced from the closer to 1 slope and lower intercepts for group VII genotypes than for VI and V (Table 10). We were not able to explain the poor fit for DP105, as evidenced by the high RMSE and intercept.

The regression analysis between actual and predicted soybean grain yield for Stonewall (Fig. 4) resulted in a lower intercept (from 300–199 kg $ha^{-1}$) and a closer to unity slope (from 0.858–0.897) from the previous stage of fitting the site characteristics to optimizing yield characteristics; however, that was not the case for every cultivar. The RMSE was marginally improved (decreased from 414–404 kg $ha^{-1}$) for Stonewall but was decreased more than 10% for three cultivars (Cook, Young, and Bryan). The generally large reduction in RMSE between measured and simulated yield was obtained by optimizing site characteristics. However, the small reductions in RMSE associated with optimizing cultivar traits suggests that realistic site input characteristics are more important for yield prediction by CROPGRO than using the "correct" genetic information.

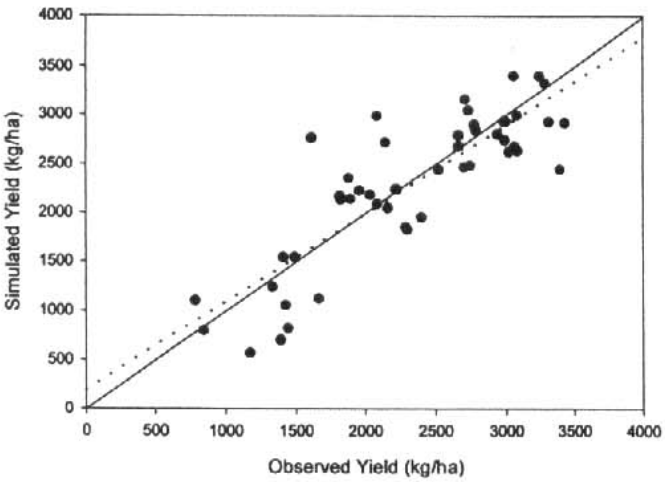


**Fig. 4. Comparison of simulated versus observed seed yield for the soybean cultivar Stonewall.**

## Stability Analysis

Stability analysis and computation of cultivar by environment interactions were pursued with both observed data and simulated responses, to evaluate how well the crop model with solved site traits and genetic (phenology and yield-influencing) traits could reproduce GE interactions. Because the original data set was not orthogonal, we selected a smaller subset of data that was orthogonal, and repeated the entire optimization procedure for site and cultivar (phenology and yield-influencing) traits. Estimated cultivar coefficients for the orthog-

**Table 11. The regression equations of observed yield for each soybean cultivar vs. the environment mean yield (Env) (kg $ha^{-1}$), and of simulated yield for each cultivar vs. the simulated environmental mean yield. Statistics include $r^2$ of the regression line; parentheses show the standard error for the slopes and intercepts.**

| Cultivar | Equation | $r^2$ |
|---|---|---|
| | **Observed yield** | |
| Stonewall = | 0.992 (± 0.125) Env − 134 (± 345) | 0.841 |
| Perrin = | 0.894 (± 0.143) Env + 75 (± 396) | 0.765 |
| Cook = | 1.051 (± 0.112) Env − 34 (± 309) | 0.881 |
| Hagood = | 1.065 (± 0.165) Env − 259 (± 457) | 0.777 |
| Bryan = | 0.990 (± 0.135) Env + 71 (± 373) | 0.818 |
| Hutcheson = | 1.064 (± 0.197) Env − 100 (± 546) | 0.708 |
| Brim = | 0.975 (± 0.160) Env + 225 (± 442) | 0.757 |
| Young = | 0.968 (± 0.126) Env + 126 (± 350) | 0.830 |
| | **Simulated yield** | |
| Stonewall = | 0.983 (± 0.035) Env − 73 (± 97) | 0.985 |
| Perrin = | 0.902 (± 0.111) Env + 113 (± 309) | 0.846 |
| Cook = | 0.982 (± 0.056) Env + 131 (± 154) | 0.963 |
| Hagood = | 0.903 (± 0.075) Env + 187 (± 208) | 0.924 |
| Bryan = | 1.053 (± 0.044) Env − 102 (± 122) | 0.980 |
| Hutcheson = | 1.050 (± 0.089) Env − 72 (± 248) | 0.920 |
| Brim = | 1.082 (± 0.084) Env − 71 (± 233) | 0.932 |
| Young = | 1.046 (± 0.064) Env − 113 (± 179) | 0.956 |

**Table 12. Observed yield, observed and simulated rankings, and percent mean difference between observed and modeled yield for each soybean cultivar.**

| Cultivar | Obs. yield | Obs. ranking | Sim. ranking | (Obs-Sim)/ Obs × 100 |
|---|---|---|---|---|
| | kg ha$^{-1}$ | | | % |
| Perrin | 2504 | 8 | 8 | 2.8 |
| Stonewall | 2561 | 7 | 7 | 1.9 |
| Hagood | 2635 | 6 | 6 | 0.6 |
| Cook | 2821 | 2 | 2 | −0.4 |
| Brim | 2904 | 1 | 1 | −0.8 |
| Young | 2757 | 5 | 5 | −0.7 |
| Hutcheson | 2789 | 3 | 3 | 0.1 |
| Bryan | 2762 | 4 | 4 | 0.3 |

onal subset in some cases differed from those estimated from the whole data set, but this was attributed to the exclusion of late planting trials from the subset.

The yield differences among the cultivars averaged over all sites, according to the Kruskal-Wallis tests, were not significant (a = 0.05) for both observed and simulated data sets. On the other hand, the yield differences among the sites for observed and modeled data were highly significant ($P < 0.001$, df = 13). The crop model was able to reproduce the larger effects of the different environments on observed soybean yield compared with the effects of genetic variability among the different cultivars.

We evaluated the ability of our approach to reproduce the actual GE interactions by regressing the observed and predicted yield of each cultivar against the means of the 14 environments (Table 11). The slopes for actual and simulated yields, according to the $P$ statistic, were not significantly different from 1.0 at a = 0.05. The intercepts were not significantly different from 0.0. No statistically significant differences were identified according to the $T$ statistic (alpha = 0.05) between the

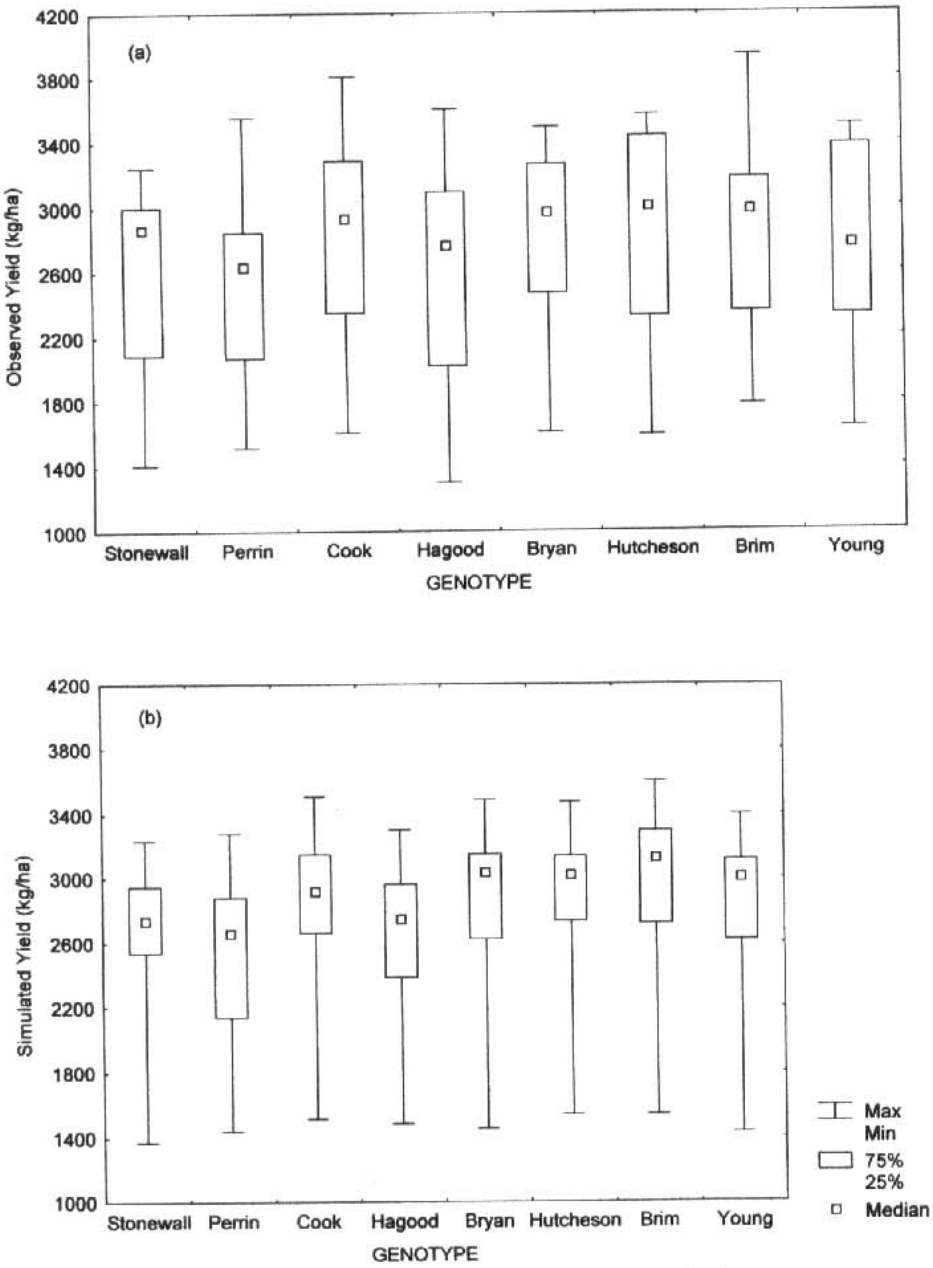


**Fig. 5. Observed (a) and simulated (b) seed yield for each soybean cultivar of the orthogonal subset.**

slopes and the intercepts of the actual and predicted yields for any cultivar. The standard errors for the slopes and intercepts were higher for the observed relationships, which is consistent with higher variability found in the actual yields.

### Cultivar Ranking and Performance

Traditionally, cultivar selection by plant breeders has been solely based on mean genotype performance across environments. To evaluate how well the crop model reproduced the observed yield-based cultivar ranking, we estimated and compared the differences in ranking of cultivars for actual and modeled yield for the orthogonal subset.

The model reproduced correctly the observed yield-based genotype ranking for each cultivar and the mean differences between measured and simulated yields averaged over all environments (Table 12). In addition, simulated yields were within 3% from the mean actual values for all cultivars.

The range of model predictions around the median yield (median yields ±25%) was smaller than that of observations. While the model correctly predicted the rankings of mean (or median) cultivar yields, it did not fully predict the range of variability in yield about the median (Fig. 5a and b). The model underestimation of the actual yield ranges was expected since the CROPGRO model does not allow or consider effects of tillage, pests, and diseases to affect yield. In the field, factors would vary with season and site. In addition, the selection of only one soil type per site and the use of average site characteristics compared with the year-to-year variation in field characteristics (i.e., actual field area moves within the station) may also have contributed to the smaller range simulated by the model. CROPGRO was able to reproduce the lowest measured yields but underpredicted the highest.

The next question is whether the predicted ranking is correct at each environment. The difference in genotype ranking between measured and predicted values was one or less for 61% of the cases (112 cases) (Fig. 6a). In an extra 12% of the simulated cases, the ranking differed from the observations by a ranking of two. The mean simulated yield for each environment was within 100 kg ha$^{-1}$ of the actual value in 20% of the cases (Fig. 6b). An additional 22% of the simulated cases were within 100 and 200 kg ha$^{-1}$ from the measured mean.

## CONCLUSIONS

Time constraints, financial resources, and large numbers of changing cultivars eliminate the possibility that scientists can carry out detailed growth analyses experiments to derive fully cultivar coefficients before culti-

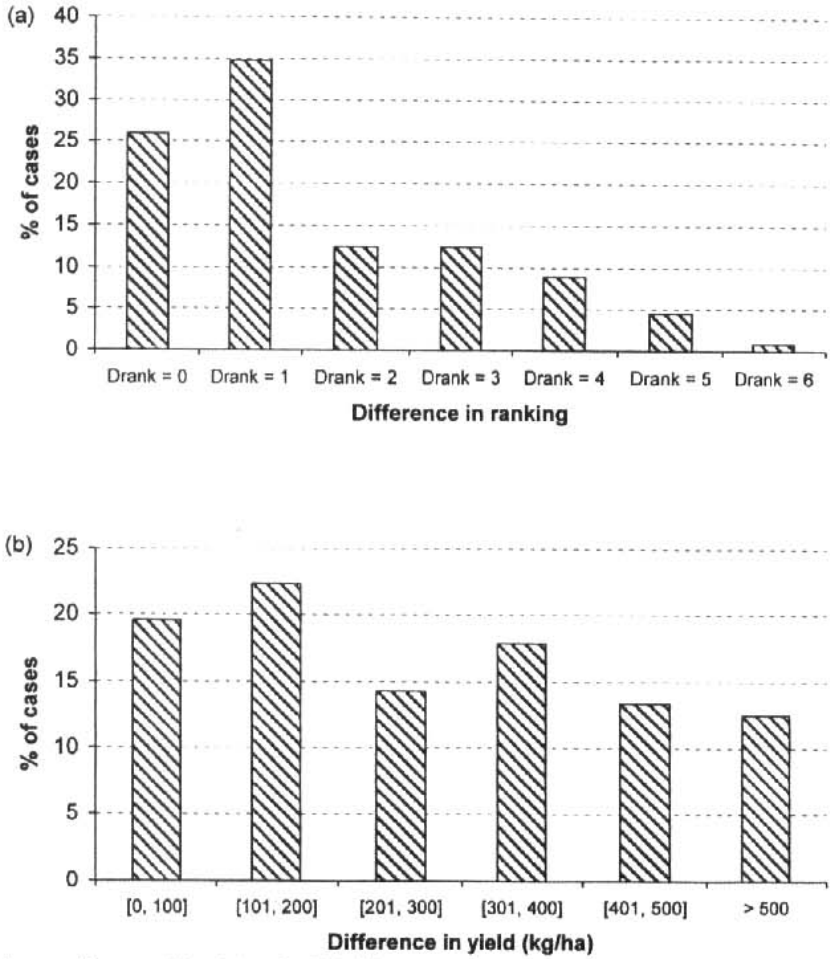


**Fig. 6. The difference in soybean cultivar ranking (a) and yield (|Obs. – Sim|) (b) between observations over all environments (112 cases) of the orthogonal subset.**

vars are released. Our results show that a large number of yield trials representing different environments can be successfully used to derive cultivar information for use in crop models. With the information available in crop performance tests and a systematic approach, it is possible to estimate and provide coefficients describing new cultivars as soon as the cultivars are released. Availability of such information will benefit farmers by allowing use of improved site-specific decision support tools. It may also give researchers new ways to compare and characterize cultivar groupings (daylength sensitivities, life cycle traits, various yield-influencing traits, etc). Knowledge about harvest index, dry matter accumulation potential, and parent-progeny relationships among cultivars could help guide a more improved parameter search. Understanding of parent-progeny relations among cultivars along with the performance prediction ability of the models will help breeders in their efforts to develop higher yielding genotypes. This synergy will become increasingly efficient over time as models evolve to incorporate the genetics that underpin and control physiological processes.

The optimization procedure we applied for site and cultivar (phenology, and yield-influencing) traits enabled the CROPGRO model to reproduce sufficiently the "mean" observed GE interactions; however, the higher variability of observed slopes and intercepts indicate that the model may not be able to simulate the actual GE interactions for the very high or lowest productivity environments. The crop model was successful in reproducing the observed yield-based cultivar ranking across environments.

The approach we described in this paper is the first step towards an efficient and non-time consuming way to derive cultivar coefficients from typical information provided by soybean performance tests. Future steps are needed to incorporate additional procedures to account for the susceptibility or resistance of each of the soybean cultivars to pests and diseases.

## REFERENCES


Allen, E.M., W.D. Batchelor, and T.S. Colvin. 1996. Validation of corn and soybean models in Iowa: Implications for precision farming. ASAE paper No. 96-1006, ASAE, St. Joseph, MI.

Boote, K.J., J.W. Jones, and G. Hoogenboom. 1998. Simulation of crop growth: CROPGRO Model. Chapter 18, p. 651–692. *In* R.M. Peart and R.B. Curry (ed.) Agricultural systems modeling and simulation. Marcel Dekker, Inc., New York.

Boote, K.J., J.W. Jones, G. Hoogenboom, and G.G. Wilkerson. 1997. Evaluation of the CROPGRO-soybean model over a wide range of experiments. p. 113–133. *In* Kropff et al. (ed.) Systems approaches for sustainable agricultural development: Applications of systems approaches at the field level. Kluwer Academic Publishers, Boston.

Boote, K.J., and M. Tollenaar. 1994. Modeling genetic yield potential. p. 533–565. *In* K.J. Boote et al. (ed.) Physiology and determination of crop yield. ASA-CSSA-SSSA, Madison, WI.

Colson, J., A. Bouniols, and J.W. Jones. 1995. Soybean reproductive development: Adapting a model for European cultivars. Agron. J. 87:1129–1139.

Curry, R.B., J.W. Jones, K.J. Boote, R.M. Peart, L.H. Allen, Jr., and N.B. Pickering. 1995. Responses of soybean to predicted climate change in the USA. Chapter 8. *In* C. Rosenzweig et al. (ed.) Climate change and agriculture: Analysis of potential international impacts. ASA Spec. Publ. 59. ASA, CSSA, and SSSA, Madison, WI.

Egli, D.B., and W. Bruening. 1992. Planting date and soybean yield: Evaluation of environmental effects with a crop simulation model: SOYGRO. Agric. Forest Meteorol. 62:19–29.

Goffe, W.L., G. Ferrier, and J. Rogers. 1994. Global optimization of statistical functions with simulated annealing. J. Econometrics 60:65–99.

Grimm, S.S., J.W. Jones, K.J. Boote, and D.C. Herzog. 1994. Modeling the occurrence of reproductive stages after flowering for four soybean cultivars. Agron. J. 86:31–38.

Grimm, S.S., J.W. Jones, K.J. Boote, and J.D. Hesketh. 1993. Parameter estimation for predicting flowering date of soybean cultivars. Crop Sci. 33:137–144.

Heiniger, R.W., R.L. Vanderlip, and S.M. Welch. 1997. Developing guidelines for replanting grain sorghum: I. Validation and sensitivity analysis of the SORKAM sorghum growth model. Agron. J. 9:5–83.

Hodges, T., D. Botner, C. Sakamoto, and Haug J. Hays. 1987. Using the CERES-Maize model to estimate production for the U.S. cornbelt. Agric. Forest Meteorol. 40:293–303.

Hoogenboom, G. 1996. The Georgia Automated Environmental Monitoring Network. p. 343–346. *In* Preprints 22nd Conference on Agricultural and Forest Meteorology with Symposium on Fire and Forest Meteorology. American Meteorological Society, Boston.

Hoogenboom, G., and D.D. Gresham. 1997. Automated Weather Station Network. p. 483–486. *In* K.J. Hatcher (ed.) Proceedings of the 1997 Georgia Water Resources Conference. Institute of Ecology, The University of Georgia, Athens, GA.

Hoogenboom, G., J.W. Jones, P.W. Wilkens, W.D. Batchelor, W.T. Bowen, L.A. Hunt, N.B. Pickering, U. Singh, D.C. Godwin, B. Baer, K.J. Boote, J.T. Ritchie, and J.W. White. 1994. p. 95–244. Crop models. *In* G.Y. Tsuji et al. (ed.) DSSAT v3. University of Hawaii, Honolulu, HI.

Jones, J.W., K.J. Boote, G. Hoogenboom, S.S. Jagtap and, G.G. Wilkerson. 1989. SOYGRO V5.42—Soybean crop growth simulation model, User's guide. Florida Exp. St. J. No. 8304. University of Florida, Gainesville, FL.

Liu, W.T.H., D.M. Botner, and C.M. Sakamoto. 1989. Application of CERES-Maize to yield prediction of a Brazilian maize hybrid. Agric. Forest Meteorol. 45:299–312.

Paz, J.O., W.D. Batchelor, T.S. Colvin, S.D. Logsdon, T.C. Kaspar, and D.L. Karlen. 1998. Calibration of a crop growth model to predict spatial yield variability. Transactions of the ASAE 41: 1527–1534.

Perkins, H.F., J.E. Hook, and N.W. Barbour. 1986. Soil characteristics of selected areas of the Coastal Plain Experiment Station and ABAC Research Farm. p. 62. Res. Bull. 346. The University of Georgia Agricultural Experiment Stations, Athens, GA.

Perkings, H.F., R.A. McCreery, G. Lockaby, and C.E. Perry. 1979. Soils of the Southeast Georgia Branch Experiment Station. p. 51. Res. Bull. 245. The University of Georgia Agricultural Experiment Stations, Athens, GA.

Perkins, H.F., R.B. Moss, and A. Hutchins. 1978. Soils of the Southwest Georgia Branch Experiment Stations. p. 34. Res. Bull. 217. The University of Georgia Agricultural Experiment Stations, Athens, GA.

Perkins, H.F., V.R. Owen, and E.E. Worley. 1983. Soils of the Northwest Georgia Branch Experiment Station (Floyd County Unit). p. 47. Res. Bull. 302. The University of Georgia Agricultural Experiment Stations, Athens, GA.

Perkins, H.F., L.M. Schuman, F.C. Boswell, and V. Owen. 1985. Soil characteristics of the Bledsoe and Beckham research farms of the Georgia Station (Griffin). p. 33. Res. Bull. 332. The University of Georgia Agricultural Experiment Stations, Athens, GA.

Piper, E.L., K.J. Boote, and J.W. Jones. 1998. Evaluation and improvements of crop models using regional cultivar trial data. Trans. ASAE 14:435–446.

Piper, E.L., K.J. Boote, J.W. Jones, and S.S. Grimm. 1996. Comparison of two phenology models for predicting flowering and maturity date of soybean. Crop Sci. 36:1606–1614.

Raymer, P.L., J.L. Day, R.B. Bennet, S.H. Baker, W.D. Branch, and M.G. Stephenson. 1994. 1993 Field crops performance tests: Soybean, peanut, cotton, tobacco, sorghum, and summer annual forages. Res. Rept. 627, University of Georgia, Athens, GA.

Raymer, P.L., J.L. Day, R.B. Bennet, S.H. Baker, W.D. Branch, and

M.G. Stephenson. 1993. 1992 Field crops performance tests: Soybeans, peanuts, cotton, tobacco, sorghum, and summer annual forages. Res. Rept. 618, University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, R.B. Bennet, R.D. Gipson, S.H. Baker, W.D. Branch, and M.G. Stephenson. 1992. 1991 Field crops performance tests: Soybeans, peanuts, cotton, tobacco, sorghum, and summer annual forages. Res. Rept. 609, University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, A.E. Coy, S.H. Baker, W.D. Branch, and S.H. LaHue. 1997. 1996 Field crops performance tests: Soybean, peanut, cotton, tobacco, sorghum, grain millet and summer annual forages. Res. Rept. 644. University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, A.E. Coy, S.H. Baker, W.D. Branch, and M.G. Stephenson. 1996. 1995 Field crops performance tests: Soybean, peanut, cotton, tobacco, sorghum, grain millet, and summer annual forages. Res. Rept. 639. University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, A.E. Coy, S.H. Baker, W.D. Branch, and M.G. Stephenson. 1995. 1994 Field crops performance tests: Soybean, peanut, cotton, tobacco, sorghum, grain millet, and summer annual forages. Res. Rept. 633. University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, C.D. Fisher, and R.H. Heyerdahl. 1989. 1988 Field crops performance tests: Soybeans, peanuts, cotton, tobacco, sorghum, summer annual forages, and sunflowers. Res. Rept. 568, University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, C.D. Fisher, and R.H. Heyerdahl. 1988. 1987 Field Crops performance tests: Soybeans, peanuts, cotton, tobacco, sorghum, summer annual forages, and sunflowers. Res. Rept. 556, University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, R.D. Gipson, S.H. Baker, W.D. Branch, and M.G. Stephenson. 1991. 1990 Field crops performance tests: Soybeans, peanuts, cotton, tobacco, sorghum, and summer annual forages. Res. Rept. 599, University of Georgia, Athens, GA.
Raymer, P.L., J.L. Day, and R.D. Gipson. 1990. 1989 Field crops performance tests. Res. Rept. 589, University of Georgia, Athens, GA.
Ritchie, J.T., D.C. Godwin, and U. Singh. 1989. Soil and water inputs for the IBSNAT models. p. 31–45. *In* Proceedings of IBSNAT Symposium: Decision Support System for Agrotechnology Transfer. University of Hawaii, Honolulu, HI.
Specht, J.E., J.H. Williams, and C.J. Weidenbenner. 1986. Differential responses of soybean genotypes subjected to a seasonal soil water gradient. Crop Sci. 26:922–934.
Tsuji, G.Y., G. Uehara, and S. Balas (ed). 1994. DSSAT version 3. University of Hawaii, Honolulu, HI.
Willmott, C.J. 1982. Some comments on the valuation of model performance. Bull. Am. Meteorological Soc. 63:1309–1313.

# Genetic Factors Influencing High Oleic Acid Content in Spanish Market-Type Peanut Cultivars

Yolanda López,* Olin D. Smith, Scott A. Senseman, and William L. Rooney

## ABSTRACT

**Increasing the ratio of oleic to linoleic acid (O/L) in peanut (*Arachis hypogaea* L.) significantly improves the nutritional and quality attributes of the crop. In currently grown cultivars, the O/L ratio ranges from 0.8 to 2.5. Variation in peanut for O/L ratio has been characterized, and the O/L ratio is digenically inherited at two loci designated as $Ol_1$ and $Ol_2$. Previous research has been conducted with Virginia- and runner-type peanut; however, there have been no reports regarding the inheritance and allele frequency at these loci in Spanish-type peanut. The objectives of this study were to determine if the inheritance of the high oleate trait in Spanish-type peanut is similar to that previously reported and to determine the allelic composition of Spanish-type peanut at $Ol_1$ and $Ol_2$. Six different Spanish-type peanut cultivars (low oleate) were hybridized with F435-2—2 (high oleate). $F_2$ and $BC_1F_1$ progenies were evaluated for the O/L ratio. Segregation patterns indicated that high oleic acid content is digenically inherited in Spanish-type peanut, but there seems to be more allelic variation both within and among these cultivars. In addition, variation within the high and low oleate ratio classes indicated that other factors may be involved in determining the precise O/L ratio.**

Peanut is a globally important protein and oilseed crop. In the USA, peanut is used primarily for food, but throughout the rest of the world, it is utilized as an oilseed crop (McWatters and Cherry, 1982). Like most edible oils, peanut oil is subject to oxidation which results in rancidity. Consequently, the shelf-life of products which contain peanut and peanut oil is limited. Improvements in quality and stability of peanut oil may improve the market value (Knauft and Ozias-Akins, 1995).

The oxidative stability and shelf-life of oil is influenced by the concentrations of specific fatty acids (Fore et al., 1953; Sanders, 1980a,b). In general, saturated fatty acids are less susceptible to oxidation than less saturated fatty acids. Two fatty acids, oleic and linoleic, comprise over 80% of the oil content of peanut. Of these, linoleic acid is less saturated and less stable than oleic acid. Holley and Hammons (1968) reported a strong negative correlation between linoleic acid content and oil stability in peanut. Robertson and Thomas (1976) reported that oil with higher ratios of oleic to linoleic acid retains its quality longer in the seed or as oil. Gorbet and Knauft (1997) found that 'SunOleic 95R', a high oleic runner cultivar, had much longer shelf-life than the traditional runner-type peanut cultivars.

In addition to increasing the stability of peanut oil, increasing the O/L ratio in peanut appears to have health benefits as well. Research has associated high oleic acid with lowered blood serum cholesterol, especially low-density lipoproteins (LDL) in humans (O'Bryne et al., 1997). Renaud et al. (1995) reported a 70% decrease in recurrent myocardial effects when oleic acid levels are increased in plasma fatty acids.

Increasing the O/L ratio in peanut seems to have positive effects on peanut quality and nutritional value. The majority of peanut cultivars average 55% oleic acid and 25% linoleic acid (Knauft et al., 1993). Norden et al.

Dep. of Soil and Crop Sciences, Texas A&M Univ., College Station, TX 77843-2474. This work was supported in part by The Texas Peanut Producer Board and The Peanut CRSP, U.S. Agency for International Development, under grant Number DAN-4048-G-0041-00. Received 21 Jan. 2000. *Corresponding author (y-lopez@tamu.edu).

Published in Crop Sci. 41:51–56 (2001).

**Abbreviations:** BHT, butylated hydroxytoluene; GC, gas chromatography; IDBLs, independent derived backcross lines; LDL, low-density lipoproteins; O/L, oleic to linoleic acid ratio; TAES, Texas Agricultural Experiment Station.