

FORECASTING COTTON YIELDS FROM FRUIT AND  
PLANT CHARACTERISTICS

by

H. F. Huddleston and W. E. Kibler

Research and Development Branch  
Standards and Research Division

Prepared for the Agricultural Economics Staff Seminar  
November 1962

Washington, D. C.

September 1963

## Forecasting Cotton Yields From Fruit and Plant Characteristics

This paper describes the development of two forecasting models which are in use in the current estimating work and will illustrate their use for a group of five States for August 1, 1962 data. The forecasting models which have been developed are based on the observed characteristics of the fruiting pattern of the cotton plant. As these models have been developed over the past several years, additional information on the fruiting pattern has been introduced into the models so that they now reflect a greater use of this kind of information than they did earlier. In addition, it is expected that the need for greater reliability will result in even more of this type of information being introduced into the models in the future.

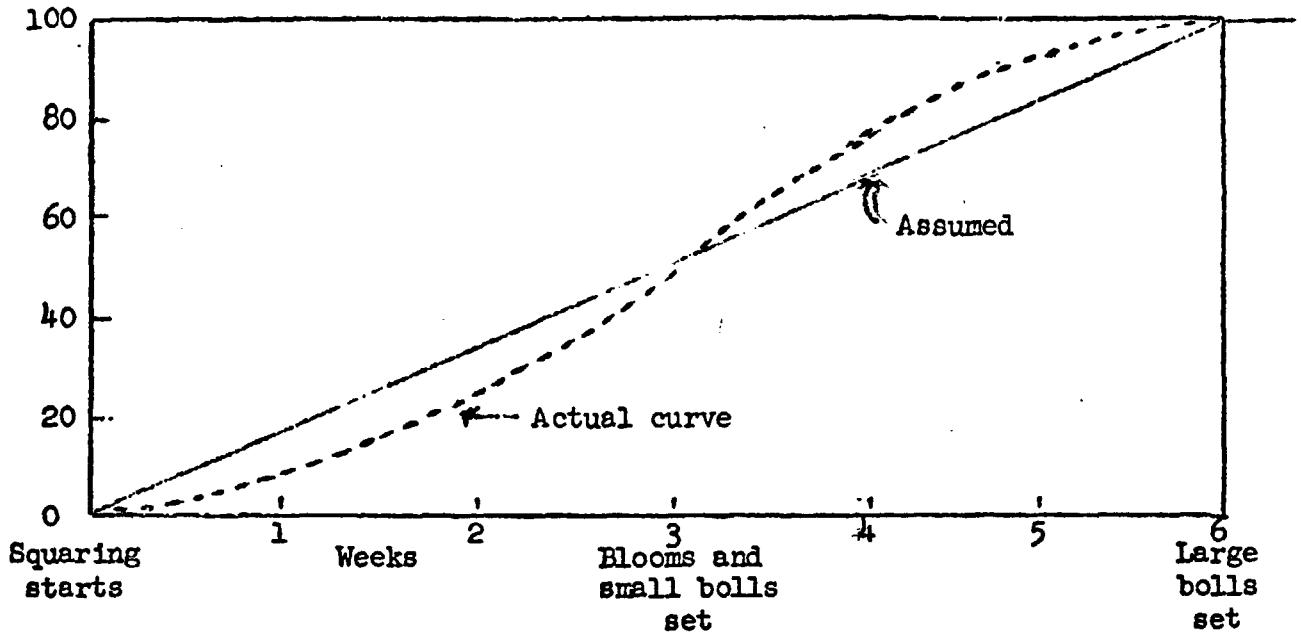
Several earlier studies were available as a basis for deriving these models. One study was made in the early twenties, another in the late thirties and early forties in connection with a crop weather project, and a third study was made in connection with the irrigation of cotton in the West. All of these studies made available detailed information on the fruiting pattern of the cotton plant by tracing the history of individual bolls through a growing season. This type of information is extremely helpful in developing reliable, predicting models. The use of this detailed data collected under experiment station conditions posed several problems, among which are: (1) the construction of the models, (2) a logical translation of the model into observable characteristics in surveys at less frequent intervals (in this case at essentially monthly intervals), and (3) the development of constants or parameters for the models which would apply to the data observed and to the forecast date specified. 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. For this, the fruiting pattern of the cotton plant is required.

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 approximately three weeks will develop into a bloom. After several days the bloom will dry up and drop from the plant and become a small boll. It takes about two and a half weeks more for the small boll to attain maximum size. In most areas, it takes a fruit bud approximately six weeks to reach its maximum size and then another three to four weeks to mature.

The first model which will be described is based upon a fruit load curve which we developed from the early studies. Here the observed fruit on the plant at any time is expressed as a percent of the maximum fruit load 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 curve line. The straight solid line represents the approximation to this curve which we use in practice. From this we see that if we consider a time when the first square appears as the zero point, three weeks later we have our first bloom and at this time we have about half of our maximum fruit load. At the end of approximately six weeks we expect our first large boll to appear and we have our maximum fruit load which the plant will carry.

Chart 1. Fruit Load Relative to Maximum

Observed fruit load as percent of maximum



Now in order to use this maximum fruit load relationship, we must first count the fruit into several categories and then determine the stage of development for each field we are working with.

#### Fruit Categories

- $X_1$  = Large bolls
- $X_2$  = Small bolls and blooms
- $X_3$  = Squares

The fruit are counted for each set of sample plots by the following categories. The large bolls one inch or more in diameter are considered one category, the small bolls and blooms as another category, and the squares as our third category. The sample plots are then classified into four stages of development based upon the type of fruit counted on the plant.

- Category 1 - 20 or more large bolls per 40 foot row
- Category 2 - Small bolls or blooms and less than 20 large bolls
- Category 3 - Squares only
- Category 4 - Fruiting has not started

In Category 1 is the group where fruiting has been going on for six weeks or more; these are those plots that have 20 or more large bolls. The significance of requiring 20 or more large bolls before we classify the plot into Category 1 is a rather simple one. If there are as many as twenty large bolls in a plot this means that on the average half of the plants will have one or more large bolls. Category 2 are those plots that have small bolls

and/or blooms but as yet do not have as many as 20 large bolls in the plots. Category 3 are those sample plots that have only squares but no older fruit such as small bolls, blooms, or large blooms. Category 4 is the case where fruiting has not started and we have no fruit. Having classified each plot into one of these four categories, we then will refer to Chart 1 and see if we can determine an average percent of a maximum load for each group. Those plots that have no fruit obviously are on the left hand portion of Chart 1 and have zero percent of their maximum load. Those plots which have squares but no older fruit on them are some place between the zero point and three weeks of age. For a large number of plots we assume that they will average a weeks and a half old and will have 25 percent of the maximum load. For those plots that have blooms and small bolls but no large bolls, we will assume again for a large number of fields that they will have an average age of four and a half weeks which on our chart implies 75 percent of the maximum load. Finally, our last category, those fields that have 20 or more large bolls, has 100 percent of their maximum load. In order to arrive at an average fraction of a maximum load, we weight each one of these fractions of a maximum load by the percent of the plots in that category.

$$\text{Average maximum fruit load} = \text{AMFL} \\ = 1.00 F_1 + .75F_2 + .25F_3 + 0.F_4$$

Where  $F_1$  = Fraction of fields in each maturity category

Maturity category	Number of samples	Fraction of samples ( $F_1$ )
1	571	.490
2	271	.232
3	247	.212
4	77	.066
Total	1,166	1.000

If we look at the above formula and table we see for the August 1, 1962 data what this classification was for a group of States which we call the late fruiting States. This group of States includes Arkansas, Mississippi, Louisiana, Oklahoma, and West Texas. In the third column, we see that .490 of the plots were in Category 1 and .232 in Category 2, .212 in Category 3 and .066 in Category 4. If we use these  $F_1$ 's in the above formula, we calculate the average maximum fraction of a full load as .717.

$$\text{AMFL} = (1.00)(.490) + (.75)(.232) + (.25)(.212) + 0 (.066) = .717$$

Now if we look at the table below we will see the actual count of fruit by the three categories which were described earlier.

Table 1. Fruit Counts Per 40 Feet of Row

Category of Fruit	Number
Large bolls	79.1
Small bolls and blooms	159.6
Squares	542.6
Total fruit	781.3

Total of all kinds of fruit on the plant on August 1 was 781.3. Now if we take a look at the following formula we see how the maximum fruit load is derived.

$$\text{Maximum fruit load} = \text{MAXF} = \frac{\sum X_i}{\text{AMFL}}$$

Where  $X_i$  = Number of fruit for the  $i$ th category

$$\text{Maximum fruit load} = \frac{781.3}{.717} = 1089.7$$

Having obtained the maximum fruit load we then calculate an average slope from the fruit load curve as shown in Chart 1. The average slope, of course, is the average rate at which fruit are increasing on the plant. It is clear, if we look at the chart, that those fields which have not started to fruit have a slope of zero. For those fields in which squaring has started during the first six weeks the slope is simply the height of the curve, which is one, divided by the length of time the fruiting has been going on, which is six weeks, so the slope during this interval,  $1/6$ .

After the fruit has reached the large boll category, the slope is then again zero. In order to get an average slope for the group, we weight the slopes of these three groups by the fraction of the samples that fell in these three groups as follows:

$$\begin{aligned} \text{Average slope} = S &= 0 \cdot F_1 + 1/6 (F_2 + F_3 + F_4) \\ &= 1/6 \sum_{i=2}^4 F_i = 1/6 (.232 + .212 + .66) \\ &= .085 \end{aligned}$$

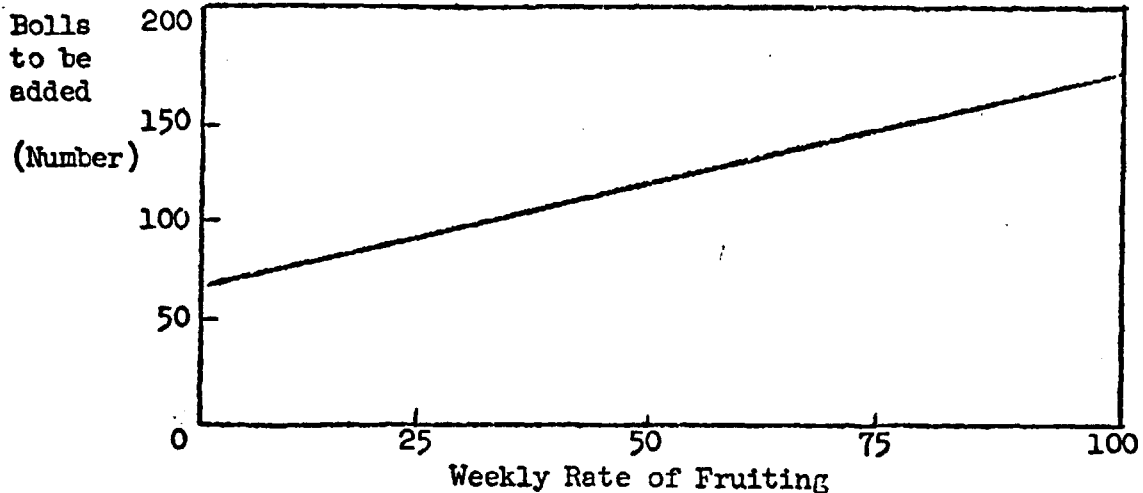
We have placed the fields with no fruit at the zero point on this curve for the reason that squaring is generally ready to start or may have already started but the fruit has shed. For situations in which there are very few fields in this category this seems like the best assumption to make on the slope for fields with no fruit. If a large fraction of the fields were in Category 4, then we should not assume that they were just ready to start fruiting but were maybe a week or so away from fruiting and should treat them as though the rate of fruiting was zero. In effect, the average slope is computed rather simply by getting the fraction of the fields that fall in Categories 2, 3, and 4 and multiplying them by  $1/6$ . If we look at the computations made above we see that the rate of fruiting was .085. Next, we convert the rate of fruiting from a relative value to an absolute total by multiplying it by the maximum fruit load. This converts the rate to absolute numbers of fruit (called Weekly Rate of Fruiting)

$$\text{Weekly Rate of Fruiting} = \text{WRF} = S \times \text{MAXF} = .085 \times 1089.7 = 92.6$$

Now having obtained the weekly rate of fruiting or the rate at which fruit are being added to the plant as of August 1, we make use of the fact that if plants are adding fruit at a rapid rate then it is likely they will continue to do so and will form large bolls in the next few weeks. The number of bolls that are added after August 1 has been related to the weekly rate of fruiting 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.

Chart 2. Bolls to be added after August 1 based on weekly rate of fruiting



For August 1, 1962, with a weekly fruiting rate of 92.6, we see that if we read on the line above 92.6 that 153 bolls will be added after August 1 which can be expected to mature. Having established the number of bolls to be added, we are then ready to forecast large bolls expected.

Now looking back to Table 1 we see that with 153 bolls to be added that the total bolls we may expect to reach maturity by this method is  $X_1$  (large bolls present), plus  $X_2$  (small bolls plus bloom) which gives a total bolls present of 238.7 for this date, plus  $\hat{Z}$  (bolls to be added after August 1.) These components give a total bolls expected to mature of 391.7.

$$\begin{aligned} \text{Total bolls expected} &= \text{TBE} = X_1 + X_2 + \hat{Z} = 79.1 + 159.6 + 153.0 \\ &= 391.7 \end{aligned}$$

By way of summary, let us recap model one. First, we wish to predict a maximum fruit load for the sample units; next we computed the average rate at which fruit are being set on the plant as of August 1, and finally we use the average rate at which bolls are being set on the plant to predict the additional bolls to be formed after August 1. This component with the count of bolls already set provides the forecast of total bolls expected at harvest time.

For model 2 we were able to glean from the same earlier studies the fact that fruit which appeared on the plant first had the best chance of surviving. Fruit which appeared late in the season have very little chance of surviving. Since the conditions under which these observations were made were highly specialized, we decided to obtain our own probability of survival for the classes of fruit that we counted on August 1. We did this by tagging the large bolls and the small bolls on August 1 with different colored tags. The only fruit that we did not tag were the squares which appeared later as the bolls without tags. After we tagged bolls and blooms for several years,

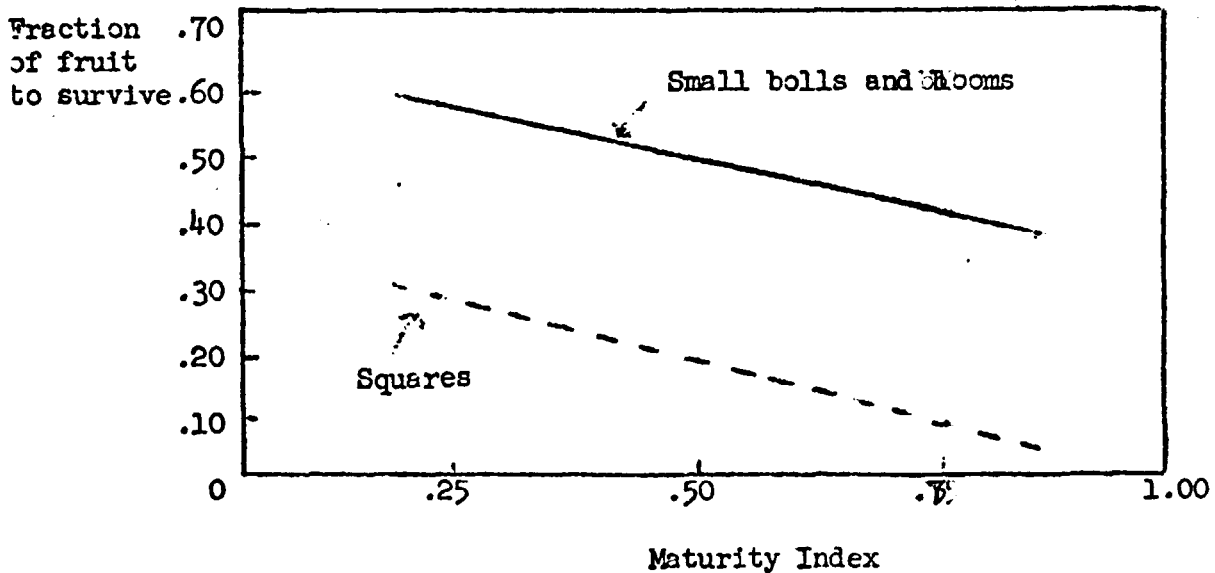
we derived fairly stable averages of the fraction of fruit that would survive in each one of the categories. We used these average probabilities for several years and for large bolls the average fraction surviving has worked quite well. However, for the small bolls, blooms, and squares, it has fluctuated somewhat from year to year. The apparent reason for this is that the large bolls having survived or been retained by the plant till August 1 are very unlikely to be shed at all. While small bolls which may be just barely beyond the bloom stage are much more likely to be shed than a small boll which is about to reach a large boll category. It should be clear that the weekly time scale used on Chart 1 is somewhat crude and for the purpose of getting the age of different kinds of bolls was not satisfactory for the purpose of indicating the age of the average fruit. We constructed a maturity index in order to have a more reliable measurement of age.

$$\text{Maturity Index} = \text{MI} = \frac{X_1}{X_1 + X_2} \quad 0 \leq \text{MI} \leq 1$$

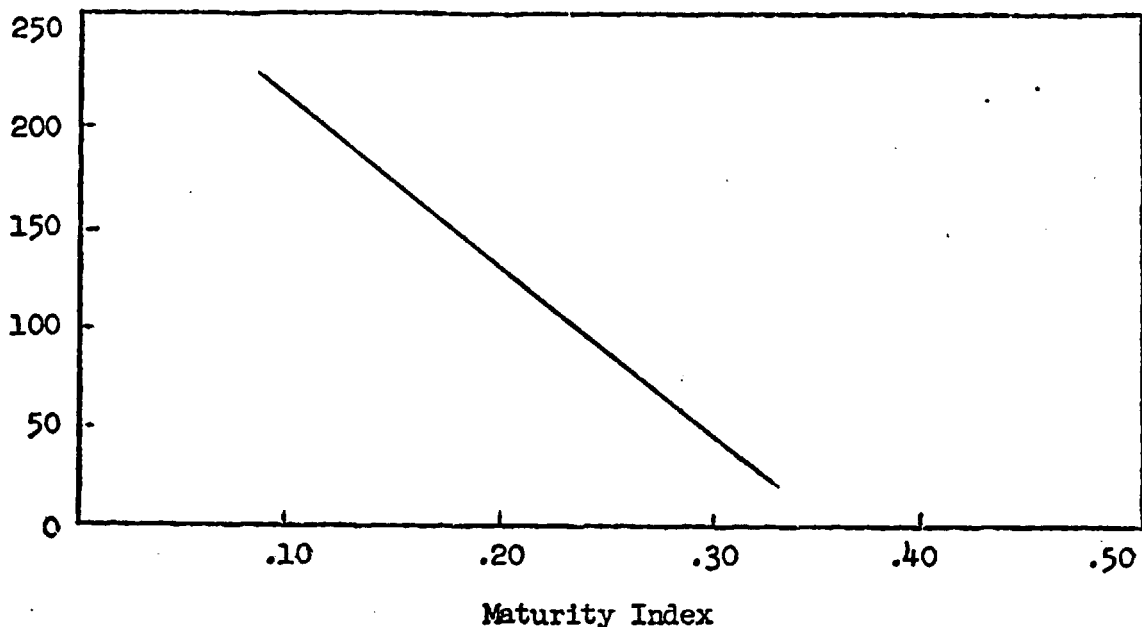
The maturity index is simply the ratio or the percent which the large bolls are of the total bolls on the plant. If a sample field has no large bolls, it is assumed to have zero or no maturity. In other words, if a plant has only one large boll on it and maybe 50 small bolls then on the average the bolls are quite young. While if it has 50 large bolls and 50 small bolls, then the bolls are substantially older. If all the bolls are large, then the field is mature. Chart 3 shows the fruit expected to survive by the small boll and bloom category and the square category as related to the maturity index.

Chart 3

(a) Survival of Small Bolls, Blooms, and Squares



(b) Large Bolls From Fruit Not Set



One additional problem remains. Since all the fruit have not been set by August 1, it becomes necessary, in most States, to make some allowance for large bolls to mature from fruit not yet set. This could be done in several ways, but we have indicated the direct relationship between large bolls to develop from fruit not yet set and related this to the maturity index. This relationship is shown by the (b) portion of Chart 3.

For August 1, 1962, the large bolls were approximately 30 percent of the total bolls on the plant. The index was calculated using the numbers of fruit given in Table 1. The value of .290 is read on Chart 3 (a & b) for the three different curves shown. The probability of survival for the squares is .281, for the small bolls and blooms .578, and the average for the large bolls is .795, which are applied to the counts found in Table 1. The large boll expected from each one of these categories of fruit is determined as follows:

Large bolls on plant expected to survive	= 62.9
Small bolls and blooms expected to survive	= 92.2
Squares expected to survive	= 152.5
Large bolls to come from fruit	= 52.9
<hr/> Total bolls expected at harvest	<hr/> 360.5

The large bolls expected from fruit not set is indicated and the totals from the various types of fruit set gives the total bolls expected of 360.5 for model 2 which we call the Probability of Survival Model.

Having calculated the bolls expected by these two methods, we then do the obvious and average these two since one seems to be performing about as well as the other. The large boll forecast as of August 1 by this procedure gives 376.1. It is not possible to forecast weight per boll directly since on August 1 practically none of the bolls are open, so we must rely on an historical average boll weight. We use the boll weight of seed cotton expressed in grams. For this group of States, the historical average was



5.009 grams. This is then multiplied by the number of bolls to obtain the number of grams of seed cotton on 40 feet of row which is the length of row in the plots. The next calculation converts the weight of seed cotton per 40 feet to a gross yield per acre using the factor of .266 to give us a gross yield of 501 pounds. Having obtained the biological or gross yield, it is still necessary to make allowance for the fact that the grower will not harvest all of the cotton in his field. The post-harvest survey, which is conducted after the farmer has completed his harvesting, for the past five or six years indicates a harvesting loss of about six percent. A harvesting loss of 30 pounds is subtracted from the gross yield to obtain a net yield of 471 pounds. The corresponding calculation for November 1, 1962 results in a net yield of 466 pounds. This comparison using only one year should not be considered as a measure of the accuracy for the August 1 forecast. On the average, the spread between the August 1 forecast and the final estimate has been somewhat larger.

The difference in number of large bolls forecasted by Model 1 and Model 2 as of August 1962 is typical for the two models in most years. In general, two models provide forecasts of large bolls with a maximum difference of about ten percent.

The forecasting models for September 1 is very similar to August 1 except the Rate of Fruiting approach is shifted to estimate the fraction of large bolls set rather than predicting bolls to be added. This change made primarily to take advantage of the 30 additional days of maturity. By September 1, most cotton fields have already set their maximum fruit load. The second component of yield, boll weight, is estimated on September 1 for the earlier States where as many as one fourth of their bolls are picked. Most early maturing bolls weigh considerably more than the late maturing bolls; consequently, weight for the first bolls picked requires a downward adjustment in order to forecast the final boll weight.