United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Applications
Division

SRB Staff Report
Number SRB-90-04

March 1990

BIAS OF SOYBEAN OBJECTIVE YIELD ESTIMATORS DUE TO VARIATION IN ROW SPACE WIDTH

David Pawel

BIAS OF SOYBEAN OBJECTIVE YIELD ESTIMATORS DUE TO VARIATION IN ROW SPACE WIDTH. By David Pawel; Research and Applications Division, National Agricultural Statistics Service; U.S. Department of Agriculture, Washington, D.C. 20250; 1990. Staff Report No. SRB-

ABSTRACT

The paper documents the theory behind field-level final objective yield (OY) estimators. The paper demonstrates why modifications in the OY program may be necessary to be consistent with current and anticipated farming practices. A model developed to identify sources of bias in field-level OY estimators exposes flaws in procedures and estimators in narrow row fields with skip row planting. The flaws may account for overestimates of yields in some fields by as much as 40%. The overall effect on the national and state-level soybean OY estimators is minor because very few farmers use skip row planting. An anticipated increase in the use of skip row planting could cause a dramatic increase in the soybean OY national and state final estimators. The paper recommends that theoretical frameworks be developed for OY estimators for all crops before empirical studies are authorized or procedures are changed.

ACKNOWLEDGEMENTS

The author thanks Ron Fecso for motivating the topic and for his valuable input. Phil Kott and Mike Palleson provided many helpful comments during the review process. Stan Hoge and George Hanuschak also made helpful suggestions. Fred Warren, who is most responsible for the original design of the soybean OY program, is especially appreciated for providing much of the historical background.

CONTENTS

LIST OF TABLES

Page

FIGURES

SUMMARY

An introductory model was developed to help identify and anticipate sources of bias in soybean field-level objective yield estimators related to variation in row space distances. Major results are:

1) Soybean OY procedures were first developed in the mid 60's, and were consistent with contemporary planting and cropping practices. NASS must anticipate future trends to keep procedures current and avoid excessive bias in yield estimators.

2) Flaws in current procedures and estimators in narrow row fields with skip row planting may account for overestimation of yields in some fields by as much as 40%.

3) Bias in national and state-level estimators due to row-space variation is probably minor. Unfortunately, this is nearly impossible to confirm.

4) The bias of national and state-level OY estimators depend on changes in the use of skip row patterns. To adequately detect and evaluate shifts in bias, NASS must establish a record of skip-row patterns for each OY field with such skip row patterns. At present, skip row planting is a rare but growing practice, and may be a dominant practice of the future. In Ohio, although skip row planting may account for only about 1% of soybean acreage, the practice may be many times more popular than it was five years ago.

5) Bias due to variation from "gross" errors in measuring four-row space measurements may be considerable, especially in narrow row fields, unless the gross errors are very rare. This bias is very difficult to evaluate. Recording of skip row patterns for each field may help detect occurrences of errors in the measurement of row-space distances.

6) Bias of soybean OY estimators is surprisingly sensitive to small changes in procedures. Accurate estimation of the impact of changes on the properties of our estimators requires theoretical analyses. The theoretical analyses should be completed and documented before procedures are changed.

7) Some procedural details and much of the theory used to justify OY procedures and estimators have not been documented. NASS should consider establishing a set of statisticians' manuals to enable operational personnel and researchers to understand the reasoning behind original procedures and methods, and their modifications.

8) Proper theoretical analysis of OY programs now may eliminate the need for more expensive and less effective empirical studies in the future.

The paper concludes with a brief description of ways to change

soybean OY procedures to reduce bias caused by row-space distance variation and to improve overall properties of yield estimators.

BIAS OF SOYBEAN OBJECTIVE YIELD ESTIMATORS DUE TO VARIATION IN ROW SPACE WIDTH

By David Pawel

INTRODUCTION

Identification and evaluation of sources of bias in soybean objective yield (OY) estimators has been a topic of National Agricultural Statistics Service (NASS) staff reports for over twenty years. While procedures for estimation of state yield and production assuming unbiased field-level estimators have been evaluated (Fecso, Francisco, and Fuller, 1986), adequate theoretical justification for field-level soybean objective yield (OY) final estimators is not documented. This paper evaluates a portion of the bias associated with this last component of the soybean estimators, the field level yield. Biases in field-level estimators are shown to vary from field to field and are dependent on skip-row patterns. In some fields, yields may be overestimated by as much as 40% of the actual yield. Although, the bias of national and state OY estimators may now be minor, the bias will increase with the popularity of the practice of using skip rows. In Ohio, only about 1% of soybean acreage exhibits skip row planting. But according to conversations with Dr. Jim Beurlein of the Ohio Extension Service, skip row planting may become a dominant planting practice within the next 20 years. NASS must anticipate future trends, such as the trend toward use of skip row planting, and do the appropriate theoretical analyses to keep procedures current and control bias in yield estimators. Theoretical analyses such as this probably cost the agency less than $50,000. Without this type of analysis, NASS would probably suffer the consequences of an unpredictable bias over several years. Perhaps NASS would then implement far more costly and less effective empirical studies.

BACKGROUND

No past published report has described a theoretical framework that could be used to determine whether field-level soybean OY final estimators are unbiased or consistent. Zarkovich (1966) provides a good theoretical description for "traditional" OY surveys, but his description does not apply directly to the survey designs used by NASS. In both the NASS and more traditional OY surveys, a fixed number of units or plots are randomly chosen within a field, and the fruit from each unit (plot) is weighed. Plots in traditional procedures are areas of predefined shape and size. The weight of the fruit in the plots is divided by the plots' area. The plots in NASS procedures, called units, may be seen as intervals of fixed length along planting rows where the estimated weight of the fruit in the units is divided by the units' "area". The area of a unit is the product of the length of the unit, a measurement of the row space width, and a conversion factor.

NASS field level OY estimators are approximately equivalent in form to ratio estimators with the unit weight of fruit in the numerator and a quantity proportional to "area" in the denominator. Classical sampling theory tells us that ratio estimators, such as the field-level soybean yield estimators, may be biased whenever the denominator value varies among the sampling units. The denominator of the NASS estimator, the area quantity, often varies depending upon the location of the unit in the field. Variation in the denominator is especially evident in narrow row soybean fields (fields where beans are planted no more than 15 inches apart) which exhibit skip row patterns. In narrow row fields, an OY procedure is used which increases the magnitude of the bias due to variation in row space distances. In narrow row fields with skip row patterns, the area measurement in the denominator is often biased downward. The effect of several common skip row patterns on bias in NASS field-level OY estimators is shown to be nonnegative and vary from 0 to over 40% of the field's actual yield.

After a brief description of soybean OY field-level procedures and the formula used to calculate the final gross yield estimators, an introductory model is developed to help identify some of the major potential sources of bias for the soybean OY final field-level estimators of gross yield. This will provide the necessary background to understand the methodology used in this paper. After the description of the theoretical model, results of analyses are presented. The paper concludes with a summary of results and a discussion of ways to reduce biases associated with skip row patterns.

THE WIDE ROW FIELD-LEVEL SOYBEAN OBJECTIVE YIELD PROCEDURE

This section offers a simplified description of NASS soybean OY field-level procedures in wide row fields. In 1988, fields were categorized as wide row fields when enumerators determined that rows were planted at least 15 inches apart. To ease the flow of presentation, many burdensome details are omitted. For example, in the actual operational procedure, the random numbers generated for the rows and paces used to locate units from any one corner are in general too small for complete coverage of each field. The enumerator is instructed to start at the most accessible corner of each field. The procedure is based upon the assumption that on average, yields relatively near to accessible corners are not systematically different from yields farther away from such corners. For a more complete description of OY procedures, the reader may consult the 1987 and 1988 Soybean OY Enumerator Manuals.

Assume for now a rectangular field as shown in figure 1. Let N be the number of rows in the soybean field, and R be the length (in feet) of each row. The enumerator is instructed to walk from one of the corners of the field, labeled O, to a preselected row, and then walk into the field a preselected number of paces. The paces would determine the starting point of a 3 foot interval labeled A1, known as the 3 foot section of row 1 of unit 1. The 3 foot section of row 2 of unit 1, labeled B1, is a 3 foot interval opposite A1 along a planting row next to A1. For simplicity, I will assume that the starting corner is closer to A1 than B1. Just before farmer harvest, the enumerator is instructed to harvest all pods and beans hanging from plants located in interval A1, and pick up all loose pods and beans on the ground in the shaded area between A1 and B1. The harvested and loose pods and beans are placed in a labeled bag and sent to a laboratory for analysis. At the laboratory, the pods with beans are weighed, counted, threshed, weighed again, and the moisture content is measured. The laboratory procedure is repeated one more time to obtain measurements from a second unit with 3 foot sections labeled A2 (row 1) and B2 (row 2). At each unit, the enumerator measures a four-row space measurement, the distance from row 1 to row 5. For our example, the starting corner is always closer to row 1 than row 5 unless the unit is located too close to the opposite boundary of the field. In such a case, the four-row space measurement is taken in a direction towards the starting corner starting from row 2. In figure 1, the four-row space distances for units 1 and 2 are labeled S_1 and S_2.

PROCEDURE FOR LOCATING UNITS IN NARROW ROW FIELDS

In recent years the operational procedure for fields categorized as narrow row fields has been different from the procedure for wide row fields. In narrow row fields, the enumerator first walks a random number of **paces**, rather than **rows**, along a path perpendicular to the rows. According to conversations with survey training personnel, the enumerator selects the row immediately in front of where the paces end. A random number of paces are then walked along the selected row.

The method of row selection for narrow row fields in the 1987 enumerator manual is alluded to on page 741:

"If the sample field has been planted in narrow rows (rows less than 18 inches wide), substitute paces for rows when counting along the edge of the field."

The same paragraph appeared in the 1988 enumerator manual, except the definition for narrow rows changed to "rows less than 15 inches wide". The difference probably does not matter because the common practice is for the enumerators to use their own judgement to determine whether a field is a narrow or wide row field before any unit is located.

The effect of substituting paces for rows is not as trivial as it may seem. In wide row fields, rows are chosen with **equal** probability. In narrow row fields, a row is chosen with probability proportional to the distance between the row and the row immediately preceding it. Thus in narrow row fields, rows following skip rows are more likely to be chosen than other rows, and four-row space measurements are less likely to include skip rows. As we shall see, the combination of these two factors creates an enormous bias in fields with certain skip-row patterns.

The 1987 and 1988 enumerator manuals do not specify that the selected row lies immediately in front of where the pacing along the edge of the field ends (row Q in figure 2). If the procedure would be changed so the selected row would be the last row the enumerator crosses along the edge of the field before the paces end (row R in figure 2), the bias of the OY estimators could be much smaller in fields with skip rows as we shall see. In any event, NASS must document which procedure it uses, so a casual change in the procedure due to, say, a new person being put in charge of OY training, would not cause an unwarranted shift in OY indications.

NASS did not always use different methods of locating units in narrow versus wide row fields. In the 1983 enumerators manual, no distinction is made between narrow and wide row fields. Changes are commonplace in NASS procedures. Most changes are adequately documented for the enumerators and other operational personnel, but not for the statisticians responsible for determining the effect the changes have on statistical properties of estimators. NASS may want to consider establishing a set of statisticians' manuals that document the statistical properties of the estimators from all OY programs. Some sections might describe the basis for the original OY programs. Other sections might describe why procedures were modified and what the effect of these changes were empirically or theoretically.

FORMULA FOR SOYBEAN FIELD-LEVEL OY GROSS YIELD ESTIMATORS

The formula for soybean field-level OY estimators changed in 1986 (see Battaglia and Nealon, 1985). This section describes the formulas used from 1986 to 1990.

Gross yield for almost all non-broadcast soybean fields is estimated using equation 1, (see for example the 1987 supervising and editing manual).

$$\hat{Y} = 1.2195 * (W_C/N_C) * (W_T M)/(W_1+W_2) * \left[\frac{W_1 N_C}{W_C S_1} + \frac{W_2 N_C}{W_C S_2} \right] \qquad \text{(Eq. 1)}$$

where

$M =$ 1 - (moisture content)/ 100),

N_C = the number of pods with beans from row 1 of the chosen unit, C = 1 or 2, [Note: it does not matter which unit is chosen, as we shall see],

W_C = weight in grams of pods and beans from row 1 of the chosen unit, (see above),

W_T = combined threshed weight of the beans from units 1 and 2,

W_1 and W_2 = the weights of the pods with beans from units 1 and 2, respectively, and

S_1 and S_2 = the four-row space widths in feet of units 1 and 2, respectively.

The constant, 1.2195, is just the result of a little arithmetic.

1.2195 = .5 * (constant associated with moisture content)
* (constant associated with number of pods with beans per 18 foot section)
* (conversion factor for grams per 18 sq. feet to bu/acre)

$$= .5 * (1/.875) * (18/(3/4)) * (43560/(18 * 453.6 * 60)) \quad \text{(Eq. 2)}$$

Since the N_C 's cancel in equation 1, the count of pods with beans has <u>no</u> effect on the final estimate, although the count is used for forecasting. The W_C's also cancel so that,

$$\hat{Y} = 1.2195 * \frac{W_T M}{(W_1+W_2)} * \left[\frac{W_1}{S_1} + \frac{W_2}{S_2} \right] . \quad \text{(Eq. 3)}$$

Equation 3 has a simple interpretation. The factor in brackets is the sum of the unit estimators of pod with beans yield, the pod and bean weights divided by the 4-row space widths. This term multiplied by a constant gives an estimator of the yield of pods with beans in bushels per acre. The other factor, the ratio of moisture adjusted threshed bean weight to the combined weight of pods with beans is needed to convert the estimate of pods with beans yield to an estimate of moisture adjusted threshed bean yield.

Let S denote the average of S_1 and S_2, that is the average four-row space measurement. If $S_1 = S_2$, equation 3 simplifies to

$$\hat{Y} = \frac{1.2195 * W_T M}{S} . \quad \text{(Eq. 4)}$$

The equation indicates that for many fields with near constant row-width, the final objective yield estimator is approximately equal to the moisture adjusted threshed bean weight divided by the average 4-row space width. The 4-row space width is proportional to the "area" attributed to each unit.

FORMULA AND PROCEDURE CHANGES FOR NARROW ROW FIELDS IN 1988

In 1988, partly because of a "narrow row" research project, some changes were made in procedures and formulas for final yield estimators for fields with narrow rows. When 4-row space measurements were less than 5 feet for both units, the beans were not combined before threshing. The threshed weight and moisture content were determined separately for each unit. The formulas for determining final field-level yield mirror the formulas used for forecasting yield before harvest. Before harvest, the eventual bean weight per pod is forecast using historical data, and multiplied by a constant times an estimate of pods per square foot to forecast final yield. Similarly, the final estimator of yield in narrow row fields in 1988 was

$$\hat{Y} = 1.2195 * \text{(bean weight per pods with beans)} * \text{(number of pods with beans per unit area)}$$

where 1.2195 is given in equation 2 and formulae for estimating bean weight per pods and number of pods with beans are given in equations 6 and 7 below.

Let $\hat{A}$ denote the estimator of bean weight per pod with developed beans, and $\hat{B}$ denote the estimator for number of pods with developed beans per square foot. Then the 1988 yield estimator is

$$\hat{Y} = \hat{A} * \hat{B}, \text{ where} \qquad \text{(Eq. 5)}$$

$$\hat{A} = .0508 * \frac{(W_{T1}M_1 + W_{T2}M_2)}{(N_1 + N_2)} \text{ and} \qquad \text{(Eq. 6)}$$

$$\hat{B} = 24 * [(N_1/S_1) + (N_2/S_2)] , \qquad \text{(Eq. 7)}$$

where W_1, W_2, S_1, S_2 are defined as in the previous section, .0508 and 24 are conversion factors (1.2195 = .0508 * 24), and

M_j = 1 - (moisture content for unit j) / 100,

N_1 and N_2 = the number of pods with developed beans from units 1 and 2, respectively, and

W_{T1} and W_{T2} = the threshed bean weights for units 1 and 2, respectively.

Substituting for $\hat{B}$ in equation 5 and through a little algebra,

$$\hat{Y} = 1.2195 * \frac{\hat{A} * N_1}{S_1} + \frac{\hat{A} * N_2}{S_2} . \qquad \text{(Eq. 8)}$$

But since N_1 and N_2 are the number of pods with beans in units 1 and 2, respectively, and $\hat{A}$ equals the threshed weight adjusted for moisture content of beans per pod for both units, the 1988 estimator for narrow row beans given in equation 8 is approximately equal to the more intuitive

$$\hat{Y} = 1.2195 * [\frac{M_1 W_{T1}}{S_1} + \frac{M_2 W_{T2}}{S_2}] \qquad \text{(Eq. 9)}$$

Table 1, a comparison of the 1988 and the equation 9 estimators, sends mixed signals. Except for four or five outliers, the estimators are almost identical.

THE EXPECTED VALUE OF THE 1987 GROSS YIELD ESTIMATOR

This section develops equations that will be used to calculate the expected value of the 1987 gross yield estimator

$$\hat{Y} = 1.2195 * \frac{W_T M}{(W_1+W_2)} * [\frac{W_1}{S_1} + \frac{W_2}{S_2}] . \qquad \text{(Eq. 3)}$$

Consideration of the effect of measurement error is given in a later section. The 1987 gross yield estimator equals

$$= \{ 1/.875 * \frac{W_T M}{(W_1+W_2)} \} * \{ 1.067 [\frac{W_1}{S_1} + \frac{W_2}{S_2}] \}$$

$$= \hat{R}_{TM} * \hat{Y}_{PB} . \qquad \text{(Eq. 10)}$$

$\hat{Y}_{PB}$ is an estimator for the field's "yield of pods with beans" where:

$$Y_{PB} = \frac{\text{wgt. of pods with beans for whole field (bu)}}{\text{area of field (acres)}}$$

$\hat{R}_{TM}$ is an estimator for the ratio of the field's moisture adjusted threshed bean weight divided by the total weight of the field's pods with beans. Call that ratio R_{TM}.

Assumptions about the statistical properties of $\hat{R}_{TM}$ and $\hat{Y}_{PB}$ must be made to calculate the bias of $\hat{Y}$. Empirical evaluation of the correlation between $\hat{R}_{TM}$ and $\hat{Y}_{PB}$ is impossible because pods and beans from the two units in each field are combined before threshing and measurement of moisture content. For now, assume a) $\hat{R}_{TM}$ is an unbiased estimator of R_{TM}, and b) $\hat{R}_{TM}$ and $\hat{Y}_{PB}$ are uncorrelated. Then

$$E\{\hat{Y}\} = E\{\hat{Y}_{PB} * \hat{R}_{TM}\}$$
$$= E\{\hat{Y}_{PB}\} * E\{\hat{R}_{TM}\}$$
$$= R_{TM} * E\{\hat{Y}_{PB}\}, \text{ and} \qquad \text{(Eq. 11)}$$

$$B = \text{Bias of } \hat{Y} = R_{TM} * (\text{Bias of } \hat{Y}_{PB}). \qquad \text{(Eq. 12a)}$$

From equation 12a, the relative bias of $\hat{Y}$,

$$\frac{B}{Y} - 1 = \frac{R_{TM} * (\text{Bias of } \hat{Y}_{PB})}{R_{TM} * Y_{PB}} - 1$$

$$= \text{Relative Bias of } \hat{Y}_{PB} \qquad \text{(Eq. 12b)}$$

(Here, the expectation operator is design based. For the sake of simplicity, we will assume that the field may be divided into units 3 feet long and 2 rows wide, and that each unit has the same probability of being selected). From equation 12, $\hat{Y}$ will be an unbiased estimator of yield (Y) if $\hat{Y}_{PB}$ is an unbiased estimator of yield of pods with beans (Y_{PB}).

For those uncomfortable with assumptions of unbiasedness and zero correlation, a weaker assumption may suffice: for large biases in $\hat{Y}_{PB}$, the bias of $\hat{Y}$ increases as the bias of $\hat{Y}_{PB}$ increases. Conclusions about bias of $\hat{Y}$ would be less specific, but their general thrust would remain the same. In fact, $\hat{R}_{TM}$ is probably not an unbiased estimator of R_{TM} and $\hat{R}_{TM}$ and $\hat{Y}_{PB}$ are probably correlated. So long as the aforementioned bias and correlation are sufficiently small, the conclusions of this report should still be valid. Otherwise, if it is unreasonable to assume that both the bias and correlation are small, the OY procedures should be modified so both can be measured. Partial information about either the bias of $\hat{R}_{TM}$ or the correlation between $\hat{R}_{TM}$ and $\hat{Y}_{PB}$ is insufficient. One must be able to specify how large the correlation is, not, for example, just that it may be positive.

THE EXPECTED VALUE OF THE GROSS YIELD ESTIMATOR IN WIDE ROW FIELDS

To evaluate B it is helpful to think of $\hat{Y}_{PB}$ as the average of unit-level estimators of "pods with bean yield", $\hat{Y}_{PB1}$ and $\hat{Y}_{PB2}$ where

$$\hat{Y}_{PBj} = 2.134 * W_j / S_j, \; j=1,2 \qquad \text{(Eq. 13)}$$

Note: 2.134 = 2 * 1.067. A proof is given in Appendix A to show that for wide row fields, the bias of $\hat{Y}_{PB}$ is

$$B_{PB} = \text{Bias of } \hat{Y}_{PB} = - \text{Cov } (\hat{Y}_{PBj}, S_j) / E\{S_j\} . \qquad \text{(Eq. 14)}$$

Then from equations 12 and 14,

$$B = \text{Bias of } \hat{Y} = - R_{TM} * \text{Cov } (\hat{Y}_{PBj}, S_j) / E\{S_j\} . \qquad \text{(Eq. 15)}$$

The bias in $\hat{Y}$ may be attributed to the variation in row space measurements. Note that if the row widths are constant, the covariance in equation 15 would vanish and the bias would be zero. From equation 15, the bias would be more severe in a field with smaller row space distances, (where $E\{S_j\}$ is smaller), than in a field with larger row space distances. This assumes the covariance between $\hat{Y}_{PBj}$ and S_j is the same for both fields.

Coincidentally, a ratio estimator may be used to test whether the bias of $\hat{Y}_{PB}$ equals zero. An unbiased estimator of the covariance term is $\hat{C}$ where

$$\hat{C} = (S_1 - S_2) * (\hat{Y}_{PB1} - \hat{Y}_{PB2}) / 2 . \qquad \text{(Eq. 16)}$$

A bias estimator, $\hat{B}$, is then

$$\hat{B} = \hat{R}_{TM} * \hat{C} / (S_1 + S_2) . \qquad \text{(Eq. 17)}$$

To estimate the bias in a state, one could calculate $\hat{B}$ for each field in the state, then average the $\hat{B}$'s. Although $\hat{B}$ may be a biased estimator of B for some fields, $\hat{B}$ is unbiased in any field with B = 0. (A proof of this last statement is given in Appendix B). Thus, a t-statistic may be used to test the hypothesis that NASS field-level estimators are theoretically unbiased for each wide row OY field. Additional insight into the performance of OY estimators in both narrow and wide row fields is given in the next three sections on the effect skip rows have on bias.

WIDE ROW FIELD ANALYSIS OF BIAS DUE TO SKIP ROW PATTERNS

Skip row systems are sometimes used to plant soybeans in narrowly spaced rows. Although only a few of these fields may be categorized as wide row fields by NASS OY procedures, we must investigate the effect of skip row systems on wide row fields a) to compare procedures in narrow and wide row fields, and b) for the sake of completeness of this analysis.

In Ohio, skip row systems often "consist of sets of 3 to 8 rows, 7 to 20 inches apart and bordered by a 20 to 30 inch wide middle to accommodate tractor tires" (1985 Ohio Agronomy Guide). Relative biases (ratio of bias divided by a field's yield) due to variation in row-space distances in wide row fields resulting from a sampler of skip row patterns are shown in Table 2. In the table, (m,n) denotes a skip row pattern where m regularly spaced rows of soybeans are followed by n skip rows. A field characterized by a (6,2) skip row pattern is shown in figure 3. For a (6,2) pattern, the spaces between neighboring skip-border rows (marked F and A) would be 3 times as large as the regular spacing. Skip border rows are defined to be rows next to skip rows. The table indicates that in wide row fields with skip row patterns yields are never underestimated and relative biases can be as large as 8%.

The methodology used to derive the biases in table 2 is given in appendix B. In constructing table 2, I assumed that within a field the unit measurements of 4-row space widths and pod with beans weights are independent. Table 3 shows what the biases might be in wide row fields if unit pod with bean weights are larger in skip-border rows, rows that are next to skip rows.

A comparison of tables 2 and 3 indicates that the relative biases caused by skip row patterns depend on whether unit pod with bean weights tend to be larger in skip-border rows. Unit pod with bean weights might be larger in skip-border rows because plants next to skip rows might benefit from more sunlight and would have less competition for nutrients and moisture from the more distant plants on the skip row side. In both tables, the effect of measurement error is not considered.

For each skip row pattern table 4 shows the smallest relative bias of the two entries in tables 2 and 3. Using the (7,1) skip row pattern as an example, the relative bias given in table 2 when pod with beans weights are uncorrelated with 4-row space widths is 1.22%. The relative bias in table 3 is 1.44%, so the entry in table 4 is 1.22%. Table 4 would show the smallest possible relative biases in wide row fields for each skip row pattern if the relative biases, as I suspect, can not be smaller than either of the entries in tables 2 and 3. Thus, the theoretical bias for wide row fields with (7,1) skip row patterns would be no smaller than 1.22%. A more complete description of the information in tables 3 and 4 is given in appendix C.

NARROW ROW FIELD ANALYSIS OF BIAS DUE TO SKIP ROW PATTERNS

Relative biases of yield estimators due to skip row patterns in narrow row fields are given in the last four columns of table 5. The biases depend on the difference in the average production inside units from skip-border rows versus the average production inside other units. If in a field with a (5,2) skip row pattern, units from skip-border rows tend to produce 10% more than other units, the relative bias of the field's OY estimator would be 16%. If a field has the same (5,2) skip row pattern, and units from skip-border rows produce 50% more than other units, the relative bias of the field's OY estimator would be 25%. In general, bias increases as the difference in yields for skip-border vs. other units increases. Assuming production per unit is on average at least as great in skip-border rows as in other rows, the relative bias in a field with a (5,2) skip row pattern can not be less than 13.3%. 13.3% is the relative bias for fields where the average skip-border row unit production equals the average production from other units. Even in fields where skip-border row units and other units are equally productive, relative biases can be as large as 25.7%, and relative biases about 10% should be common. Depending on skip row patterns and differences in average border row unit production versus average production from other units, relative biases in field-level estimators can range from 0 to about 40%.

The use of skip row patterns could become popular, and if so, the bias due to skip row patterns would become very large. If in ten years, a) the OY estimator bias in a typical narrow row soybean field with skip row patterns is 10%, b) only 20% of all soybean acreage is narrow row with skip rows, and c) the average narrow row soybean yield is 40 bu/acre, the bias due to skip row planting in the national OY estimator would be .8 bu/acre. In some states, the bias would probably be much larger. Without careful adjustment of field-level procedures, NASS could only assess the impact skip row patterns have on yield and production estimators through the recording of skip row patterns for each field. The current validation survey is not designed to measure with sufficient accuracy the impact of this one error component.

Sometimes, even when the level of indications from a survey program is incorrect, survey results may still be used to make year-to-year comparisons. If skip row planting becomes popular, meaningful estimates of year-to-year changes in yields would be very difficult to make using soybean OY indications. Table 5 shows properties of OY estimators can be extremely sensitive to changes in planting practices and conditions. NASS can measure changes in soybean acreage in narrow row fields. However, NASS has not collected data to measure changes in use of skip row patterns, or to determine whether the ratio of production in skip-border row units over production in other units changes when crop conditions change.

Changes in soybean OY procedures without sufficient documentation of potential effects of these changes also greatly complicate estimation of year-to-year changes in yield. In at least one instance, a seemingly minor but important procedural detail was not mentioned in OY enumerator manuals. Table 6 shows what the OY estimator bias due to skip row planting might be if enumerators, in locating a unit would, after counting paces along the edge of a field, select the last row crossed instead of the row immediately in front of the enumerator. Table 6 assumes skip-border row units are 10% more productive than other units. The biases in table 6 are much smaller than in table 5, showing that biases can be extremely sensitive to the method of row selection.

Although the soybean program seems robust, NASS must solve problems related to changes in farming practices and OY procedures **before** the OY program loses credibility. In 1990, NASS must ensure that in 1995, when different procedures or planting practices are used, NASS will be able to reasonably estimate the biases in 1995 and what the biases might have been in 1990. Similarly, NASS must, in 1990, act to reduce uncertainty about bias related to changes in procedures and planting practices so in twenty years, NASS may be able to use OY data to compare yields and production between 2000 and 2010 versus the 1990's. Of course, the uncertainty about bias related to changes can not be completely eliminated. The validation survey, which can measure the combined effect of changes that affect the bias of estimators, is a necessary component of the soybean OY program.

The next section uses an example to describe the methodology used to calculate the relative biases in tables 5 and 6.

METHODOLOGY FOR CALCULATING BIASES IN NARROW ROW FIELDS

From equation 12b, the relative bias of the gross yield estimator $\hat{Y}$ equals the relative bias of $\hat{Y}_{PB}$. Most of this section describes how to calculate the relative bias of $\hat{Y}_{PB}$ for fields with a (6,2) skip row pattern. The same method works for other skip row patterns.

Figure 3 shows a (6,2) skip row pattern. Based on location in relation to nearest skip rows, there are six types of rows, marked A, B, C, D, E, and F. It seems reasonable to assume that units from the four row types, B, C, D, and E would tend to have equivalent yields, and units in the border row type A would tend to be as productive as units in border row type F. Let μ_1 denote the average weight of pods with beans in border row (marked A and F) units and μ_2 denote the average pod with bean weights for all other units (from rows B, C, D, and E). There are eight row spaces between successive rows of type F. The probability that a unit is from row type A is 3/8, and the probability for row types B, C, D, E, or F is 1/8. Let x denote the regular spacing between neighboring rows, so the distance between border rows A and F is 3x and the distance between any other neighboring rows is x. If a unit is from rows A or B, the four row space measurement is 4x, otherwise if the unit is from rows C, D, E, or F, the four row space measurement is 6x. Recall that, since $\hat{Y}_{PB}$ is the average of the two unit level yields, $\hat{Y}_{PB1}$ and $\hat{Y}_{PB2}$,

$$E(\hat{Y}_{PB}) = E(\hat{Y}_{PBj}) \text{ where}$$

$$\hat{Y}_{PBj} = 2.134 * W_j / S_j, \; j=1,2. \qquad \text{(Eq. 18)}$$

For convenience sake, let k = 43560 / (453.6 * 60), the conversion factor from lb/sq. ft. to bu/acre. The reader may confirm that 2.134 = 4/3 * k. The expected value of $\hat{Y}_{PBj}$ is

$$\begin{aligned} E(\hat{Y}_{PBj}) = (4/3)k * (&(\text{probability that unit is from row A}) * \mu_1/4x \\ &+ (\text{probability that unit is from row B}) * \mu_2/4x \\ &+ (\text{probability that unit is from row C}) * \mu_2/6x \\ &+ (\text{probability that unit is from row D}) * \mu_2/6x \\ &+ (\text{probability that unit is from row E}) * \mu_2/6x \\ &+ (\text{probability that unit is from row F}) * \mu_1/6x). \end{aligned}$$

$$= (4/3)k * \{3/8 * \mu_1/4x + 1/8 * \mu_2/4x + 3 * (1/8 * \mu_2/6x) + 1/8 * \mu_1/6x\}$$
$$= (4/3)k * \{(3/32 + 1/48)\ \mu_1/x + (1/32 + 1/16)\mu_2/x\}$$
$$= (.1528\mu_1 + .125\mu_2)k/x. \qquad \text{(Eq. 19a)}$$

Let $r = \mu_1/\mu_2$. Then

$$E\{\hat{Y}_{PBj}\} = (.1528r + .125)\mu_2 k/x. \qquad \text{(Eq. 19b)}$$

In a rectangular field of length L (across rows) and row width R, assume each row is divided into 3 foot intervals, and that each interval defines a unit that may be chosen for enumeration. The field-level yield is

$$Y = k * \frac{(\#\text{ of skip-border row units})\mu_1 + (\#\text{ of other units})\mu_2}{RL} \qquad \text{(Eq. 20)}$$

$$= k * \frac{(2/8)(L/x)(R/3)\mu_1 + (4/8)(L/x)(R/3)\mu_2}{RL} \qquad \text{(Eq. 21)}$$

$$= \frac{k}{3x} * [(2/8)\mu_1 + (4/8)\mu_2] \qquad \text{(Eq. 22)}$$

$$= (.0833r + .1667)\mu_2 k/x \qquad \text{(Eq. 23)}$$

From equations 19b and 23, the relative bias in $\hat{Y}$ is

$$E\{\hat{Y}_{PBj}\}/Y - 1 = (.1528r + .125)/(.0833r + .1667) - 1 \qquad \text{(Eq. 24)}$$

If average production for skip-border row units equals the average production in other units, r = 1 and the relative bias is (.2778/.25) - 1 = 11.1% ; if skip-border row units are 10% more productive, r = 1.1 and the relative bias is (.2931/.2583) - 1 = 13.5%, and so on.

Much of the bias is caused by the expected value of the four-row space measurement being too small. In all skip row patterns, the rows that immediately follow skip rows (rows A in figure 3) are chosen with higher probabilities than all other rows. When rows that follow skip rows are chosen, the four-row space measurements are smallest. The first column of table 5 shows how for various skip row patterns, the four-row space distance underestimates the average spacing between rows.

General formulas for the expected value of estimators for all (m,n) skip row patterns are given in appendix D.

Suppose the method of unit location changes. The enumerator is instructed to choose the last row crossed after pacing along the edge of the field ends. (Currently, enumerators choose the row directly in front of where the pacing ends). This change in the method of unit selection changes the probabilities of row selection. For a (6,2) skip row pattern, row A is now selected with probability 1/8 and row F is selected with probability 3/8. The probabilities associated with the other rows do not change. Under this procedure, (compare to equation 19a and b),

$$E\{\hat{Y}_{PBj}\} = (4/3)k * 1/8 + \mu_1/4x + 1/8 * \mu_2/4x + 3\ (1/8 * \mu_2/6x) + 3/8 * \mu_1/6x\}$$

$$= (4/3)k * \{(1/32 + 3/48)\ \mu_1/x + (1/32 + 1/16)\mu_2/x\}$$
$$= .125(\mu_1 + \mu_2)k/x. \qquad \text{(Eq. 25)}$$

The relative bias would be

$$E\{\hat{Y}_{PBj}\}/Y - 1 = (.125r + .125)/(.0833r + .1667) - 1\ . \qquad \text{(Eq. 26)}$$

If r = 1.1, the relative bias would be only 1.6%.

WHY FOUR-ROW SPACE MEASUREMENTS?

Four-row space measurements seem to be the root of so many potential problems. Why did NASS decide to use four-row space measurements to estimate field level yields? A brief review of soybean OY history is revealing.

The soybean OY procedures, modelled after corn OY procedures, were developed in the mid to late 1960's. It is not clear whether at first, one-row space measurements from row middle to row middle, or four-row space measurements were used to estimate areas associated with units. Bias in one-row space measurements would be due to difficulties in defining row middles, round-off errors in measurement, measuring the distance in a direction not perpendicular to rows, and so on. Variation in one-row space measurements would be due to these reasons and the variation in actual spacing between rows. It is possible that four row space measurements were introduced initially as a check on the accuracy of the one-row space measurements. Eventually, four-row space measurements were used in the field-level yield estimation formulas, and one-row space measurements were used as a check.

In the 60's, farmers would use the same equipment to plant soybeans and corn. The planters would usually plant 2 or 4 rows per pass. Thus, in each field, the row space distances would be almost constant within sets of four rows planted during the same pass, but irregular spacings would be possible between rows planted through successive passes. Four-row space measurements might be more reliable than one-row space measurements because of an averaging effect, and because within any four-row space both the irregular spacings from successive passes and the regular spacings from the same pass would be correctly represented. The effect on (say) bias of yield estimators due to round off errors in measurement, difficulties in locating (defining) row middles, incorrect determination of direction of measurement, and so on, for four-row space measurements should be approximately one-quarter of the effect for one-row space measurements. Five, six, or seven-row space measurements might be better than one-row space measurements, but unlike four-row spaces, the irregular spaces due to successive passes would not be correctly represented within a five, six, or seven-row space.

Denominators of yield estimators based upon four row space measurements would show relatively little variation. Thus, not only could four-row space measurements be used to detect measurement errors, but four-row space measurements probably minimized the theoretical bias in yield estimators due to row space variation. The use of four-row space measurements in the 60's showed how knowledge of planting practices could be intelligently incorporated in an OY design.

Soybean planting practices have changed since the 60's, and will continue to change. The practicality of using four-row space measurements deserves review. Four row space measurements worked because farmers used planters that usually planted two or four rows per pass, but now most planters can plant more than four rows per pass, and new practices such as skip row planting are becoming popular. The analysis has already demonstrated that even well-conceived statistical programs must be periodically reviewed to certify that sampling procedures remain consistent with evolving farming practices.

EDIT CHECK DATA

The previous section stated that four-row space and one-row space measurements were made for furnishing edit check data to detect errors in row space measurements. Edit check data are used to detect errors associated with collection of data. As a general philosophy consistent with much of quality control thinking, edit check data should be used with a scheme that measures the overall effect of all nonsampling and sampling errors. Edit check data should be used to detect errors in either groups of observations or specific observations, but preferably in specific observations. Once errors in specific observations have been identified, the observations may sometimes be edited to either reduce or eliminate the effect of the errors, and a validation program would then indicate the effect of the edits and other adjustments on overall levels of the estimator. Careful elimination and reduction of errors associated with data collection may help to reduce the effect or number of sources of variation that plague both raw estimators and estimators of bias from validation programs. Survey organizations must always be careful to evaluate the effect of any change in editing procedures on time-series properties of estimators, such as statistical properties of year-to-year changes in estimators.

The four-row and one-row space measurements probably provided a more powerful edit check in the 60's than now. To see this, consider a state where within each field the spacings between rows are practically constant. For any unit, a four-row to one-row space measurement ratio that differs from four would be a strong indication that an enumeration error had been made in at least one of the two measurements. For most fields, ratios may not equal four because of skip row patterns, "dead" spots in the field, rounding errors in enumeration, measuring a three or five instead of a four row space distance, and so on. In the 90's, drawing conclusions about the quality of enumeration based upon the same ratio should be more difficult than in the 60's. In the 60's, farmers rarely if ever planted in narrow row fields or used skip row patterns, and the effect of many types of enumeration errors may have been smaller. In the 90's, farmers will undoubtedly continue to plant in narrow row fields, where rounding errors should be more serious. At present, more information must be collected for the four-to-one ratio to allow for reasonably certain identification of fields with incorrect four-row space measurements. NASS already encourages enumerator feedback to improve detection of row space measurement errors, but a more formal approach would be appropriate. Perhaps the enumerator should be asked to complete a coded question whenever the ratio is not within prespecified error limits (eq. 3.7 to 4.3). Possible choices may include a) the four-row space measurement includes a "dead" spot and b) farmer used a skip-row pattern. The final form of such a question should be seriously discussed, for the question might provide a record of vital missing information for more complete OY data analyses.

At present, plots such as figures 4 through 9, which show graphs of one and four row space measurements for narrow row OY fields in Ohio in 1987 and 1988, show that in some states at least one of two hypotheses must be correct. Either a) enumeration error in measuring four row space measurements may be a serious problem or b) there is much more variation in actual row space measurements than can be attributable to skip row patterns. In either case, yield estimators might be biased. Figures 4 and 5 are plots of unit 1 four-row space measurements versus unit 2 four-row space measurements. Figures 6 through 9 are plots of (for either unit 1 or 2) one-row space measurements versus one-quarter of the four-row space measurements. Somewhat subjectively, I identified observations (shown in bold type) showing discrepancies between the two measurements displayed in each plot that can not be attributed to round-off error, triangularization (to be explained shortly), and other minor measurement errors.

Round-off error in one-row and four-row space measurements should never exceed one-tenth of a foot and only in rare circumstances one-twentieth of a foot, since enumerators use rulers that measure to the nearest one-tenth foot. Triangularization occurs when row space distances are measured along a direction not perpendicular to rows. We will use the Pythagorean Theorem and Figure 10 to show that triangularization probably has little effect on individual row space measurements. Suppose an enumerator measures the four row distance between rows A and B as the distance from points P to Q, instead of the correct P to R. The segment PR is perpendicular to the rows. Suppose PR is 10 feet long, and QR is 1 foot long. The distance from P to Q would be the square root of $(10^2 + 1^2)$ or about 10.05 feet. The measurement would be off by about .05 feet or 1/2 of 1 percent of the correct four-row space distance. If instead the enumerator walks 2 feet parallel to the rows instead of 1 foot, so QR is 2 feet, PQ would be the square root of $(10^2 + 2^2) = 10.2$ feet, and the four-row space measurement would be 2 percent larger than the correct distance. In general, if an enumerator strays 1 foot in a direction parallel to the rows for every 10 feet, the measurement will be off by about 1/2 percent; in the rare case where the enumerator strays 2 feet, measurements will be off by about 2 percent. Triangularization may have a minor but nonnegligible effect on state-level estimators, because triangularization can only increase four-row space measurements, and decrease field-level yield estimators. Thus, underestimation of yield due to triangularization in one field can never be compensated by a triangularization error in another field.

Table 8 provides another way to summarize information from four row and one row space measurement edit check data. Table 8 displays further evidence that either a) row space measurement error may be a problem or b) within field row space distance variation may substantially bias estimators. For each unit in each nonbroadcast field, I calculated the difference d = a-b, where a = .25 * (the 4-row space measurement) and b = (the 1-row space measurement). "a" would be a four-row space measurement estimate of a field's average row space distance. An average of the two differences associated with each unit was calculated for each field, and then the field-level average differences were averaged again over all fields in each state. Table 8 shows that in almost every state in 1987 and 1988, the average row spacing implied by the four-row space measurements is small compared to the average one-row space distance. For many of the states, a t-statistic rejects the hypothesis that the average four row space measurement is four times the average one-row space measurement in either 1987 or 1988. Overall, the average four-row space measurement appears to be about 2% less than four times the average one-row space measurement. (Please note: this does not imply a 2% bias in four-row space measurements, and a 2% bias in four-row space measurements would not necessarily cause a 2% bias in yield estimators). The analysis is clouded somewhat because, as table 7 demonstrates, four-row space measurements (divided by four) and one row space measurements sometimes have different expectations.

In crude terms, due to the way units are located in narrow row fields, four-row space measurements (divided by 4) and one-row space measurements do not always measure the same thing. Unfortunately, we can not be sure whether the difference in the four-row space measurements and (four times) the one-row space measurements is due to measurement error or to theoretical differences in expectation. Thus, without modification in NASS procedures, the four-to-one row space measurements can not be simply used in states with narrow row fields to determine whether row space measurement error bias exists.

MEASUREMENT ERROR

Regardless of whether four-row space measurements are unbiased, measurement error would add to the variation in four-row space measurements, and may cause bias in the field-level yield estimator. The results in table 9 indicate that unless the imprecision in the four-row space measurements is unexpectedly high, the resulting bias due to the additional variation should be minor.

To understand the results, suppose that because of measurement error, the ratio,

$$\frac{\text{measured four row space distance}}{\text{actual four-row space distance}},$$

is uniformly distributed from .95 to 1.05. For a field with a constant four-row space distance of 10 feet, the enumerator would measure the four-row space distance to be 9.5, 9.6, 9.7, ... , or 10.5 feet with equal probabilities. Table 8 indicates the resulting bias would be .08% of the field's yield. If the enumerator would be more imprecise and would measure the four-row space distance to be anywhere from 9.0 to 11.0 feet with equal probabilities, table 9 indicates the bias would still be only .35% of the field's yield.

In a more realistic scenario, the distribution of the measured to actual four-row space distance ratio might follow a triangular distribution. A graph of this and other hypothetical measurement error distributions is given in figure 4. Suppose the distribution is triangular from .9 to 1.1. In the same field with a constant four row space distance of 10 feet, the enumerator could still measure the four-row space distance to be anywhere from 9 to 11 feet, but would be twice as likely to measure the distance to be 10 feet as 9.5 feet, and much more likely to measure the distance at 9.5 feet as at either 9.1 feet or 10.9 feet. From table 9, the resulting bias would be only .17%.

Biases due to various distributions of the ratio of the measured to actual four-row space distance are given in table 8. The results can be used in a creative fashion to evaluate biases under different scenarios. For example, suppose in 9 out of 10 fields, an enumerator makes only small mistakes in measuring the four-row space distance, and the measured to actual ratio follows a symmetrical triangular distribution from .9 to 1.1, but in 1 out of 10 fields, the enumerator makes a gross error by either including extra rows or omitting rows from the four-row space measurement. Suppose the gross mistakes can be modeled by a uniform distribution from .5 to 1.5. The resulting bias would be .1 * 9.86% + .9 * .17% = 1.01%.

It seems that the cumulative effect on bias of "small" errors is minor, but the effect of gross errors on bias may be of concern if their occurrence is sufficiently frequent. Gross errors may be more likely to occur in narrow row fields with skip row patterns where it is most difficult to detect errors using ratios of four-row to one-row space measurements.

CONCLUSIONS AND RECOMMENDATIONS

An introductory model was developed to help identify and anticipate sources of bias in soybean field-level objective yield estimators related to variation in row space distances. Major results of the paper are:

1) Soybean OY procedures were first developed in the mid 60's, and were consistent with contemporary planting and cropping practices. NASS must anticipate future trends to keep procedures current and avoid excessive bias in yield estimators.

2) Flaws in current procedures and estimators in narrow row fields with skip row planting may account for overestimation of yields in some fields by as much as 40%.

3) Bias in national and state-level estimators due to row-space variation is probably minor. Unfortunately, this is nearly impossible to certify.

4) The bias of national and state-level OY estimators depend on changes in the use of skip row patterns. To adequately detect and evaluate shifts in bias due to skip row planting, NASS needs data on skip-row pattern usage for each state. At present, skip row planting is a rare but growing practice, and may be a dominant practice of the future.

5) Bias due to variation from small but frequent errors in measuring four-row space measurements have little effect on bias of OY estimators. Bias due to variation from gross errors in measuring four-row space measurements may be considerable unless the gross errors are very rare. This bias (due to gross errors) is very difficult to evaluate. Additional information needs to be collected to determine reasons for row space measurement variation

within each field.

6) Bias of soybean OY estimators may be surprisingly sensitive to small changes in procedures. Accurate estimation of the impact of changes on the properties of our estimators requires theoretical analyses. The theoretical analyses should be completed and documented before procedures are changed.

7) Data indicate that sources of row space variation other than skip row planting may affect bias and other statistical properties of yield estimators.

8) Some procedural details and much of the theory used to justify OY procedures and estimators has not been documented. NASS should consider establishing a set of statistician's manuals to enable operational personnel and present and future researchers to understand the reasoning behind original procedures and methods, and their modifications.

9) Proper theoretical analysis of OY programs now may eliminate the need for more costly and less effective empirical studies later.

NASS may decide to modify soybean OY procedures. The next two paragraphs sketch two approaches, one more radical, one conservative.

A more radical way to change the procedure in narrow row fields would be to locate units by counting groups of rows between skip rows, instead of paces or single rows. The procedure is illustrated in figure 12. The enumerator is instructed to locate a unit 2 row groups and 4 paces into a field with a (3,2) skip row pattern. The numbers 2 and 4 would have been randomly generated using a computer program. Treating the edge of the field as a set of skip rows, the second row group is defined to be the third group of (in this case 3) rows between skip rows. The unit would then extend from the first row of the third row group (labeled A) to the first row of the next row group. The enumerator would measure this distance (designated the row space distance). The unit forms a rectangle, a specified number of feet wide, and with a length that depends on the skip row pattern and the spacing between rows. NASS may decide to reduce the width of the unit from 3 feet, or to enumerate only one unit in narrow row fields to allow for easier enumeration. At harvest, the enumerator would harvest all pods from plants from rows A, B, and, C, and would pick up all loose beans within the shaded area. An advantage to this method is that the row space distance would not depend on the location of the unit. Of course, the method would only work for the type of simple skip row patterns discussed in this paper. The method might have to be modified to be practical in fields with many rows between skip rows. The example should show that with more information about planting practices, NASS might appropriately tailor its OY procedures to eliminate bias without unnecessary complication.

The more conservative approach would change the method of unit location in narrow row fields as suggested by results in table 6. The enumerator could be instructed to choose the last row crossed after pacing along the edge of the field ends. This approach might work well in states with many fields using skip row patterns similar to ones given in table 6.

Of course, NASS may decide simply to collect more information about planting practices, and then use the information to adjust yield estimators.

Before deciding on whether and how basic OY procedures should be modified, I recommend that NASS:

1) Incorporate into its OY program a way to gather information about planting practices as soon as possible.

2) Increase (doubling might be about right) the size of validation samples of narrow row fields in states with large proportions of narrow row fields. This may allow for a crude assessment of bias in narrow row fields.

3) Change the soybean OY procedures and estimators **after** reliable information about planting practices is gathered. Theoretical properties of estimators should be documented before accepting any future proposed change.

REFERENCES

1. Battaglia, R. and Nealon, J., Simplifying Soybean Objective Yield Estimation, U.S. Department of Agriculture, NASS, Staff Report YRB-85-02, 1985.

2. Beuerlein, James, Skip Row Soybeans, Research Bulletin 1151, Cooperative Extension Service, Ohio State University, 1985.

3. Fecso, R., Francisco, C., and Fuller, W., An Assessment of Crop Production Estimators Used With the Objective Yield Surveys, U.S. Department of Agriculture, NASS, Staff Report YRB-86-08, 1986.

4. National Agricultural Statistics Service, 1988 Soybean Objective Yield Survey Enumerator's Manual, U.S. Department of Agriculture, NASS, 1988.

5. National Agricultural Statistics Service, 1987 Soybean Objective Yield Survey Enumerator's Manual, U.S. Department of Agriculture, NASS, 1987.

6. National Agricultural Statistics Service, 1987 Soybean Objective Yield Survey Enumerator's Manual, U.S. Department of Agriculture, NASS, 1983.

7. National Agricultural Statistics Service, 1988 Objective Yield Supervising and Editing Manual, U.S. Department of Agriculture, NASS, 1988.

8. National Agricultural Statistics Service, 1987 Objective Yield Supervising and Editing Manual, U.S. Department of Agriculture, NASS, 1987.

9. Ohio State University, Ohio Agronomy Guide, Department of Agronomy, Ohio State University, 1985.

10. Sukhatme, P.V., and Sukhatme, B.V., Sampling Theory of Surveys with Applications, Iowa State University Press, Ames, Iowa, 1970.

11. Zarkovich, S.S., Quality of Statistical Data, Food and Agricultural Organization of the United Nations, p. 331-365, 1966.

APPENDIX A: THE EXPECTED VALUE OF AN ESTIMATOR OF YIELD OF PODS WITH BEANS

I will derive the expected value of

$$\hat{Y}_{PB} = 1.0670 * \left[\frac{W_1}{S_1} + \frac{W_2}{S_2} \right] , \qquad \text{(Eq. 1)}$$

an estimator of a field's yield of pods with beans (unadjusted for moisture content) defined by:

Pods with beans yield

$$= \frac{\text{weight of pods with beans for entire field (bushels)}}{\text{area of field (acres)}}.$$

Please refer to figure 1. We will assume that the field is rectangular. R denotes both the length of each row and the width of the field. We will assume the rows in the field are chosen using a table of random numbers from 1 to N, the number of paces is determined using a table of random numbers from 0 to (L/3 - 1), and the length of each pace is equal to 3 feet. In addition, the enumeration procedures are followed as intended, without measurement or recording errors.

Let

Y_{PB} = yield of pods with beans in bushels per acre,

w = field's total weight in grams of pods with beans,

w_{ij} = weight of pods with beans from plot j paces into row i,

d = the length of the field,

d' = the distance between the first and last rows,

s^*_{Ri} = distance in feet between row i and row (i+1),

s_{Ri} = four-row space measurement for row i,

$$s^* = \left(\sum_{i=1}^{N-1} s^*_{Ri} \right) / (N-1) \text{ , and}$$

$$s = \left(\sum_{i=1}^{N} s_{Ri} \right) / N \text{ .}$$

One should verify that

$$Y_{PB} = (43560/(453.6 * 60)) * w/(Ld)$$
$$= 1.6005 * w/(Ld) \quad \text{(Eq. 2)}$$

s^* and s represent the "average" one-row and four-row space distances. For most fields, (N-1)/N is approximately equal to 1, $4s^*$ is approximately equal to s, and

$$d' = (N-1) * s^*.$$

Thus d' is approximately equal to

$$d'' = N * s/4 \quad \text{(Eq. 3)}$$

Let S_1 and S_2 be the unit 1 and 2 four-row space measurements. Since each row has an equal chance of selection, the random variables

$$D'_k = N * S_k/4,\ k = 1,2, \quad \text{(Eq. 4)}$$

would be unbiased estimators of d''. Since there are L * (N/3) plots in the field, and each plot has an equal chance of being selected, the expectation of W_1 and W_2, the weights of pods with beans from units 1 and 2, is given by

$$E_p\{W_k\} = \frac{w}{(L/3) * N},\ k = 1 \text{ or } 2. \quad \text{(Eq. 5)}$$

The subscript p in the expectation operator specifies that the expectation is made in relation to the probabilities of each plot's selection. In this case, since each plot has an equal chance of selection, the expectation is a simple average of the plot weights of pods with beans. We can now show that $\hat{Y}_{PB}$ in equation 1 approximates a method-of-moments estimator of

$$Y_{PB} = 1.6005 * w/(Ld) \quad \text{(Eq. 6)}$$

Since for most fields the ratio (d'/d) is sufficiently close to 1,

$$Y_{PB} = 1.6005 / 3 * \frac{w}{(L/3) * N} * \frac{1}{d'/N} \quad \text{(Eq. 7)}$$

Substituting $E_p\{W_k\}$ for the first factor after the constant, and $E_p\{D_k\}$ for d',

$$\hat{Y}_{PB} = .53351 * E_p\{W_k\} / E_p\{(S_k/4)\},\ k=1,2$$
$$= 2.134 * E_p\{W_k\} / E_p\{S_k\},\ k=1,2. \quad \text{(Eq. 8)}$$

Corresponding "ratio" estimators are

$$\hat{Y}_{PBk} = 2.134 * W_k / S_k, \ k=1,2 \ . \qquad \text{(Eq. 9)}$$

$\hat{Y}_{PB}$, already given in equation 1, is just the average of $\hat{Y}_{PB1}$ and $\hat{Y}_{PB2}$, that is

$$\hat{Y}_{PB} = .5 * (\hat{Y}_{PB1} + \hat{Y}_{PB2}) \qquad \text{(Eq. 10)}$$

$$= 1.0670 * [\ (W_1/S_1) + (W_2/S_2)\ ] \ . \qquad \text{(Eq. 1)}$$

From equation 9, the bias of $\hat{Y}_{PB}$ is equal to the bias of the estimators from each unit, $\hat{Y}_{PB1}$ and $\hat{Y}_{PB2}$. Both $\hat{Y}_{PB1}$ and $\hat{Y}_{PB2}$ are ratios of random variables that are unbiased estimators of the numerator and denominator of a ratio equal to Y_{PB}:

$$\hat{Y}_{PBk} = 2.1340 * W_k / S_k \ , \ k=1,2 \ , \text{ and} \qquad \text{(Eq. 11)}$$

$$Y_{PB} = 2.134 * E_p\{W_k\} / E_p\{S_k\}, \ k = 1,2. \qquad \text{(Eq. 12)}$$

The bias of ratio estimators, such as $\hat{Y}_{PB}$, has been extensively studied (Sukhatme & Sukhatme, 1970, p.136-192):

$$B_{PB} = \text{Bias of } \hat{Y}_{PB} = - \text{Cov}\ (\hat{Y}_{PBk}, S_k) / E_p\{S_k\} \ . \qquad \text{(Eq. 13)}$$

APPENDIX B: METHODOLOGY FOR CALCULATING BIASES IN TABLE 2

I will describe the methodology used to calculate the biases in table 2. I will assume that within a field the unit measurements of 4-row space widths and pods with beans weights are independent.

Let $k = 1.067$. Then the 1987 formula for estimating yield is

$$\hat{Y} = k * R_{TM} * \{(W_1/S_1) + (W_2/S_2)\}, \qquad \text{(Eq. 1)}$$

where

$$Y = 2 * k * R_{TM} * \frac{E\{W_i\}}{E\{S_i\}}, \; i = 1,2.$$

Suppose $S_i = s$ if the 4-row space includes no skip rows, and
$= rs$ if the 4-row space includes skip rows.

Let $p = P\{$ 4-row space does not include skip rows$\}$, so that

$$\begin{aligned} S_i &= s \text{ with probability } p \\ &= rs \text{ with probability } 1-p. \end{aligned} \qquad \text{(Eq. 2)}$$

Both p and r may be easily determined for any of the skip row patterns in table 2. Suppose a field is characterized by the (6,2) skip-row pattern shown in figure 3. Since the 4-row space would not include skip rows if the first row of the unit is A or B, but would contain skip rows if the first row were C, D, E, or F, p would be 2/6 or 1/3. When the 4-row space includes the 2 skip rows, the 4-row space would include 6 "regular" row spacings, otherwise only 4 "regular" row spacings. r would be equal to 6/4 or 1.5.

The next equations will demonstrate that the relative bias of $\hat{Y}$ depends solely on p and r.

$$\begin{aligned} E\{\hat{Y}\} &= k * R_{TM} * E\{(W_1/S_1) + (W_2/S_2)\} \\ &= 2k * R_{TM} * E\{(W_i/S_i)\}, \; i = 1,2. \end{aligned} \qquad \text{(Eq. 3)}$$

Then (1 + Relative Bias of $\hat{Y}$) $= \dfrac{E\{Y\}}{Y} = \dfrac{E\{S_i\}}{E\{W_i\}} * E\{(W_i/S_i)\}$

$$= \frac{E\{S_i\}}{E\{W_i\}} * E\{W_i\} * E\{1/S_i\}$$

$$= E\{S_i\} * E\{1/S_i\}. \qquad \text{(Eq. 4)}$$

But,

$E\{S_i\} = ps + (1-p)rs$, and (Eq. 5a)

$E\{1/S_i\} = p/s + (1-p)/rs$. (Eq. 5b)

$$E\{S_i\} * E\{1/S_i\} = (p + (1-p)r) * (rp + (1-p)) / r$$

$$= 1 + (\text{Relative Bias of } \hat{Y}). \qquad \text{(Eq. 6)}$$

For a (6,2) skip-row pattern, the relative bias would be
$(1/3 + 2/3(1.5)) * (.5 + 2/3) / 1.5 - 1 = .037$.

APPENDIX C: IF BORDER-ROW UNITS GROW MORE BEANS

In appendix B, I assumed that within a field, the four-row space measurements and unit pod with beans weights are independent. Perhaps a more realistic model would be:

$W_j = \mu_W + \beta_K + e_j$ if row 1 of unit j is bordered by a skip row
$W_j = \mu_W + \beta_N + e_j$ otherwise, (Eq. 1)

where

W_j = weight of pods with beans from unit j,
$\mu_W = E\{W_j\}$,
$\beta_K - \beta_N$ = difference between the average unit weight of pods with beans for units with row 1 next to a skip row (border units) versus units where row 1 does not border a skip row.
e_j = unit j random error in weight of pods with beans.

It is natural to assume that the weight of pods with beans would be greater for border units than for other units, or that $\beta_K > 0 > \beta_N$.

To investigate how bias depends on β_K and β_N, I made substitutions in equation 1 for $\beta_K - \beta_N$. Let T_j be the two row space distance between the rows that border row 1 of unit j (see figure 13). If the farmer used any of the skip row patterns in table 2, we could define t_K and t_N so that

$T_j = t_K$ if the two row space includes skip rows, (Eq. 2)
$= t_N$ otherwise.

Let μ_T = P(two-row space includes skip rows) * t_K
+ P(two-row space does not include skip rows) * t_N.

Then b could be defined so that

$$\beta_K - \beta_N = b * (t_K - t_N) \ (\mu_W/\mu_T).$$

Dividing both sides of the last equation by μ_W, we find that the
Relative change in expected unit weight of pods with beans
$= (\beta_K - \beta_N)/\mu_W = b * (t_K - t_N)/\mu_T$.
= b * Relative change in two-row space distance. (Eq. 3)

Normally, b should be between 0 and 1. If pods with beans weights are independent of the 4-row space measurements, b would equal 0, and $\beta_K = \beta_N$ would equal 0. The resulting biases in yield have already been given in table 2. On the other extreme, if b = 1, a relative change in the two-row space distance would cause the same expected relative change in the unit weight of pods with beans. Biases in yield due to variation in row space measurements are shown in table 3 for b = 1. Linear interpolation may be used to calculate biases in yield for values of b between 0 and 1.

To gain a more complete understanding of these results, let us reconsider a fundamental theoretical result of the paper:

$$B = \text{Bias of } \hat{Y} = - R_{TM} * \text{Cov}\ (\hat{Y}_{PBj},\ S_j)/\ E(S_j)\ . \qquad \text{(Eq. 4)}$$

If unit pod with beans weights are independent of 4-row space measurements, the pod with bean yield estimator, $\hat{Y}_{PBk}$, will tend to decrease when the four-row space measurements include skip rows. For instance, if the four-row space includes skip rows, the numerator of Y_{PBj} (weight of pods with beans) will show no tendency to change, but the denominator of Y_{PBj} (the four-row space measurement) will be larger. The covariance term will be negative, and the bias will be positive. If unit pod with beans weights increase as 4-row space measurements increase, both the numerator (weight of pods with beans) and denominator (four-row space measurement) will (tend to) increase when the four-row space measurement includes skip rows. In this case, the covariance between Y_{PBj} and the four-row space measurement is unpredictable. Finally, if the unit pod with beans weights decrease as the four-row space measurements increase, Y_{PBj} will tend to decrease when the four-row space measurements include skip rows, the covariance term will be negative and the bias will be positive. The bias will be greatest in the last case where the pods with beans weights tend to decrease when the 4-row space measurements include skip rows. Unfortunately, this would be expected for most of the skip row patterns in tables 2 and 3. Why?

Concerning the location of units, there are only four possibilities we need consider.

1) The four row space measurement includes skip rows, and row 1 of the unit is bordered by a skip row.

2) The four row space measurement includes skip rows, but row 1 of the unit is not bordered by a skip row.

3) The four row space measurement does not include skip rows, but row 1 of the unit is bordered by a skip row.

4) The four row space measurement does not include skip rows, and row 1 of the unit is not bordered by a skip row.

We would expect the pods with beans weights to be larger when row 1 of the unit is bordered by a skip row. When possibilities 1 and 4 occur more frequently and possibilities 2 and 3 less frequently, pods with bean weights will tend to be positively correlated with four-row space measurements. More specifically, the pods with beans weights will be positively (negatively) correlated with the four-row space measurements if the product of the probabilities of possibilities 1 and 4 minus the product of the probabilities of possibilities 2 and 3 is positive (negative). This only occurs for skip-row patterns (3,1), (3,2), and (3,3).

For example, consider the (6,2) skip row pattern shown in figure 2. If row 1 of the unit is A, possibility 3 occurs, if the first row is B, possibility 4 occurs, if the first row of the unit is C, D, or E then possibility 2 occurs, and if F is chosen to be the first row of the unit, possibility 1 occurs. The probabilities of possibilities 1, 2, 3, and 4 would be 1/6, 1/2, 1/6, and 1/6. Put in a contingency table format, the probabilities look like:

	Bordered by Skip Row	Not Bordered by Skip Row
4-row space includes skip rows	probability = 1/6	probability = 1/2
4-row space includes no skip rows	probability = 1/6	probability = 1/6

and the difference between the cross-products (1/6)(1/6) - (1/2)(1/6) is negative. If a (6,2) skip row pattern is used, the four row space measurements will tend to have a negative correlation with the unit pods with beans weights. Our analysis of equation 15 indicates that this would tend to increase the bias of the field-level yield estimators. This would not occur if NASS would revise its OY procedures by pushing back the 4-row space measurement one row space, as shown by S^* in figure 3, so the four-row space measurements would include the one-row spaces on either side of row 1 of the unit. This would ensure that the four-row space measurement would include a skip row, whenever row 1 of the unit borders a skip row.

In table 4, I have combined the information from tables 2 and 3 to show the minimum relative biases due to variation in row space measurements for some skip row patterns.

APPENDIX D: FORMULAS FOR CALCULATING BIAS IN NARROW ROW FIELDS

Let Y denote the actual yield in a field with skip rows,

k = 1.6005 (conversion factor)
L = length of a rectangular field,
R = length of each row,
(m,n) be the skip row pattern,
x = the shortest distance between any two rows,
μ_1 = the production per unit for rows next to skip rows, and
μ_2 = the production per unit for other rows.

$$Y = k * \frac{(2/(m+n))(L/x)(R/3)\mu_1 + ((m-2)/(m+n))(L/x)(R/3)\mu_2}{RL}$$

$$= (k/3x) * [\ 2/(m+n)\mu_1 + ((m-2)/(m+n))\mu_2\ ] \qquad \text{(Eq. 1)}$$

The formula for $E\{\hat{Y}\}$ depends on m.

If m = 3,

$$E\{\hat{Y}\} = (4/3)k * [\ \frac{(n+1)\mu_1/(n+3)}{(4+n)x} + \frac{\mu_2/(n+3)}{(4+n)x} + \frac{\mu_1/(n+3)}{(4+2n)x}\ ] \quad . \qquad \text{(Eq. 2)}$$

In equation 2, (n+1)/(n+3) is the probability that the unit is within a row that follows skip rows. For units in rows between border rows and rows that precede skip rows, the probability is 1/(n+3). The four row space measurements for the three types of units is either (4+n)x or (4+2n)x depending on whether the measurement includes 1 or 2 skip row groups.

If m = 4,

$$E\{\hat{Y}\} = (4/3)k * [\ \frac{(n+1)\mu_1/(n+4)}{(4+n)x} + \frac{2\mu_2/(n+4)}{(4+n)x} + \frac{\mu_1/(n+4)}{(4+n)x}$$

$$= \frac{4k}{3x(n+4)^2} + [(n+2)\mu_1 + 2\mu_2\ ] \ . \qquad \text{(Eq. 3)}$$

If m is greater than 4,

$E\{\hat{Y}\}$

$$= \frac{4k}{3x} * [\ \frac{(n+1)\mu_1/(m+n)}{4} + \frac{(m-5)/\mu_2(m+n)}{4} + \frac{3\mu_2/(m+n)}{4+n} + \frac{\mu_1/(m+n)}{4+n}\ ]$$

or

$$E\{\hat{Y}\} = \frac{k}{3x(m+n)} * [(n+1+\frac{4}{4+n})\mu_1 + (m-5+\frac{12}{4+n})\mu_2] . \qquad \text{(Eq. 4)}$$

The relative bias can be calculated using equation 1, either equation 2, 3, or 4, and

$$\text{Relative bias} = E\{\hat{Y}\}/Y - 1 . \qquad \text{(Eq. 5)}$$

Table 1 - Comparison of 1988 operational gross yield estimator for narrow row fields with gross yield estimator of equation 9.

State	Sample Size	Op. Gross Yield Mean	Op. Gross Yield Standard Error	Eq. 9 Yield Mean	Eq. 9 Yield Stand. Error	Oper. Less Eq.9 Yield Mean Dif.	Oper. Less Eq.9 Yield Stan. Error
Ark.	4	41.3	18.6	41.4	18.5	-.1	.2
Ill.	29	32.2	13.5	32.2	13.5	+.0	.1
Ind.	16	32.1	16.0	32.1	16.4	-.0	.2
Iowa	8	35.4	16.6	35.3	16.3	+.1	.5
Ky.	8	42.0	34.6	42.2	31.3	-.2	15.0
La.	8	36.2	13.8	36.2	13.9	+.0	.2
Minn.	28	35.8	13.9	34.8	13.3	1.0	5.1
Ms.	7	30.4	5.8	30.4	5.7	+.0	.1
Mo.	30	42.5	20.5	41.3	20.2	1.2	6.6
Neb.	8	36.2	26.2	36.2	26.2	+.0	.1
N.C.	26	28.0	20.7	28.0	20.7	+.0	.2
Ohio	47	38.3	23.7	38.5	24.4	+.1	1.0
Tenn.	7	19.2	11.3	19.3	11.3	+.1	.1

Operational Minus Equation 9 Yield Outliers:

2 observations in Kentucky where operational yield minus equation 9 yield are -27.0 and 29.1, 1 observation in Minnesota with difference (operational minus eq. 9) -27.1, 1 observation in Missouri with difference -36.0, 1 observation in Ohio with difference 7.0.

Table 2 - Relative biases caused by skip row patterns in wide row fields when 4-row space widths and pods with bean weights are independent.

Skip-row pattern	Relative bias (in percent)
(3,1)	.74
(4,1)	.00
(5,1)	.80
(6,1)	1.11
(7,1)	1.22
(8,1)	1.25
(3,2)	1.85
(4,2)	.00
(5,2)	2.66
(6,2)	3.70
(7,2)	4.14
(8,2)	4.17
(3,3)	2.86
(4,3)	.00
(5,3)	5.14
(6,3)	7.15
(7,3)	7.87
(8,3)	8.04

Table 3 - Relative biases caused by skip row patterns in wide row fields when 4-row space widths and pods with bean weights are correlated and b is 1.

Skip-row pattern	Relative bias (in percent)
(3,1)	.00
(4,1)	.00
(5,1)	2.00
(6,1)	1.66
(7,1)	1.44
(8,1)	2.50
(3,2)	.00
(4,2)	.00
(5,2)	4.66
(6,2)	5.55
(7,2)	4.82
(8,2)	4.17
(3,3)	.00
(4,3)	.00
(5,3)	12.86
(6,3)	10.71
(7,3)	9.18
(8,3)	8.04

Table 4 - Minimum relative biases in wide row fields caused by skip row patterns.

Skip-row pattern	Relative bias (in percent)	Adj. factors
(3,1)	.00	1.00
(4,1)	.00	1.00
(5,1)	.80	.99
(6,1)	1.11	.99
(7,1)	1.22	.99
(8,1)	1.25	.99
(3,2)	.00	1.00
(4,2)	.00	1.00
(5,2)	2.66	.97
(6,2)	3.70	.96
(7,2)	4.14	.96
(8,2)	4.17	.96
(3,3)	.00	1.00
(4,3)	.00	1.00
(5,3)	5.14	.95
(6,3)	7.15	.93
(7,3)	7.87	.93
(8,3)	8.04	.93

Table 5 - Relative biases in estimators of average row space distances and field-level yields in narrow row fields with skip row patterns.

Skip Row Pattern	Row Space Distance Estimator Rel. Bias	Yield Estimator Relative Bias when Production in Border Rows is x% higher than Production in Other Rows and x = 0	10	20	50
(3,1)	-1.6%	2.2%	2.9%	3.5%	5.0%
(4,1)	0	0	1.0	1.8	4.0
(5,1)	-2.8	4.0	5.4	6.7	10.0
(6,1)	-2.0	3.4	4.5	5.6	8.6
(7,1)	-1.6	2.9	3.9	4.9	7.5
(8,1)	-1.2	2.5	3.4	4.3	6.7
(3,2)	-4.0	5.6	6.8	7.8	10.4
(4,2)	0	0	1.6	3.0	10.7
(5,2)	-8.2	13.3	16.0	18.5	25.0
(6,2)	-6.3	11.1	13.5	15.6	21.4
(7,2)	-4.9	9.5	11.6	13.5	18.8
(8,2)	-4.0	8.3	10.2	11.9	16.7
(3,3)	-6.3	8.6	10.2	11.6	15.0
(4,3)	0	0	2.1	3.9	8.6
(5,3)	-14.0	25.7	29.7	33.3	42.9
(6,3)	-11.1	21.4	24.9	28.1	36.7
(7,3)	-9.0	18.4	21.4	24.3	32.1
(8,3)	-7.4	16.1	18.8	21.4	28.6

Table 6 - Relative biases in estimators of field-level yields in narrow row fields under an alternate OY procedure.[1]

Skip-row pattern	Relative bias (in percent)
(3,1)	-1.7
(4,1)	1.0
(5,1)	1.1
(6,1)	1.0
(7,1)	.8
(8,1)	.7
(3,2)	-4.7
(4,2)	1.6
(5,2)	1.9
(6,2)	1.6
(7,2)	1.4
(8,2)	1.2
(3,3)	-7.5
(4,3)	2.1
(5,3)	2.5
(6,3)	2.1
(7,3)	1.8
(8,3)	1.5

[1] Production in border row units is assumed to be 10% greater than in other units. Units are located by choosing the last row crossed by enumerator instead of the row directly in front of enumerator after pacing along the edge of a field ends.

Table 7 - Comparison of relative bias in one-row and four-row space measurements for selected skip-row patterns in narrow row fields.

Skip row pattern	Relative bias in one row space measurement[1]	Relative bias in four row space measurement[2]
(3,1)	**-6.3%**	**-1.6%**
(4,1)	**-4.0%**	**0**
(5,1)	-2.8	-2.8
(6,1)	-2.0	-2.0
(7,1)	-1.6	-1.6
(8,1)	-1.2	-1.2
(3,2)	**-16.0**	**-4.0**
(4,2)	**-11.1**	**0**
(5,2)	-8.2	-8.2
(6,2)	-6.3	-6.3
(7,2)	-4.9	-4.9
(8,2)	-4.0	-4.0
(3,3)	**-25.0**	**-6.3**
(4,3)	**-10.7**	**0**
(5,3)	-14.0	-14.0
(6,3)	-11.1	-11.1
(7,3)	-9.0	-9.0
(8,3)	-7.4	-7.4

[1] as an estimator of the average 1-row space distance
[2] as an estimator of the average 4-row space distance

Table 8 - Average of (4-row space measurements/4) minus 1-row space measurements.

	1987			1988			
State	N	Mean	Stand. Error	N	Mean	Stand. Error	Ratio
Alabama	73	-.009	.031				-.00
Arkansas	115	-.003	.015	109	.020	.019	+.00
Georgia	70	-.015	.015	69	-.010	.014	-.01
Illinois	169	-.035	.009	172	-.028	.010	-.02
Indiana	94	-.051	.013	105	-.034	.009	-.02
Iowa	151	-.007	.013	149	-.033	.012	-.01
Kentucky	77	-.025	.025	77	.001	.021	-.01
Louisiana	71	-.060	.027	80	-.001	.022	-.02
Minnesota	89	-.020	.015	108	-.036	.014	-.01
Miss.	97	-.082	.025	82	-.031	.022	-.03
Missouri	136	-.021	.012	132	-.047	.013	-.02
Nebraska	70	-.072	.031	78	-.015	.018	-.02
N. C.	65	-.007	.023	83	-.008	.019	-.00
Ohio	107	+.000	.012	113	-.016	.011	-.00
Tennessee	75	-.041	.019	70	-.028	.016	-.02

t-statistics are significant at the .05 level in Illinois (1987, 1988), Indiana (1987, 1988), Iowa (1988), Minnesota (1988), Mississippi (1987), Missouri (1988), Nebraska (1987), and Tennessee (1987).

The ratio given in the last column is the ratio of the average mean difference for 1987 and 1988 over 1.5. This is meant to give some idea of how large the discrepancy is between one and four row space measurements compared to a "typical" row width space.

Table 9 - Effect of imprecision of measurement of four-row space measurement on bias of field-level yield estimators.

Distribution and Range of the Ratio of the Measured over Actual Four-Row Space Distance		
Distribution	Range	Relative Bias
Uniform	.95 to 1.05	.08%
Uniform	.9 to 1.1	.35%
Uniform	.8 to 1.2	1.37%
Uniform	.5 to 1.5	9.86%
Symmetrical Triangular	.95 to 1.05	.04%
Symmetrical Triangular	.9 to 1.1	.17%
Symmetrical Triangular	.8 to 1.2	.67%
Skewed Triangular (L)	.95 to 1.07	.06%
Skewed Triangular (R)	.93 to 1.05	.07%

Figure 1: Field-level soybean OY procedure.

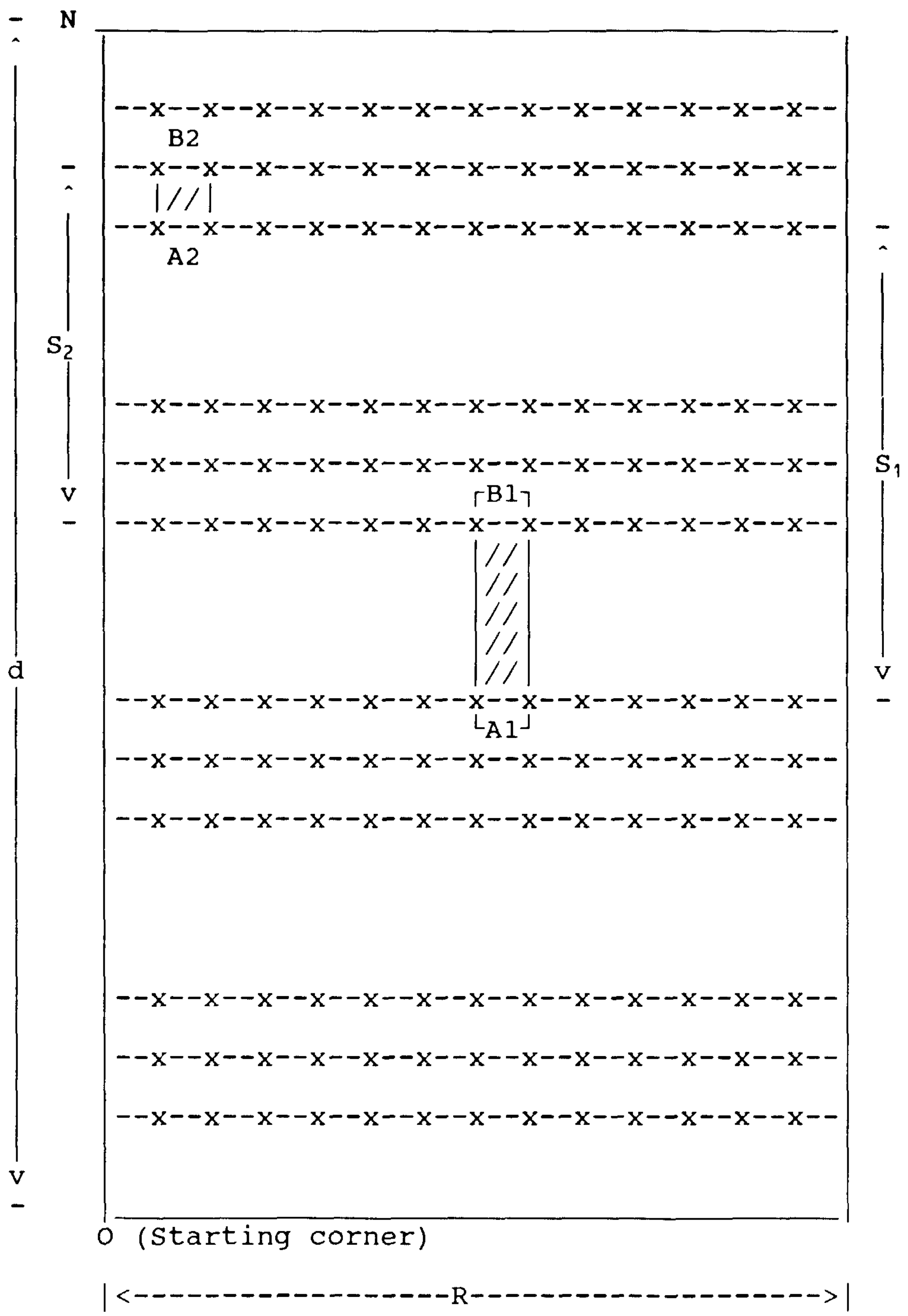


Figure 2: Location of units in narrow row fields.

```
 ____________________________________________
|                                            |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
|                                            |
|                                            |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
|                                            |
|                                            |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
|           |//| <--Unit                     |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |  Q
|                                            |
|                                            |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |  R
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
| --x--x--x--x--x--x--x--x--x--x--x--x--x--  |
|                                            |
|____________________________________________|
O (Starting Corner)
```

Figure 3: A (6,2) Skip row pattern.

```
            |                                          |
     F      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     E      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     D      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     C      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     B      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     A      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
            | ........................................ |
            |                                          |
            | ........................................ |
            |                                          |
     F      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     E      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     D      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     C      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     B      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     A      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
            | ........................................ |
            |                                          |
            | ........................................ |
            |                                          |
     F      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     E      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     D      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     C      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     B      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
     A      |--x--x--x--x--x--x--x--x--x--x--x--x--x--|
            |                                          |
            | ........................................ |
 Starting  /
  corner  /
      |   |
      v   |
```

Figure 4: Plot of unit 1 vs. unit 2 4-row space measurements from 1987 Ohio OY data from narrow row fields

```
Unit   |
  2    |
Four   |
Row    +
Space  |
Mea-   +
sure-  |
ment   +
       |
   6.0 +
       |
   5.5 +                                           A   +
       |                                             +
   5.0 +                                           C
       |                                       AA+A
   4.5 +                                       B A
       |                                     A   A
   4.0 +                                   +
       |                                 +
   3.5 +                             B +         A
       |                          A  + B              A
   3.0 +                         BAA A
       |                     AB B+      A
   2.5 +                     DGCAC      A
       |                    ABD      A
   2.0 +                   A
       |                 +
   1.5 +               +
       |             +
   1.0 +           +
       |         +
   0.5 +       +
       |     +
   0.0 +   +
       |
        ---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
          0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5 7.0
```

Unit 1 Four Row Space Measurement

Legend: A = 1 obs, B = 2 obs, etc.
+ = diagonal: unit 1 measurement = unit 2 measurement

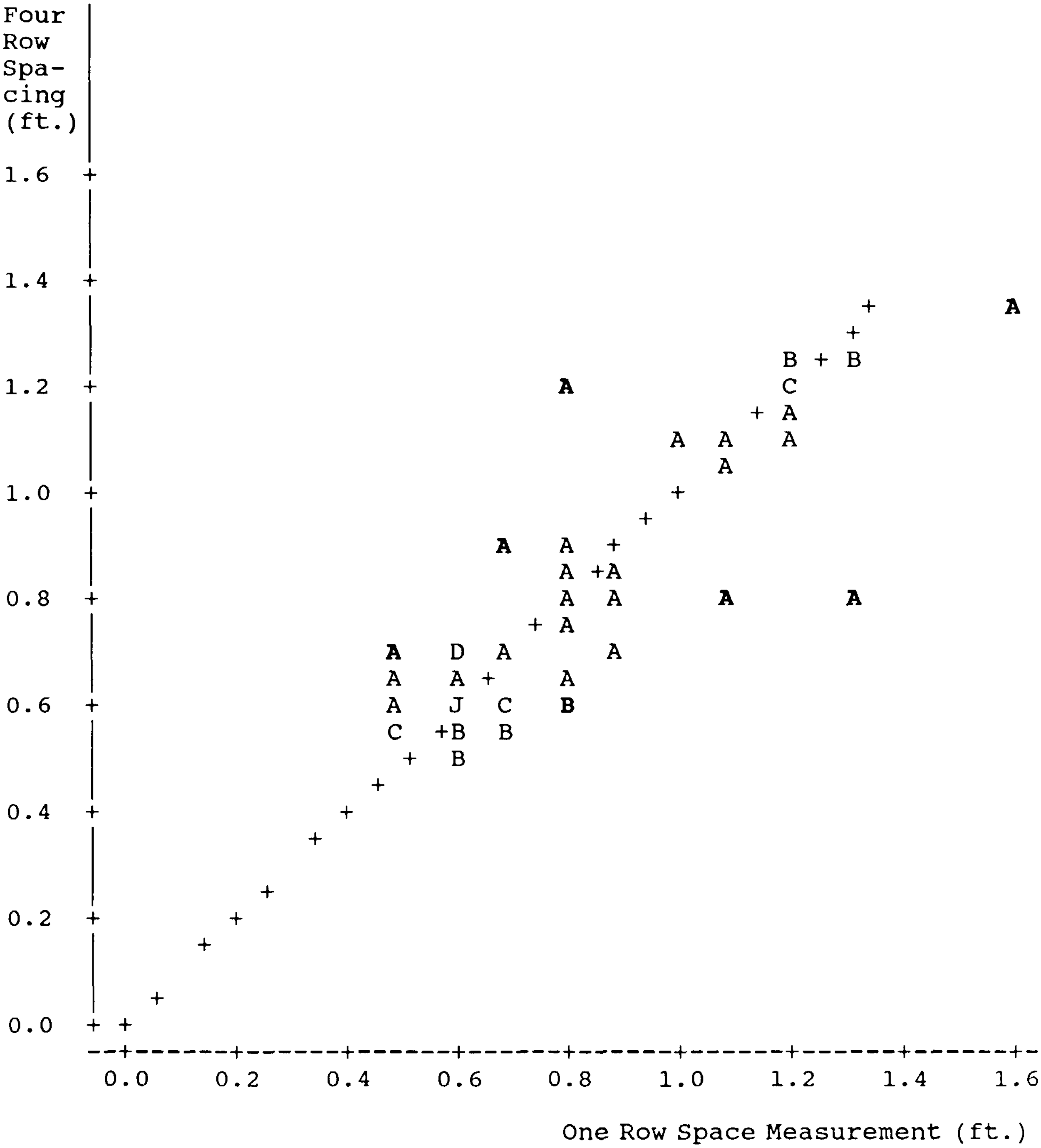
Figure 5: Plot of unit 1 vs. unit 2 4-row space measurements from 1988 Ohio OY data from narrow row fields

```
Unit  +
  2   |
Four  +
Row   |
Space +
Mea-  |
sure- +
ment  |
(ft.) +
      |
  6.0 +
      |
  5.5 +
      |                                            +
  5.0 +                                       A BC
      |                                      AA+ A
  4.5 +                                      +   A
      |                                    A
  4.0 +                                A A
      |                                +       A
  3.5 +                              B
      |                           AB
  3.0 +                     A    +  A     A
      |                    AB ACC
  2.5 +                    FCCDA  A
      |                    DBA
  2.0 +                A +
      |                +
  1.5 +              +
      |            +         A
  1.0 +          +
      |        +
  0.5 +      +
      |    +
  0.0 +  O
      |
      ---+---+---+---+---+---+---+---+---+---+---+---+---+---+-
        0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5 4.0 4.5 5.0 5.5 6.0 6.5

                    Unit 1 Four Row Space Measurement (ft.)
```

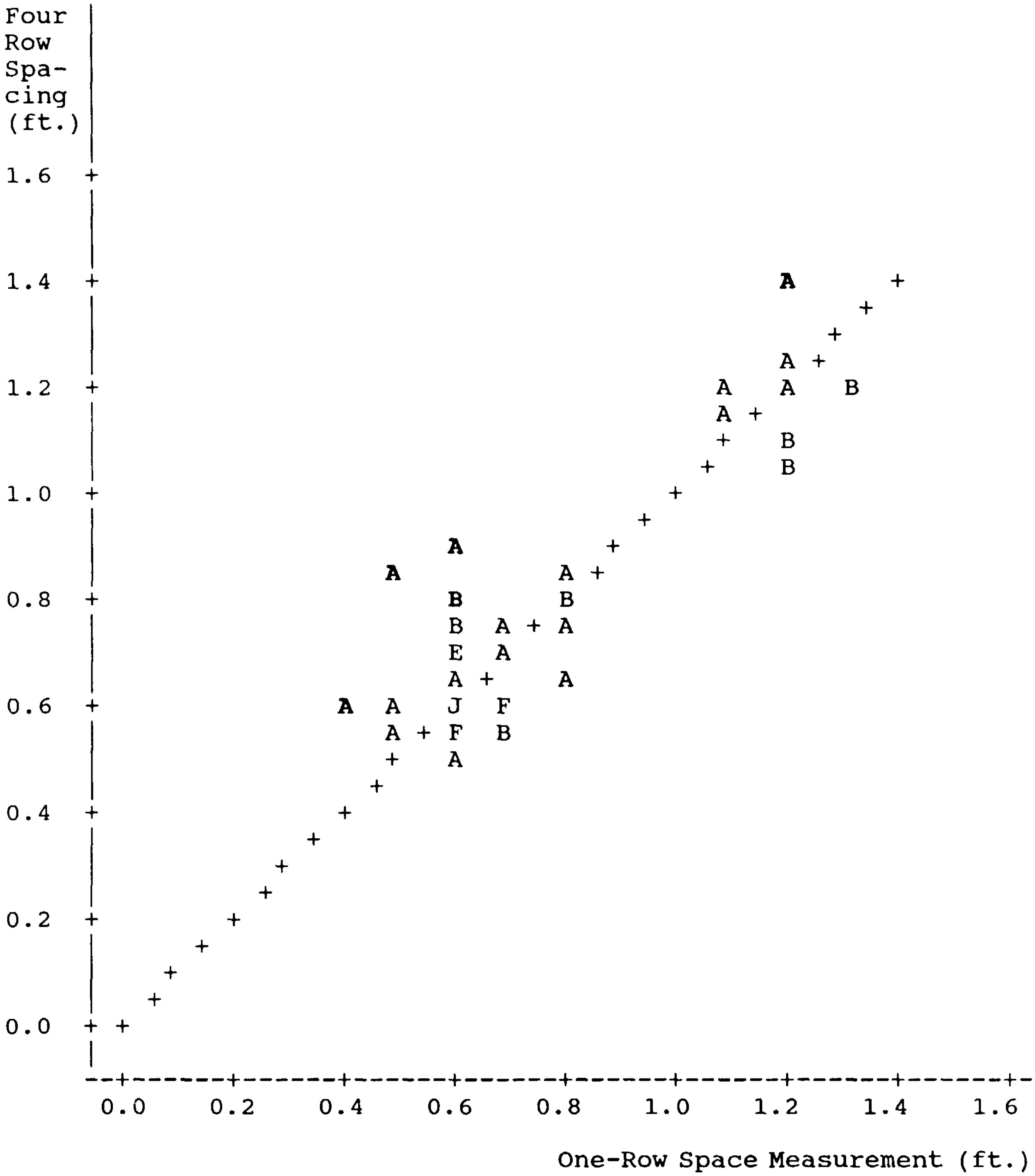
Legend: A = 1 obs, B = 2 obs, etc.
+ = diagonal: unit 1 measurement = unit 2 measurement

Figure 6: Plot of unit 1 1-row vs. 4-row space measurements from 1987 Ohio OY data from narrow row fields



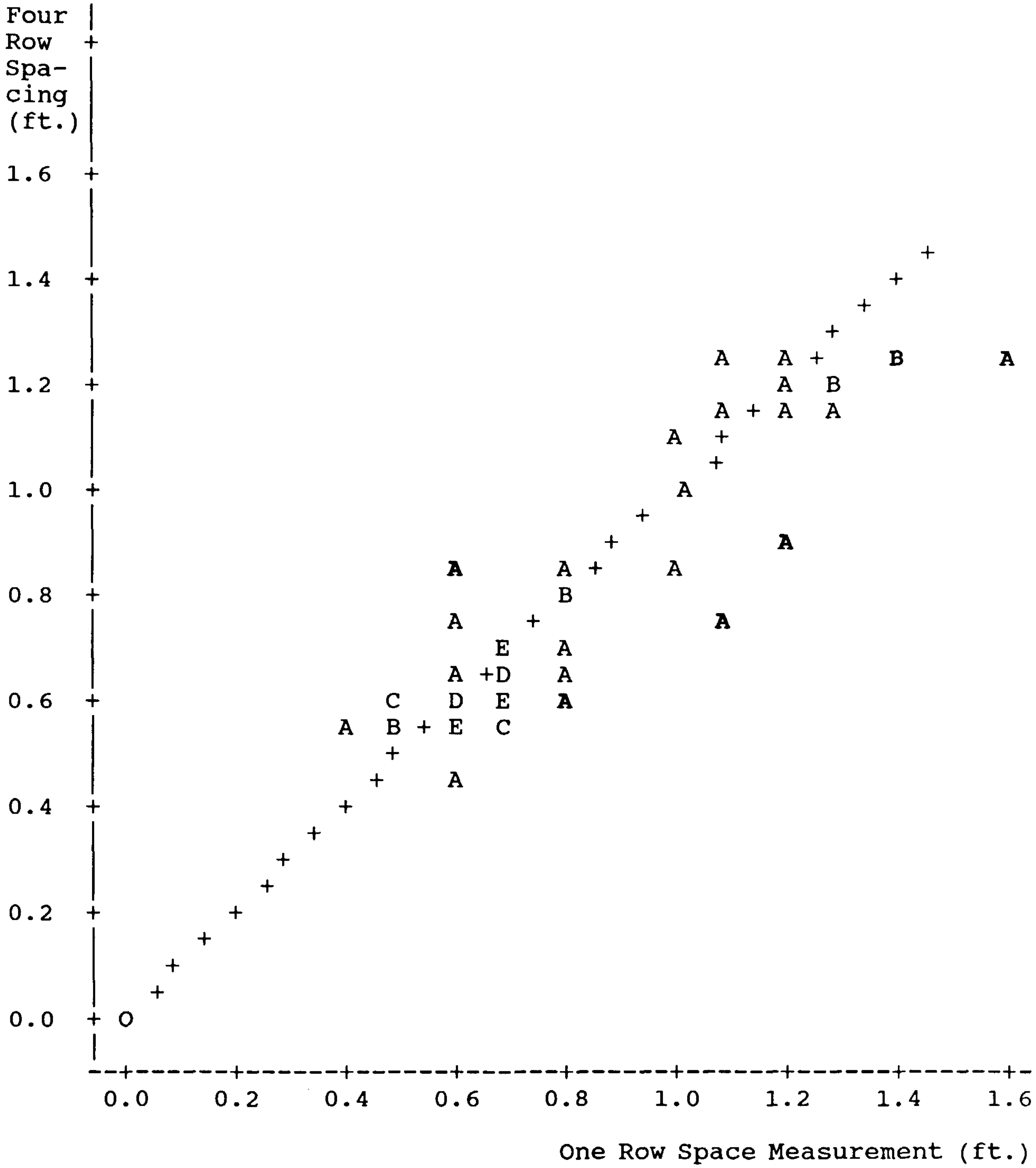
Legend: A = 1 obs, B = 2 obs, etc.
+ = diagonal: four-row spacing = one row space measurement

Figure 7: Plot of unit 2 1-row vs. 4-row space measurements from 1988 Ohio OY data from narrow row fields



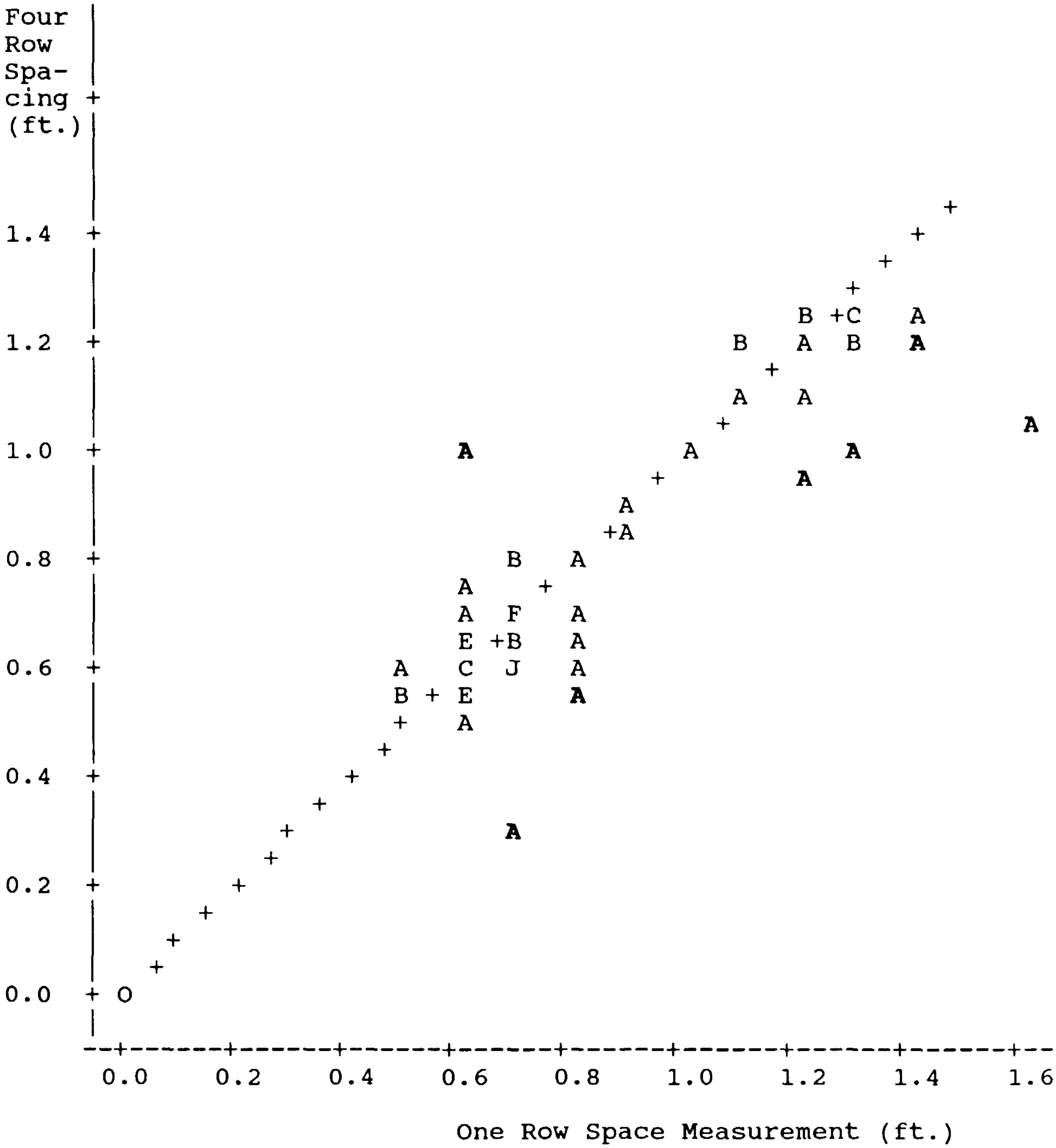
Legend: A = 1 obs, B = 2 obs, etc.
+ = diagonal: four-row spacing = one row space measurement

Figure 8: Plot of unit 1 1-row vs. 4-row space measurements from 1988 Ohio OY data from narrow row fields



Legend: A = 1 obs, B = 2 obs, etc.
+ = diagonal: four-row spacing = one row space measurement

Figure 9: Plot of unit 2 1-row vs. 4-row space measurements from 1988 Ohio OY data from narrow row fields



Legend: A = 1 obs, B = 2 obs, etc.
+ = diagonal: four-row spacing = one row space measurement

Figure 10 - Triangularization

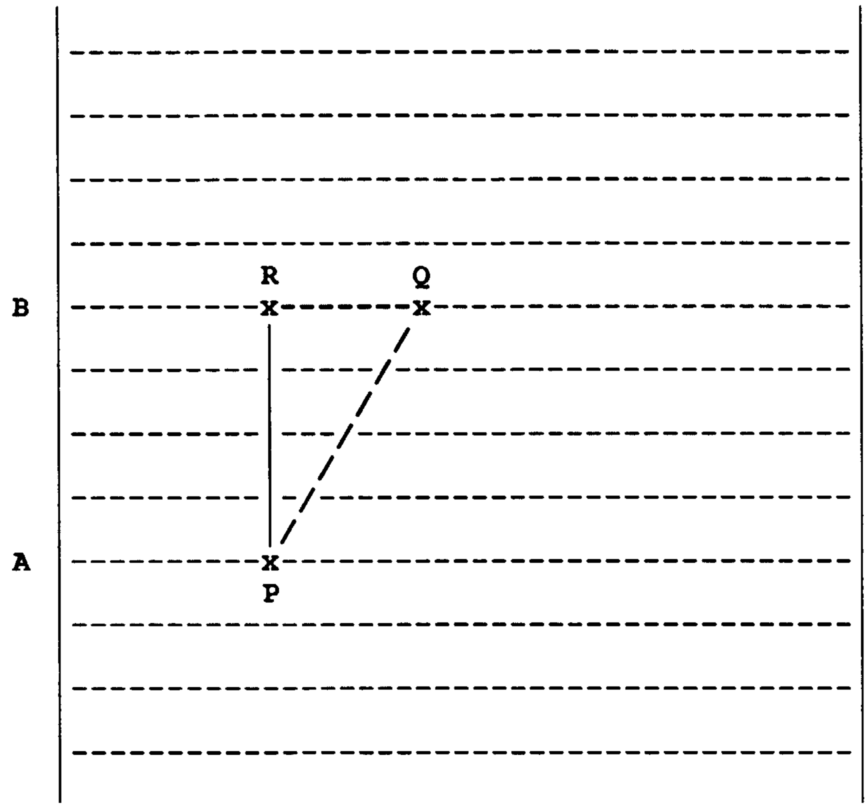
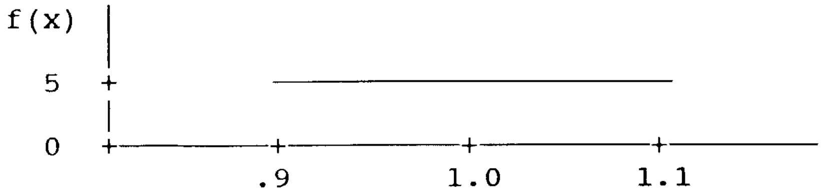
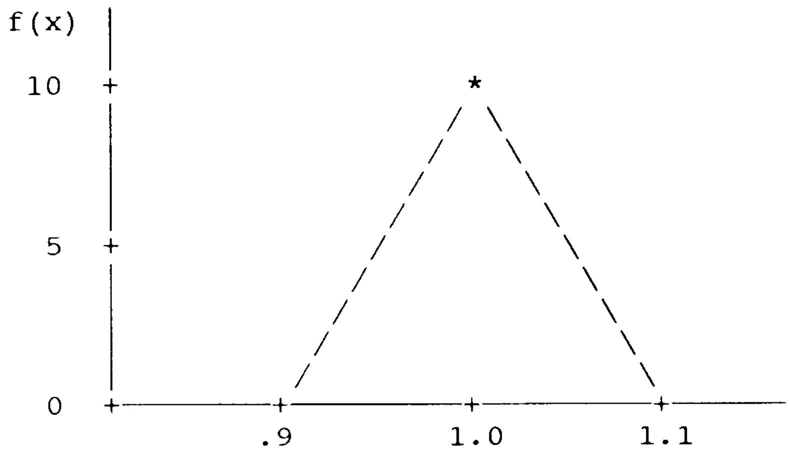


Figure 11: Distributions used to calculate effect of measurement error in table 9.

a) Uniform (.9 to 1.1)



b) Symmetrical Triangular (.9 to 1.1)



c) Skewed Left Triangular

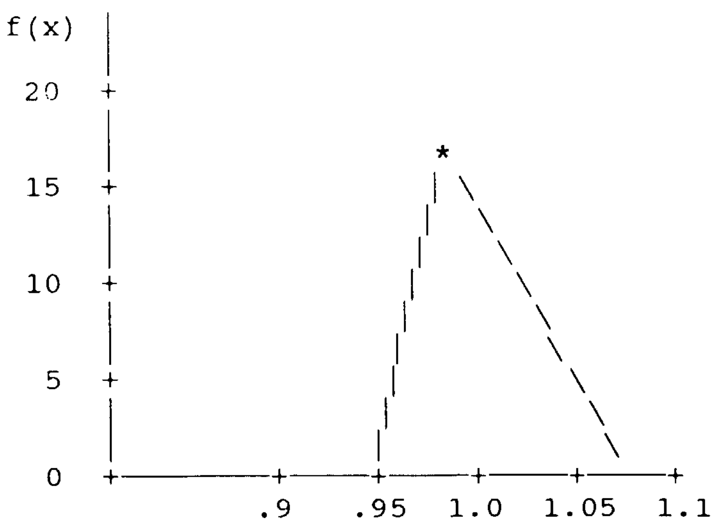
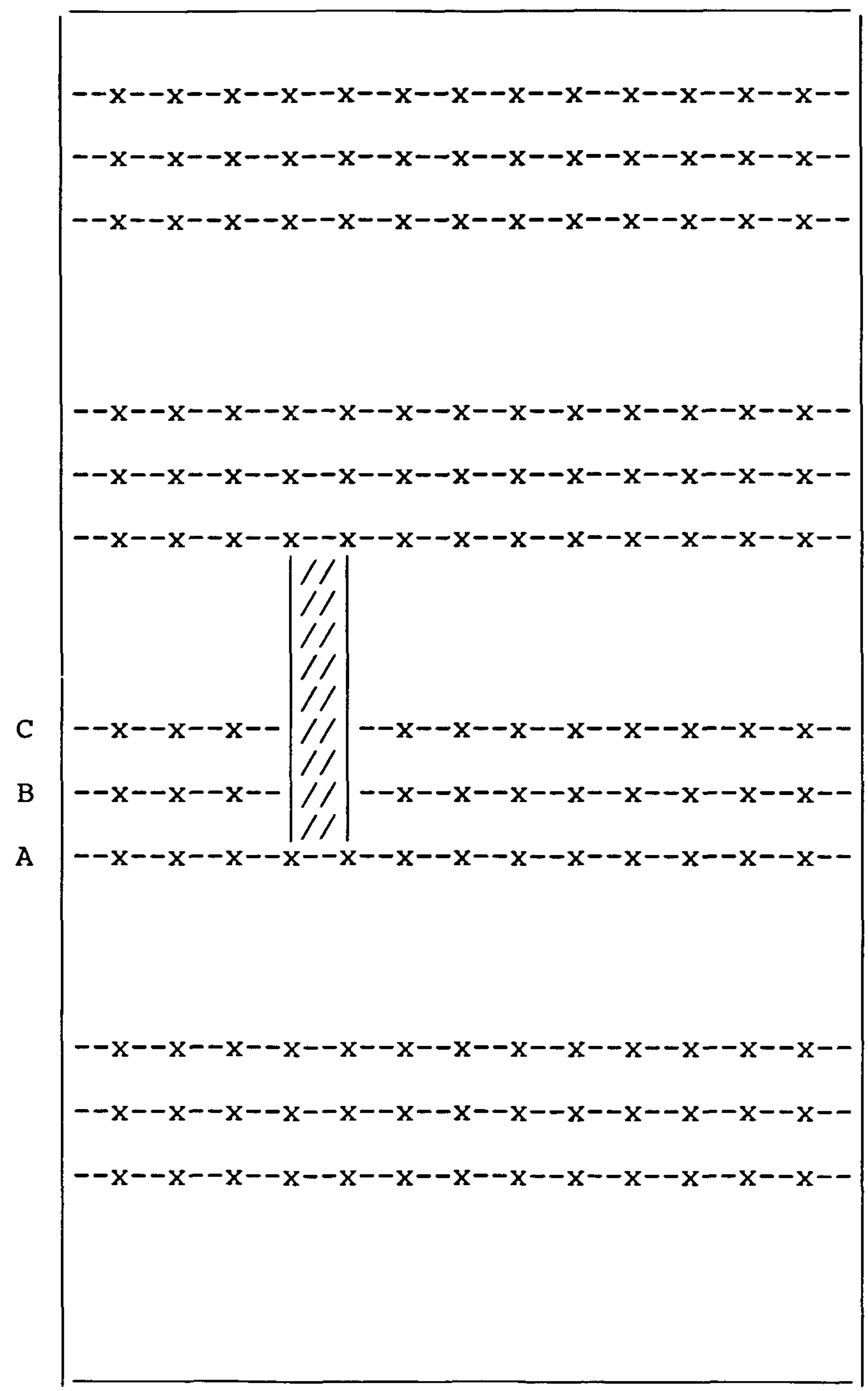
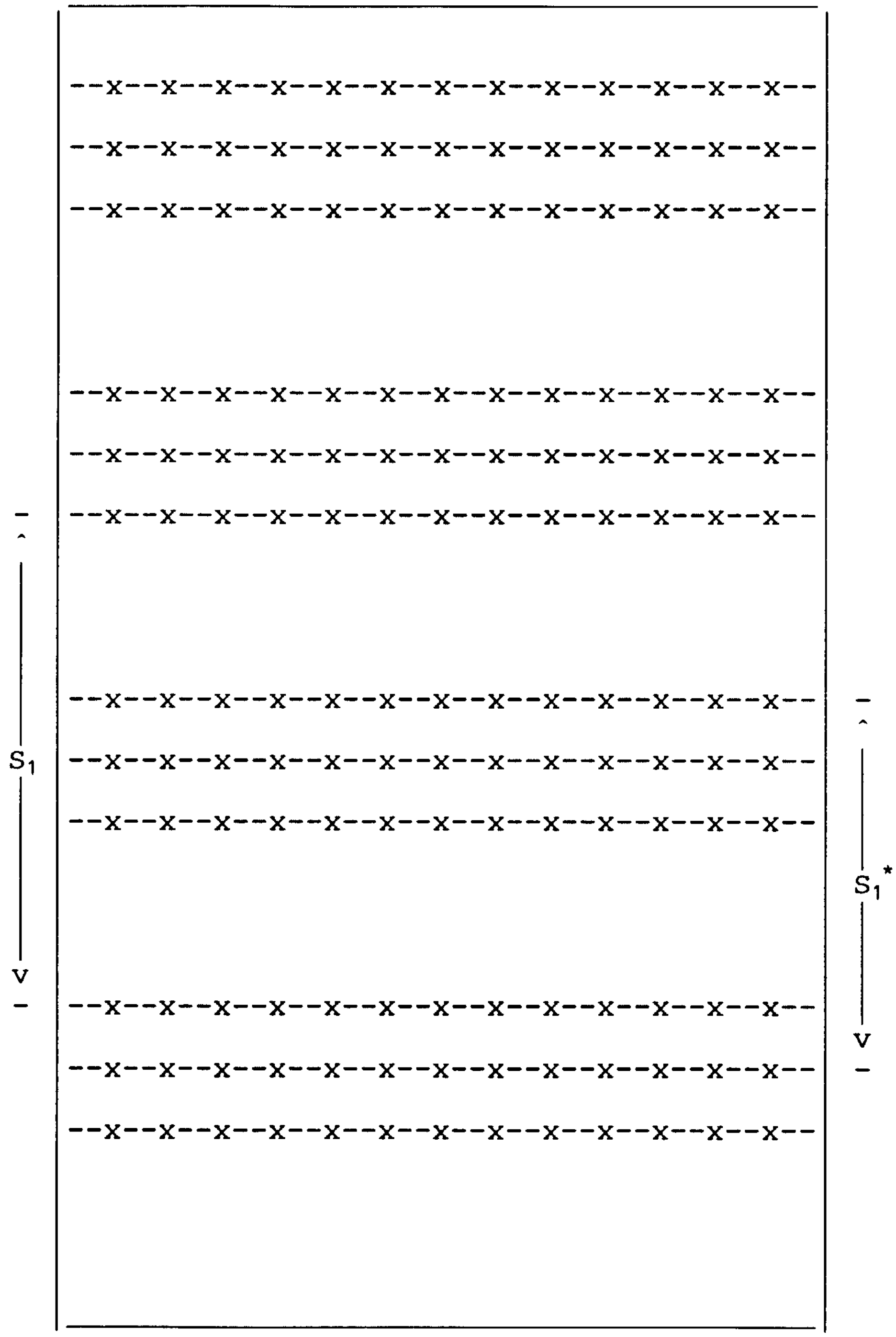


Figure 12: Proposed method for eliminating bias due to skip row planting.



O (starting corner)

Figure 13 - Illustration for appendix C.



☆ U.S. Government Printing Office : [illegible] - 261-461 20431