PREDICTING SOYBEAN WEIGHT PER POD

Timothy P. Keller

Michael Schmidt

Catherine Schmidt

PREDICTING SOYBEAN WEIGHT PER POD

Timothy P. Keller
National Agricultural Statistics Service - Research Division

Michael Schmidt
Department of Plant, Soil and General Agriculture
Southern Illinois University, Carbondale

Catherine Schmidt
Department of Plant, Soil and General Agriculture
Southern Illinois University, Carbondale

ABSTRACT

The National Agricultural Statistics Service (NASS) makes predictions of soybean yield at both the state and national level. Experience indicates that the key to improving yield predictions is improving the prediction of soybean weight pod. Towards this end, NASS has entered into a three year cooperative agreement with the Department of Plant, Soil and General Agriculture at Southern Illinois University at Carbondale (SIU-C) to investigate predictive models for soybean weight per pod based on a measurement of pod width and a count of the number of seeds per pod. Aside from the scientific issues involved in this effort, the practical issues of eventually developing procedures which can be implemented in the context of the NASS Objective Yield Survey, which is an immense data collection effort, must be addressed by this cooperative effort. The results obtained from the three years of this cooperative effort will provide a useful starting point for further work in this direction.

1. Introduction

The NASS Soybean Objective Yield Survey: Brief Overview

The National Agricultural Statistics Service (NASS) conducts the Soybean Objective Yield Survey in an Objective Yield Region comprising the eight states: Arkansas, Illinois, Indiana, Iowa, Minnesota, Missouri, Nebraska, and Ohio. In a typical year a total of over 1300 samples are allocated to the Region. Forecasts of acreage, yield, and production are made monthly from the August 1 Crop Report through the November 1 Crop Report, with final estimates published in January.

Sample fields for the Soybean Objective Yield Survey are selected from farms reporting soybeans for harvest in the area frame of the June Agricultural Survey. The sample fields are selected with probability proportional to area, consequently the sample design is a self-weighting sample with respect to the area of soybeans planted. Data are collected from each sample at monthly intervals starting in late July and continuing through December, or until the sample has been harvested. Each month during the Objective Yield Survey, data

collected from the sample fields are used to produce indications of planted acres (August only), acres for harvest, and yield.

A sample consists of two independently located units, each of which consists of two parallel 3.5 foot sections of row partitioned into a 3 foot section and a 6 inch section. Field enumerators use a random number of rows along the edge of the field and a random number of paces into the field to locate each unit. At harvest, the beans from the sample plots are weighed to determine the final yield from that sample. Plant counts are made in the full unit while detailed fruit counts are limited to the 6 inch section at the end of each row, which usually contains 1 to 4 plants. All beans from the 3.5 feet of each row are picked and weighed at harvest to establish gross yield. The yield is reported in bushels per acre (by definition a bushel of soybeans is 60 pounds at 12.5 percent moisture by weight). Harvest loss is measured in separate units located near the monthly yield plots.

Weight per Pod and Yield

Soybean seed yield for a given area may be written as:

seed yield = plants/area × pods/plant × seed weight/pod

The accuracy of yield predictions is most severely limited by the difficulty in making accurate early season predictions of seed weight per pod. At the level of the Objective Yield Region, or a state within the region, NASS currently uses the average number of lateral branches per acre as the independent variable to predict weight per pod on August 1; and the average number of pods per acre as the independent variable to predict weight per pod on September 1. For October through December, the average weight per pod of the soybeans harvested to date are used to predict final harvested weight per pod. At the level of an individual Objective Yield sample, weight per pod is predicted via the 5 year moving-average of the corresponding state level weight per pod. None of these independent variables provide very good predictors of weight per pod; and there is no scientific basis for expecting that they should.

The NASS Objective Yield Survey measures variables *endogenous* to the plant, e.g. pods per plant, or plants per acre,--- as contrasted with variables *exogenous* to the plant, e.g. rainfall, or growing degree days, --- as the basis for predictions of yield. A natural extension to this approach is to measure variables endogenous to a fruiting form as a basis for predictions of fruit weight. For soybeans, measurements of plant characteristics alone are probably inadequate for an accurate August 1 prediction of weight per pod at harvest. However, measurements on pods in the R5 to R6 stage of development, which under full season production corresponds to middle or late August, may provide the basis for a significantly improved September 1 prediction of seed weight per pod.

Predicting Weight Per Pod via Seeds per Pod and Weight per Seed

Seed weight per pod may be written as:

Applied Statistics in Agriculture

weight/pod = seeds/pod × weight/seed.

If one has unbiased estimators S and W of seeds/pod and weight/seed, respectively, then the basic equation

$$E[SW] = E[S] \times E[W] + cov(S, W)$$

implies that if S and W are uncorrelated, SW is an unbiased estimator of weight per pod. Even if S and W are correlated, the bias will be acceptably small for a sufficiently large sample size.

Overview of the Cooperative Research Agreement With Southern Illinois University

NASS and the Department of Plant, Soil and General Agriculture at Southern Illinois University Carbondale (SIU- C) have entered into a cooperative agreement to investigate the potential for predicting harvested seeds per pod via a count of seeds per pod on unshelled pods, and for predicting harvested weight per seed via a measurement of the across-the-seams width of soybean pods. This project comprises the three growing seasons covered by the period from July 1, 1999 to December 1, 2001.

2. Materials and Methods

Each of the three years an indeterminate Maturity Group IV soybean variety will be selected and planted for full season soybean production. The planted area designated for this study will be sufficiently large to accommodate 30 plots; each plot consisting of 12 rows, each 24' long, planted at 30" row spacing. Individual plots will be configured as depicted in Figure 1. Rows 1 and 12 will serve as border rows. Rows 2, 5, 8, and 11 will be used for collection of plant samples. Rows 3 and 4, 6 and 7, 9 and 10 will be harvested as pairs of rows for seed yield after end trimming to 20 feet. Plots will be established at flowering (growth stage R1). On each of three sample dates, eight individual plants will be collected from each plot, two plants per sample row. To insure a random selection of plants and maintain a constant border effect on plants within the row, plants closest to 36" and 144" from the beginning of the plot will be sampled for the first sample date. For successive sample dates, plants will be sampled at 24" intervals from those sampled on the first date. The sample dates will begin during growth stage R5, be positioned at approximately weekly intervals, and occur prior to August 25. This last date was selected to allow for a September 1st yield prediction. This design allows for 30 replications, each having an eight plant sub-sample for estimation of plant characteristics.

Characteristics to be measured on a whole plot basis are:

1) Plant stand at flowering, at the first sample date, and at harvest. Plant stand will be calculated based on the mean number of plants counted in the center 10 feet of row from rows 4, 7 and 10.

2) Seed yield based on the mean seed weight of the 3 pairs of harvested rows, each 20' in length.

3)   Seed size based on a mean of three 1000 seed samples, with samples to be taken from each pair of harvested rows.

Characteristics to be measured on a sub-sample plant basis are:

1) The number of trifoliolate nodes on each plant.

2) The number of pods, at least ½ cm in length, on each plant.

3) The width, to the nearest 0.1 mm, of the most fully developed pod at each node on each plant, with measurement to be taken at the seed position on the pod nearest the stem.

4) The number of seeds in each pod selected for the pod width measurement.

5) Pods per node will be calculated from number of nodes and number of pods.

These characteristics will be measured for plants removed from the field. The unit of analysis for both predictor and predicted variables will be the plot. In other words, key relationships will be investigated using variables averaged to the plot level. This sampling technique may be described as *partial destructive sampling* in that, while individual plants will be removed from the field for subsequent measurements, the unit of analysis, which is a plot, remains substantially undisturbed.

3. Factors Affecting Relationships between Model Variables

The problem of predicting soybean weight per pod has been split into the dual problems of predicting seeds per pod and predicting weight per seed. Our approach to the first problem is based on a visual and tactile determination of seeds per pod in unshelled pods. Our approach to the second problem, motivated by simple geometry, is based on a measurement of across-the-seams pod width. To complete the journey from these simple ideas to a model sophisticated enough to reflect the subtleties of reality, it is necessary to identify and investigate the factors influencing the relationship between predictive and predicted variables. Some key factors are:

(1) Plant architecture.
(2) Growing conditions.
(3) Time at which measurements are taken.
(4) Plant diseases.

These factors are so highly interrelated that a linear presentation, factor by factor, is difficult. With this in mind, let us examine what this investigation reveals about the effects of these factors and their interactions.

Variation of Pod Width and Seeds per Pod with Nodal Position and Other Factors

Applied Statistics in Agriculture

Figure 2 exhibits the variation of pod width with nodal position. In an operational setting it would be impractical to make measurements on pods from all nodes. This is particularly true if measurements were to be made *in situ*. Hence one of the goals of our investigation is identify a small set of nodes from which to select the pods on which measurements are taken. Clearly, the choice of nodes makes a difference.

Figure 2 also demonstrates the importance of time to the model. Seeds are developing rapidly at this point in the season. It is particularly striking that pod width was increasing so rapidly in spite of the inadequate rainfall received in 1999. (See Table 1.)

Figure 3 shows that the number of seeds per pod has achieved its value at harvest by the time of sampling event 1. The measurements from the 2000 season indicate that seeds per pod, as well as pod width, varies significantly with nodal position.

A Comparison of the 1999 and 2000 Seasons

The amount of rainfall received during the 2000 season was much more conducive to soybean growth than that received during the 1999 season. (See Table 1.) Better growing conditions and better germination rates in 2000 led to higher plant densities: 96,500 plants per acre in 2000, as compared with only 58,750 plants per acre in 1999. (Row spacing was 30" for both seasons.) It is well known that the development of lateral branches is inversely related to plant density. This was borne out by the observed data: the fraction of measured pods from lateral branches was 12.46% in 1999 and 8.89% in 2000. In this study, nodal position is used to refer both to pods along the main stem and also to pods from lateral branches. For both 1999 and 2000, development of lateral branches were almost entirely confined to nodal positions 1 to 6. The data suggest that the differences in both pod width and number of seeds per pod for pods from lateral branches and as compared with pods from the main stem are not large enough to be of practical significance, especially at the plant densities which maintain in present day commercial soybean production. In both 1999 and 2000 the bulk of the measured pods came from nodal positions 1-20. In 1999 the pods tended to be from lower nodal positions than in 2000. This may be related to the fact that the average height of the 240 samples plants was 93 cm in 2000, but only 74 cm in 1999. Table 2 summarizes the development of some key plant characteristics by event.

The Effect of SDS

Sudden Death Syndrome (SDS), a plant disease caused by the soil borne fungus *Fusarium solani*, made its appearance in the research field in 2000. Each of the 30 plots was evaluated and assigned a disease index, DX, reflecting both the intensity and spatial prevalence of the disease symptoms. Values of $DX \leq 20$ are considered to correspond to only "mild" to "moderate" levels of SDS. Nevertheless, Figure 4-A shows significant reduction in yields due to SDS. The sample correlation coefficient computed using these data was -0.8544.

Fusarium solani produces a toxin, called monorden, that is translocated throughout the plant.

Secondary veins of infected plants are plugged by the toxin causing a reduction in the water supply to the leaves. This reduces yield, by reducing seed size (Figure 4-B), and, at higher levels of the disease, by reducing pod numbers. The evidence for an associated decrease in number of seeds per pod is more equivocal. (Figure 5.) The discovery of the toxin monorden in described in [2]; some of the mechanisms by which the toxin induces SDS are described in [3] .

There is also evidence, as exhibited by Figure 6, that the *relationship* between seed weight and pod width is altered by the incidence of SDS. The slope of the regression line corresponding to the 9 plots with DX $\geq$ 10 was not significantly different from zero, with a p-value of 0.88 associated with the standard test of the null hypothesis that the slope is zero. (See Section 11.4 of [1] .) But the slope of the regression line corresponding to the 21 plots with DX < 10 was significantly larger than the slope of the regression line for the 9 plots with DX $\geq$ 10, with a p-value of 0.034 associated with the standard test of the null hypothesis that the slope of one regression line exceeds the slope of a second regression line. (See Section 11.6 , pages 288 - 290 for a discussion of the standard statistical methodology.) One conjectures that under conditions of adequate soil moisture the relative concentration of photosynthate delivered to the developing seeds from plants effected by SDS is lowered; or, to put it another way, the fraction of seed weight lost in dry down is greater. SDS may also change the *mix* of photosynthates delivered to the developing seeds, which would change the density of the seeds. Soybeans from the 2000 season test field were observed to weigh in at significantly less than the standard 60 pounds per bushel (as a unit of volume), which supports this conjecture. With all this in mind, the analysis of pod width as a predictor of weight per seed for the 2000 season will use only the 21 plots with values of the SDS index DX which are less than 10.

4. Predicting Weight Per Seed

We make the usual assumptions for a fixed effect linear model relating harvested seed weight, H, and pod width, w. That is, we assume that $H = f(w) + \varepsilon_w$, where for all values of w, ε_w is a normal random variable with zero mean and variance constant with respect to w; and for all pairs w, z , w $\neq$ z , ε_w and ε_z are independent.

A soybean at harvest is more or less spherical; hence geometry suggests that a third degree polynomial is a reasonable choice for the function f. However, in all of the applications subsequently investigated, computations show that the parameters corresponding to the quadratic and cubic terms are not significantly different from zero. Calculating pod width at the level of a plot as the independent variable gives a narrower range of values than that obtained using pod width at the level of a plant, and much narrower than the range of values obtained measuring pod width for individual pods. Thus there may be higher order effects not detected by the present analysis: by definition, *any* differentiable function f (t) is approximately linear for values of t in a short interval. In the following discussion, we assume the function f has the form $f(t) = \alpha + \beta t$.

From the standpoint of (future) operational practicality, it seems reasonable to investigate using pod widths calculated using measurements on pods chosen from just four consecutive nodes. Using one or two nodes significantly lessens the predictive power of the model; using five or six nodes gives very little, if any, improvement in predictive power. Tables 3-A&B show how the

coefficient of determination varies with the four nodes chosen to compute plot level pod width. The coefficient of determination can sometimes increase significantly with the exclusion of a few "outliers". Thus the values of R^2, presented in Tables 3-A&B do not give a complete picture of how well the pod width based a given set of nodes performs as a predictor of weight per seed . For example, based on a naive outlier analysis, nodes 1- 4 were judged to be the best choice of nodes at event 3 in 1999, even though the largest value of R^2 in the right column of Table 3-B corresponds to nodes 7-10. One of our research goals is to develop a technique of outlier analysis based on an understanding of the relevant biological and physical relationships.

Table 3-A also shows how the predictive power of a pod width measurement increases at times closer to physiological maturity. Time is best thought of, not in terms of calendrical days, but in terms of how much photosynthetic activity lies ahead. The relatively better values of R^2 obtained in 1999, even at event 1, than in 2000 reflect the fact that under the drought conditions of 1999 the seeds were much closer to their final seed weight at the same date. In Figure 7&8, the plots of seed weight versus pod width for the "best" four nodes, together with the associated regression lines are presented. ($p < 0.05$)

5. Predicting Seeds Per Pod

Justification for Counting Seeds at the Most Developed Pod at each Node

Informal experimentation shows that the pod width measurements obtained using digital readout calipers, such as those employed in this study, produce values which are consistent from person to person. Such experimentation also shows that a subjective judgment of which pod is the most developed pod at each node that employs both the visual *and* tactile senses does indeed correspond to the widest pod at a given node. From the standpoint of statistical efficiency of estimation alone, it would be better to base an estimate of the number of seeds per pod on an average over *all* the pods at every node of every plant chosen to be sampled. From a practical standpoint, this would require a prohibitive amount of effort. It is also impractical to expect human judgement to select a pod at each node in a truly "random" manner. Hence we have chosen to base our prediction of mean number of seeds per pod on a count of seeds per pod for those closed pods an enumerator would be most likely to choose if given no specific instructions in the matter: the most developed pod at each node.

Modeling Seeds per Pod

It is apparent that a count of seeds for the most developed pod at each node gives an overestimate of the mean number of seeds per pod for the population of all pods. A simple way to "model down" to a reasonable estimate of mean number of seeds per pod is to assume that the "most developed" pod at each node is also the pod at that node with the maximum number of seeds per pod at that node. Secondly, assume that the number of seeds per pod for the pods at a given node of a randomly chosen plant are independent random variables. Soybean pods have 1, 2, 3 or 4 seeds; but the proportion of pods with 4 seeds is generally much less than 1%, hence it is satisfactory to assume a randomly chosen pod has 1, 2 or 3 seeds. The majority of pods have 2 or 3 seeds, and it is hard to accurately estimate the probability a pod has only a single seed using the

sort of data collected in this study, which are essentially order statistics for the number of seeds per pod at a node. Hence, letting S denote the number of seeds per pod of a randomly chosen pod, we define the parameter $\beta \equiv P(S = 1 \mid S \leq 2)$, to be set by "expert judgment."

Letting $p = P(S \leq 2)$, one has

$$P(S = 1) = \beta p, P(S = 2) = (1 - \beta)p, P(S = 3) = 1 - p$$

Then one finds

$$(1) \quad E[S] = 3 - (1 + \beta)p$$

Given that there are k pods at a given node, with the corresponding number of seeds per pod denoted $S_1, S_2, \ldots, S_k$, let $U = \max\{S_1, S_2, \ldots, S_k\}$. Then under the assumptions of the model:

$$(2) \quad E[U] = 3 - (1 - \beta^k)p^k$$

From equations (1) and (2) it follows that for a fixed value of k:

$$(3) \quad E[S] = 3 - (1 + \beta)[\frac{3 - E[U]}{1 - \beta^k}]^{\frac{1}{k}}$$

Recognizing that the number of pods at a node is a random variable, call it K, equation (3) means

$$(4) \quad E[S \mid K = k] = 3 - (1 + \beta)[\frac{3 - E[U]}{1 - \beta^k}]^{\frac{1}{k}}$$

and hence

$$(5) \quad E[S] = E_K[E[S \mid K = k]] = E_K[3 - (1 + \beta)[\frac{3 - E[U]}{1 - \beta^K}]^{\frac{1}{K}}]$$

To calculate the mean number of seeds per pod using (5) requires knowledge of the distribution of K, the number of pods at a node. A cruder estimate of E[S], requiring less information, is obtained by replacing k on the right side of equation (4) by an estimate of E[K]. One may verify that for $k \geq 1$ the expression on the right side of equation (4) is a convex function of the variable k; hence by Jensen's Inequality:

$$(6) \quad E[S] \geq 3 - (1 + \beta)[\frac{3 - E[U]}{1 - \beta^{E[K]}}]^{\frac{1}{E[K]}}$$

Our estimator for E[S] is obtained by taking the right side of (6) and replacing E[K] by the plot level average number of pods per node, and replacing E[U] by the plot level average number of seeds per pod for the most developed pod at each node. In both cases the average is computed across all three sampling events, since pods per node and seeds per pod have attained their final value by the first sampling event. (See Table 2.)

Applied Statistics in Agriculture

Under the assumptions of our model, equation (6) suggests that this estimator for E[S] is negatively biased. However, the right side of (4) is a fairly flat function of k for k ≥ 1, so the size of the bias should be acceptable.

For both 1999 and 2000 the value of β was taken to 0.20. Our estimate of E [S] is relatively insensitive to changes in the parameter β : taking $\beta = 0$ only changes the value of the estimate at the field level by about 4%. The values of the estimate obtained, which are listed in Tables 5-A&B, are a priori quite reasonable; however, more work is needed to evaluate the performance of our estimator for seeds per pod.

6. Computing a Yield Prediction

The accuracy of a yield prediction based on the product identity

$$\text{seed yield} \ = \ \text{plants/area} \times \text{pods/plant} \times \text{seeds/pod} \ \times \text{weight/seed}$$

depends on the accuracy with which each term in the product is predicted. The focus of our research has been on predicting the last two terms of this product: there is little point in "reinventing the wheel" given the relative success in predicting plant density and pods per plant which NASS already enjoys. Nevertheless, for the sake of completeness of exposition, the values of the principal model variables at the plot level, and the corresponding plot level yield predictions at event 1 of the 1999 season, and event 3 of the 2000 season, are presented in Tables 5-A&B, respectively.

The values of predicted yield for both 1999 and 2000 overestimate the true yield. The bias in predicted yield seems to be positively correlated with number of pods per acre (see Figures 9-A&B) in both 1999 and 2000. Recall that the number of pods per plant which is being recorded is the number of pods ½ cm or more in length. This is an overestimate of the number of pods per plant which will develop beans which are *large enough to be harvested*. Moreover, the competition among pods as the number of pods per acre increases as the number of pods per acre increases, which means that as the number of pods per acre increases, so also does the fraction of pods which will not contribute beans at harvest. This accounts for the observed upward bias in the yield prediction which increases with number of pods per acre. (Even the observation that at very *low* values for pods per acre the bias in the yield prediction is slightly *negative* can be accounted for by the possibility that a few of the pods ½ cm or more in length have not been counted.) As further evidence that the bias in yield arises from a bias in the estimate of number of pods per acre, a predicted yield is computed using the plot level weight per seed observed at harvest, — the closest thing to a "true" value of weight per seed available. We note, referring to Figures 9-A&B again, that this predicted yield has the essentially the same pattern of bias as the prediction using the modeled values for plot level weight per seed.

Relationship of Other Variables with Plant Density and Pods Per Plant

The foregoing discussion leads one to the question of whether either the measured pod widthor the observed number of seeds per pod are related to plant density or number of pods per plant.

As indicated by Table 4, the measured pod width does not seem to be correlated with number of pods per plant. The data from the 1999 season do not indicate a correlation of pod width with plant density; but the 2000 data did produce a statistically significant correlation. Number of seeds per pod at the most developed pod was not observed to have a statistically significant correlation with either plant density or number of pod of pods per plant in either 1999 or 2000.
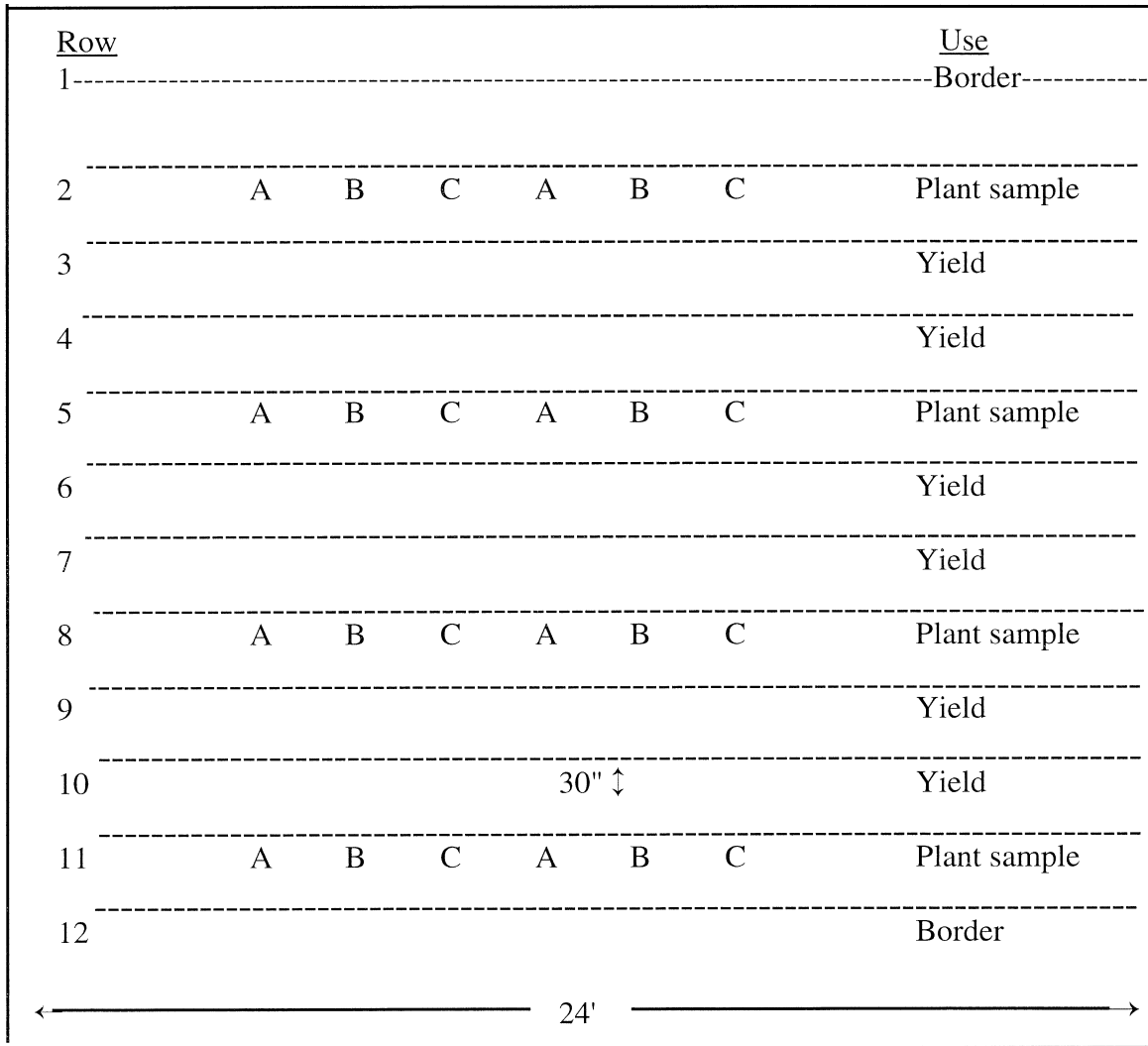
7. Summary

Simple linear regression of plot level weight per seed on plot level pod width gives a better fit than we had initially expected. At a sampling rate of only eight plants per plot per sampling event, we had feared that we might obtain data plots very similar to the pattern of a shotgun blast. Using measurements on eight plants per plot to estimate plot level averages means one has the case of regression for which the independent variable is observed with error. Such an experimental design produces a "downward distortion" [4] in the observed value of R^2. (The reader is referred to [4] for a statistical analysis of the general phenomenon.) Eight plants was the most we could do, under the constraints of a fixed budget, and the necessity to have a reasonably large number of plots.

On the other hand, we suspect that the problem of adequately predicting the number of seeds per pod at harvest might be more difficult than anticipated. We are currently working on finding the resources to collect data at harvest with which to make an estimate of number of seeds per pod at the plot level for the 2001 season, and thereby obtain an objective measure of how well our model predicts the number of seeds per pod. Given sufficient resources, a count of pods per plant at harvest, and a count of pods 1 cm or more in August would also be of interest to obtain.

By the end of the three year cooperative agreement between NASS and SIU-C we hope to have acquired enough basic knowledge about seed development and the practical issues surrounding data collection to begin thinking about developing operational procedures.

References

1. Brownlee, K. A., Statistical Theory and Methodology in Science and Engineering, John Wiley and Sons, Inc., (1960) 288-290

2. Jin, H., G. L. Hartman, C. D. Nickell, J. M Widholm, Phytotoxicity of culture filtrates from *Fusarium solani* isolates, the causal agent of sudden death syndrome of soybean, *Plant Dis.* **80** (1996) 922-927 .

3. Jin, H., G. L. Hartman, C. D. Nickell, J. M Widholm, Characterization and purification of a phytotoxin produced by *Fusarium solani* isolates, the causal agent of sudden death syndrome of soybean, *Phytopathology* **86** (1996) 277-282

4. Snedecor, G. W., Cochran, W. G., Statistical Methods, The Iowa State University Press, Ames, Iowa, Sixth Edition, (1967) 164-166

Row							Use
1---Border-----------
2	A	B	C	A	B	C	Plant sample
3							Yield
4							Yield
5	A	B	C	A	B	C	Plant sample
6							Yield
7							Yield
8	A	B	C	A	B	C	Plant sample
9							Yield
10				30" ↕			Yield
11	A	B	C	A	B	C	Plant sample
12							Border

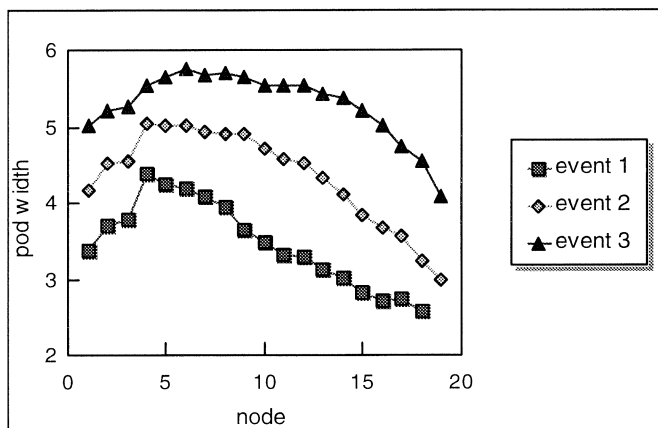←————————————————— 24' —————————————————→

A = location of plants samples at first date.
B = location of plants sampled at second date.
C = location of plants sampled at third date.

FIGURE 1. Single plot design for soybean weight per pod prediction study.

Plot Level Pod Width by Node and Event
2000 Season
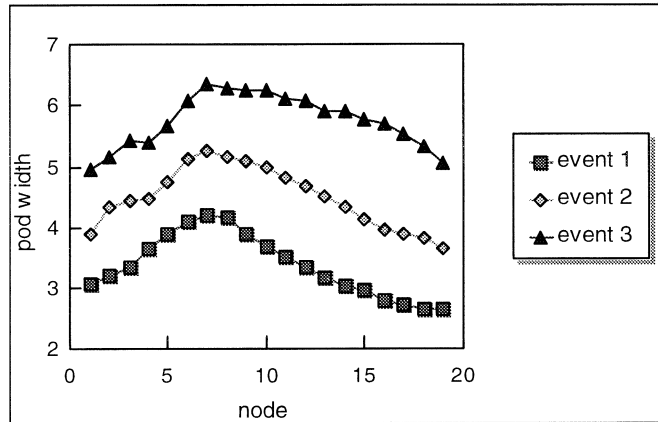
Plot Level Pod Width by Node and Event
1999 Season

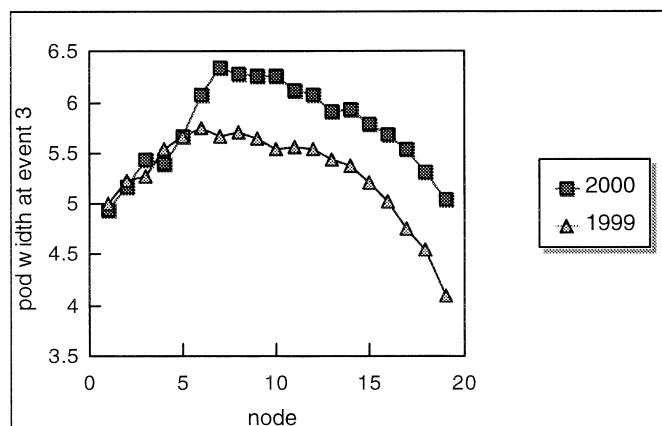

Comparison of Plot Level Pod Width: 1999 vs. 2000

FIGURE 2

Rainfall (inches) and Temperature Agronomy Research Center - Carbondale					
	1999 Rainfall	2000 Rainfall	30 Yr Average Rainfall	1999 Temp (°F)	2000 Temp (°F)
Jan	5.78	N/A	2.92	34.8	N/A
Feb	2.6	N/A	2.65	42.9	N/A
Mar	2.43	N/A	4.4	41.8	N/A
Apr	7.5	2.36	4.7	59.3	55
May	2.72	4.49	5	65.1	68.2
Jun	4.78	12.51	4.5	73.5	72.4
Jul	1.85	2.86	3.24	80.1	75.6
Aug	1.7	2.6	4.1	73.7	77.5
Sep	0.7	1.8	3.11	67.3	67.1
Oct	3.44	1.1	3.35	57.2	60
Nov	0.61	N/A	4.67	51.8	N/A
Dec	5.45	N/A	3.78	39.1	N/A
Total (May -Sept.)	11.75	24.26	19.95		

| *Some Key Dates :*	**1999**	**2000**
Research Field Planted	May 25	May 17
Sampling Event 1	August 10-11	August 7
Sampling Event 2	August 18-19	August 14
Sampling Event 3	August 25-26	August 21
Research Field Harvested	September 10	October 11

TABLE 1

MEAN VALUES OF SOME KEY PLANT CHARACTERISTICS BY EVENT

	Event Means : 1999				Event Means : 2000			
	Number of Seeds:Most Developed Pod	Predicted Number of Seeds	Pods per Plant	Pod Width Most Developed Pod (mm)	Number of Seeds: Most Developed Pod	Predicted Number of Seeds	Pods per Plant	Pod Width Most Developed Pod (mm)
Event 1	2.94	2.50	59.1	3.73	2.67	2.18	51.30	3.52
Event 2	2.89	2.30	84.0	4.40	2.72	2.20	55.81	4.63
Event 3	2.88	2.31	75.4	5.30	2.71	2.18	56.06	5.90

TABLE 2
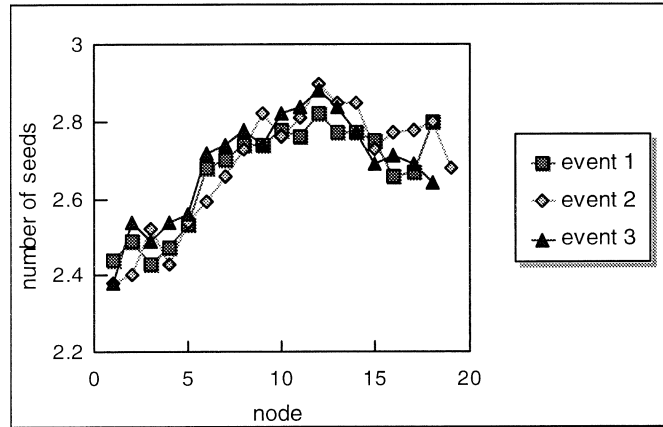
**Plot Level Seeds per Pod for Most Developed Pod at Each Node
2000 Season**
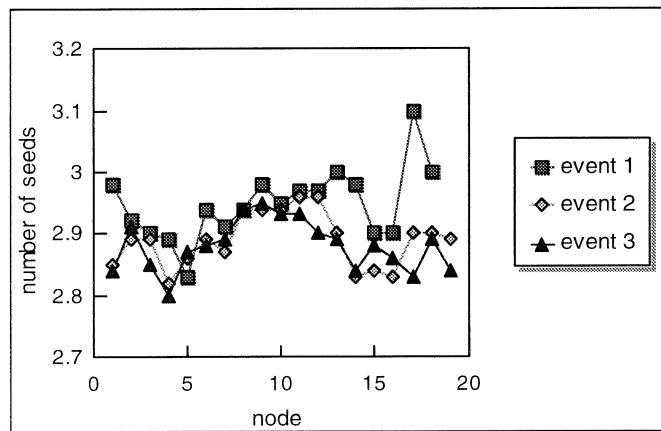
**Plot Level Seeds per Pod for Most Developed Pod at Each Node
1999 Season**
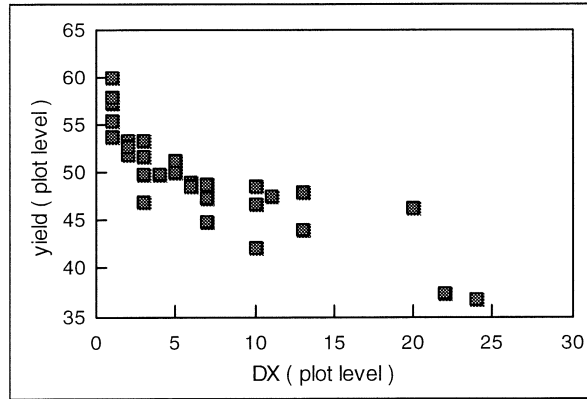
FIGURE 3

The Effect of SDS on Yield

FIGURE 4-A

The Effect of SDS on Weight Per Seed (Grams)

FIGURE 4-B

The Effect of SDS on Seeds per Pod for the Most Developed Pod (All Events)
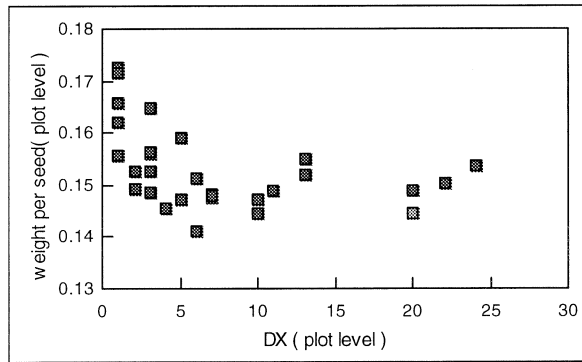
FIGURE 5

Plot Level Pod Width at Most Developed Pod vs. Seed Weight
Event 3 - 2000 Season
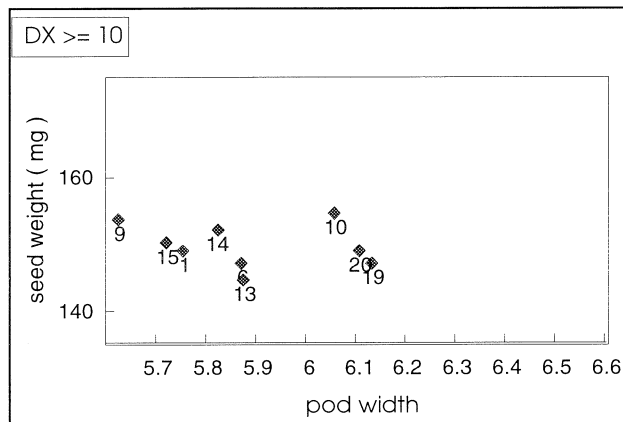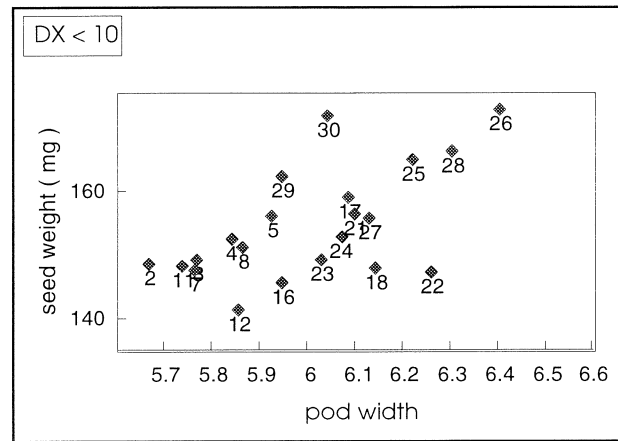
The plots with DX ≥ 10 are plots 1,6,9,10,13,14,15,19, and 20.
(Points in both charts are labeled with the corresponding plot numbers.)

FIGURE 6

Values of R² for Simple Linear Regression of Seed Weight on Pod Width ;
Plot Level Pod Width Computed Using Nodes i to i + 3 .

2000 Season

		EVENT 3	EVENT 2	EVENT 1
i =	1	0.2119	0.0653	0.0184
	2	0.2165	0.1011	0.0431
	3	0.1870	0.1359	0.0311
	4	0.2702	0.1629	0.0009
	5	0.2661	0.1217	0.0068
	6	0.2412	0.0790	0.0017
	7	0.2884	0.1167	0.0248
	8	0.2649	0.1287	0.0271
	9	0.2700	0.1951	0.0486
	10	0.3047	0.2511	0.0165
	11	0.3552	0.2048	0.0019
	12	**0.3900**	**0.2503**	**0.0265**
	13	0.3666	0.1950	0.0250
	14	0.3120	0.1223	0.0440
	15	0.2547	0.1059	0.0676
	16	0.1421	0.0411	0.0721
ALL NODES		**0.3062**	**0.2132**	**0.0703**

TABLE 3- A

1999 Season

		EVENT 3	EVENT 1
i =	1	0.1614	0.2012
	2	0.1849	0.3215
	3	0.1649	0.3701
	4	0.1779	0.3257
	5	0.1751	0.3333
	6	0.1131	0.2853
	7	0.1872	0.2349
	8	0.1160	0.2007
	9	0.1531	0.0101
	10	0.1703	0.0556
	11	0.0956	0.0259
	12	0.0677	0.0139
ALL NODES		**0.2166**	**0.2902**

TABLE 3- B

Simple Linear Regression of Plot Level Pod Width (Nodes 3,4,5,6) for Most Developed Pod on Weight per Seed

1999 Season - Event 1
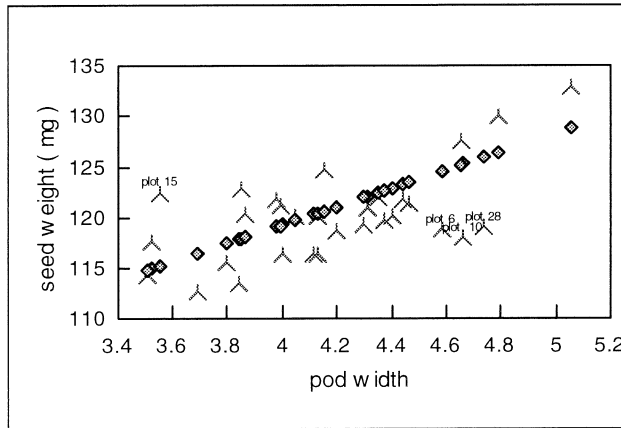
Values of R^2:

Using all 30 data points :	0.3701 .
Excluding plots 6,10 and 28 :	0.5312 .
Excluding plots 6,10, 28 and 15 :	0.6350

Simple Linear Regression of Plot Level Pod Width (Nodes 1,2,3,4) for Most Developed Pod on Weight per Seed
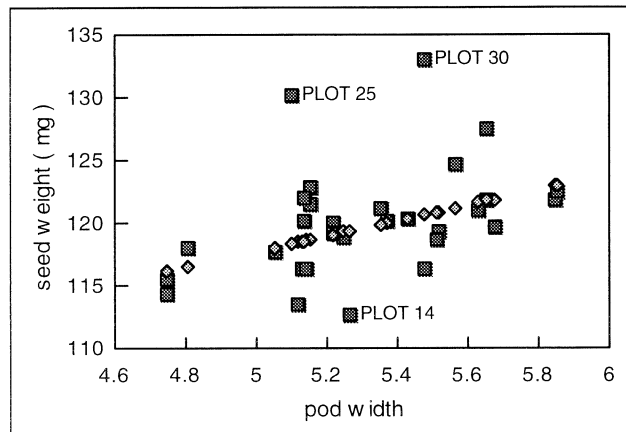
1999 Season - Event 3

Values of R^2:

Using all 30 data points :	0.1614 .
Excluding plots 14,25 and 30 :	0.3672 .

FIGURE   7

**Simple Linear Regression of Plot Level Pod Width (Nodes 12,13,14,15) for
Most Developed Pod on Weight per Seed**
2000 Season - Event 3

Values of R^2:   Using all 21 data points with DX < 10 :   0.3900 .

                        Excluding plot 22 :   0.5079 .

**Simple Linear Regression of Plot Level Pod Width (Nodes 12,13,14,15) for Most
Developed Pod on Weight per Seed**
2000 Season - Event 2

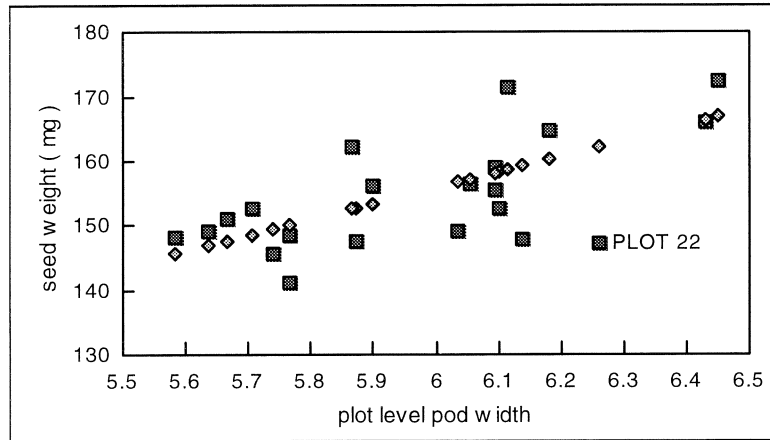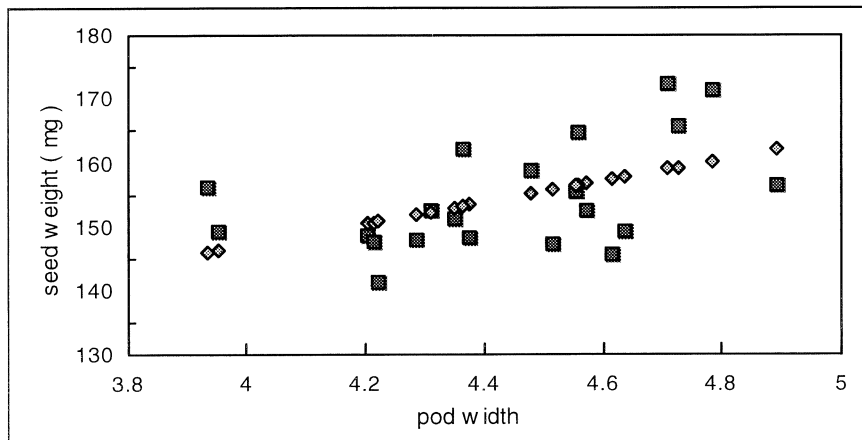Value of R^2 (Using all 21 data points with DX < 10) : 0.2503

FIGURE 8

Correlation Table : 1999 - Event 3

	Pod Width	Seeds per Pod	Plant Population	Pods Per Plant
Pod Width[*]		- 0.0885 (0.6420)	-0.0416 (0.8721)	0.0577 (0.7621)
Seeds per Pod			0.2417 (0.1981)	0.0282 (0.8826)
Plant Population				-0.5637 (0.0012)

Key :   Value of Sample Correlation Coefficient
(Prob | r | > | Value | given $\rho = 0$)

[*] Pod Width is the plot level width of the most developed pod at each node
for nodes 1 - 4.

Correlation Table : 2000 - Event 3

	Pod Width	Seeds per Pod	Plant Population	Pods Per Plant
Pod Width[*]		0.3670 (0.1017)	0.6153 (0.0030)	-0.0250 (0.9142)
Seeds per Pod			0.3618 (0.1071)	0.2047 (0.3733)
Plant Population				-0.4827 (0.0267)

Key :   Value of Sample Correlation Coefficient
(Prob | r | > | Value | given $\rho = 0$)

[*] Pod Width is the plot level width of the most developed pod at each node
for nodes 12 - 15.

TABLE 4

Applied Statistics in Agriculture

1999 PLOT LEVEL VALUES OF PRINCIPAL VARIABLES

plot	seeds per pod most developed pod	pods per node	estimated seeds per pod	event 1 pod width (mm)	predicted seed wt. (mg)	seed wt. (mg)	pods per plant	plants per acre	yield (bu /acre)	pred. yield (bu /acre)
1	2.90	3.01	2.45	3.84	117.14	113.56	49.1	63560	33.5	32.9
2	2.95	3.83	2.45	4.00	118.81	116.41	63.7	56563	35.6	38.4
3	2.90	3.22	2.41	3.85	117.24	122.78	56.8	67059	36.7	39.6
4	2.95	3.60	2.47	3.52	113.84	117.63	71.6	60644	38.0	44.9
5	2.90	3.53	2.37	3.80	116.66	115.57	60.5	77555	29.9	47.7
6	2.87	3.27	2.35	4.59	124.83	118.81	56.5	60061	35.6	36.6
7	2.88	3.63	2.33	4.44	123.33	121.86	62.2	55396	38.6	36.4
8	2.88	3.24	2.39	4.40	122.90	120.06	53.9	53647	36.3	31.1
9	2.87	4.25	2.26	4.12	119.96	116.40	78.2	66476	35.1	51.8
10	2.93	3.31	2.45	4.66	125.61	118.03	56.4	62977	36.4	40.2
11	2.88	3.33	2.36	4.47	123.58	121.44	57.7	58895	32.6	36.5
12	2.86	4.29	2.24	3.98	118.51	121.83	77.3	48399	37.2	36.5
13	2.81	3.45	2.26	4.13	120.08	116.34	61.4	61228	31.1	37.5
14	2.86	4.40	2.23	3.69	115.59	112.72	83.9	47816	30.3	38.0
15	2.92	3.77	2.39	3.56	114.17	122.52	67.0	62977	35.1	42.3
16	2.89	4.14	2.29	4.05	119.26	120.21	72.5	45483	25.8	33.1
17	2.88	5.16	2.20	4.13	120.13	120.22	96.5	25074	35.1	23.5
18	2.91	4.35	2.30	4.35	122.42	122.00	82.3	60644	32.6	51.6
19	2.91	4.68	2.28	4.31	122.00	121.03	81.2	40818	29.8	33.8
20	2.96	4.09	2.46	4.16	120.37	124.72	74.5	68225	36.2	55.4
21	2.90	3.78	2.35	4.30	121.83	119.40	66.3	50731	27.4	35.3
22	2.91	4.03	2.34	3.99	118.68	121.11	70.9	47233	33.1	34.2
23	2.91	3.63	2.38	4.37	122.60	119.67	65.4	82220	32.8	57.7
24	2.86	3.18	2.36	3.87	117.40	120.38	53.2	59478	30.8	32.2
25	2.90	3.61	2.37	4.79	126.92	130.18	63.6	64726	30.7	45.5
26	2.89	3.08	2.41	4.66	125.55	127.52	50.5	73473	33.0	41.3
27	2.92	4.08	2.35	4.20	120.78	118.62	73.5	58895	32.7	45.0
28	2.94	3.69	2.44	4.74	126.40	119.14	68.5	67642	33.0	52.4
29	2.89	3.59	2.36	3.51	113.71	114.29	66.0	53647	30.3	34.8
30	2.92	3.36	2.43	5.06	129.68	133.00	53.8	61228	30.7	38.2
mean	**2.90**	**3.75**	**2.36**	**4.18**	**120.67**	**120.25**	**66.5**	**58759**	**33.2**	**40.1**
s.d.	**0.03**	**0.49**	**0.07**	**0.38**	**3.93**	**4.32**	**10.8**	**10854**	**3.0**	**7.7**

Notes: (1) Pod width is computed using nodes 3-6.
        (2) All measured variables, except pod width, are averaged across all three events.
        (3) Predicted seed weight is obtained using the regression model
            developed with plots 6,10,15 and 28 omitted.

TABLE 5-A

## 2000 PLOT LEVEL VALUES OF PRINCIPAL VARIABLES

plot	seeds per pod most developed pod	pods per node	estimated seeds per pod	event 3 pod width (mm)	predicted seed wt. (mg)	seed wt. (mg)	pods per plant	plants per acre	yield	pred. yield (bu /acre)
2	2.64	3.07	2.14	5.77	150.06	148.53	52.6	64275	47.0	39.9
3	2.67	3.11	2.16	5.64	146.83	149.13	53.0	73181	53.4	45.2
4	2.70	3.41	2.16	5.71	148.58	152.53	60.3	70857	51.8	50.4
5	2.72	3.51	2.16	5.90	153.40	156.13	60.6	68921	51.7	50.9
7	2.65	2.94	2.16	5.87	152.70	147.53	50.4	108416	48.8	66.2
8	2.66	2.93	2.17	5.67	147.57	151.17	47.3	109965	49.0	61.1
11	2.68	3.37	2.15	5.58	145.56	148.27	55.0	84409	47.4	53.3
12	2.73	3.07	2.22	5.77	150.06	141.23	51.7	104931	48.6	66.2
16	2.78	3.25	2.25	5.74	149.44	145.50	53.9	102995	49.7	68.6
17	2.73	3.27	2.20	6.09	158.22	158.97	55.7	116547	51.3	83.0
18	2.69	3.01	2.19	6.14	159.32	147.97	50.6	110352	44.9	71.6
21	2.74	3.14	2.22	6.05	157.25	156.40	54.5	100672	49.8	70.3
23	2.77	3.02	2.26	6.04	156.77	149.23	50.3	100672	52.7	66.0
24	2.63	3.06	2.13	6.10	158.37	152.63	51.8	99123	53.3	63.7
25	2.71	3.44	2.16	6.18	160.39	164.80	56.7	94477	53.4	68.3
26	2.75	3.01	2.25	6.45	167.07	172.43	53.5	116547	59.9	86.0
27	2.70	3.18	2.18	6.09	158.22	155.53	56.2	111126	57.3	79.0
28	2.71	2.94	2.21	6.43	166.53	165.93	52.3	115773	53.9	82.1
29	2.66	3.03	2.17	5.87	152.55	162.07	53.0	97574	55.5	62.8
30	2.77	3.13	2.25	6.11	158.68	171.57	54.8	104157	57.8	74.8
mean	**2.70**	**3.15**	**2.19**	**5.96**	**154.88**	**154.88**	**53.7**	**97748**	**51.9**	**58.3**
s.d.	**0.04**	**0.17**	**0.04**	**0.24**	**6.02**	**8.45**	**3.2**	**16184**	**3.8**	**22.9**

Notes: (1) Pod width is computed using nodes 12-15
(2) All measured variables, except pod width, are averaged across all three events.
(3) Predicted seed weight is obtained using the regression model
developed with all plots with DX < 10 , except plot 22.

TABLE 5-B

COMPANION FIGURES FOR 2000 SUMMARY TABLE (TABLE 5 - A)
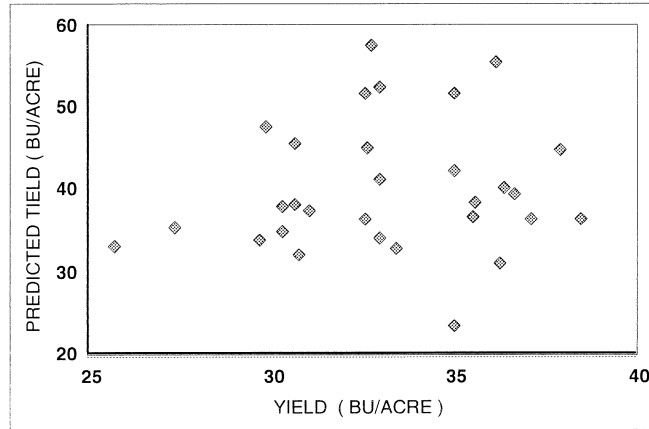
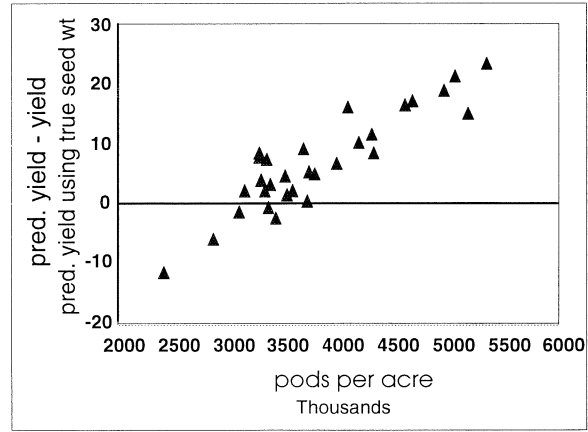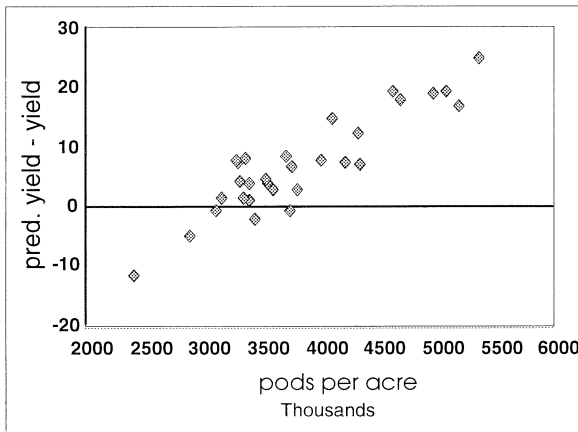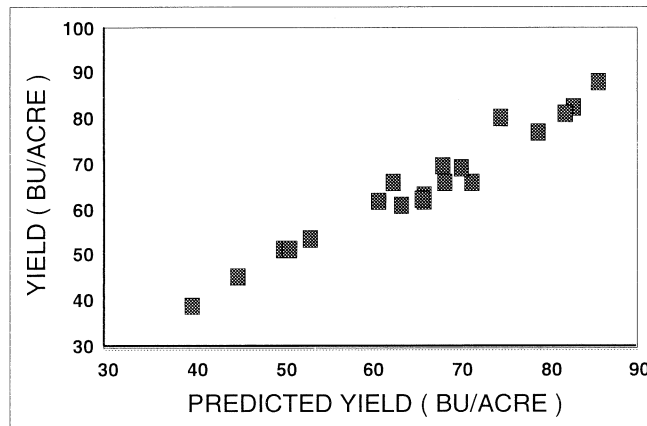Sample correlation coefficient : 0.1318 (PLOTS 6,10,28 and 15 excluded)

FIGURE 9-A

COMPANION FIGURES FOR 2000 SUMMARY TABLE (TABLE 5 - B)

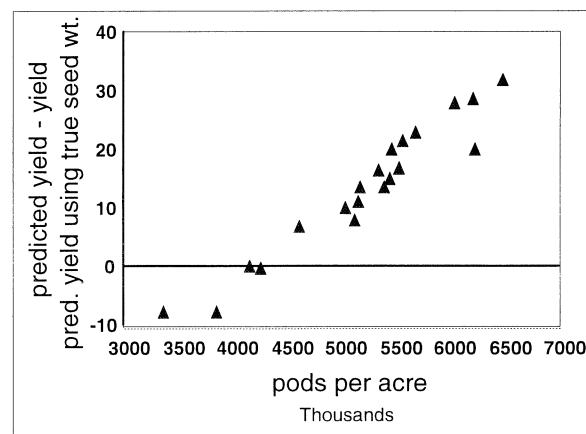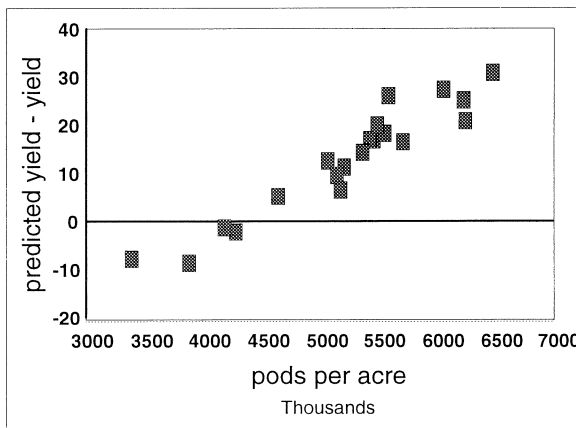Sample correlation coefficient : 0.4391 (PLOTS with DX ≥ 10 excluded) .

FIGURE 9-B