

## PREDICTING YIELDS FROM OBJECTIVE COUNTS AND MEASUREMENTS

Rapid changes in agriculture and related industries during recent years have brought about a very critical need for accurate, current predictions of final crop yields during the growing season. Predicting yields from objective counts and measurements is a new technique being developed by the Department to improve its predictions of crop yields. Predictions for nearly a century have been based almost entirely on mailed reports submitted by farmers who voluntarily appraise crop prospects during the growing season. The predictions and estimates generated from mail surveys have generally been good over the years. However, the marked increases in crop yields, greater specialization and larger sized farm units are but a few of the factors that have reduced the ability of a non-probability sample to supply accurate information about crop prospects. This development has led the Department of Agriculture to initiate a limited program of surveys based upon probability samples, the limitation being imposed by relative lists and available resources.

Presently the Department is conducting two general purpose probability sample surveys each year based upon an area sampling frame. The first, centered about June 1, is to obtain planted acreages, farm numbers, livestock inventories, poultry and dairy statistics, farm labor and various items of economic information associated with the farm headquarters. This survey provides an early season base for, planted acreages of crops, farm numbers,

*As presented by William E. Kibler,  
Convention of the Association of Southern  
Agricultural Workers - February 1964.*

and livestock inventories, as well as a listing of crops by fields that can be used for the random selection of fields in which to make objective yield counts and measurements for several crops. The second general purpose survey is conducted about December 1 and is similar to the June Survey except emphasis is placed upon harvested acres, production and livestock inventories. In conjunction with these two surveys objective yield surveys are conducted during the growing season, usually at monthly intervals, for corn, cotton, wheat, soybeans, tobacco, and for a number of tree crops which include oranges, lemons, peaches, pears, sour cherries, walnuts, filberts, and almonds.

Techniques for estimating and predicting yields have been developed which are based upon actual field counts and measurements rather than grower appraisals. Predicting yields by objective counts and measurements has many advantages over those based upon judgment estimates of crop condition or probable yield. The various fruit counts and maturity data collected from a probability sample provide a more precise picture of crop conditions and progress through the entire growing season. Pronounced changes in crop prospects due to improved varieties, cultural practices, and more favorable weather tend to be reflected more accurately by objective rather than subjective appraisals particularly in a year following a below average yield. With the final harvest estimate obtained from the objective yield survey the actual predicting error can be measured. Too, the estimate of yield along with

the acreage estimate based on the Enumerative Survey provides an estimate of production with known precision. Objective measures of yield, however, are much more expensive to collect than subjective.

The predicting models established to date for most crops are generally still in their infant stages. This is a new field of endeavor and many new relationships are yet to be discovered that will serve to improve current forecasting procedures. The principal disadvantage is that these predicting models cannot forecast the weather for a season. They merely reflect what can be expected on the average. Consequently abnormal freezes, droughts, or wet seasons occurring after the monthly prediction, which was the case with the August 1, 1962 prediction of large bolls, can only be accurately evaluated as they happen or after they have occurred. Another deficiency is that the precision of the prediction cannot be measured at the time each month's prediction is made. Most of the work to date in the field of objective counts and measurements has been of the very simplest type and has served only to lay the foundation for a future program that will enable the Department to more accurately appraise crop prospects during the growing season.

When any crop is mature and ready for harvest estimating yield is principally a sampling problem. The harvesting and weighing of the final crop production at harvest time, from a well designed sample of suitable sized sampling units, can produce estimates of yield that are as precise as desired provided biases associated

with small plots can be controlled. Experience to date in making objective counts and measurements indicates that most of these inherent biases can be controlled effectively by well defined field procedures and by intensive enumerator training, supervision, and quality checks. The following table indicates the size of samples (allocated to produce both State and national estimates) and the standard error, at the obtained level, for the four field crops included in the program.

TABLE 1: OBJECTIVE YIELD AND MEASUREMENT SAMPLING INFORMATION FOR WINTER WHEAT, CORN, SOYBEANS, AND COTTON

Crop	: Approximate: : size of : sample plots: : per field (acres)	: Number : of : fields (number)	: Total : aggregate : sample area: : to be harv. (acres)	: Standard error : of the : estimated : yield
Winter Wheat	.00021	2,300	.48	.25 bu.
Corn	.00455	3,300	15.02	.70 bu.
Soybeans	.00085	1,900	1.62	.30 bu.
Cotton	.00300	2,600	7.80	7.50 lb. lint

The number of fields in this table represents an objective yield program that is expected to be in operation by 1966. The yield of winter wheat produced in the nation's 33 million acres will be estimated with a coefficient of variation of less than 1 percent by the sample of plots aggregate areas of which are less than  $\frac{1}{2}$  an acre. For the 57 million acres of corn, the same precision is obtained from a sample of 3,300 corn fields (from which about 15 acres of actual corn would be harvested). Likewise, for cotton and soybeans the indicated sample size will result in a comparable sampling error for the yield of those crops. Estimates produced

for most of the individual State yields as an intermediate step in deriving national estimates have sampling errors of 3 to 5 percent for most crops. These yield per acre estimates at harvest time for the various crops and the probability survey data for acreages provide unbiased estimates of production that have measures of precision and reliability. The final objective counts and measurements from sample plots at harvest also serve to measure the accuracy of predictions that were made for the unit during the growing season and to determine relationships between early season plant characteristics and the final yield components.

Predicting yield for a crop during the growing season, before it reaches maturity, is a more difficult task than estimating yield at harvest time. Since the final yield has not materialized it is necessary to discover plant characteristics which may be used to predict the components of yield. The development of objective yield predicting formulas for yield components, must be based upon observable plant characteristics and a comprehensive knowledge of the fruiting behavior of the plant. These equations must translate plant characteristics observed on any date into accurate indications of various yield components.

Predictions currently being made can be divided into two categories, (1) tree crops and (2) field crops. Forecasting yield for tree crops is different from field crops in that by the date of the first forecast for tree crops, the bloom has occurred and all fruit for the season actually set. The predicting problem is then confined to estimating the fruit present and predicting droppage and sizing. However, for many

field crops the first forecast is often made before all the young fruit is set which requires forecasts for additional components. To illustrate this difference, Florida citrus will be used as an example of a tree crop and cotton as an example of a field crop.

#### PREDICTING CITRUS PRODUCTION

Florida oranges bloom in February or March. There is a heavy drop the following June, and sometimes an additional bloom in June. Early varieties begin to ripen in late September or early October with later maturing varieties ripening as late as March of the following year. The first forecast of production, for which yield per tree is needed, is made October 1. Fruit numbers are estimated by a sample survey conducted each year in August and September. The sample consists of about 4,000 randomly selected trees in 1,000 orchards. One of the more difficult tasks with objective fruit counts is that of counting fruit on the sample trees. Counting fruit on a sub-sample of limbs for the sample tree is a technique recently developed that is efficient and precise. Sample limbs and branches are selected on sample trees by a random process in which the probability of selection of a limb or branch is governed by size. The practical application of the technique is fairly simple. First the primary branches of the tree are assigned probabilities of selection based upon their cross-sectional

areas and one of these branches selected at random based on the probabilities assigned. The secondary branches emanating from the branch selected by the above method are treated in the same fashion, and the process is repeated until a branch of suitable size for counting is reached. The process of selecting a sample branch ends when a branch representing close to 10 percent of the total tree is reached. All fruit on the selected branch are then counted. An unbiased estimate of the total fruit on the tree is given by multiplying the number of fruit counted on the selected branch by the reciprocal of the product of the probabilities assigned at each stage of the branch selection as shown in formula (1).

$$(1). \quad X_i = \frac{1}{P_{p_i} \times P_{s_i}} C_i$$

$X_i$  = Estimate of number of fruit for ith tree

$P_{p_i}$  = Probability of selection for primary limb ith tree

$P_{s_i}$  = Probability of selection for secondary limb ith tree

$C_i$  = Count of fruit on sample limb for the ith tree

This procedure from a 4,000 tree sample produces an estimate of fruit per tree with a standard error of about 2 percent.

A fruit size and droppage survey to collect information for projecting size and droppage to date of harvest is conducted monthly from October 1 through the following May in about 600 randomly selected orchards with 2 sample trees per orchard. There is evidence that fruit growing on the south side of a tree

tends to mature first. For this reason, it is important that the point on the circumference of the tree where fruit is selected for measurement be chosen at random, or at least in such a manner as to give the sampler no opportunity to introduce a bias by letting him select the position by judgment. For the size measurement survey, which is made to provide a basis for projecting harvest sizes, the 10 closest fruit to a predetermined point on sample trees are calibrated to establish circumference and then converted into spherical volumes and analyzed. The relative increase in fruit size between September 1 and harvest has been surprisingly constant from year to year. Since the growth rate is of interest, size data are plotted against time on a semi-logarithmic scale. That curve provides a good basis for projecting fruit size at the time of measurement to size at harvest when the stage of maturity of the fruit at the time of measurement is taken into account. From fruit size measurements obtained from packing house bins, a relationship was worked out between the size of fruit and the number of fruit per box, and this relationship is used to convert projected size at harvest to the expected number of fruit per box.

For the droppage estimate sample limbs are chosen, in a manner similar to those for estimating fruit numbers, that represent about 2 percent of the total tree. Sample limbs are tagged and the fruit on these counted monthly during the growing season.



Total droppage to harvest is projected from the droppage data collected to date each month.

The estimate of fruit per tree in terms of boxes is given by formula (2) below.

$$(2) \quad Y = X (1.0 - d) \times \frac{1}{N}$$

Y = Estimate in boxes per tree

X = Average number of fruit per tree in September limb count

d = Estimated droppage rate from September through harvest

N = Number of fruit required per box based on September or later size projected to a harvest size

This estimate Y (boxes of fruit per tree) when multiplied by number of trees of bearing age gives an unbiased estimate of total production. A ratio estimate which relates the current counts and measurements to a base (usually from the previous year) and is expanded by production from the base year, has provided good predictions of final production. New research work continues annually with the citrus objective counts to find more effective ways of translating these measurements and counts into more precise predictions.

## PREDICTING COTTON YIELDS

Predicting cotton yields by objective counts and measurements is a more difficult task than making fruit predictions. Predictions of the cotton yield must be made August 1, September 1, and October 1. At the time of the first prediction (August 1), cotton in many areas has just begun to set young fruit and neither of the two components which determine final yield are present for making projections of survival and development. These two components are (1) the number of mature bolls per acre and (2) the average weight per boll, both of which can only be measured accurately at harvest time. Two predicting models have been developed, to project fruiting patterns that might enable these components to be predicted. Refinements in these models are made annually as additional information has become available. This paper will discuss the development and techniques used for only one of these models. This model, termed "The Rate of Fruiting Model," makes use of known fruiting characteristics of the cotton plants and permits plant observations collected during the growing season to be translated into indications of final yield in logical fashion.

In crop seasons when cotton is late and fruiting is not well advanced this model is generally more reliable. When fruiting is well advanced on August 1 the second model based on the "Rate of Survival" is usually better.

On August 1 the cotton plant is setting fruit rapidly and a substantial part of the fruit is still to be set. Thus it is necessary to predict the fruit to come. Fortunately the fruiting pattern of the cotton plant follows an orderly procedure. When the plant is three to four weeks old it begins to set fruit. The fruit first appears as a bud or square which in about three weeks will develop into a bloom. After several days the petals dry up and drop from the plant and the bud becomes a small boll. It then takes about two and a half weeks more for the small boll to attain maximum size. In most areas it takes a fruit bud about six weeks to reach its maximum size and then another three to four weeks to mature. For the "Rate of Fruiting Model" the observed fruit on the plant at any time is expressed as a fraction of the maximum which the plant will carry. This is then related to the length of time which has elapsed since squaring started. Chart 1 illustrates the nature of this relationship as shown by the dashed curved line. The straight solid line represents the approximation to this curve which is used in practice. This curve shows that the appearance of the first square constitutes the starting of the fruiting curve at the zero point, then 3 weeks later we have our first bloom and at this time about half the maximum fruit load. At the end of about

six weeks we expect our first large boll to appear at which time the plant has set the maximum fruit load which it will carry.

#### CHART 1

In order to use this maximum fruit load relationship, the fruit counts must be made and the fruit present classified into several categories to determine the stage of development for each sample field. Objective counts and measurements are currently being made in about 2,100 sample fields in 10 States. The sampling unit for cotton consists of a plot 2 rows wide and 10 feet long. Two such plots are layed out in a sample field. A typical cotton State will have about 200 randomly selected sample fields chosen with probability proportional to size. The sample plots are randomly located in each field in late July, the time of the first visit. The fruit are counted for each set of sample plots by the following types.

$X_1$  = Large bolls

$X_2$  = Small bolls

$X_3$  = Blooms

$X_4$  = Squares

To be classified as a large boll a boll must measure one inch or more in diameter, small bolls are fruit for which two or more days have elapsed since blooming but whose diameter is less

than one inch. Blooms consist of one and two day old flowers. Squares are fruiting buds, whose triangular shapes have reached a minimum size of 1/8 inch at the base.

After the fruit for the sample plots have been counted and classified according to type, each sample field is classified into one of the following categories:

- |            |   |
|------------|---|
| Category 1 | 20 or more bolls per 40 feet of row   |
| Category 2 | Small bolls or blooms present but less than 20 large bolls per 40 feet of row |
| Category 3 | Squares only fruit present  |
| Category 4 | Fruiting has not started  |

Category 1 represents fields in which fruiting has been going on for six weeks or more. The reason for requiring 20 or more large bolls to classify the field into Category 1 is that on the average about one-half the plants would have at least one large boll. Category 2 represents fields that have been fruiting for three weeks or more but generally less than six weeks.

Category 3 consists of fields that have started fruiting but have not been setting fruit for three weeks. Category 4 is comprised of fields where fruiting has not started and no fruit is present to be counted.

Having classified each sample field into one of these four categories, the average fraction of a maximum load can be determined

for each category. Those fields on the left hand portion of Chart 1 with no fruit present have set none of their maximum fruit load. Fields which have squares but no other fruit are at some point between zero and three weeks of age and for a large number of fields are assumed to have been fruiting on the average, about one and a half weeks and have about one-fourth the maximum load. Similarly for Category 3 (fields with small bolls or blooms and fewer than 20 large bolls) over a large number of fields it is assumed that will have an average age of about four and a half weeks which implies about three-fourths of the maximum load. Finally for the more mature category (those fields with 20 or more large bolls) their maximum load has already been reached. To arrive at an average fraction of a maximum load (AMFL) each of the fractions of a maximum load is multiplied by the fraction of fields in the corresponding category as shown by formula (3).

$$(3) \quad AMFL = 1.00F_1 + .75F_2 + .25F_3 + .00F_4$$

$F_i$  = Fraction of fields in the  $i$ th maturity category

$i = 1, 2, 3, 4$

With the average fraction of fruit load derived and the fruit counts determined, an actual maximum fruit load (MAXF) can be computed.

$$(4) \quad \text{MAXF} = \frac{\sum_{i=1}^4 X_i}{\text{AMFL}}$$

$X_i$  = Number of fruit of the  $i$ th type

AMFL = Average fraction of a maximum fruit load

Having obtained the maximum fruit load the next step required is to calculate an average slope from the fruiting curve shown in Chart 1. The average slope, of course, is the average rate at which fruit are being added to the plant. Those fields which have not started to fruit have a slope zero. From the time squaring begins until large bolls are set (or for about six weeks) the slope is simply the height of the curve, which is one, divided by the length of time fruiting has been going on, which is six weeks. Hence, the slope for this interval is  $1/6$ . After a sample field has reached the large boll maturity category, the slope is then again zero. To arrive at an average slope formula (5) is used which weights the slope for each maturity category by the fraction of samples falling into each category.

$$(5) \quad S = 0 \cdot F_1 + 1/6 (F_2 + F_3 + F_4)$$

$F_i$  = Fraction of fields in  $i$ th category

$S$  = Average slope

Sample fields with no fruit set have been placed at the zero point on the curve since by late July squaring is generally ready to start or may have already started but the fruit was

shed. When there are very few fields in this category, this seems to be the most valid assumption to make. If a large fraction of the fields were in category 4 then it could not be assumed that these fields are ready to start fruiting but that they are likely a week or so away from fruiting and should be treated as though the rate of fruiting was zero.

To convert the average rate of fruiting or average slope from a relative value to an absolute total the average slope (S) is multiplied by the derived maximum fruit load (MAXF). This converts the rate to absolute numbers of fruit termed, "Weekly Rate of Fruiting" (WRF).

$$(6) \quad WRF = S \times MAXF$$

Having obtained the weekly rate of fruiting (or the rate at which fruit is being added as of August 1) the model makes use of the fact that if plants are adding fruit at a rapid rate than it is likely they will continue to do so, and will form large bolls in the next few weeks. The number of bolls added after August 1 has been related to the weekly rate of fruiting (WRF) as a means of predicting the additional bolls that are to be added to the plant after August 1 and which can be expected to mature in the manner shown in Chart 2 below.

CHART 2



The estimated bolls to be added (EEA is a linear function of the weekly rate of fruiting as follows:

$$(7) \quad EEA = b_1 + b_2 (WRF)$$

EEA = Estimated bolls to be added

$$b_1 = 18.0*$$

$$b_2 = 1.640*$$

With an estimate for the number of bolls to be added established, the prediction of total bolls expected (TEE) at harvest time is made using formula (8).

$$(8) \quad TEE = X_1 + X_2 + X_3 + EEA$$

TEE = Total bolls expected

$X_1$  = Count of large bolls

$X_2$  = Count of small bolls

$X_3$  = Count of blooms

EBA = Estimated number of bolls to be added

On August 1 the large bolls present have not begun to open and consequently cannot be picked and weighed. Hence, the predicted boll weight used at that time is the historic average over a period of about 5 years. This has proved to be a reasonable assumption in that the between year boll weight variation over a group of States is generally less than 2 percent. Having the second component needed in the model established, a net yield prediction can be made by this model using formula (9).

---

\* Parameters developed from data collected from 1956-1962.

$$(9) NY = \frac{(A)(TBE)(G)(BW)}{(GP)(40)(RS)} - HL$$

NY = Net yield at harvest time (pounds of lint per acre)

A = Square feet per acre = 43,560

TBE = Total bolls expected

G = Pounds of lint cotton per pound of seed cotton

BW = Average boll weight in grams

GP = Grams per pound = 453.99

RS = Average row space

HL = Estimated harvesting loss

When the above constants are inserted the equation reduces to:

$$NY = \frac{(43,560)(.380)(TBE)(BW)}{(453.99)(40)(RS)} - HL = \frac{.888 (TBE)(BW)}{RS} - HL$$

The following table gives a comparison of the August 1 predictions for the two components and the final estimate made at harvest time for 1962 and 1963.

TABLE 4: COMPARISONS OF COMPONENTS PREDICTIONS ON AUGUST 1 FOR RATE OF FRUITING MODEL AND FINAL HARVEST ESTIMATE FOR OBJECTIVE YIELD 1962 AND 1963

Item	August 1 Prediction		Final Estimate	
	1962	1963	1962	1963
Number of bolls	375	361	336	376
Average boll weight	4.92	4.89	4.82	4.94

In general the August 1 prediction has had a forecast error of 4 to 5 percent over the past 3 years.

BY PRODUCE DATA GENERATED BY NEW PROGRAM

The objective counts and measurements made during the growing season for use in the predicting models for the various crops should be of considerable value in numerous other areas of Agriculture such as Economics, Agronomy, Agricultural Policy, and Regulatory Work. For example corn yields in the South have been changing quite rapidly during recent years and have increased about 28 percent during the past 5 years. The following data gathered in making objective counts and measurements during this period provide a good picture of some of the changes that have occurred. During this period farmers have increased the number of stalks per acre from about 6,800 to 8,400, an increase of 24 percent. The number of ears harvested per stalk has increased from 1.07 to 1.11. An average ear of corn today can be expected to yield about .322 (South) pounds of grain at 15.5% moisture. This ear characteristic has been almost identically the same over the 5 year period.

The increased stalk population has been the result of closer planting in the drill as today's average row width is only .03 of one foot narrower. Gleanings from harvested fields indicate that close to 8 percent (or more than 3.5 bushels per acre) of the crop produced is lost or missed in harvesting. Farmers in the South on the average applied 60.9 pounds of nitrogen (N), 34.9 pounds of phosphate ( $P_2O_5$ ), and 37.1 pounds of potash ( $K_2O$ ) to each acre of corn in 1963. This represented increases of 9.6, 0.8, and 2.6 pounds per acre from the previous year for nitrogen,

phosphate, and potash, respectively. Similar type information is available for most crops for which objective counts are being made. There are also many other possible uses for information gathered on the final visit to sample fields. During the final pre-harvest visit to samples to make objective counts and measurements, which is usually a week to 10 days ahead of actual harvest, small amounts of grain, cotton, etc., are harvested, weighed, and used for certain laboratory determinations. These small amounts could yield very important information about crop quality prior to farmer harvest. These are but a few of the residual uses that can be made from the data the Department is currently collecting for use in strengthening predictions of yields during the growing season.

Our experience with objective yields so far has been encouraging. That precision sample estimates of yield just prior to harvest can be obtained by relatively small samples has been demonstrated. Early season predictions of yield based upon counts and measurements have been fairly satisfactory even with the forecasting procedures in their present state of development. The prospects for further improvement seem good, but with the degree of precision that seems desirable at this time, there appears to be little prospect of any substantial reduction in sample size or the costs associated with collecting the objective observations. Data collection costs are a severe limitation to the applicability of these methods and their use is likely to be restricted to important or high-valued crops for which accurate supply information is needed. For crops of lesser importance,

predictions of yield will continue to be based upon grower estimates obtained by the less costly mailed survey. For all crops estimated by SRS, the mailed survey will continue to play an important part in establishing yields for States and local areas. However, workable methods based upon objective counts and measurements have been devised and perhaps the limitations imposed by costs may be overcome and a wider application result.

CHART I: FRUIT LOAD RELATIVE TO MAXIMUM

Fraction of  
Maximum Fruit  
Load

1.00

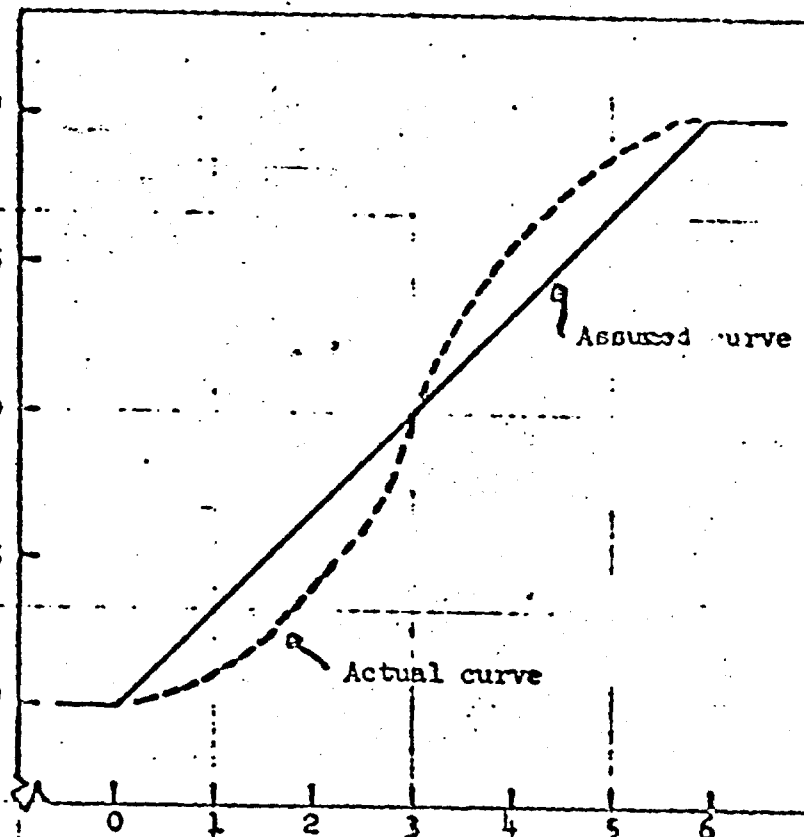
.75

.50

.25

0

Weeks of Fruiting



Assumed curve

Actual curve