

# This document is discoverable and free to researchers across the globe due to the work of AgEcon Search. 

## Help ensure our sustainability. Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

Papers downloaded from AgEcon Search may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.

## Historic, archived document

Do not assume content reflects current scientific knowledge, policies, or practices.

## SAMPLING TECHNIQUES FOR MEASURING AND FORECASTING CROP YIELDS



LEConomics satiticts And
U.S. DEPARTMENT OF AGRICULTURE

ESCS
No. 09

# SAMPLING TECHNIQUES FOR <br> MEASURING AND FORECASTING CROP YIELDS 

## By

Harold F. Huddleston
Economics, Statistics, and Cooperatives Service
U.S. Departmentof Agriculture

001017

August 1978

## PREFACE

The purpose of this manual is to call attention to some of the sampling techniques for estimating crop yields. Many of the important changes that have occurred in the techniques of measuring and forecasting crop yields during the past 30 years have been introduced into practice, some of them in countries with moderate resources.

This manual assembles information on mathematical modeling concerning crop yields in a single document for domestic and foreign users of crop statistics. In providing technical assistance to countries in the collection of agricultural data, it has been clear that measuring crop yields is extremely important for decisions affecting imports and exports as well as recommendations for improved crop techniques. Frequently, techniques have been attempted by or recommended for countries which require a historical base of data that is nonexistent. Consequently, yield and production information derived under these circumstances can be quite unreliable for many years and generate little factual information about crops.

In this manual, major emphasis is placed on forecasting of currentyear yield per acre prior to harvest, since both market and crop management problems necessitate time to formulate strategies or plans. It is hoped this document will serve as a basis for training courses as well as a reference manual for countries developing or modifying agricultural data systems. However, it is necessary to emphasize that this manual is not expected to serve as a training module without an instructor or consultant experienced in crop sampling and yield modeling. Also, participants or agricultural officials are assumed to have had or will receive training in sampling and data collection, since all techniques assume inferences are to be made with respect to a specific crop and population of units.

In presenting these techniques, there are three major topics which emerge: (1) determining the yield at harvest, (2) predicting yield from plant characteristics observed during the growing season,
and (3) predicting yield from environmental factors observed during the growing season. The first chapter is devoted largely to topic (1), but this topic is also related to the discussions in sections 2.3, 2.5.2, $2.7 .3,3.4,3.5 .4,3.7 .8$, and 3.8 .2 . The second major topic is discussed and illustrated in chapters 2 and 3 , sections 2.5.3, 2.5.5, 2.6, $2.7,3.5,3.6,3.7$, and 3.9 . The third topic is covered in chapters 2 and 3 , sections $2.4,2.6$, and 3.8 .

An alternative presentation of this material by these three topics would have been logical. However, yield forecasting techniques used for large geographic areas require a means of measuring harvested yields (or final yields) and data sets that are appropriate for estimating and verifying the model parameters. For these reasons, it is believed these topics should be interwoven rather than considered separately in developing forecasting techniques. Likewise, the data collection task needs to combine or include the different concepts to insure that valid data sets are obtained in order to develop reliable models for commercial fields.

It is hoped that readers will obtain a better understanding of the importance of measuring yields accurately at maturity as a prerequisite for yield forecasting, yield projections, and historical analyses of agricultural production.
CHAPTER 1 - A REVIEW OF YIELD MEASUREMENT TECHNIQUES
1.1 Introduction ..... 1
1.2 Grower-Reported Yields ..... 3
1.3 Market- or Processor-Reported Production ..... 9
1.4 Determination of Harvested Yields by Crop Cutting ..... 14
1.4.1 Sample Selection ..... 15
1.4.2 Plot Size and Location ..... 18
CHAPTER 2 - MODELS FOR FORECASTING YIELDS BASED ON PLANT RESPONSE
2.1 Introduction ..... 21
2.2 Mathematical Models ..... 23
2.3 Grower Subjective Appraisal Systems ..... 28
2.4 Crop-Weather Relations for Predicting Yields ..... 34
2.4.1 Introduction ..... 34
2.4.2 Joint Precipitation and Temperature Effects ..... 34
2.4.3 Agrometeorological Forecasting of Crop Yields ..... 42
2.4.4 Auxiliary Environmental Variables and Yields ..... 45
2.5 Estimating Crop Yields From Plant Characteristics ..... 49
2.5.1 Introduction ..... 49
2.5.2 Preharvest Measurement of Yields ..... 51
2.5.3 Forecasting Corn Yields Based on Plant Parts ..... 55
2.5.4 Forecasts Based on Growth Models for Yield Components ..... 67
2.5.5 Forecasting Methodology for Citrus Yields ..... 74
2.6 Simulation Models Based on Plant Physiology ..... 89
2.6.1 Introduction ..... 89
2.6.2 The Model ..... 89
2.6.3 Data Input ..... 97
2.6.4 Example of Model Results for Observed Plant Data ..... 98
2.7 Forecasting Yields for Small Geographic Areas ..... 100
2.7.1 Introduction ..... 100
2.7.2 Sampling Methodology ..... 100
2.7.3 Yield Estimation Model ..... 101
2.8 Summary of Yield Modeling for Forecasting ..... 105
CHAPTER 3 - DATA COLLECTION CONCEPTS USED IN FORECASTS FOR SPECIFIC CROPS
3.1 Introduction ..... 112
3.2 Sample Design Considerations ..... 113
3.2.1 Introduction ..... 113
3.2.2 Selection of Farm Holdings and Fields ..... 114
3.3 Determining Land Area in Yield Surveys ..... 118
3.3.1 Introduction ..... 118
3.3.2 Deriving Net Area From Gross Area Planted ..... 118
3.3.3 Deriving Net Area When Planted Area Is Not Known ..... 119
3.4 Yields From Crop-Cutting Surveys ..... 120
3.5 Forecasting Corn Yields From Plant Parts ..... 123
3.5.1 Listing of Crop Fields for Area Sampling Units ..... 127
3.5.2 Selection of Sample Fields ..... 129
3.5.3 Selection of Units Within Field ..... 130
3.5.4 Concepts for Collection of Plot Data ..... 132
3.5.5 P1ant Growth Models ..... 143
3.5.6 Corn Yield Forecasts ..... 147
3.5.7 Corn Yield Forecast Based on Within-Year Growth Model ..... 148
3.6 Grower Yield-Appraisal Models ..... 151
3.6.1 Introduction ..... 151
3.6.2 Dry Bean Yield Based on Growers' Appraisals ..... 155
3.7 Forecasting Walnut Yields ..... 158
3.7.1 Introduction ..... 158
3.7.2 Block and Tree Selection ..... 158
3.7.3 Measurement of Tree Spacing ..... 159
3.7.4 Limb Selection ..... 161
3.7.5 Counting Nuts ..... 164
3.7.6 Selecting Subsamples of Nuts for Sizing and Weighing ..... 166
3.7.7 Nut Measurements ..... 167
3.7.8 Forecast Based on Tree Data and Market Pro- duction Records ..... 175
3.7.9 Forecast Based on Objective Tree Data ..... 178
3.8 Forecasting Yields from Historical Crop and Weather Data ..... 179
3.8.1 Introduction ..... 179
3.8.2 Corn Yield Forecast ..... 179
3.9 Forecasting Citrus Yields ..... 182
3.9.1 Introduction ..... 182
3.9.2 Block and Tree Selection ..... 182
3.9.3 Limb Selection ..... 183
3.9.4 Fruit Drop Surveys ..... 183
3.9.5 Size of Fruit ..... 185
3.9.6 Florida Citrus Forecast ..... 189
3.9.7 Costs of Objective Yield Surveys, 1967-68 ..... 191
3.10 Conclusions ..... 192

### 1.1 Introduction

In recent years, there has been a renewed interest in the modeling of crop yields. This is the result of the great importance of food and feed crops in meeting the needs of an increasing world population, as well as coping with inflated prices and imbalances in supply. Under these conditions, there has been considerable emphasis on forecasting yields, and knowing harvested yields for model building. An unusual amount of attention has been given to those techniques which employ secondary or environmental data that can be related to harvested yields based on previous years' data, without proper recognition of the fact that harvested yields must be measured as a prerequisite. This consideration is also important where the emphasis is on making yield projections a year in advance.

For some developing countries, no efforts are made to measure harvested areas and yields on a reliable and timely basis because of lack of resources. This circumstance may severely limit the choice of models which can be employed. In other countries, harvested yield data are subject to moderate errors at the country level, and even large errors for geographic regions within the countries. In addition, available secondary and environmental data do not relate to the same units as the yield data, which can lead to biases in the model parameters being estimated for the forecasting or projection of yields. Greater attention must be given to this modeling problem as well as the population being sampled in order to properly evaluate and reduce forecasting errors.

For a long time the accurate measurement of the production of crops was believed possible only for those crops which were completely marketed or processed off the farm. In general, this was true for only relatively few crops in those countries with highly organized and modern means of crop handling and processing. However, the development and use of sampling theory in the last 35 years have made it possible to accurately estimate production of most crops based on sample surveys of crop acreage (or hectarage) and yield per acre.

Accurate annual estimates (i.e., with known sampling errors) of crop acreage and yield per acre are dependent only on possession of sufficient financial resources and adequately trained personnel. In many countries, this goal has been achieved for major crops and production areas. Unfortunately, accurate annual food and feed production estimates have not existed for many countries when improved forecasts of yields have been sought. Where acreage and yields have been measured annually, economic planners and others have employed various techniques to project acreage, yields, and production one to five years in advance of harvest. These projections are dependent on various scenarios which seem appropriate to the analysis and to the existence of acreage and yield data measured accurately over a period of years as a basis for projections.

This manual does not propose to discuss or evaluate these techniques of projecting yields over years but rather to examine methods of measuring yields for individual crop years that are needed in developing the historical basis for yield projections.

For many crops, estimates of harvested areas and yields do not exist, and only forecasts based on opinions of a panel of agricultural officials are available. The ability to evaluate crop growth conditions prior to harvest can be useful in crop management for evaluating optimum planting date, fertilizer application rates and timing, irrigation amounts and scheduling, insect control, and choosing varieties or alternative crops. Crop yields also affect market management. Yield forecasts can affect the price and sales policies of agricultural commodities, associated storage, and handling requirements on farms as well as at national and international terminal points and the cost of transporting or shipping to markets.

The principal yield-measurement techniques in common usage for mature or ripe crops are: (1) grower-reported yields, (2) marketed or processed quantities divided by area planted or harvested, and (3) cropcutting surveys. These techniques are discussed in this chapter.

### 1.2 Grower-Reported Yields

Annual yield data are generally obtained by sampling farms or fields which are known to grow the crop(s) of interest based on land use or acreage surveys conducted during the crop season. Probability acreage surveys immediately after crop planting and up to harvest provide a basis for selecting subsamples of farms or fields for crop yield surveys. Nonprobability surveys of farmers or fields are sometimes used to obtain yield data based on the assumption that biases in reported yields will be small either because the yields do not vary greatly within an area or the nonrepresentativeness (i.e., bias) of the sampling procedure is not important. Nonprobability surveys for yields are not likely to be satisfactory unless independent yield or production data become available after the crop has been marketed to adjust the yields for biases or to verify the assumption of little variability in yields over the area. Reports by volunteer growers, participation of farms in improvement programs, and sampling of fields along roads are data-collection techniques widely used in nonprobability surveys.

Probability surveys of farms or crop fields provide the only satisfactory direct means of insuring accurate and unbiased methods of measuring crop yields. Even though a probability survey of farms growing the crop of interest is the only method of data collection which can provide a direct estimate for the agricultural population of concern, there are many factors under the heading of nonsampling errors which may result in biased estimating or reporting techniques.

Growers may not know their yields even after harvest or may not report accurately for various reasons, including: (1) fear of taxation, (2) fear of confiscation of part of their crop, (3) desire to affect price (cash-crop bias), (4) desire to impress persons with their success in growing the crop, and (5) desire to establish a high production base in event of production controls. Despite these possible limitations, growers are probably the most reliable source of data on yields after harvest if independent check data (i.e., yield or production) are available on a periodic basis for adjusting for biases.

Even without check data, farmers' reports of harvested yields based on quantities taken from their fields are fairly reliable when based on probability surveys (nonsampling errors or biases are no greater than sampling error for moderate-size samples, $100 \leq \mathrm{n} \leq 400$ ), even though counter examples have been cited based on sampling from inappropriate but convenient populations by reporters or officials usually using nonprobability sampling techniques. Surveys of local governmental officials, bank officers, and locally informed cooperators do not constitute samples of the population being estimated for and can, at best, only provide opinions on yields or production for their locality.

Growers should be asked to report on individual fields, parcels, or farms under their management. The reporting basis used depends on the number of fields per farm. If other types of crop data are desired, such as the area interplanted with other crops, the reporting basis will depend on the detail with which the farmer is familiar for the particular crop.

The content of the reported data from these surveys will vary depending on whether acres harvested, yield per harvested acre, or total production for harvested acreage is sought. The yield-per-acre data may be reported directly or may be derived from harvested acres and production. For most crops, yields reported by growers are based on a volume measurement in terms of an available commercial-size container rather than on weight, because scales are seldom available. In addition, the use of different kinds or sizes of containers leads to some inaccuracy in the tabulated yields as well as some fuzziness in the definition of the yield. The users of yield data frequently change the volume units to corresponding weights based on generally accepted trade or industry conversion factors.

For some crops which are marketed at elevators, or processed by gins or oil crushers, the yield (or production) can be obtained on a weight basis from the growers after they obtain a delivery ticket or crop payment based on weight. Yield surveys which seek crop information derived from delivery tickets or payment records generally are quite accurate. However, these yields tend to be in terms of marketed
volumes or weights, or total monetary value after allowances for grade, moisture, or foreign material rather than quantities harvested in the field by the grower.

Frequently, the concept of the yield may differ because of the harvesting equipment or method used and/or the marketing practice for the crop. Consequently, it may be necessary to obtain information on various possible utilizations the grower may have for the crop, such as: used for seed, destroyed to comply with marketing quotas, fed to animals, stored in field or on plant, used as household food, or sold to other farmers or dealers, if total crop yield (or production) is desired. Crops for which weight information could be obtained in major producing countries are: wheat, soybeans, oil crops, cotton, rice, tobacco, sugar, coffee, and a few fruit and vegetable crops.

The differences in yields reported by a volunteer sample of farmers and by a probability sample of farmers can be moderately large. For several years, large samples of both types of surveys were available in the U.S. for corn, which is a crop with poor independent market check data. The nonprobability sample yields were 6 percent below the probability sample yields on the average, but the results varied by regions. In the Midwest, the difference was about 5 percent, but in the Southeastern States the differences were close to 15 percent. The probability sample of farmer-reported yields averaged 3 to 4 percent below crop-cutting yields (after adjustment for harvesting losses) for the same farmer fields, but there were important regional variations. In the Midwest, the farmer-reported yields were about 4 percent below crop-cutting yields, and in the Southeastern States the farmer-reported yields averaged about 4 percent above crop-cutting yields.

In other situations, the yield cannot be measured accurately after maturity, because of planting or harvesting practices. In some countries or primitive agricultural societies, the area of land planted to a crop may not be known by the farmer. The farmer can merely identify the field or area cleared for planting of crops. In some cases, the amount of crop harvested will depend on the needs of the household or
farm animals. Consequently, the crop may be harvested only as needed with the unharvested portion being stored on the plant in the field. Under these circumstances, the grower may not be able to report accurately the total yield per area.

Grower-reported yields are used largely for market management purposes, since the data do not provide information on crop characteristics and become available too late for current-year crop management decisions. Table 1 summarizes some of various yield measurement concepts which may be used in reports from growers.

Exhibits 1 and 2 are examples of questionnaires sent by mail or left with growers to secure data on harvested quantities of a crop, along with the purpose of harvest and crop utilization. A few additional crop-related questions may be desirable to insure that the statistical quantity to be estimated is reported consistently or, if necessary, can be derived from several questions.

Table l--Grower Concepts Involved in Yield Measurements (Column concepts are not necessarily related horizontally)

| Area |
| :--- |
| Planted |
| Harvested |
| Contracted |
| Gov't. allot- <br> ment <br> Number of trees <br> Interplanted <br> area <br> Equivalent solo <br> planted (or <br> harvested) |


| Production <br> or field- <br> reporting <br> units |
| :---: |
| Standard volume <br> container <br> Weight basis <br> Number of bunches <br> Number of heads <br> (or fruit) <br> Sized fruit or <br> head |


| Harvested <br> form of <br> crop |
| :---: |
| Husked heads (or <br> bean, berry) <br> Unhusked heads <br> (or bean, berry) |
| Threshed grain <br> Whole leaves <br> Brushed roots or <br> tubers <br> Whole fruit <br> Stalks or whole <br> plant |


| Use |
| :---: |
| Hauled from <br> field to farm <br> Delivered to <br> market <br> Sales <br> Consumed as feed <br> or food <br> Destroyed or <br> "dumped" <br> Processed <br> Seed |

## ACREAGE AND PRODUCTION OF CROPS - 197

INSTRUCTIONS: Report for the land you are operating, including land rented from others. In reporting acres harvested and total production, include acres that still remain to be harvested and probable production.

## REPORT FOR CROPS GROWN IN 197

Give the information as accurately and completely as possible. Where acreages and production are not definitely known, make careful estimates.

## FIELD CROPS

1. Corn planted for all purposes
2. Corn harvested and to be harvested for grain
3. Corn cut for silage.
4. Corn cut for fodder, pastured and hogged down (without husking)
5. Corn abandoned (will not be harvested or pastured)................
6. Soybeans planted for all purposes
7. Soybeans harvested and to be harvested for beans
8. Sovbeans used for hay, silage, pasture only, plowed under or abandoned.
9. Wheat planted for all purposes last fall and this spring
10. Wheat harvested for grain
11. Wheat used for hay, silage, pasture only, plowed under or abandoned.
12. Barley planted for all purposes last fall and this spring
13. Barley harvested for grain
14. Barley used for hay, silage, pasture only, plowed under or abandoned.


## SOYBEAN INQUIRY

| REPORT FOR THE FARM YOU OPERATE | Answer <br> here <br> $\downarrow$ |
| :---: | :---: | :---: |

## 1973 CROP PRODUCTION AND PURCHASES

| 1. Soybeans HARVESTED for beans <br> on this farm, last year's crop......... Bushels |  |
| :---: | :---: |
| 2. Soybeans BOUGHT FOR SEED <br> to plant the 1974 crop................... Bushels |  |
| 3. TOTAL harvested and bought (sum of items 2 and 3)................... Bushels |  |
| USE AND SALE OF ABOVE SOYBEANS |  |
| 4. $\begin{aligned} & \text { Soybeans SOLD AND TO BE SOLD between } \\ & \text { Sept. 1, } 1973 \text { and Sept. 1, } 1974 \ldots . . . . \text { Bushels }\end{aligned}$ |  |
| 5. Soybeans USED FOR SEED on this farm for planting the 1974 crop............... Bushels |  |
| 6. Soybeans FED AND TO BE FED <br> to livestock on this farm (beans <br> fed whole or ground) between <br> Sept. 1, 1973 and Sept. 1, 1974........ Bushels |  |
| 7. O1d-crop soybeans expected to be on hand Sept. l this year................. Bushels |  |
| 8. TOTAL (sum of items $4,5,6$, and 7 should equal item 8)......................... Bushels |  |
| 9. 1973-CROP SOYBEANS SOLD in each of the following months: <br> Sold in 1973 <br> September. <br> Bushels |  |

### 1.3 Market- or Processor-Reported Production

For crops which are marketed or processed through commercial channels, government or trade sources frequently report quantities handled monthly by elevators, gins, mills, oil processors, or crushers. Accurate data on the volume or weight delivered are available when the crop marketing is complete. While this is too late for either current-year market or crop management, the information is very useful in verifying the crop production, which serves as a basis for adjusting or revising crop acreage and/or yield estimates that are used in future yield forecasts and planning decisions.

The crop area harvested, in practically all cases, is estimated from grower-reported data, or in some instances from land contracted for specific crops by processing or marketing firms. In some cases, the acreage is based on production guidelines established by a governmental agency. Data on planted crop areas based on politically prescribed or suggested guidelines are usually unreliable. The yield is obtained preferably by dividing the market production by the growerreported harvested acreage. The existence of these marketing data generally results in development of reliable yield data for historical crop series.

However, the yield concept is frequently altered, when these data are used, to refer to reported marketed quantities rather than to amounts harvested by the farmer for all purposes. Such yield series may be useful for determining marketed quantities, but may fall considerably short of measuring total quantities harvested. For crops consumed as food or feed without commercial processing this difference can be important. For crops where utilization information from farmers can be obtained, it is possible to determine accurately the total yield harvested by combining the two sources of information.

The following table 2 presents some examples of the reporting of quantities marketed or processed through government or trade sources in various countries.

Table 2--Some Crops Marketed or Processed (Column concepts are not necessarily related horizontally)
\(\left.$$
\begin{array}{|l|l|}\hline \text { Crop } \\
\text { Cotton } \\
\text { Soybeans } \\
\text { Rice } \\
\text { Coffee } \\
\text { Oranges } \\
\text { Grapes } \\
\text { Therries }\end{array}
$$\left|$$
\begin{array}{l}\text { Data } \\
\text { Source }\end{array}
$$\right| \begin{array}{l}Gins <br>
Crushers <br>
Mills <br>
Exporters <br>
Gov't. inspec- <br>
tion and <br>
grading <br>
Wineries <br>
Private pro- <br>
cessors to <br>
trade assoc. <br>
Private auc- <br>
tions <br>

Flour mills\end{array}\right]\)| Sugar facto- |
| :--- |
| ries |


| Units <br> reported |
| :---: |
| No. bales, gross <br> or net weight |
| Oil, cake or |
| meal |
| Milled |
| Roasted |
| Juice, fresh |
| fruit |
| Tons crushed |
| for wine |
| Containers |
| packed |
| 'Hands" |
| Milled |
| Tons of brushed |
| roots, or |
| sugar |

Frequency

Monthly

For season

Seasonal

Semimonthly

Exhibits 3 and 4 are reports used by processors in reporting harvested quantities of cotton and soybeans to a governmental agency. Exhibit 5 is a summary from weekly reports designed for state inspectors and graders of citrus. The individual weekly totals are accumulated to give a running total for the season to date. This type of crop data is extremely valuable in checking the overall validity of yield and production models.
. a. Total number of bales of cotton ginned from this crop prior to October 1
b. Total weight of the bales reported in item la above

$\longrightarrow |$|  |  |
| :--- | :--- |
|  | Total bales <br> Total weight |
|  | lbs. |

c. The weight reported above is: $\square$ NET (Excludes bagging and ties)
$\square$ GROSS (Includes bagging and ties)
Average weight of bagging and ties used per bale
Enter the AVERAGE weight of bagging and ties used per bale here $\rightarrow$
lbs.
If you are unable to report the total weight of bales ginned in item 1 above, please read the following instructions and enter the necessary information in the columns below. Be sure to check above the column headings whether the weights reported for each bale are NET or GROSS.
a. If you ginned less than 1,000 bales:
List each bale bearing tag numbers ending with 5 in column (a) and enter bale weight in column (b).
b. If you ginned between 1,000 and 5,000 bales: List each bale bearing tag numbers ending with 15,35 , 55, 75, and 95, in column (a), and enter the bale weight in column (b).
c. If you ginned thore than 5,000 bales:
List each bale bearing tag numbers ending with 15 or 65 in column (a) and enter the bale weight in column (b).
he bale weights listed below are:


NET
GROSS

| Line <br> No. | Bale <br> number <br> (a) | Bale <br> weight <br> (Pounds) <br> (b) | Bale <br> number <br> (a) | Bale <br> weight <br> (Pounds) <br> (b) | Bale <br> number <br> (a) | Bale <br> weight <br> (Pounds) <br> (b) | Bale <br> number <br> (a) | Bale <br> weight <br> (Pounds) <br> (b) | Bale <br> number <br> (a) | Bale <br> weight <br> (Pounds) <br> (b) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 |  |  |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |  |

EXHIBIT 4: SOYBEANS MONTHLY REPORT OF PRIMARY PROCESSORS

OILSEEDS, BEANS, AND NUTS

PRELIMINARY WEEKLY PROGRESSIVE REPORT OF FRUIT RECEIVED AT PROCESSING PLANTS

| Week Ending | March 27 |  |  |  |  |  |  |  | units of | -3/5 bu.) |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Grapefruit | $\begin{gathered} \text { Early-Mid } \\ \text { Oranges } \\ \hline \end{gathered}$ | Late Oranges | Navel <br> Oranges | Tangerines | Temples | Tangelos | K Early | Honey Tangerines | Total |
|  | 435,359 | 188,110 | 616,710 | 140 | - | 13,296 | 3,220 | - | 9,193 | 1,266,028 |
|  | 461,125 | 151,990 | 761,736 | - | - | 14,558 | 141 | - | 8,859 | 1,398,409 |
|  | 345,970 | 75,391 | 651,119 | - | - | 5,275 | 279 | - | 5,280 | 1,083,374 |
|  | 337,623 | 110,431 | 601,549 | - | - | 6,332 | 273 | - | 5,035 | 1,061,243 |
|  | 489,835 | 106,828 | 732,804 | - | - | 6,281 | - | - | 4,134 | 1,339,882 |
|  | 470,171 | 154,744 | 1,193,716 | - | - | 16,625 | 59 | - | 13,078 | 1,848, 393 |
| Total | 2,540,083 | 787,494 | 4,557,634 | 140 | - | 62,367 | 3,972 | - | 45,579 | 7,997,329 |
| Previous total | 18,581,654 | 106,436,990 | 6,998,148 | 415, 120 | 983,328 | 2,310,907 | 2,496,868 | 137,862 | 847, 012 | 139,207,889 |
| GRAND TOTAL | 21,121,737 | 107,224,484 | 11,555,782 | 415,260 | 983,328 | 2,373,274 | 2,500,840 | 137,862 | 892,591 | 147,205,218 |
| Corresponding total |  |  |  |  |  |  |  |  |  |  |
| last season | 21,346,681 | 91,761,636 | 3,500,654 | 360,538 | 1,036,844 | 2,960,135 | 3,252,563 | 127,098 | 672,322 | 125,018,471 |
| Week Ending | March 27 | PRELIMINARY | WEEKLY PROG | ESSIVE REPO | OF FRUIT | EIVED AT P | CKING HOUS |  | units of | (5 bu.) |
|  | Grapefruit | $\begin{gathered} \text { Early-Mid } \\ \text { Oranges } \\ \hline \end{gathered}$ | Late Oranges | Nave1 <br> Oranges | Tangerines | Temples | Tangelos | K Early | Honey Tangerines | Total |
|  | 121,175 | 575 | 63,287 | - | - | 654 | 21 | - | 5,992 | 191,704 |
|  | 146,129 | 937 | 53,038 | - | - | 18 | - | - | 2,334 | 202,456 |
|  | 209,078 | 370 | 45,705 | - | - | 1,159 | - | - | 1,515 | 257,827 |
|  | 192,599 | - | 39,106 | - | - | - | - | - | 2,521 | 234,226 |
|  | 190,282 | - | 54,831 | - | - | 777 | - | - | 7,345 | 253,235 |
|  | 165,783 | - | 44,113 | - | - | - | - | - | 1,198 | 211,094 |
| Total | 1,025,046 | 1,882 | 300,080 | - | - | 2,608 | 21 | - | 20,905 | 1,350,542 |
| Previous total | 23,276,523 | 7,997,408 | 1,576,450 | 2,894,293 | 4,443,968 | 1,967,601 | 4,235,430 | 842,445 | 502,990 | 47,737,108 |
| GRAND TOTAL | 24,301,569 | 7,999,290 | 1,876,530 | 2,894,293 | 4,443,968 | 1,970,209 | 4,235,451 | 842,445 | 523,895 | 49,087,650 |
| Corresponding total |  |  |  |  |  |  |  |  |  |  |
| last season | 27,829,266 | 9,308,093 | 2,721,786 | 2,247,688 | 4,526,250 | . $4,221,881$ | 4,158,512 | 438,453 | 2,076,323 | 57,528,252 |

### 1.4 Determination of Harvested Yields by Crop Cutting

The techniques of crop cutting vary greatly in different parts of the world. The techniques used are dependent upon a number of factors. These factors include the administrative setup, type and size of field staff, farmer cooperation, crop practices, and harvest conditions. Consequently, it is not possible (nor desirable) to lay down a single uniform approach for crop-cutting surveys.

However, all crop-cutting surveys do have one element in common. One or more plots (or groups of plants) are chosen as samples from commercial fields. The plots comprise only a small fraction of the total area in the field. Therefore, it is not possible to estimate the yield in an individual field with acceptable statistical precision unless many plots are selected. The yields calculated from one or two plots in a field are not highly correlated with the yield for the entire field because the mean of all plots in a field is statistically independent of the individual plots. Where it is desired to estimate or compare yields for individual fields, the number of plots needs to be large. For instance, small field plots consisting of less than 200 square feet have a within-field coefficient of variation of approximately 20-25 percent for yield per acre. Therefore, an estimate of yield for an individual field would require around $20-25$ units per field to achieve a standard error of the mean equivalent to a coefficient of variation of 5 percent.

Costs and sampling variability considerations always indicate a survey design for crop cutting with (1) as many fields on as many farms as possible and (2) only one or two plots per field, if the survey objective is to obtain yield statistics for the country or a major region of the country.

In general, measuring yields annually by crop cutting for small political or many administrative districts within a country is too costly. However, attempts have been made to employ auxiliary data or double sampling involving a large number of fields as a basis for adjusting a smaller crop-cutting survey to obtain current yields for
small geographic regions. The auxiliary data needs to be acquired quite cheaply and to be highly correlated with the yield from the crop-cutting plots. Typically, eye estimates of yield per acre are made for many fields (or trees) and a random subsample of fields for crop cutting is taken. Under favorable costs and moderate-to-high correlation between the two data sources, annual crop-cutting surveys can provide yield statistics for small areas. However, the number of instances where these techniques have been successfully employed for small-area yield statistics is very small, because costs and correlations of the two data sources have not been favorable.

Yield measurement by crop cutting has been largely confined to major food or export crops in India, Europe, and the United States. In the United States, industry marketing programs for specialty fruit and nut crops have employed crop-cutting techniques for yield information.

### 1.4.1 Sample Selection

The measurement of yields by crop cutting involves the selection of a representative (probability) sample of fields or blocks of trees. The process of plot selection within the field also requires very careful location, measurement of plot size, delineation of the plants associated with the plot as well as careful handling of the plant parts that are used to derive the yield per area. The following table illustrates the major steps required in the selection process for a field crop.

Table 3(a)--Sample Field and Plot Selection

| Selection Step | Information Needed |
| :--- | :--- |
| 1. Random selection of farms | List of farms having crop for which <br> yield is to be estimated |
| 2. Random selection of <br> fields | Number of fields or area of each to <br> determine probabilities of selection <br> for individual fields |
| 3. Subdivision of field |  |
| into plots |  |$\quad$| Dimensions of field or number of |
| :--- |
| crop rows in field, used to deter- |
| mine plots of a given size and shape |

A corresponding table for a tree crop would be as follows:

Table 3(b)--Sample Block and Tree-Part Selection for Data Collection

| Selection Step | Information Needed |
| :--- | :--- |
| 1. Random selection of farms | List of farms or commercial plant- <br> ings with tree crop |
| 2. Random selection of <br> blocks of trees | Number of trees, age, variety for <br> all blocks for deriving probabili- <br> ties for selecting individual blocks |
| 3. Random selection of trees | Rows of trees and trees per row are <br> used to determine selection proba- <br> bilities |
| 4. A random selection of a |  |
| small portion of tree is |  |
| to be made since complete |  |
| harvesting is costly |  |$\quad$| The main trunk and primary-limb |
| :--- |
| sizes and number, as basis for |
| selection probabilities |

### 1.4.2 Plot Size and Location

Variations in plot size are primarily dependent upon costs and the magnitude of variance components between and within fields. In some countries the ability of the workers to lay out and harvest plant materials in plots according to specifications is an additional factor which is considered in choosing the plot size. The smallest plot sizes for field crops are used in the U.S. where an area as small as 0.0001 acre (approximately) has been used. Much larger plot sizes are found in India where plot sizes as large as 0.1 acre have been used.

Table 4 gives some examples of plot sizes and shapes which have been used throughout the world. Table 5 lists some of the crops in various countries where crop-cutting surveys have been employed. Neither table is complete, but they do suggest the wide application of this technique.

Table 4--Size and Shape of Plots Used for Field Crops

| Plot size | Shape |
| :---: | :---: |
| 2, 4, 5, 8 ft diameter | Circular |
| 3 meters diameter | Circular |
| 5 ft 3 in . (1/2000 acre) | Circular |
| ```33 ft x 16\frac{1}{2}}\textrm{ft}(1/80 acre (50 x 25 (links))``` | Rectangular |
| $16^{\frac{1}{2}} \mathrm{ft} \times 16 \frac{1}{2} \mathrm{ft}$ ( $1 / 160 \mathrm{acre}$ ) | Rectangular |
| $33 \times 16$ (1/80 acre) | Rectangular |
| 5 x 10 meters | Rectangular |
| 1.5 sq meters | Rectangular |
| . 3 sq meter | Rectangular |
| 15 ft x 2 rows | Rectangular |
| 7 x 7 yd ( $1 / 100 \mathrm{acre}$ ) | Square |
| 6 ft 7 in. (1/1000 acre) | Square |
| 33 ft | Triangular (Equilateral) |
| 16 ft 6 in . | Triangular (Equilateral) |
| 8 ft 3 in. | Triangular (Equilateral) |
| $24 \text { in. x } 26.136 \text { in. } \underset{\text { acre })}{(1 / 10,000}$ | U-shaped frame |
| 21.6 in. x 3 rows | Length-of-row frame |
| 1 sq meter | Square frame with closing bar |
| Entire field | In terraced areas where very small parcels are seeded |

Table 5--Crop-Cutting Surveys by Countries

| Crop | Country |
| :---: | :---: |
| Wheat | India, U.S., W. Germany |
| Rice | India |
| Cotton | India, U.S. |
| Sugarcane | India |
| Coconuts | India |
| Almonds | U.S. |
| Walnuts | U.S. |
| Citrus | U.S. |
| Peaches | U.S. |
| Pears | U.S. |
| Lemons | U.S. |
| Grapes | U.S. |
| Cherries | U.S. |
| Cranberries | U.S. |
| Soybeans | U.S. |
| Tobacco | U.S. |
| Corn | U.S., Basutoland |
| Sorghum | U.S., Basutoland |
| Peas | Basutoland |
| Barley | Basutoland |
| Oats | Basutoland |
| Beans | Basutoland |
| Rye | W. Germany |
| Potatoes | W. Germany, U.S. |

### 2.1 Introduction

This chapter presents a variety of techniques which have been used with varying degrees of success in forecasting yields. Some of the models have been discarded since they were first introduced because: (1) the cost of data acquisition was too high, (2) the need (or timing) for the forecast changed, or (3) the model performed poorly and a new technique was adopted.

However, this chapter is not intended to be complete catalog of techniques, but rather to indicate the diversity of approaches which have been found "useful" in yield estimation and to focus on the data requirements for the different models. Many of the techniques were devised to make use of available data rather than to provide a deliberate effort to systematically model crop yields; this serves as an important distinction. Recently, efforts have been made to identify the concepts needed to model the crop yield and gather the required data. The data collection methods or sampling schemes have a profound influence on the validity of a forecast just as the choice of model has.

It should be understood that the sampling concepts are important even though the concepts are only briefly discussed here. It is assumed that proper training has been or will be obtained in sampling so that valid inferences can be made to the desired population of units. It is hoped that a broad exposure to yield determination techniques and their data requirements will assist agricultural program managers in choosing a suitable yield estimation method, or, at least, in narrowing the alternatives to be considered. The usefulness of the various techniques will also be dependent on other factors, such as: the crop, length of growing season, environment, and date the information on yields is needed.

The models described in this chapter are based on data available from the time the crop is planted. However, the purpose is to model the yield at maturity and not the plant development, during the plant
life unless this is necessary to model the yield at maturity. Several different models are discussed in sufficient detail so the reader will be able to grasp the data collection and modeling concepts.

In some cases, the examples cited may provide a basis for starting new work on the same or similar crops. An acreage inventory survey is assumed to have been completed after planting so sample farms or fields may be selected for observation. Likewise, the acreage sample is expected to provide validation of harvested yields or yield components as well as permit the derivation of production based on yield and acreage. Most of the models presented were developed on a farm, field, plot, or plant basis. For some yield models, especially those involving a historical series of data, averages derived from several discrete locations are attributed to large geographic areas rather than individual fields or plots.

Grower observations on reporting units are generally in terms of yield per harvested area for either the farm or individual fields. In some instances, public-minded growers may be willing to cooperate by observing plots or plants for governmental or industrial organizations. Models using plant counts and yield-component-measurement techniques which are carried out by volunteer or paid cooperators usually are on a plot or plant basis. The models based on plot or individual plant data are expressed in terms of a standard unit for conversion to a per acre or per hectare basis by the sponsoring agency for publication.

The choice of model is a basic forecasting step. In general, the techniques commonly used do not consider the data as a time series from which forecasts are made, but as a series of independent data points where a new observation(s) is generated each year; neither is it likely that purely mathematical rules can be found which will be adequate to describe the phenomena.

The models rarely describe the real world, owing to random or natural variation shown by most data from commercial crop plantings and plots. Thus, the forecasting methods that have been developed are either statistical in nature or require statistical estimates of key parameters for successful implementation. Some of the models are deterministic, but these generally require statistical estimates of some of the model parameters for implementation in large areas. In addition, the models are generally incomplete because some important factor has been omitted due to either our incomplete understanding of the phenomena or the cost of including it in the model. Often we use the models, not in the belief that they describe exactly the underlying structure of the situation, but in the faith that, at least for the recent past and the near future, they give a reasonable description of the underlying situation.

We consider several situations. In the first situation, the structure is regarded as highly stable over years and the chosen model represents the underlying structure of the data. The model in this case will be referred to as a between-year or global model.

In the second situation, the structure is believed to be stable in the short run but not necessarily in the long run. Slow changes in the model structure or parameter values may occur which will not affect the data adversely enough to invalidate the forecast for only one year ahead (or a short period). In this case, the model will be referred to as a transitory or local model.

In a third situation, the structure may be unstable over the short run. The model in this situation is referred to as a within-year or individual crop-year model.

Experience suggests that using transitory models often leads to better forecasts, because we have many more replications in time for evaluation of the method, while the between-year or global model may be viewed as a single observation of the process or phenomenon. The withinyear or individual crop year phenomenon is recognized, but too often there are little data available to model the situation. Frequently, there is no difference in the mathematical or statistical formulation of these models, but the differences lie in the way in which we make use of or interpret the parameters represented in the models.

Several basic statistical models are described before examining techniques which have been developed and put in use. The simplest statistical model is the constant-mean model:

$$
x_{t}=\mu+\varepsilon_{t}
$$

where

$$
\begin{aligned}
\mathrm{x}_{\mathrm{t}}= & \text { past data for the } t^{\text {th }} \text { period (usually years) for } \\
& \text { a yield characteristic } \mathrm{x}_{1}, \mathrm{x}_{2}, \ldots, \mathrm{x}_{\mathrm{t}} \\
\varepsilon_{t}= & \text { the normal random error for time } t \\
\mu= & \text { a constant mean }
\end{aligned}
$$

$$
\text { and we wish to forecast the characteristic for time } t+\mathrm{k} \text {. }
$$

The forecast for time $t+k$ is given by the sample mean

$$
\bar{x}_{t+k}=\frac{1}{t}\left(x_{1}+x_{2}+\ldots+x_{t}\right)
$$

The model might be appropriate for weight of grain per head, or weight of grain per kernel where $x_{t}$ is for a series of years; that is, an overyears model might be appropriate for certain characteristics of the plant even though it might not be appropriate for yield per area.

Another formula for the constant mean which might be used when a transitory model is appropriate is that which assigns weights to the data points as follows (for computation of coefficients see page 102):

$$
\bar{x}_{t+k}=(1-a)\left(x_{t}+a x_{t-1}+a^{2} x_{t-2}+\ldots\right)
$$

where "a" is a number between 0 and 1. Typical values of "a" for yield work would be between $1 / 3$ and $2 / 3$. This model has the effect of always giving the greatest weight (or importance) to the last observed data point. The above formula can be rewritten so it is more convenient to use for calculation purposes, as

$$
\bar{x}_{t+k}=(1-a) x_{t}+a \bar{x}_{t-1, k}
$$

This is a type of moving average, but gives variable weights to the years, in contrast to the simple moving average, which gives an equal weight to each year. Again, this model might be appropriate for certain plant characteristics or yield per area.

Where neither a between-year nor transitory model is appropriate, a within-year or logistic-type growth model may represent the data approximately:

$$
x_{t}=\frac{\alpha}{1+\beta \rho}+\varepsilon_{t}
$$

where

$$
\begin{aligned}
& x_{t}=\text { given data value for time } t \text { in a sequence of times } \\
& \quad \text { during crop season for a yield characteristic } \\
& \alpha, \beta, \rho,=\text { constants or model parameters } \\
& \varepsilon_{t}=\text { random error for time } t \\
& \text { and we wish to forecast the characteristic for time } t+k .
\end{aligned}
$$

Some of the other models commonly encountered are as follows:
Linear trend: $\quad x_{t}=\alpha+\beta t$ for all $t$ (i.e., the time variable).
Linear regression: $x_{t}=\alpha+\beta z_{t}$ where $z_{t}$ is another variable.

$$
x_{t}=\alpha+\beta x_{t-1}
$$

where $x_{t-1}$ is the previous value of $x$.
Exponential growth: $x_{t}=\alpha \varepsilon^{\beta t}$ for all $t$.
First-order moving
average: $\quad x_{t}=\varepsilon_{t}-\theta \varepsilon_{t-1}$

The linear-trend model to be employed can be either global or local. The ideas are similar to those in the constant-mean model in that the least squares line can be altered by assigning different weights (or importance) to the errors to be minimized in estimating the model parameters. This has the effect of forcing the trend line to fit the most recent data points more closely. Similar ideas, likewise, carry over to the linear-regression model; however, the regression model also requires attention to the selection of the other variable. In most of the models the forecast time is $t+1$, except for the growth model, where $\mathrm{t}+\mathrm{k}$ is "quite large" compared with t .

During the past few years, a major emphasis has been given to developing yield models in which the parameters are derived from the current year for use prior to harvest. That is, a deliberate effort has been made to make the techniques less dependent on a historical series of data as a prerequisite to being able to forecast the yield. Models that achieve this independence are referred to as "within-year models" and are considered to be more desirable than between-year models if each year is different from the preceding years or there are technological changes taking place which cannot be evaluated. The fact that these models do not require a historical series of similar information before yield forecasts can be started is considered quite important when starting work on a new crop or developing a system for a country without a crop-forecasting system.

However, the models which do not depend on historical series of yields require greater understanding of the relations of plant responses or growers' knowledge of harvested yields. This type of model has been
considered for yield forecasts based on both grower subjective yield forecasts or appraisals as well as objective yield methods. It is helpful to start with a look at grower yield appraisals (or probable yields), which are used for many crops.

The fact that relatively few crops have been included in yield forecasting, based on plant characteristic or crop-cutting programs for countries with official published series, suggests that this approach should be examined carefully. In addition, opportunities for use of grower appraisals exist in technical assistance work when starting current statistical programs in crop-yield and crop-production forecasting.

The remainder of this chapter is devoted to a discussion of various techniques which have been tried. In general, no attempt has been made to evaluate each method or compare it with all competing models, since the necessary information for doing this was not available. However, it is hoped that by the end of the manual the reader will recognize some of the differences in the model assumptions, data needs, and the ability to validate the forecasts and model parameters as factors to take into consideration when comparing forecast models.

### 2.3 Grower Subjective Appraisal Systems

A common approach used by governmental agencies and private forecasters is the charting or deriving of relations between grower forecasts of probable yields and harvested yields obtained at the end of the season. This approach is based on the relations over years, being the same for a period of $5-10$ years, but is frequently put in use after yields have been collected for only 3-5 years. In most cases, yield charts or relationships are based on voluntary reports from growers or cooperative agents who report by mail, telephone, radio, or messenger. Consequently, the reported probable yields frequently may not be representative of the population and/or the reporters may not be able to forecast the crop accurately for their village, district, region, or some area with vaguely defined boundaries. In either case, the probable yields require adjustment or correction for various kinds of unknown biases. Frequently, there appear to be different relations indicated for different periods of years. The dashed lines in Figure 1 indicate approximately the nature of two different regressions, and the solid line the least squares regression line over both periods. This chart illustrates some of the common problems associated with between-year or global regression lines. There may be a strong trend and neither the representativeness of the sample nor the ability of the growers to forecast their yields is measured or known. The same information is frequently analyzed by employing a time trend chart and plotting the residuals or deviations from the forecasted yields against time.

Table 6 indicates the correlation and nature of the linear relation between growers' forecasts and their reported harvested yields for several crops. The relations found for cotton and soybeans in both years in adjoining States are similar, but the relations for corn are different in each of the years in adjoining States. In general, the ranges in the average yields over years based on probability surveys of growers and crop-cutting surveys agree closely, but the levels of the growers' average yield are several percent lower.


Table 6--Correlation Coefficient and Regression of Farm Operators' Reported Yield (Y) after Harvest on Farm Operators' Projection of Yield (X) at the Beginning of Fruit Setting (for probability samples)
(a)

1972

| State \& Crop | n | r | Linear Regression Model |
| :--- | :---: | :---: | :---: |
| Arkansas/Cotton | 128 | .330 | $\mathrm{Y}=.410+.578 \mathrm{X}$ |
| Mississippi/Cotton | 151 | .468 | $\mathrm{Y}=.481+.491 \mathrm{X}$ |
| Illinois/Corn | 56 | .627 | $\mathrm{Y}=36.11+.724 \mathrm{X}$ |
| Iowa/Corn | 35 | .411 | $\mathrm{Y}=68.93+.482 \mathrm{X}$ |
| Illinois/Soybeans | 71 | .621 | $\mathrm{Y}=14.95+.659 \mathrm{X}$ |
| Iowa/Soybeans | 9 | .384 | $\mathrm{Y}=13.66+.507 \mathrm{X}$ |

(b)

1973

| State \& Crop | n | r | Linear Regression Model |
| :--- | :---: | :---: | :---: |
| Illinois/Corn | 38 | .174 | $\mathrm{Y}=86.24+.220 \mathrm{X}$ |
| Iowa/Corn | 49 | .517 | $\mathrm{Y}=14.18+.796 \mathrm{X}$ |
| Illinois/Soybeans | 68 | .446 | $\mathrm{Y}=14.58+.535 \mathrm{X}$ |
| Iowa/Soybeans | 70 | .640 | $\mathrm{Y}=12.38+.666 \mathrm{X}$ |

Sometimes a different approach is needed to overcome shortcomings due to trend, changing relations over time, or even the influence of previous crops on the current year's appraisal. An approach will be discussed which provides at least partial answers to some of these problems. The method is referred to as the "grower-graded yield appraisal." It seeks to determine the following: (1) What does the grower expect the yield of a specific planting of a crop to be?
(2) How does the grower rate (or evaluate) the expected yield of this planting of the crop according to five descriptive categories? The acreages (or areas) planted are then summarized by the five categories and the average or expected yield (or expected production) is derived by weighting the yields with the acreages or percent of acreages reported by categories.

The descriptive ratings provided by the growers are assumed to be distributed normally, as in the grading system commonly used by teachers when a large number of students are to be graded. Thus, the name "grower-graded yield appraisal" is given to the method, since the grower, in effect, "grades" his own yield appraisal. This grading scheme and its relation to the normal distribution is illustrated by Chart 1 on page 33.

Some experience with this approach in Central America has indicated that the growers do grade their yields in approximately this manner. That is, 40-50 percent of the acreage is reported by growers early in crop season to have an expected yield which is "average." The remaining expected yields are either one category above or below the average. These results suggest most growers report an average yield early in the crop season. The interpretation of the expected yield as prophesizing the harvested yield may be in serious error in any year that is not average or normal. Stated another way, many growers may not be skillful forecasters or do not wish to forecast a yield different from the average for purposes of reporting to public agencies. The most useful information comes from those growers who report a yield which is not average.

The procedure for reporting yield prospects to user agencies, private or public, for the coming harvest is as follows: (1) From land use surveys, the estimated acreage is summarized as the percent of acreage reported for the grade categories used; (2) The growers are asked to report their expected yield; and (3) The within-year average yield in (2) is derived from the categories by the percentage of the acreage in (1). The rationale behind this approach is that it may be desirable to provide the grower's expected yield, the descriptive appraisals, and the derived within-year average yield so that the users may review this data along with other information that they may have from other sources and years. Expected production can also be reported to the user in place of yield if this is preferable. If the within-year derived average yield differs from the grower's last year's average yield (or a five-year average), the user is aware of this difference and may wish to place a somewhat different interpretation or evaluation on crop prospects.

For application to specific crops, the normal distribution may be skewed slightly if a portion of the crop is grown on either dryland or irrigated land (this may be handled by altering the tail probabilities and $X$-scale values of the model). When a large fraction is grown on both irrigated land and dryland, a separate yield forecast should be made for the acreage of each. In the Dominican Republic, coffee and rice are expected to have crop failures less frequently, and outstanding crops more frequently, than shown in Chart 1 . This is the result of increased management inputs, established trees or areas in the case of coffee, and availability of water for rice. Consequently, the probability in the right-hand tail was increased. In contrast, corn and beans are two crops which would be expected to have their distributions skewed in an opposite manner from coffee and rice.

Chart 1: Grower-Graded-Yield-Appraisal Curve for a Large Number of Fields


Grade Scale
F
D
C
B
A

| ```Possible crop descrip- tion corres- ponding to grades``` | Crop failure | $\begin{aligned} & \text { Below } \\ & \text { average } \end{aligned}$ | Average | Above average | Much above average |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Very poor crop | Poor crop | Normal crop | $\begin{aligned} & \text { Good } \\ & \text { crop } \end{aligned}$ | Very good crop |
|  | No harvest |  |  |  | Outstanding <br> Excellent |
| Standardized (uniform) yield scale |  |  |  |  | 2.0 |
| Midpoint of interval $X_{i}$ | . 2 | . 5 | 1.0 | 1.4 | 1.8 |

The range of the yield scale is 0 to 2.0 and each of the 5 grades covers one-fifth of the $X$ axis (uniform scale).

$$
E(X)=\frac{5}{\Sigma} p_{i} X_{i}=1.00 \text { (average yield) }
$$

Alternative scale values developed for use in the Dominican Republic were based on the approximate center of the probability assigned to the interval rather than on a uniform $X$ scale. The merits of alternative scales to the uniform scale have not been fully verified but the ones proposed have given acceptable results.

| Center of <br> probability $i n$ <br> interval $\mathrm{Z}_{j}$ | .08 | .32 | 1.00 | 1.68 | 1.92 |
| :---: | :---: | :---: | :---: | :---: | :---: |

### 2.4.1 Introduction

Crop-weather relations have been studied by many investigators as a means of forecasting crop yields. This approach is based on the premise that a network of weather stations has been recording temperature and precipitation for a number of years and data on harvested yields are available for the same period. In most cases, the yields have no known measure of accuracy available, and the technique is largely heuristic.

In some instances the network of weather stations coincides with important regional population centers rather than being distributed geographically to coincide with the crop acreage. Under these circumstances, the crop-weather relations may be distorted and not well suited to forecasting of individual crops, unless the weather variables are rather uniform over broad areas so that a special network of stations providing paired observations is not needed. The utility of these techniques depends on the climate being critical at one or more phenological stages of the crop for the area or country. Many of the applications of this technique involve crops which also have marked technological trends that explain a portion of the year-to-year variations in yields, while the weather variables account for departures from expected yields. Generally, little or no phenological information on the crop is available.

### 2.4.2 Joint Precipitation and Temperature Effects

One of the problems in crop-weather research is that of measuring the joint effect of various weather factors simultaneously. For example, the effect of an inch of rain on the final yield of a crop depends to a large extent on the temperature and other weather factors associated with that rainfall during a critical stage of development.

One part of a crop-weather project in the U.S. was the attempt by Hendricks and Scholl to develop approaches to measuring the joint effects of several weather factors. The method involved the use of monthly temperature and precipitation data as an indicator of the departure of the yield of corn from the expected yield. The use of monthly averages may be unsatisfactory without a model parameter or factor which incorporates the occurrence of unusual short-duration events of the variables having a critical impact on yield. In these cases, the error term in the model will drastically understate the expected error. Modification of the model values for the weather variable must frequently be based on special controlled experiments, since these phenomena occur infrequently and their effects on yield are difficult to measure quantitatively. The parameters should provide for modification by an event multiplier such as $E=(1+\theta)^{n}$, where $|\theta| \ll 1$ (i.e., much less than 1) is the effect of a single occurrence of the event and $n$ is the number of times the event occurs in the month or period averaged. Generally, the event E is assumed to occur infrequently over years and only once or twice a period, so that $n$ is a small integer. In general, the occurrences of unfavorable events are better known, because the critical growth stages occur early in the development of the crop and the events are better reported by the press and agricultural industry.

The charts (pages 38-41) for the State of Illinois illustrate the techniques developed in 1951 by Scholl and Huddleston for an area where the climatic factors are generally not critical but technology is important. Following is a brief description of how the method was developed. The method was first used in graphic form, but later was expressed in equations.

The first step is that of computing the 10 -year moving averages (other periods could have been used) of corn yields (Chart 2) to eliminate the effects of all nonweather factors (i.e., "technology") on yields so that the net effects of weather
could be better evaluated. Obviously, one disadvantage of using 10-year averages is the necessity of projecting the trend or normal yield so that it may be used currently for forecasting.

The next step involves constructing the isograms on a chart for each month during the critical period of crop growth (June, July and August). These charts are prepared by plotting the monthly rainfall (i.e., daily precipitation accumulated for the month) data on the X axis, and temperature values (i.e., daily mean temperatures averaged) on the $Y$ axis. The departures of the final annual yields from the $10-y e a r$ moving average were inserted at these points. For example, assume a monthly temperature of 75 degrees and rainfall of 3.00 inches for one of the June months in the series; also, assume a departure of yield from the trend line of +5.0 bushels for this particular year. The line coinciding with 3 inches of rainfall on the horizontal scale of the June Weather Chart (No. 3) is followed up until it intersects the line coinciding with 75 degrees on the vertical scale. At this point the figure +5.0 is entered. This is repeated for each June in the series of years. Isograms which best represent equal departure values of yields are then drawn on the chart. Obviously, judgment or subjectivity is involved in drawing these lines. It even may be necessary to ignore partially some of the individual data points in drawing the isograms. A period of 40 years was used in the study.

In drawing these isograms it is assumed that the most radical departures in final yields are the accumulated results of weather during several months, since a crop failure has never been experienced in any major geographic area of Illinois. Therefore, the full amount of such departures should not be allowed for in any individual month. It appears that perhaps no more than half the extreme departures should be indicated by the isograms for an individual month. For example, the isograms on the July chart might indicate a range from -6 to +6 bushels; whereas, the actual departures for some individual years are considerably larger.

The same types of joint relations between rainfall, temperature, and yields were also investigated more rigorously through mathematical models, such as:

$$
\begin{equation*}
Y=a+b T+c R+d(T R) \tag{1}
\end{equation*}
$$

or $\quad Y=a+b T+c R+d(T R)+g T^{2}+h R^{2}$
where $\quad T=$ average monthly temperature
$\mathrm{R}=$ monthly rainfall
and
$\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{g}$, and h are regression parameters.
The individual month1y charts giving the estimated joint effects of temperature and rainfall, after removal of trend, are shown as Charts 3, 4, and 5 for equation (1). These charts were generated by a computer plotter.

In order to limit the effects on yields attributable to an individual month, the departures from the mean yield for each month might be divided by two or three, as was done for the graphic approach. This is equivalent to dividing the calculated slope parameters (b, c, $g$, h) for a month by 2 or 3 in the alternative form of the regression equation (1).

$$
\begin{equation*}
Y_{t}=\bar{Y}+\frac{b}{2}\left(T_{t}-\bar{T}\right)+\frac{c}{2}\left(R_{t}-\bar{R}\right)+\frac{d}{2}(T R-\overline{T R}) \tag{3}
\end{equation*}
$$

where $\bar{Y}=$ is the normal yield based on trend (or base-period average yield if no trend is present)
$\overline{\mathrm{T}}, \overline{\mathrm{R}}$, and $\overline{\mathrm{TR}}$ are the averages for the base period $T_{t}, R_{t}$, and $T R$ are the monthly values for year $t$.

An alternative way of adjusting the slope parameters for a month is to multiply by the correlation coefficient squared, $R_{i}^{2}$, divided by $\sum_{i=1}^{3} R_{i}^{2}$, where $R_{i}^{2}$ is the multiple-correlation coefficient squared for $i=1$
an individual month. However, June and July were the key months. The relation for August was the least important, since after corn tasseling in July the plant is fully developed and soil moisture is less important.
CHART 2 - ILLINOIS CORN - TE.J-YEAR SIMPLE MOVING
averages of yield per acre

YIELD PER ACRE

CHART 3 - YIELD DEPARTURE ISOGRAMS BASED ON JUNE RAINFALL AND TEMPERATURE REGRESSION EQUATION: $\mathrm{Y}^{\prime}=173.801-43.275 \mathrm{R}-2.475 \mathrm{~T}+0.6208 \mathrm{RT}$


CHART 4 - YIELD DEPARTURE ISOGRAMS BASED ON JULY RAINFALL AND TEMPERATURE REGRESSION EQUATION: $Y^{\prime}=89.939-23.66 R-1.263 T+0.3397 R T$



In the USSR great attention has been paid to the scientific investigations aimed at finding the relations between the productivity of basic crops and the agrometeorological conditions. Methods have been developed by Ulanova and other workers for the agrometeorological forecasting of crop yields and the preparation of outlook guidance for the yields of crops. The relations discovered between the cereal crop productivity and agrometeorological conditions also are used to divide the territory of a state or entire country into agrometeorological areas in estimating the extent of favorable climatic resources for the growth of a crop. Relations have been found for the basic cereal crops, spring and winter wheat, as well as for corn.

Quantitative relations have been found between the yield of winter wheat and the soil water storage in spring. It was found that the main inertial factors for the future winter-wheat yield in the black-earth zone are the water storage in the upper onemeter layer of the soil and the number of stems of winter wheat per square meter in the spring. Summer precipitation is of less importance, and the dependence of the winter-wheat yield on the summer precipitation (without taking into account the soil moisture and winter-wheat state) is low.

The temperature during the spring-summer period in the blackearth regions of the USSR is completely sufficient (i.e., not critical) for the winter wheat.

The analysis of a long series of data shows that, although winter-wheat yields in the Ukraine and the North Caucasus depend mainly on spring water storage during many years, a good forecasting relation between crop yields and spring water storage can be found by taking into account the number of stems that survived the winter.

It is known that the number of stems of the winter wheat during the period of spring-summer vegetation does not remain constant, but the number of stems in spring may be considered as an indicator of the probable number of eared stems in the future.

As a result of field observations of winter wheat carried out by hydrometeorological and agrometeorological stations, a rather close relation between the number of eared stems of waxen ripeness (mature heads) ( $Y$ ) and the number of stems in spring (X) of the different kinds of winter wheat was found.

For the winter wheat of Belotserkovskaya 198 kind (i.e., variety), the equation of the relation is:

$$
\mathrm{Y}=0.22 \mathrm{X}+199.0 \quad \mathrm{r}=0.75
$$

And for the winter wheat of Bezostaya 1 kind

$$
\mathrm{Y}=0.24 \mathrm{X}+241.2 \quad \mathrm{r}=0.79
$$

In winter wheat of Odesskaya varieties 3,12 and 16 , the quantitative relations between spring-effective soil moisture supply and the number of stems in spring are given below for highquality agrotechniques on the same fallow in black soils of steppe and forest-steppe zones of the Ukraine and the North Caucasus.

The equations are given for most probable crop yields (Y) to be expected and also for the highest ( $Y_{h}$ ) and the lowest ( $Y_{L}$ ) yields that are predicted from the soil moisture (X) in millimeters in the top meter of soil during April, May, and June.

The regression equations of winter-wheat yield on spring moisture supply in years of favorable autumn-winter conditions when the number of stems of winter wheat in spring was 1,000 to 2,000 per square meter, have the following outlook:
(a) lowest crop yields ( $Y_{L}$ ) under unfavorable weather conditions of April, May, and June:

$$
Y_{L}=0.24 X-16.0
$$

(b) highest yields $\left(Y_{h}\right)$ under the most favorable weather conditions of April, May, and June:

$$
Y_{h}=0.24 X-4.4
$$

(c) the most probable winter-wheat yields (Y) in a particular year:

$$
Y=0.24 X-10.2
$$

The coefficient of correlation of this relation is $r=0.86$. An error of the equation of regression is $S_{y}= \pm 3.4$ centner/ha.

The relation of winter-wheat yield of Odesskaya 3, 12, and 16 to spring supply of moisture in years of unfavorable autumn or winter conditions with a small number of stems in spring (400-900 per square meter) is presented by the following equations:
(a) the lowest yields ( $\mathrm{Y}_{\mathrm{L}}$ ) under unfavorable conditions of weather of April, May, and June are as follows:

$$
Y_{L}=0.2 X-15.0
$$

(b) the highest yields ( $Y_{h}$ ) under the most favorable weather conditions of April, May, and June have an outlook:

$$
Y_{h}=0.2 X-7.2
$$

(c) the most probable expected yields (Y) have the outlook:

$$
\mathrm{Y}=0.2 \mathrm{X}-11.1
$$

The coefficient of correlation is $r=0.89$. An error of the equation of regression is $S_{y}= \pm 2.9$ centner/ha. In the equations $X$ is the productive moisture supply (mm) under winter wheat in a one-meter soil layer at a mean daily air temperature of $+5^{\circ}$ in the spring, where all the equations act in the range of the values of spring moisture supply from 100 to 200 mm . The technique is based on forecasting yield from the soil water during April, May, and June after "conditioning" yield on the expected number of stems per square meter.

### 2.4.4 Auxiliary Environmental Variables and Yields

Variables such as hourly or daily temperature, rainfall, solar radiation, minutes of sunshine, dew point, and others are used to derive new parameters directly identifiable with plant growth processes. The physical and physiological variables which are commonly derived are photosynthesis, available soil water, evapora-tion-transpiration, light interception, albedo, and canopy temperature. While it would be possible to measure some of these variables directly, the cost of instrumentation and data collection for an extensive network of locations is beyond the normal budgetary means of most users of crop yield data. Consequently, most of these variables are estimated or approximated through relations with weather data normally collected by an established experiment station or meteorological network. However, these networks are generally too sparse or the location of equipment is not representative of the plant environment for a widely dispersed commercial crop. These two factors introduce errors into the "independent" or predictor variables which lead to bias in the estimated parameters in the model, as mentioned earlier.

The idea of relating crop yields to derived variables such as evapotranspiration is not new. One model is presented, but there have been many attempts during this century to employ evapotranspiration. The basic assumptions are that (1) water is the major limiting factor in most crop production situations and (2) as transpiration is decreased by water stress, photosynthesis is proportionally decreased and thus affects yield. Hence, pertinent transpiration relations should reflect relative photosynthate production (yield).

A versatile and effective ET model has been described by Kanemasu. This model has been adapted and tested for winter wheat across Kansas with some success, and applied to soybeans with better results. The yield (actual) and ET model data were available for several sites for the crop years 1974-75 and 1975-76. Selected
sites were used as calibration points, and regression analyses of various model formulations for yield prediction were evaluated. Wheat yield differences were related to the number of days in each growth stage--the greater yields occurring in lengthened seasons.

The model most physically acceptable that gave reasonable $\mathrm{R}^{2}$ values between the observed yield and predicted yields was as follows:

$$
Y=A \underset{n=1}{3}\left(\Sigma\left(T / E_{0}\right)_{n}\right)_{n}^{\lambda_{n}}
$$

where $Y=$ bushels winter wheat per acre

$$
n_{1}=\text { period from emergence to jointing }
$$

$$
\mathrm{n}_{2}=\text { period from jointing to heading }
$$

$$
n_{3}=\text { period from heading to soft dough }
$$

$$
\lambda_{\mathrm{n}}=\text { growth-stage weighting factor }
$$

$$
T=\text { actual transpiration (daily) }
$$

$$
E_{0}=\text { potential evapotranspiration from a wet soil (daily) }
$$

A = multiplier constant

The fitted model is as follows:

$$
Y=2.856\left(\Sigma\left(T / E_{0}\right)\right) 1_{1}^{172} \cdot\left(\Sigma\left(T / E_{0}\right)\right) 2_{2}^{104} \cdot\left(\Sigma\left(T / E_{0}\right)\right) 3_{3}^{.646}
$$

and $R^{2}=.54$.

However, the yields for $75-76$ appear to be at a slightly higher level than 74-75, which suggests some other factor(s) of importance has been omitted from the model. A graph of predicted values versus observed values across Kansas is given in Chart 6. It can be seen that prediction follows the range of observed values reasonably well.

CHART 6 - OBSERVED VERSUS PREDICTED WINTER-WHEAT YIELDS, USING ET MODEL


A second model based on derived weather indices and management inputs is illustrated for wheat in Turkey. Weather variables used in the yield equations required mean monthly temperatures and monthly precipitation for January, February, May, and June from the Ankara weather station. Monthly aridity indices were found according to $I=12 P /(T+10)$, where $P$ represents precipitation in millimeters and $T$ represents temperature in degrees Celsius. For example, the January 1970 temperature of $4.2^{\circ} \mathrm{C}$, with precipitation of 47.5 millimeters, gives $570 / 14.2$ or 40.1 . By the same method an index value of 49.4 is obtained for February. For May and June 1970, the indices are 6.9 and 12.0 , respectively. In combining the months, the monthly values are weighted by the ratio of their variances, which for January and February is approximately 2.5:1. For May and June, the ratio is $2: 1$. These ratios are assumed not to change from year to year. Thus, the January-February index is 45.4-$(2.5 \times 47.5+40.1) / 3.5=45.4$. Similarly, for May and June, the value is $8.6--(2 \times 6.9+12.0) / 3=8.6$. By the same method, the 1969 January-February index is 85.7 and the May-June index is 21.4. These values are now used in the estimation equations.

If an estimate is desired before June data are available, a June index value calculated from the long-term average temperature of $20.0^{\circ} \mathrm{C}$ and precipitation of 30.6 mm can be used, since they are the expected values based on the historical series. The resulting June index of 9.2 can then be used until June data are available. The complete yield model is as follows:

$$
Y=9.18+0.00098 \mathrm{~F}-0.0148 \mathrm{JF}+0.0706 \mathrm{MJ}
$$

and

$$
R^{2}=0.70 \quad S D=1.074
$$

where
$Y$ = quintals of wheat per hectare
F = fertilizer used (in 1,000 -metric-ton units)
JF = Jan. -Feb. De Martoneau aridity index for Ankara
MJ = May-June De Martoneau aridity index for Ankara
SD = standard error of estimate at the historical means of the predicting variables
$R^{2}=$ correlation coefficient.

### 2.5 Estimating Crop Yields From Plant Characteristics

### 2.5.1 Introduction

The prediction of crop yields from plant counts and measurements in lieu of farmers' reports on expected yield and crop condition has at least two major advantages: (1) by-product information that is available, cr obtainable by making minor modifications in data collection, and (2) greater objectivity in the data and yield concepts. Possible useful by-product information includes crop quality characteristics as well as trends in crop techniques (i.e., components or attributes of yield over time) and comparisons among varieties or cultural practices. With regard to objectivity, forecasts based on human judgment, such as farmers' appraisals, tend to be conservative in that they are too high in poor years and too low in good years. That is, the appraisals reflect the average of past yields to too great an extent. Also, farmers' appraisals may not include an accurate current reflection of the impact of changes in varieties or cultural practices. Although changing farm practices may alter the parameters in the models based on plant characteristics, the impact of such changes on fruit counts and size are measured currently. A forecasting system based on plant counts and measurements is generally believed to be more responsive to changes in farm practices than are farmers' appraisals.

Within each sample field small plots are selected, essentially by use of random coordinates. These plots are frequently marked off in a system whereby the same plots may be visited from time to time during the growing season to obtain the data needed for relating plant characteristics to harvested yield components designed to forecast yields. The plots are harvested as soon as the crop is mature for purposes of estimating the yield of the entire geographic area. Immediately after harvest, the fields are again visited, using another sample of plots, in order to measure commercial harvesting losses, that is, the amount left in the field.

In the development of statistical models, three periods might be considered, because each presents a different problem. The first is the short period just prior to harvest called preharvest (after physiological maturity) when the problem is limited to developing appropriate sampling and estimation techniques, as forecasting is not involved. These techniques were discussed earlier under the general heading of crop cutting. Not all fields may mature at the same time, therefore the dates for this preharvest period can vary from field to field, which requires advance knowledge about when each sample field is likely to be harvested. This is known from the observations taken during the growing season on stage of development and from contacts with, or information supplied by the farmers in the local areas.

The second time period might be called late season or drymatter accumulation in the "fruit."" It begins with the date when all fruit has been set or the time when, if any additional fruit is to be set, the probability of its contributing to the yield is zero for practical purposes. Hence, for the second period as just defined, the problem can be stated as that of predicting the survival of the fruit and predicting the average size or weight of fruit (or the fruit parts) of commercial interest at the time of harvest.

The third or early-season period is the time after the plant has developed a portion of its leaf structure up to bud flowering or ear silking. This period is the active vegetative period when the plant structure is being established, and hence plant survival is no longer in doubt due to natural causes. There is a fourth period immediately after sowing or prior to spring greenup of winter grains (in the colder climates) which is beset

In this paper "fruit" is used in a botanical sense and includes the developing parts that have potential for contributing to the product for harvest.
with too many uncertainties to develop a reliable relation based primarily on plant vegetative characteristics that are helpful in predicting the size or weight of yield components.

### 2.5.2 Preharvest Measurement of Yields

As already indicated, the problem of preharvest estimates is essentially one of sampling and estimation, not forecasting. Nevertheless, experience has indicated that preharvest yield estimates (adjusted for harvesting losses) may be on a different level than estimates derived from reports from farmers. Which, if either, is correct? Since potential biases are inherent in any procedure, it is important that provision be made for ascertaining the validity of the preharvest sampling and estimating techniques. The probability of selection of each plot is very small, so an unusual amount of attention must be given to avoidance of nonrandom errors. Field workers may not be completely objective in the process of locating sampling plots. Or, if plots are subsampled for certain characteristics, there may be opportunity for bias in the techniques of subsampling. Also, instances may occur where the definition of the fruit to be harvested is replaced by a worker's own personal definition or interpretation. Thus, workers must be trained to develop an attitude toward the work such that the execution of operational tasks conforms to the model and an unbiased estimate of the parameter can be derived.

In the U.S., the Statistical Reporting Service had in operation beginning with the 1965 crop season a program of preharvest sampling for winter wheat, corn, cotton, and soybeans, as summarized in Table 7. In addition, measurements for forecasting are taken on May 1, June 1, and July 1 for winter wheat and on August 1, September 1, and October 1 for the spring-planted crops (corn, cotton, and soybeans).

Table 7--Preharvest Sampling in 1965

| Crop | Number <br> of <br> sample <br> fields | Approxi- <br> mate size <br> of plot <br> in acres <br> 1/ | Approximate size <br> of population <br> Acres <br> in <br> millions | Percent <br> of U.S. <br> total | error of <br> estimated <br> yield <br> per acre |
| :--- | :---: | :---: | :---: | :---: | :---: |
| Winter wheat | 2,300 | 0.0001 | 31.4 | 91 | 0.25 bu |
| Corn | 3,000 | 0.0023 | 54.5 | 95 | 0.70 bu |
| Soybeans | 1,900 | 0.0004 | 27.2 | 95 | 0.30 bu |
| Cotton | 2,600 | 0.0015 | 13.9 | 97 | 7.50 lb |

1/ Two plots are selected in each field.

There are various ways of getting a valid yield check, depending upon the crop. Take corn as an example. Farmers generally do not have weight measurements of the amount harvested and often have only approximate measures of volume, thus a new and more rigorous concept of yield per acre is being defined and introduced for checking purposes. To obtain a good independent check, special arrangements might be made with a small number of selected farmers for getting the total weight and other relevant measurements for the entire crop harvested from particular fields. Sample plots in these fields should be selected and harvested, using procedures identical to those used in the survey. The number of plots per field and in total would need to be large enough to give estimates having low sampling errors, so that any appreciable bias could be detected even at the field level. Adjustments may be necessary for such factors as differences in moisture percentage at the time of the preharvest sampling and at the time of harvest. Also, when comparing yield estimates and actual yield from the entire harvested field, one should be alert for inconsistencies in concepts of acreage. One of the problems arises from the possible difference between the actual land area of the
field from which sample plots are selected and what a farmer reports as the acreage in a field. However, the introduction of a yield concept based on weight is unlikely to present problems for a country without prior official yield series, since a change in concept is not involved. Table 8 shows some results of a validation study for preharvest fields of corn. The validation study suggests that the preharvest crop-cutting procedure results in yields which are slightly higher than the regular commercial harvesting procedure. However, except for the one State, the differences are within the sampling errors. There is no substantial evidence to suggest the reason for this difference, but the most likely yield component is the weight of grain per ear. A strong suspicion is that the difference is due to the amount of grain recovered per ear, or the scales possibly getting out of adjustment in the crop-cutting operation. A greater difference would have been found in the yields if the acreage had not been measured, since the same area was used to derive both the crop-cutting and commercially harvested yield. In the U.S. Corn Belt, the difference between grower reported acreage figures and measured net acreage occupied by the plants is about 2 to 3 percent. Earlier results from Sweden indicate the agreement between the biological yield adjusted for waste and farmers' estimates of yields to be quite good.
Table 8--Validation Study for Corn

| State | Number <br> of fields | Harvested acres (measured) | Preharvest Survey |  |  |  |  |  |  | Commercial <br> harvest of grain with 15.5\% $\mathrm{H}_{2} \mathrm{O}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  | Pairs of plots per field | Ears per plot (60' row) | Grain weight per ear with 15.5\% $\mathrm{H}_{2} \mathrm{O}$ | $\mathrm{H}_{2} \mathrm{O}$ content of grain in field | Gross yield of grain with $15.5 \%$ $\mathrm{H}_{2} \mathrm{O}$ | Net yield of grain with $15.5 \%$ $\mathrm{H}_{2} \mathrm{O}$ | Standard error net yield |  |
|  |  |  |  |  | (1b) | (pct) | (bu) | (bu) | (bu) | (bu) |
| Illinois | 13 | 340 | 18 | 57.4 | . 420 | 24.3 | 92.3 | 86.9 | 1.5 | 85.3 |
| Indiana | 16 | 245 | 18 | 66.0 | . 519 | 25.4 | 134.4 | 123.5 | 1.2 | 117.1 |
| Iowa | 16 | 325 | 18 | 60.2 | . 399 | 22.2 | 91.9 | 84.7 | 1.1 | 84.6 |
| Missouri | 16 | 271 | 18 | 53.0 | . 414 | 18.8 | 84.2 | 74.9 | 1.0 | 72.4 |

### 2.5.3 Forecasting Corn Yields Based on Plant Parts

The development of objective yield forecasting formulas that apply to specific forecast dates usually rest upon observable plant characteristics and sufficient knowledge of the fruiting behavior of the plant, so that plant characteristics observed on any date can be translated into an indication of yield. The studies reported here relate to the forecast dates August 1, September 1, and October 1. Field observations, in each instance, were taken during the previous week. In general, the techniques can be applied to most grain crops with minor variations. The basic yield models are the same. The yield per area is defined in terms of its components:
$\mathrm{Y}=\mathrm{plants}$ per acre x fruit per plant x grains per fruit x weight per grain (adjusted for commercial harvesting loss)

Each component in this model would be based on a specific set of linear or nonlinear prediction equations or, in computer terminology, different subroutines. Alternative models for corn, not based on yield components but on plant characteristics, could also be considered. Such a model might include the plant characteristics of basal area of stalk, height of plant at tasseling, leaf number and size. It is likely that within-year correlations of these characteristics with harvested yield might be moderately high, but would differ by varieties and areas.

Early in the season, "ears" (ear shoots) that may already be present can be counted in sample plots. But when counts are made before all "ears" have had time to emerge, other observable plant characteristics must be used which will indicate the rate of "ear" emergence or silking.

As the crop matures, ears attain their maximum length before the dough stage, so that the average size of the ears that will be harvested can be ascertained by direct measurement. The average quantity of ripe grain that will be produced per ear is closely
related to the length (or size as indicated by length times diameter) of ear at maturity. Maximum ear length is attained well before the grain is ripe. In order to ascertain whether an ear has reached its maximum length on a given forecast date, the stage of maturity or age of the ear must be considered. Many studies on corn show that ears in the milk stage have reached their maximum length. Consequently, measurements of ears in the milk stage were used to forecast the average weight of grain per ear at harvesttime.

When corn is already ripe on a forecast date, sample ears can be harvested, weighed, and subjected to laboratory analysis to compute the average weight of grain per ear at a standard moisture content.

On August 1, not all "ears" have appeared in all fields in the main region growing corn in the U.S. An August 1 model must first provide a forecast of the number of ears that will be present at harvesttime. However, the ears which have not appeared by August 1 contribute very little to harvested production in most years. It is also necessary to forecast the quantity of grain that will be produced per ear.

By September 1, all ears that have a chance of reaching maturity are present and most are well developed, so the presence of grain is discernible. But in many fields the ears have not yet laid down all the dry matter in the grain. The kernel weight levels off at a maximum by the time the moisture content of the kernel has decreased to 30 percent, or 60 to 70 days have elapsed since silking.

By October 1, virtually all ears have attained the dry-matter content of grain that can be expected at harvest, except in the very latest maturing fields. In parts of the Northern States, the accumulation of dry matter may be stopped by killing frost before the full yield potential is realized. If frost occurs late enough, the ears may still be harvested for grain, but the grain will be
lighter than if development had not been arrested by frost. If frost occurs earlier, the ears may be so immature that the crop must be diverted to uses other than for harvest as grain. If this occurs, the contribution of these ears to the total production of grain may be zero or unimportant for yield, since the acreage is now for another use.

### 2.5.3.1 Relations for the August 1 Yield Forecast

The relations which are set forth in this and subsequent sections are intended to illustrate approaches which have been tried and found useful at different stages in a program that has been operating since 1960 over large geographic areas. If an optimum model was desired for each individual State or small area, separate parameters would be required for the small area. The forecast of number of ears to be produced is considered first. An observable ear or ear shoot is defined as one that has already developed to the stage where some silks are protruding from the husk. By August 1 , all ears or ear shoots that have a chance of maturing are already present on the plants in the Southern States. In a few fields in the more northern portions of the Southern States, and in the North Central States, the ears and ear shoots present are less than the number that will be found at maturity.

The plant observations were made in two double 15-foot row sections in each sample field. If the ears in these small plots have already reached the milk stage, there is little chance of any additional ears appearing later. The ear count represents all ears that will be formed. But if no ears have yet reached the milk stage, the total number of ears to be formed must be forecast. Two methods of making this forecast are described.

The first approach involves counting the stalks in the measured plots and assuming a constant number of ears per stalk from year to year. The second approach assumes a fixed linear relation between the fraction of stalks with ears on August 1 and the ratio
of ears already present to the total number of mature ears that will be produced. Both approaches gave about the same results, and the same type of relations appears to hold in both regions. The second approach might be preferable if the number of ears produced per stalk were subject to greater variation from year to year. However, the introduction of new varieties or marked changes in plant density per area may well invalidate the parameter values for both approaches. In this case, a transitory or indi-vidual-year model would be indicated for this component.

Data collected over a period of years showed that the number of mature ears produced in 60 feet (i.e., two plots) of a row is related to the August 1 stalk count, as shown in Table 9. The data in this and several subsequent tables are based on freehand regression lines drawn on scatter diagrams in which the original data and group averages were plotted.

Table 9--Number of Mature Ears Produced per 60 Feet of Row, and Relation to August 1 Stalk Count

| August 1 <br> stalk <br> count | Mature <br> ears <br> produced | August 1 <br> stalk <br> count | Mature <br> ears <br> produced |
| :---: | :---: | :---: | :---: |
| 10 | 10 | 45 | 45 |
| 15 | 16 | 50 | 50 |
| 20 | 21 | 55 | 55 |
| 25 | 26 | 60 | 59 |
| 30 | 41 | 65 | 64 |

On the average, about 1.05 mature ears are produced in the Southern Region for each stalk counted on August 1. In the North Central Region, where yields are higher and the ears larger, the
average is 0.98 . This difference is not inconsistent with the relation in Table 9, which holds for both regions.

When the fraction of stalks that have ears or ear shoots on August 1 is used to forecast the number of mature ears that will be produced, relations in the South differ somewhat from those in the North Central Region. The ratio of ears and ear shoots counted on August 1 in the Southern Region to mature ears produced is about 1.4 times the fraction of stalks having ears or ear shoots on August 1. For the North Central Region, the relation is as shown in Table 10.

Table 10--Ratio of "Ears" Counted August 1 to Mature Ears Produced, in Relation to Stalks with "Ears" on August 1, North Central States

| Stalks <br> with <br> "ears" <br> August 1 | Ratio of <br> August 1 <br> ear" count <br> to mature <br> ears produced | Stalks <br> with <br> "ears" <br> August 1 | Ratio of <br> August 1 |
| :---: | :---: | :---: | :---: |
| "ear" count |  |  |  |
| to mature |  |  |  |
| ears produced |  |  |  |

Whenever the August 1 percentage of stalks with ears is very low and ears have emerged in only a few fields, it is preferable to assume a fixed number of ears per stalk (1.05 in the South or 0.98 in the North Central States), rather than to use the observed August 1 "ear" count and divide by the appropriate ratio shown in

Table 10. In practice, it is desirable to consider fields in which no ears have yet emerged separately from those in which some ears have emerged. If there are fewer than 20 sample fields in the second group, Table 10 may fail to give a good indication of fruiting potential, even for the fields in that group.

The weight of grain produced per ear did not vary much from year to year during the period in which these initial studies were conducted. But a method of forecasting weight per ear early in the season might be desirable, since it may provide a clue of a departure from average. In much of the South, most ears have reached the milk stage, and their maximum length, by August 1. The length of the entire ear, or of the part of the ear that is covered by kernels, can then be used to predict the average weight of grain per ear at maturity. It is more convenient and quicker to measure the length of the entire cob over the husk. This procedure also avoids damage to the ear, but may not work very well if the kernel-row length is quite variable. In this case, pulling back the husk is preferable.

For ears that have reached their full length, but are not ripe, the linear regression of weight of grain produced per ear on length of ear, measured over the husk, was given by:

$$
\begin{equation*}
Y=0.0854 X-0.304 \tag{4}
\end{equation*}
$$

In this equation, $X$ is the total length of cob in inches, measured over the husk, and $Y$ is the weight of grain produced in pounds, adjusted to 15.5 percent moisture content.

For ears that are already mature (maximum dry matter attained), the regression equation becomes:

$$
\begin{equation*}
Y=0.0886 X-0.310 \tag{5}
\end{equation*}
$$

The difference in the two equations that arises is believed to be from ears shrinking slightly by the time they ripen due to drying out.

An alternative approach is to consider the weight of the grain predicted in some other way. A relation between the number of mature ears produced in 60 feet of row and weight of grain was used. If the planting system in any area is relatively unchanged from one year to another, variations in ear counts reflect differences in growing conditions. Favorable growing conditions are conducive to good stands and the formation of large numbers of ears. These conditions are also conducive to good development of the ears. This view is consistent with the behavior of other crops that were studied in the research program on objective yield forecasting. The data in Table 11 indicate that this is also true for corn. As the number of mature ears expected can be forecasted fairly well, this offers some chance of predicting the change in the quantity of grain to be produced.

Table 11 is used directly to forecast the weight of grain when the number of ears per 60 feet of row is known. However, the curve describing the relation is at a different level for different States; consequently, it is more accurate to use the table to indicate change from a previous year if small-area or State yield estimates are desired. If the number of mature ears per 60 feet of row and the weight of the grain are known for a previous year, the change in the weight associated with the change in the number of ears as indicated by Table ll can be applied to the grain weight for the previous year.

Table 11--Relation of Weight of Grain per 60 Feet of Row to Number of Ears with Grain

| Ears with grain per 60 feet of row | Weight of grain at 15.5\% moisture |  |
| :---: | :---: | :---: |
|  | North Central States | Southern States |
| (no.) | (1b) | (1b) |
| 5 | 1.0 | 0.8 |
| 10 | 2.0 | 1.6 |
| 15 | 3.7 | 3.0 |
| 20 | 5.7 | 4.5 |
| 25 | 8.0 | 6.4 |
| 30 | 10.5 | 8.5 |
| 35 | 13.2 | 11.0 |
| 40 | 16.0 | 13.7 |
| 45 | 18.5 | 16.4 |
| 50 | 21.5 | 19.1 |
| 55 | 25.0 |  |
| 60 | 28.2 |  |
| 65 | 31.5 |  |
| 70 | 34.8 |  |

By September 1, the ears that will produce grain can be identified and counted. If a few fields have not reached the milk stage, the total number of mature ears expected can be predicted as for the August 1 forecast. But, as a practical matter, it is simpler and just as satisfactory to assume that the average number of ears per stalk producing grain will be about the same for these fields as for the fields that are already more mature. The weight of the grain that will be produced can be estimated from the length of the cob, measured over the husk, as for the August 1 forecast.

A slightly more accurate indication can be obtained by considering only the length of the part of the cob that is covered by kernels (i.e., average length of kernel rows). The average weight of grain per ear is related to this length by the equation:

$$
\begin{equation*}
Y=0.0890 X-0.215 \tag{6}
\end{equation*}
$$

As in equations (4) and (5), the weight per ear is in terms of pounds of grain at 15.5 percent moisture, and the length of kernel rows is measured in inches.

When fields are fully mature the sample ears can be weighed in the field, the shelled grain weighed in the laboratory, and moisture tests made. But even for such fields, ear-size measurements give an accurate weight indication much more quickly. In most States, the percentage of fields that have matured fully by September 1 is small.

The fraction of total dry matter already present in the kernels can be estimated from the ratio of dry-kernel weight to wet-kernel weight, as shown in Table 12. This ratio can be compared with the dry-matter fraction laid down at maturity, or used for adjusting grain weights when sample ears are harvested and weighed too early. It is also useful for estimating the reduction in yield caused by frost before ears reach full maturity. The data in Table 12 are average figures derived from laboratory
studies for the North Central States during the early 1960's. Table 12 gives the relation between averages for large numbers of ears. Although any one ear for which the ratio of dry-kernel weight to wet-kernel weight is 70 percent will have already laid down all of its dry matter, a group of ears for which the average ratio is 70 percent must obviously include some ears for which the ratio is less than 70 percent. For this reason, the data in Table 12 indicate a slightly different relation than would have been observed for individual ears. Tables 13 and 14 are based on individual ear data.

Table 12--Relation Between Ratio of Dry-Kernel Weight to WetKernel Weight and Fraction of Total Dry Matter Laid Down

| Average ratio <br> of dry-kernel <br> weight to wet- <br> kernel weight | Average <br> fraction of <br> total dry <br> matter <br> laid down | (pct) <br> (pct) <br> of dry-kernel <br> weight to wet- <br> kernel weight | Average <br> fraction of <br> total dry <br> matter <br> laid down |
| :---: | :---: | :---: | :---: |
| 10 | 5 | (pct) <br> 20 | (pct) |
| 30 | 30 | 60 | 70 |
| 40 | 50 | 80 | 85 |

2.5.3.3 Characteristics Useful in Forecasting Kernel and Ear Weight

Tables 13 and 14 show some correlation between various characteristics for corn. The tables suggest some of the possible atlernative approaches which could be considered for corn on any of the three dates, based on stage of development.

Table 13--Typical Kernel Characteristics and Correlation With Dry-Ear Weight by Days After Silking

| Days after silking | Number kernels per ear $1 /$ | Dry weight per kernel | Wet weight per kernel | Correlation between dry-ear weight and |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  |  |  | $\begin{gathered} \text { Number } \\ \text { of } \\ \text { kernels } \end{gathered}$ | Dry- <br> kernel weight |  |
|  |  | (gm) | (gm) |  |  |  |
| 15 | 790 | . 048 | . 198 | . 45 | 2/ | 2/ |
| 25 | 760 | . 120 | . 280 | . 45 | 2/ | 2/ |
| 40 | 610 | . 225 | . 385 | . 90 | . 55 | . 42 |
| 55 | 605 | . 255 | . 385 | . 80 | . 40 | . 30 |
| 70 | 600 | . 260 | . 360 | . 84 | . 40 | . 30 |
| 85 | 600 | . 260 | . 355 | . 80 | . 25 | . 10 |

1/ Based on a count of kernel with evidence of fluids or coloring appropriate to stage of development.

2/ Too few data points.

Table 14--Typical Correlation of Ear Characteristics With DryEar Weight by Days After Silking

| Days <br> after <br> silking | Maximum <br> circum. <br> (after <br> husk <br> removed) | Kernel <br> surface <br> area 1/ <br> (husk <br> removed) | Total <br> ear <br> weight <br> (wet) | Cob <br> length <br> over <br> husk | Surface <br> area 2/ <br> (over <br> husk) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 15 | .70 | .75 | .70 | .60 | .65 |
| 25 | .50 | .75 | .75 | .65 | .75 |
| 40 | .50 | .85 | .85 | .75 | .75 |
| 55 | .65 | .85 | .90 | .75 | .85 |
| 70 | .65 | .90 | .90 | .82 | .92 |
| 85 | .65 | .90 | .90 | .70 | .85 |

1/ Average kernel-row length times maximum circumference.
2/ Cob length times maximum circumference.

### 2.5.3.4 Relations for an October 1 Yield Forecast

By October 1 all dry kernel weight has been laid down in most fields. But, in a few fields, the weight of grain per ear must be estimated by ear-size measurements or other means.

The most accurate indication can be obtained by weighing sample ears and applying the relations in Table 12 to adjust the observed grain weight to a weight at maturity. But if the production of dry matter is halted by a killing frost before the ears have a chance to reach maturity, an allowance must be made for the resulting reduction in yield. When the moisture content is known, Table 12 can be used for this purpose.

### 2.5.4 Forecasts Based on Growth Models for Yield Components

Within-year growth models for forecasting components have been investigated. These methods rely on plant data only from the current year. As such, they have the opportunity of reflecting unique characteristics of the crop year for which the forecast is desired. However, the same type of growth model is assumed each year and statistical estimates of the model parameters are derived for the current year.

Within-year models depend on relating a response (the component to be forecast) to values of a second variable which has a known value at maturity. Various measures of time or a variable which reflects the aging of the component provide a suitable independent variable for this purpose. In fact, Table 12 is an example of a relative growth model based on percent of dry matter as a time or aging variable.

In modeling the average weight of grain per ear per plant for corn, "days since silking" has been considered as the time variable. Note the uniformity of weight after physiological maturity in Table 13. The model provides an estimate of grain weight at any given time after silk emergence. The forecast is dependent on how well the model represents the actual situation and on our ability to know what value of time corresponds to maturity and how to measure these accurately. In this case, the time value at maturity is any value in the plateau region.

Within-year models for survival (the complement of the growth model) of the fruit, ears, or other characteristics may also be developed. The forecast of the number of ears is then combined with the growth model for weight per ear. The dependence of a survival ratio on days since a base date for plants with ears per acre is used to forecast number of ears at maturity. The base date of plants with ears per acre is defined to be day zero for the survival ratio. The forecasted survival ratio at maturity is multiplied by the base estimate of plants per acre with ears to adjust it to
number of units at harvest. Under conditions in the U.S., this ratio is about . 98 to .95 for corn ears and has no known relation to the yield estimates.

Research on both growth and survival models has also been found applicable to several crops. For example, survival models have been investigated to forecast the portion of papayas set each week surviving to harvest (some five to six months later), as well as used as a growth model to forecast weight per grape. Because previous year data are required for developing over-the-year models, within-year methods may be more applicable in developing and implementing objective yield forecast procedures for new crops. However, it is assumed that the basic models are known from other studies or research.

The general form of the logistic growth model which is commonly utilized is:

$$
\hat{Y}_{t}=\frac{1}{\hat{\alpha}+\hat{\beta}(\hat{p})}
$$

An alternative but equivalent form is:

$$
\hat{\mathrm{Y}}_{t}=\frac{\hat{\alpha}^{-}}{1+\hat{\beta}^{-}\left(\hat{\rho}^{\prime}\right)}
$$

This is a nonlinear model, where $t$ is the independent time variable, $Y_{t}$ is the dependent variable, and $\alpha, \beta$ and $\rho$ are the parameters which can be estimated from data sets of the form:


In the application discussed here, $\hat{Y}_{t}$ is the estimated mean dry weight of corn per ear or per kernel at time $t$. The variable $t$ is the time (days) after one of the phenological events: tassel emerged, silks emerged, silks starting to dry, silks finished drying; or the "time" variable can be dry matter fraction of the grain when sampled.

The basic model uses repeated observations from the current year to estimate the parameters needed to predict the dependent variable (dry weight of grain per ear, per plant or per kerne1) at maturity. The model parameters may be updated at various times during the growing season as more data become available for later stages of growth. The same type of model is used each year, but the parameters derived from the data would relate to: (1) the current year, and (2) a given cutoff time within the growth period. Since three parameters are to be estimated, at least three observations within a season are required.

The role of the three parameters in the growth model can be described in terms of various phases of growth.

1. The initial phase or base weight is at $t=0$. Since $\hat{\rho}$ (whatever its value) raised to the power $t=0$, is 1 , $\hat{Y}_{0}=\frac{1}{\hat{\alpha}+\hat{\beta}}$ estimates the base weight or initial weight.
2. The final phase or mature forecast weight of the dependent variable is the most important in forecasting final corn yield. Assuming that $0<\hat{\rho}<1$, we see that the forecast harvest weight is $\hat{Y}_{m}=\lim _{t \rightarrow \infty} \hat{Y}_{t}=\frac{1}{\hat{\alpha}}$. That is, for large values of $t, \hat{Y}_{t}$ depends upon $\hat{\alpha}$. Therefore, the parameter $\alpha$ is termed the primary parameter. For the alternative form of the logistic growth model, $\hat{Y}_{t}=\frac{\hat{\alpha}^{-}}{1+\hat{\beta}^{-} \hat{\rho}^{-t}}$, the point estimate, $\hat{\alpha}^{\prime}$, is the forecast of dry-kernel weight per plant at maturity.
3. The intermediate phases of growth follow the initial phase and continue until maturity, when the large values of $t$ are reached. The value of $\hat{\rho}$ reflects the rapidity of the weight increase from $\hat{\mathrm{Y}}_{0}$ to $\hat{\mathrm{Y}}_{\mathrm{m}}$ as t increases. For $0<p<1$ the model is indeed a growth model and $\rho$ can be termed the rate-of-growth parameter. If $\hat{\rho}$ is near zero the growth is very rapid. If $\hat{\rho}$ is near unity, growth at a gradual rate is indicated. The ratio of $\frac{\hat{Y}_{m}}{\hat{Y}_{0}}=1+\frac{\hat{\beta}}{\hat{\alpha}}$ determines the range of the $Y_{t}$ scale.

The computer programs utilized to derive the parameters from the data require approximate starting values, since fitting nonlinear equations is based on iterative algorithms. For the dependent variable dry weight per ear, the starting values used for the data set were $\hat{\alpha}=.006, \hat{\beta}=.08$, and $\hat{\rho}=.87$. For $Y_{t}=d r y$ weight per kernel at time $t^{\prime}$, the initial or starting values used were $\alpha=3.8$, $\hat{\beta}=130$, and $\hat{\rho}=.87$. Normally, the values from a previous year could be used to start the iterative algorithm.

Each set of parameter estimates defines a specific model at a given time. For example, for 1973 in Central Iowa, $Y_{t}=$ dry weight (gm) per ear at $t$ days after silks begin to dry, and we have the following parameter estimates for data sets available after various field visits given in Table 15.

Table 15--Estimates of Model Parameters Based Upon All Data Available After Various Field Visits

| Parameter | No. of visits (during season) |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | Four <br> (IV) | Six <br> (VI) | Seven <br> (VII) | Eight <br> (VIII) | Nine <br> (IX) |
| $\alpha$ | .0059597 | .0061557 | .0062934 | .0063149 | .0063487 |
| $\beta$ | .12777 | .12930 | .14616 | .14869 | .15380 |
| $\rho$ | .88271 | .88108 | .87514 | .87428 | .87267 |

Thus, the specific model based upon data obtained on field visits I-VII is

$$
\hat{Y}_{t}=\frac{1}{.0062934+.14616(.87514)^{t}}
$$

where $\hat{Y}_{t}$ is the estimated dry weight (gm) of grain per ear at $t$ days after silks began to dry. For $\hat{Y}_{t^{\prime}},=$ estimated dry weight (gm) per kernel at $t^{\prime}$ days after silks emerged, based upon data from visits I-VII, the model is

$$
\hat{Y}_{t^{\prime}}=\frac{1}{3.8654+333.95(.87113)^{t^{\prime}}}
$$

Numerical values of the dependent variables for various values of the two time variables are shown in Table 16 for these two models.

Table 16--Estimated Dry Weight Per Ear and Per Kernel Related to Different Time Variables

| Time after <br> silks <br> started <br> to dry <br> $(t)$ | Estimated dry- <br> grain weight <br> per ear <br> $\left(\hat{Y}_{t}\right)$ | Time <br> after <br> silks <br> emerged <br> $\left(t^{\prime}\right)$ | Estimated dry- <br> grain weight <br> per kernel <br> $\left(\hat{Y}_{t},\right)$ |
| :---: | :---: | :---: | :---: |
| (days) | (grams) | (days) | (grams) |
| 0 | 6.56 | 0 | .0030 |
| 10 | 22.32 | 10 | .0114 |
| 20 | 60.82 | 20 | .0400 |
| 30 | 111.52 | 30 | .1088 |
| 40 | 142.90 | 40 | .1921 |
| 50 | 154.34 | 50 | .2380 |
| 60 | 157.67 | 60 | .2531 |
| 70 | 158.57 | 70 | .2573 |
| 80 | 158.81 | 80 | .2583 |
| 90 | 158.87 | 90 | .2586 |
| 100 | 158.90 | 100 | .2587 |
| 110 | 158.90 | 110 | .2587 |
| 120 | 158.90 | 158.90 | $\infty$ |
| $\infty$ |  |  | .2587 |

Two methods of evaluating the performance of the logistic growth model for various time variables and as data become available for later stages of growth are:
(1) The magnitude and sign of the departure of the forecast from actual mean dry weight at maturity.
(2) The magnitude of the relative standard deviation of the "primary" parameter, $\alpha$.

Mean dry weight at maturity was estimated from a large sample of plants with mature ears. The mean was for the population of plants sampled during the entire growth period for which the time variable in the model being evaluated was defined. That is, the
model forecast and estimated mean weight make valid inferences about the same subpopulation. The relative standard deviation is the estimated standard deviation divided by the estimate of the primary parameter $\quad\left(\stackrel{{ }_{\alpha}}{n}\right)$.
$\alpha$

For the two examples previously discussed, departures of the forecast from the actual mean dry weight and the relative standard deviations are shown below.

Table 17--Percentage Difference Between Forecast and Harvest Weight and Between Relative Error in Primary Model Parameters

| Dependent variable | ```Inde- pendent time variable``` | Data from visits | Departure of forecast from actual mean dry weight | Relative $\qquad$ standard deviation of estimate of primary parameter ( $\alpha$ ) |
| :---: | :---: | :---: | :---: | :---: |
|  |  |  | (pct) | (pct) |
| Dry | Days | I (only) | (No convergence to model) |  |
| weight | after | I \& II | +22.0 | $34.46$ |
| of | silk | I - III | +0.7 | 6.96 |
| grain | starting | I - IV | +7.8 | 4.32 |
| per | to dry | I - V | +8. 5 | 2.74 |
| ear | (t) | I - VI | +4.4 | 1.74 |
| $\left(Y_{t}\right)$ |  | I - VII | +2.1 | 1.29 |
|  |  | I - VIII | +1.7 | 1.16 |
|  |  | I - IX | +2.2 | 1.02 |
|  |  | I - X | +1.2 | 0.92 |


| Dry | Days | I (only) | -89.6 | 16.07 |
| :--- | :--- | :--- | ---: | ---: |
| weight | after | I \& II | -71.4 | 12.09 |
| of | silk | I - III | -37.3 | 10.10 |
| grain | emerged | I - IV | +4.6 | 6.61 |
| per | $\left(t^{\prime}\right)$ | I - V | -6.0 | 2.24 |
| kernel |  | I - VI | -1.8 | 1.59 |
| $\left(Y_{t-}\right)$ |  | I - VII | +0.4 | 1.37 |
|  |  | I - VIII | +0.5 | 1.15 |
|  |  | I - IX | +0.9 | 0.98 |
|  |  | I - X | +1.2 | 0.86 |

### 2.5.5 Forecasting Methodology for Citrus Yields

### 2.5.5.1 Introduction

The program for estimating the citrus crop in Florida was developed in the late 1950's. The yield estimating portion of the program as it was originally developed and put into operation is discussed in this section. Most of the methodology and data concepts have remained unchanged.

The inventory of trees by type, age, and location is very important in the forecasting of current yield and production because of the dynamics of the industry. It is needed to provide a complete and efficient sampling frame of trees for sample surveys designed to estimate the number of fruit per tree. The initial yield survey each year is used to estimate the average number of fruit per tree. This survey begins August 1 and continues to September 15. It is referred to as the "limb count survey."

### 2.5.5.2 Estimating Average Number of Fruit Per Tree

Number of fruit per tree varies considerably due to different ages and locations. Most citrus trees start bearing about 3 to 4 years after planting. Production increases rapidly for about 10 years, tapers off, and reaches maximum about 25 to 30 years after planting. These tree characteristics and the vital knowledge of tree numbers by age and area allow considerable reduction in estimator variances by using a stratified sample design. Prior knowledge of fruit counts by age of tree was used to construct strata.

Stratum

1
2
3
4
$\frac{\text { Age of Tree }}{\text { (years) }}$
4- 9
10-14
15-24
25 and older

The relatively small counts on trees in stratum 1 and the smaller variances of these counts combined with the large influx of young trees into the universe allow increased efficiency by using optimum allocation of sample to age strata.

Since the age-type blocks are too large to be feasible units for counting fruit, the groves are subsampled to obtain a cluster of trees. From variances on complete tree mappings (i.e., censuses on individual trees), it was determined that a limb of area equivalent to 10 to 20 percent of the main trunk area could be counted and the count expanded to obtain a fairly efficient estimate of fruit population for the total tree. The sample sizes of number of groves and number of trees per grove were determined from expanded counts made on randomly selected limbs which constituted approximately 10 percent of the main trunk area as determined by measuring the circumference with a tape. Data were summarized using analysis-of-variance techniques for a hierarchical classification. Computed variances were used for optimum allocation of sample to age strata.
B. W. Kelly conducted the pilot survey work on 50 trees in 1956, providing estimates of variance components, required sample size, and optimum allocation. The results are presented in Table 18. Subsequent analyses of variance on estimated fruit per tree from the limb count surveys indicated the pilot survey to be relatively accurate.

An aerial tree census is the source of the list of all blocks of each major type of citrus in the State from which the blocks are selected. A block of citrus is not defined by ownership but rather is defined as being a relatively homogeneous planting with at least 90 percent of the trees being of the same age and citrus type. The block identification, tree numbers, and accumulated tree numbers are listed by county and by date of planting for each type of fruit (a type consists of one or more similar varieties). A sample of blocks is selected for each type of citrus.

Table 18--Estimated Limb-Count Variance Components, 1956

| Type of fruit | Components of variance $1 /$ (nested design) |  |  |  | $\begin{array}{\|c} \text { Indicated }{ }^{2 /} \\ \text { sample } \\ \text { size } \end{array}$ | Indicated optimum trees per grove |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | County | Age | Grove | Tree |  |  |
| Oranges |  |  |  |  |  |  |
| Midseason | 0 | 43 | 118 | 360 | 519 | 3.5 |
| Late | 7 | 84 | 162 | 93 | 463 | 1.5 |
| All |  |  |  |  | 499 |  |
| Grapefruit |  |  |  |  |  |  |
| Seedy | 12 | 0 | 20 | 218 | 294 | 6.5 |
| Seedless | 20 | 3 | 69 | 152 | 418 | 3.0 |
| All |  |  |  |  | 370 |  |

1/ Variance components for number of fruit per tree estimated by limb count method. Variance components rounded to nearest thousand.

2/ Indicated number of groves required for a maximum of 4 percent sampling error (coefficient of variation at .95 level of confidence), assuming 4 sample trees per sample grove.

After the sample groves are selected, a "pivot tree" is chosen in each sample grove. The pivot tree in each case specifies two sample clusters of four trees each; clusters are rotated to minimize the effects of working in the trees to make fruit counts. The procedure used to designate pivot trees allows the proper proportions of outside trees to be selected. Due to demise, or to improper age or type, it is sometimes necessary to substitute for a sample tree using a predetermined substitution pattern.

The third and final stage of sampling pertains to selection of a portion of the tree on which the fruit is to be counted. Counts are made on sample limbs selected by the random-path technique. When this multiple-stage process terminates, the selected limb (branch or group of branches) has a probability of selection proportional to limb cross-sectional area (c.s.a.). The reciprocal of this probability of selection affords an unbiased method of
expanding sample counts to estimated total fruit on the tree and, due to the positive correlation between c.s.a. measurements of limb size and number of fruit, is a fairly efficient method of sampling. Proof of the unbiasedness of the estimator, $\left(x_{i} / p_{i}\right)$, and derivation of the probability, $\left(p_{i}\right)$, are given elsewhere.

Application of the random path selection method is fairly simple. Branches of the primary tree scaffold (first major branching) are measured with a tape which shows c.s.a. in square inches. The c.s.a. and cumulative c.s.a. square inches are recorded for each limb on the field sheet where "limb" is defined as being a branch or grouping of adjacent branches totaling 10 percent or more of the cumulative total c.s.a. at the first scaffold level. A number selected from a random-number table determines the individual portion selected. A logical alternative to the 10 -percent sample limb would be two 5-percent limbs. However, smaller limbs appear to have a lower correlation between c.s.a. and fruit count.

The principle involved in the "1imb count" selection is depicted in Figure 2 on page 84. The procedure by stages includes measurement of the first scaffold c.s.a. to determine that approximately a 19 -inch limb ( 10 percent of 190 square inches) is needed to provide the sample unit. The route toward the sample limb is determined by a random number from 1 to 190 and the accumulated c.s.a. measurements. In the example, the 100 -inch limb was selected by the random number. This limb had a probability of selection of $100 /(100+90)$. At the second scaffold the illustrated selection was the 20 -inch limb, and the 187 fruit on that limb were counted. The probability of selection at the second stage was the first-stage probability times the second-stage probability, given that the first-stage selection is known. In the example, then, the probability of the 20 -inch limb's being the sample limb is:

$$
\frac{100}{100+90} \times \frac{20}{20+40+50}=\frac{100}{190} \times \frac{20}{110}=\frac{20}{209}
$$

The sample count of 187 is expanded by the reciprocal of the probability to give the estimate of 1954 fruit on the tree (187 x 209/20 $\doteq 1954$ ) .

Counts of fruit on each "10 percent" limb are made by categories based on the major bloom cycles. Categories are determined by size of fruit at limb-count time as shown in Table 19.

Table 19--Fruit Size Classifications Used in Limb-Count Surveys

| Type <br> of <br> citrus | Diameters of fruit size classifications |  |  |
| :--- | :---: | :---: | :--- |
|  | "Regular" bloom | "First late" bloom | "Second late" bloom |
| Grapefruit | (in.) | (in.) |  |
| over $11 / 14$ | $13 / 16-11 / 4$ | less than $13 / 16$ |  |
| Oranges 1/ | over 1 | $11 / 16-1$ | less than $11 / 16$ |
| Tangerines | over $11 / 16$ | $5 / 16-11 / 16$ | less than $5 / 16$ |

1/ Same sizes used for tangelos and Temples.

Many of the trees have branches which, due to dead limbs or major pruning, carry much less bearing surface than indicated by c.s.a. at the scaffolding. Therefore, in the limb selection process, a reduced c.s.a. obtained by measuring branches beyond major prunings is accepted for determining probability of branch selection. Dead limbs are not measured. If this is limited to major reductions, it is a worthwhile method of reducing the variance of the estimator.

After the sample limb is selected, it is divided into smaller units for counting purposes. Two separate fruit counts are made, each by a different member of the survey crew. If the two counts do not agree within a specified tolerance, additional counts are made.

A random selection of one of the 10 -percent limbs in a $10-$ percent random subsample of limb-count groves is made as a quality check of the original counts. These quality checks indicate that the method provides a fairly consistent undercount of about 1 percent.

### 2.5.5.3 Forecasting Fruit Drop

A measure of fruit mortality prior to harvest must be introduced into computed crop forecasts, because initial estimates of the average number of fruit per tree are established from counts in August and September. Natural loss of fruit, from August until the month in which each type of fruit is considered mature, is measured by a sequence of monthly surveys. Maturity is considered to be reached in predetermined cutoff months which precede the heaviest harvest period. Cutoff months are: December for tangelos and tangerines, January for early and midseason oranges, February for Temples and grapefruit, and April for late-season oranges.

The sample trees for droppage surveys are drawn from a special or a restricted portion (blocks along roads) of the frame used for the limb count. Blocks along this route frame are readily accessible for monthly observations. This sample frame consists of all bearing commercial groves fronting on a $1,500-m i l e$ route which traverses producing areas of the most important counties. This microcosm of the citrus population provides a satisfactory base for sampling drop and other relatively uniform characteristics.

The sample for each variety is stratified into four areas (homogeneous county groupings) and the four age groups previously discussed. The sample size within strata is based on productivity in a base year.

A sample limb approximately two percent of the trunk c.s.a. is selected near shoulder height, on a designated side of the tree. This limb is tagged and all fruit beyond the tag are counted during
successive surveys. The differences between the initial survey counts and later survey counts indicate the droppage to the time of the survey. The average drop for each age-area is computed and then combined by production weights into the average drop for the State. The sample counts are weighted, because groves are selected with probability proportionate to production and the "two percent" limb sample survey tends to put a disproportionate part of the sample in older, more productive trees.

The monthly drop rates are adjusted by the estimated proportion of total crop harvested by the survey date. The accumulated fruit drop represents only those groves not yet harvested. The adjusted monthly droppage is projected to the cutoff month to estimate seasonal drop rate for use in the forecast models.

The 2,000-tree sample in 1966-67 indicated the proportion of oranges remaining for harvest with a maximum error of three percent at the .95 level of confidence. The sampling errors of the drop survey are expressed as the coefficient of variation for the proportion of fruit remaining to be harvested, since this is the error contribution to the production forecast.

Prior to the 1970-71 season, monthly projections of fruit loss expected to occur prior to the cutoff month were made by graphic interpretation of charts similar to those in Figure 3 on page 85. Although this procedure was satisfactory during years in which loss of fruit was within the normal range, experiences in recent seasons suggested that visual interpretation was not sufficient, particularly when the rate of drop was much higher or lower than usual. Starting in 1970, multiple-regression formulas have provided additional means of estimating total fruit loss.

The fruit-size survey coincides with the drop survey. Moreover, the same subsample of trees in sample groves drawn from the route frame is used for both sets of monthly observations. In the size survey, 10 sample fruit per tree are measured from a two-tree cluster per sample grove. Frequency distributions of standard fresh-fruit sizes and the estimated average size are obtained each month.

The fruit to be measured are determined by a "random grab" or point on the tree about shoulder height. This point on the tree is tagged and, for each survey, horizontal circumferences are measured on the 10 regular bloom fruit nearest the tag.

These circumference measurements are entered as a tally on the 240-cell field form. Summarization is done in volume, which is linearly correlated with weight and, therefore, is additive. The weight-to-volume relation has a correlation coefficient squared ( $r^{2}$ ) of .96 , which is pertinent to a production estimate, since most of the citrus crop is received or purchased on a weight basis.

Figure 4 on page 86 depicts the growth rates of two citrus types. The dates shown are the months in which surveys were conducted; usually surveys were near the third week of each month. The annual growth curves generally parallel each other, thereby allowing these relationships to be a fairly effective tool in forecasting size at maturity. It should be noted that fruit measured on-tree does not reflect harvest size. (Early observations are of immature fruit, and measurements for forecasts usually cease prior to the main or volume harvest.) The size of fruit at maturity is defined as the average size of fruit in groves in a specific month. These cutoff months are the same as in the drop surveys. Prior to the cutoff month, it is necessary to estimate the average size that fruit will attain in the cutoff month.

A regression using three variables is used to forecast size (volume per fruit) at the cutoff month:

$$
\begin{aligned}
X= & 4.34+.964 X_{1}-.159 X_{2}-.002 X_{3} \text { (for early-mid oranges } \\
& \text { on the October } 1 \text { forecast date) }
\end{aligned}
$$

and $r=.95$, where the three variables are (1) current month's average size in cubic inches, (2) growth during the preceding month, and (3) average number of fruit per tree for that type. The multiple regression has provided a sounder indication of final size than a subjective evaluation of the importance of these factors in arriving at a forecast size. In 1967-68 a subsample of fruit on 1,200 sample trees used in size surveys provided a maximum error at the .95 level of confidence of about 1.5 percent in average fruit size for all oranges.

The citrus check data, with which the forecast must be compared, is the number of certified boxes harvested--90-pound boxes for oranges, tangelos and Temples; 95-pound boxes for tangerines; and 85 -pound boxes for grapefruit. The forecasted average volume per fruit is converted to number of fruit constituting a box by graphic means, as shown in Figure 5 for grapefruit. This number depends upon type of fruit, size of fruit and whether the fruit is sold for the fresh market or is used in processing. Curvilinear relationships were also fitted by equations of the form $S=a+b X+\frac{c}{X}$, where $S$ is the average number of fruit per box and X is the average size of fruit. For early-mid oranges the equation is:

$$
S=53.77-1.696 X+\frac{2239.5}{X}
$$

Coefficients for the fresh and processed lines are then weighted together by utilization of the crop (previous season's proportion) to provide a basis for converting average volume of each type to "fruit per box." This method of converting volume to fruit per box also compensates for the deviation from spherical shape in converting circumference to volume.

### 2.5.5.5 Forecasting Yield Per Tree

Two models have been used to combine the components which determine citrus yields: A direct expansion estimator and the relative change estimator. Only the direct expansion estimator is given; that is,

$$
Y=\frac{F \times H}{S}
$$

where

$$
\begin{aligned}
& Y=\text { yield per tree in boxes of fruit } \\
& F=\text { number fruit per tree at time of limb-count survey } \\
& H=\text { proportion of fruit to be harvested } \\
& S=\text { harvest size of fruit expressed in fruit per box }
\end{aligned}
$$

The relative importance of the factors contributing to changes in production is shown in Figure 6 on page 88.

Figure 2: Random Limb Selection With Probability Proportional to CrossSectional Area


Estimated Fruit per Tree
Fruit Count $\times \frac{1}{\mathrm{P}_{1}} \times \frac{1}{\mathrm{P}_{2}}=187 \times \frac{100+90}{100} \times \frac{20+40+50}{20} \doteq 1954$

## Figure 3: Fruit Drop Curves

Extreme Years and Average of 1963-1969 Seasons



Figure 4: Fruit Growth Curves
Extreme Years and Average of 1963-1969 Seasons


Figure 5: Converting Volume to Fruic per Box, Seedless Grapefruit

## Fruit

per Box


Figure 6: Relative Importance of Factors Affecting Average Annual Change in Florida Citrus Production 1960-61 to 1967-68


Early and Midseason Oranges


Seedless Grapefruit

### 2.6.1 Introduction

Crop-growth simulation models which consider the soil-plantatmosphere continuum have only recently been introduced. The impetus to develop crop-growth models involving the plant environment resulted from the successful modeling of photosynthesis. To date such models have been developed for corn, sorghum, cotton, alfalfa, barley, and wheat. The utility of these models has been as crop management and research tools. However, the modification of these deterministic models to forecasting crop yields for large areas requires knowing the plant environment for each day of the entire growing season, as well as detailed knowledge of the plant and how it functions in this environment. These relations are based on how the major plant parts respond each day to their environment.

A brief account is presented of an approach due to Arkin, Vanderlip, and Ritchie for calculating the daily growth and development of an average sorghum plant in a field stand. The appearance of leaves, their growth rate, and the timing of these events are growth characteristics incorporated in the model. It should be clear that the objective is to model the entire plant cycle and not just the reproductive phase of the plant's life. Consequently, the adaptation of these models to forecasting yield requires very exact modeling of the yield components and realistic simulation of the daily climatic inputs for the entire growing season.

### 2.6.2 The Mode1

The physical and physiological processes of light interception, photosynthesis, respiration, and water use are independently modeled and used as submodels. The accumulated dry weight (or yield) for the crop is the product of the plant population and the modeled weight for the "average" plant. Most of the equations describing these processes are empirically derived from field and controlled

A. For Mode1 Without Feedback Data

Plant Data
Leaf number: total number of leaves produced
Leaf area: maximum area of each individual leaf, $\mathrm{cm}^{2}$
Planting Data
Planting date: month, day, year
Plant population: plants/ha
Row width in cm
Row direction in degrees
Climatic Data (daily from planting to maturity)
Maximum temperature, ${ }^{\circ} \mathrm{C}$
Solar radiation, ly/day Rainfall, cm/day

## Soil Data

Available water-holding capacity, cm Initial available water content, cm

Location Data
Latitude in degrees
B. For Model With Plant Feedback Data - for Specific Date(s) During Growing Season

Number of leaves fully expanded
Number of leaves emerged but not fully expanded
Leaf weight
Culm weight
Head weight
Grain weight
Root weight
Soil water
Leaf area index
Leaf area of individual leaves
experiment measurements. The model operates on a daily basis, and, therefore, only daily climatic inputs are required. Other inputs are initialized (i.e., assumed) at the outset of the modeling run. A generalized flow diagram of the model is given in Figure 7. The inputs required are shown in Table 20.

### 2.6.2.1 Seedling Emergence

Seeds imbibe water at very low soil-water contents. Therefore, seedling emergence as calculated is assumed to be dependent on temperature only. Mean air temperature is used in the computations of days until emergence. It was determined that approximately $10^{\circ} \mathrm{C}$ is the threshold soil temperature below which seedling emergence will not occur. The relationship between heat units above the threshold derived from average temperature (i.e., (max $+\min ) / 2$ ) and day of emergence is linear.

### 2.6.2.2 Leaf Number and Area Development

To determine the amount of light (photosynthetically active radiation (PAR) in the .4 to $.7 \mu$ wave band) intercepted by the grain sorghum plant canopy, the leaf area per plant must be known, since the amount of intercepted light is primarily dependent on leaf area. In turn, plant dry-matter accumulation (weight gain), mainly a consequence of photosynthesis, is light dependent. Leaf area per plant is also needed for calculating transpiration when the plant canopy provides only a partial ground cover.

Leaf area development was modeled from inputs of number of leaves produced by the hybrid planted and the maximum area of a leaf. Both field and phytotron studies have shown that the rate at which leaves appear out of the whorl on the grain sorghum plant and the rate at which leaves expand out of the whorl are related to mean daily temperature (or heat units) when plants are adequately watered.

Leaf appearance rate is calculated by summing daily heat units above a base temperature of $7^{\circ} \mathrm{C}$. A new leaf is initiated each time 50 heat units are accumulated. Leaf extension rate is computed in a somewhat similar fashion. Daily leaf area is calculated by summing the new leaf area each day for the expanding leaves and the leaf area of the plant computed the day before. Leaf senescence (death or "firing") results in a reduction of leaf area.

### 2.6.2.3 Canopy Light Interception

Leaves on the plant overlap one another and neighboring plants may shade one another. Thus, not all of the plant's leaf area is actually intercepting light. Shading in the plant canopy is dynamic and changes with the sun's altitude and azimuth and with plant size. To account for these interactions, a mathematical model for computing light interception in a grain sorghum plant canopy was developed. The light interception by a plant in the canopy is computed using a modification of the Bouguer-Lambert equation (commonly referred to as Beer's Law).

### 2.6.2.4 Potential Net Photosynthesis

Potential net photosynthesis, defined as the net $\mathrm{CO}_{2}$ fixed during the daylight hours on a ground area basis for nonlimiting water and temperature corditions, was calculated using relations developed from data obtained from a canopy gas-exchange chamber and simultaneous light interception measurements.

### 2.6.2.5 Daily Net Photosynthesis

A series of efficiency functions which reflect the effects of nonoptimum ambient temperature and soil-water conditions on plant growth are used in the model. Each efficiency parameter is a dimensionless fraction with a value from 0 to 1 . A multiplicative relation is developed for computing net photosynthesis. This expression for net photosynthesis was based on the hypothesis
that limiting variables in the environment proportionately reduce photosynthetic rate regardless of the value of the other limiting variables. Each efficiency parameter represents a particular environment constraint on the photosynthetic rate. Net photosynthesis is computed by multiplying potential photosynthesis by the efficiency parameter for temperature and soil water, and then subtracting nighttime respiration losses.

Mean ambient temperature was used to approximate the crop temperature, because plant temperatures rarely are available and net photosynthesis is relatively insensitive to small differences between leaf and air temperatures.

Reductions in net photosynthesis because of scarcity of soil moisture were considered to be proportionate to the reduction in plant evaporation resulting from limited water availability. Plant evaporation is not affected until a threshold extractable soil water value is reached. The threshold value is dependent on the particular soil and crop under consideration. Extractable soil water is determined daily, using a soil water balance model (based on a modified Penman equation). When approximately 80 percent of the extractable soil water has been depleted by evapotranspiration, net photosynthesis is reduced, because the efficiency parameter becomes less than 1. This relation is speculative and net photosynthesis may be affected more if it is limited by soil-water status more than is evapotranspiration.

### 2.6.2.6 Stage of Development

There are 10 stages which have been found useful in describing the plant development: (1) emergence, (2) three-leaves, (3) fiveleaves, (4) growing-point differentiation, (5) flag leaf visible, (6) boot stage, (7) half bloom, (8) soft dough, (9) hard dough, (10) physiological maturity.

Three stages in particular are important in determining what plant parts are increasing in weight: growing-point differentiation (GPD), half bloom (HB), and physiological maturity (PM). Because leaf appearance and expansion were simulated in the grain sorghum model, phasic development was defined with respect to the appearance of leaves. For example, GPD normally occurs about midway between five leaves fully expanded and flag leaf visible in the whorl. The date GPD occurs was defined as the midpoint between the computed date that leaf 5 (counting from the base) reaches maximum area and the computed date that the flag leaf emerges.

### 2.6.2.7 Daily Dry-Matter Gain

Net photosynthesis (p) is computed and converted to dry matter, using the following relation:

$$
\mathrm{DM}=\frac{12}{44} \times \frac{1}{0.4} \times \mathrm{p}
$$

where DM is dry matter, $12 / 44$ is the ratio of molecular weights of C and $\mathrm{CO}_{2}$ respectively, and 0.4 is the proportion of the plant dry matter which is carbon. The proportions allotted to each organ were empirically derived. However, the absolute amount of dry matter apportioned to a particular organ was dependent on the amount of photosynthate produced that day. The daily allocation of the plant dry matter to the various plant parts is shown in Figure 8 on page 96.
Figure 8: Dry Matter Partitioning to Plant Parts
STAGE OF DEVELOPMENT


It is clear from Table 20 that rather detailed and exacting input data are required to run a crop simulation model. Moreover, the daily climatic data is needed for an entire crop season. Consequently, a forecast of the four daily climatic variables is needed for each day after the seed is planted. These values can be generated by simulating daily values from the empirical distributions of these four variables for a base period of 10 to 25 years. Or, if long-range forecasts of weather are available, the empirical distributions can be modified by shifting the mean vector for forecasted departures from the base period and then used for simulation. In either case, the simulated weather data are substituted for actual climatic data which are yet to occur and the model is run from the date of the forecast to maturity or harvest. In order to reduce the model's reliance on the historically derived parameters in simulating the response characteristics (plant parts), the observed plant parts can be used at key times during the season. Likewise, the observed plant characteristics may be useful in adjusting or correcting the model to agree with the average plant in the commercial field stand for the current year by inputting actual plant data at several times during the growing season. This leads to an additional subroutine in the flow diagram of Figure 7, referred to as "plant feedback" subroutine, which can be made as detailed as it is possible to observe or measure plant parts for an average plant in a field. Some key plant inputs for this purpose are as follows: (1) dry weight of plant, (2) dry weight of head, (3) dry weight of grain, (4) number of leaves, and (5) size of individual leaves. The inputted value replaces the model-computed value for the date of the observation, and the model is restarted on the following day and the daily growth is continued until physiological maturity. Likewise, the daily weather variables which need to be inputted must either be forecast
or simulated from historical records to obtain the yield per plant. These types of modifications in the Arkin, Vanderlip, and Ritchie model were undertaken to make the model more useful for large-area yield estimating.

### 2.6.4 Example of Mode1 Results for Observed Plant Data

A feedback subroutine and the simulation of daily weather variables were developed to aid in forecasting yields for sorghum grown over a rather extensive area. A weather generating model enabling simulation of probable daily weather during the growing season was employed. The generated weather data were derived by a procedure that reproduces the observed historical weather data prior to the current season.

Average field observed plant characteristics for an individual field were used for grain sorghum growth simulation from the date of the feedback (i.e., date plants were observed) to physiological maturity. A sample of the use of the feedback submodel is given in Table 21.

Note that on June 7, the following ground-truth information was fed back to the model: 14 leaves full grown, LAI $^{*}=2$, plant dry weight $=20.05$ grams, head dry weight $=3.69$ grams. The model then simulated both the total plant dry weight and the head dry weight and computed the date of physiological maturity. The observed plant and modeled plant characteristics are shown for comparison. This forecast was made approximately one month before physiological maturity and two months before harvest. LAI was always overestimated, because the senescence submodel was not responsive to limited soil-water conditions.

[^0]Table 21--Plant Characteristics Observed and Simulated by Model

| Date <br> and <br> characteristic | Observed <br> plant <br> data | Model <br> with no <br> feedback | Model <br> with <br> feedback |
| :---: | :---: | :---: | :---: |


| May 3 |  |  |
| :--- | :--- | ---: |
| \# Leaves full | 8 | 14 |
| LAI | 0.83 | 3.35 |
| Plant dry wt (gm) | 2.36 | 16.16 |
| Head dry wt (gm) | 0.00 | 2.22 |

May 18
\# Leaves full
$10 \quad 14$
LAI
1.51
3.16

Plant dry wt
6.03
29.94

Head dry wt
0.00
7.05

June 7

| \# Leaves full | 14 | 14 |
| :--- | ---: | ---: |
| LAI | 2.00 | 2.00 |
| Plant dry wt | 20.05 | 20.05 |
| Head dry wt | 3.69 | 3.69 |

June 24
$\begin{array}{lrr}\text { LAI } & 2.40 & 2.59 \\ \text { Plant dry wt } & 44.92 & 46.44\end{array}$
$\begin{array}{ll}\text { Head dry wt } 21.27 & 17.25\end{array}$

Phys. Maturity

| Day | July 3 | June 3 | July 10 |
| :--- | :---: | :---: | :---: |
| LAI | 1.40 | 2.95 | 2.43 |
| Plant dry wt | 50.70 | 50.05 | 50.04 |
| Head dry wt | 35.70 | 31.93 | 33.05 |


| Emergence | March 15 | March 11 | March 15 |
| :--- | ---: | ---: | ---: |
| Anthesis | June 7 | May 10 | June 7 |

### 2.7.1 Introduction

In Chapter 1 it was pointed out that attempts to employ auxiliary data or double sampling for adjusting crop-cutting surveys to obtain current yields for small geographic regions have been largely unsuccessful. However, the combining of satellite information with appropriate plant or field response data may offer a basis for developing statistical estimators with measurable standard errors for small areas. The technique illustrated is potentially cost effective, because the satellite coverage is for large geographic areas and the field data required are increased only marginally over that needed for largearea yield estimates. Similarly, small-area acreage estimates may be obtained, so that production is derived as a product of yield times acreage. A procedure is described for obtaining corn yield estimates by counties in Illinois during the 1975 crop year.

### 2.7.2 Sampling Methodology

A subsample of corn fields was selected, based on a probability area sample, for the source of the individual corn fields. A subset of these corn fields based on pixels (i.e., picture element equal to approximately 1.1 acre) classified as corn using a quadratic discriminant function was used to develop the yield relations. That is, the crop classification of all pixels is completed first and the yield relation is based on the data for corn fields classified as corn fields by the discriminant function. These fields were then located on the LANDSAT digital tapes and a mean vector derived from the four spectral-channel values for all pixels in each corn field and paired with the forecasted and harvested yield based on objective yield data for these same fields.

[^1]In addition, a mean vector for the four channels was derived for all pixels classified as corn in each county, as well as the mean vector for all pixels classified as corn in the total analysis district of 10 counties. That is, the entire population of pixels is classified by crops for each county as well as the group of counties on the LANDSAT scene. The pixels in the sample corn fields which are also classified as corn are a subset of all acres classified as corn. The mean vectors for the spectral data were obtained from the LANDSAT imagery for August 4, 1975, while the plant and field data relate to a 10 -day period centered on August 28 , 1975. Categorical data from the classified tape were matched to the unclassified tape with the spectral values to derive the LANDSAT information needed for the yield-estimation procedure.

### 2.7.3 Yie1d Estimation Mode1

The yield models are the same as those used for acreage, except the independent variable is now a vector of four channel values. The estimation of the yield for a county or any small area was achieved through a double-sampling regression estimator using the LANDSAT data and a probability sample of fields for the large area comprising the LANDSAT frame. Consequently, it was possible to derive a double-sampling regression estimator using individual fields over a large area and apply the relation to individual counties. Several possible regressions were developed to correspond to variations of the component yield model for several dates. One regression relates yield based on plants per acre on August 1 as the principal variable to the four spectral values from LANDSAT, while the second regression relates the yield based on number of ears with grain per acre as the principal variable to the same four spectral values from LANDSAT. For an early forecast of yield, the grain per plant or per ear would be based on a short-term moving average. The estimated average number of plants (or stalks) per acre derived from a regression for each county is then multiplied by a historical weight of grain per plant to obtain the gross
or biological yield per acre. A second model considered was based on the estimated average number of ears per acre on September 1 for each county, which was multiplied by a historical weight of grain per ear to obtain the gross yield. The average grain weight used was on a per stalk and per ear basis. In this model, stalks and plants have a slightly different meaning, because suckers were counted as stalks. These weights were derived using a transitory moving-average model truncated after five years for Illinois with $a=\frac{1}{2}$. That is, the formula for weight per ear is:

$$
\bar{w}_{E}=\frac{\frac{1}{2} w_{1}+\frac{1}{4} w_{2}+\frac{1}{8} w_{3}+\frac{1}{16} w_{4}+\frac{1}{32} w_{5}}{\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\frac{1}{16}+\frac{1}{32}}
$$

where $w_{1}, w_{2}, \ldots w_{5}$ were the weights per ear for 1974 back to 1969 and $\bar{w}_{E}=.340$.

If $a=\frac{1}{3}$, the weight per ear is .332 and the corresponding formula is:

$$
\bar{w}_{E}=\frac{\frac{2}{3} w_{1}+\frac{2}{9} w_{2}+\frac{2}{27} w_{3}+\frac{2}{81} w_{4}+\frac{2}{243} w_{5}}{\frac{2}{3}+\frac{2}{9}+\frac{2}{27}+\frac{2}{81}+\frac{2}{243}}
$$

However, an alternative weight per ear for individual fields was derived by using a weight estimator based on current-year ear length measurements on September 1 and multiplied by the number of ears per acre to obtain a forecast yield per acre for each field. This yield per acre was actually used to derive the county yield estimator for counties as of the September 1 date.

While a number of different variables or combinations of variables based on the field mean vectors and variance vectors were investigated using the August 1975 imagery in western Illinois,
only two sets of spectral variables gave statistical significance consistently: (1) means of channel 2 and channel 4, and (2) means of channel 2 and channel 4 plus variances of channel 2 and channel 4. The regressions based on data set (1) for September 1 yield forecast and final harvest yield for the 10 -county area within the LANDSAT frame of August 4 are as follows:

September 1 forecast: $\hat{\mathrm{Y}}_{\mathrm{S}}=\overline{\mathrm{y}}_{\mathrm{S}}-8.68\left(\overline{\mathrm{x}}_{2}-\overline{\mathrm{X}}_{2}\right)-2.16\left(\overline{\mathrm{x}}_{4}-\overline{\mathrm{X}}_{4}\right)$

$$
R=.56
$$

Harvest yield:

$$
\begin{aligned}
& \hat{\mathrm{Y}}_{\mathrm{h}}=\overline{\mathrm{y}}_{\mathrm{h}}-10.68\left(\bar{x}_{2}-\overline{\mathrm{x}}_{2}\right)-.56\left(\bar{x}_{4}-\overline{\mathrm{x}}_{4}\right) \\
& \mathrm{R}=.49
\end{aligned}
$$

where $\quad \hat{\mathrm{Y}}_{\mathrm{S}}=$ forecasted corn yield per acre for geographic area on September 1
$\hat{Y}_{h}=$ corn yield per acre for the geographic area at harvest
$\bar{y}_{S}=$ forecasted corn yield per acre for a sample of fields on September 1
$\bar{y}_{h}=$ crop-cutting corn yield per acre for the sample fields at harvest
$\bar{x}_{2}=$ mean spectral value for channel 2 on August 4 for all classified corn pixels in county
$\bar{x}_{4}=$ mean spectral value for channel 4 on August 4 for all classified corn pixels in county
$\overline{\mathrm{X}}_{2}=$ mean spectral value for channel 2 on August 4 for the entire geographic area of 10 counties

$$
\begin{aligned}
& \overline{\mathrm{X}}_{4}= \\
& \text { mean spectral value for channel } 4 \text { on August } 4 \\
& \text { for the entire geographic area of } 10 \text { counties }
\end{aligned} \mathrm{B}_{2} \text { and } \mathrm{B}_{4}=-8.68 \text { and }-2.16 \text { or }-10.68 \text { and }-.56=
$$

The gain in information by use of spectral data for yield estimation may be computed, based on the ratio of variances. For corn yields these information gains are in the range of 1.27 to 1.42. Based on these data sets for western Illinois in 1975, the potential information gain is much less than that for acreage estimation. However, the relation could be improved (i.e., correlation increased) by increasing the number of plots per field. However, the use of the LANDSAT spectral data for both acreage and yield would result in an information gain of approximately $7.0 \times 1.3=9.1$ for estimation of corn production for a single frame or group of 10 counties.

The techniques discussed in this chapter can be grouped into six categories based on the source and type of data employed, as follows: (1) grower opinions or appraisals, (2) plant components and characteristics, (3) agrometeorological relations based on plant and weatherdependent factors, (4) historical ciimate-yield relations, (5) auxiliary environmental variables and yields, and (6) plant growth models.

In general, categories (3) and (4) place greater reliance on historic data over years while (2), (5), and (6) rely on increasingly detailed data and the observance of plant responses within years. Category (1) can be relatively free of both between-year and withinyear relations when the growers are fairly skillful forecasters of the yiel.d of a crop. However, acquiring the needed data to implement the forecasting model in each case can be the key criterion in selecting a technique to employ. Several questions concerning data acquisition need to be answered before making a selection: (1) What is the cost of the data to be collected? (2) Can the needed data be collected in a timely manner to meet the forecast date(s)? (3) What agency(s) has responsibility for data collection? (4) Are the basic "relations" or are trial values of the necessary parameters now available for evaluating the technique? (5) What type of training or staff is needed? (6) Are the variables needed simple data collection tasks or is instrumentation needed in order to use the concepts?

Based on the alternative forecasting techniques discussed, guidelines for these techniques may be set forth when a new program is to be started or a major change is to be made in an existing program. In general, techniques which require long historical data sets are not well suited to a changing or highly competitive agricultural situation. Consequently, systems are preferred that will be valid for forecasting in future years where the dependence can be confined to a short period consisting of the last 3 to 4 years or key parameters can be observed each year. If there are no major trends in yield, then a system which can use information from a series of years to forecast the current year is likely to be valid.

Systems based on techniques falling under (1) or (2) above are preferred as a starting point for estimating and forecasting yields, since harvested yields need to be measured as a basis for evaluating forecasting techniques. Consequently, one of these two will probably be required for this purpose (i.e., determining harvest yields from grower reports or crop cutting). If growers are reasonably skillful in forecasting crop yields and know their production by farms or fields, the use of grower reports from a probability sample can be expected to produce fairly accurate forecasts at reasonable costs in a timely manner. However, there is frequently a tendency to discount a technique based on crop appraisals, because of technical shortcomings reported in some studies due to sampling the wrong population, or no sampling frame, defining the wrong populations to be estimated, lack of agreement with existing production data, and fears that growers are not truthful in their reporting. In many cases, harvested-yield reports have been found to be satisfactory, but production data were unsatisfactory because the harvested area was not known accurately due to biased or erroneous estimates of area planted or harvested. Sometimes the inference is also made that, because local officials or leaders cannot provide timely or reliable reports on yields, growers also cannot provide useful data. Methods that rely on probability samples of growers who report yield and area by fields are probably not used enough. However, if growers cannot provide reliable yield data, then forecasts relying on mature plant or harvested plant components are preferred. Another reason for preferring category (2) in this situation is that modifications can be easily introduced into the model which will utilize weather or environmental variables suggested for categories (3) and (5). Techniques employed from this category generally require training and advance preparation for field work as well as careful derivation of the parameters for the forecasting models. Such a system can serve the market-management needs and provide a basis for measuring changes in crop techniques as shown by changes in yield components in the model over years. For a system to also provide information on the response of the plants during
the crop season to water and other factors, a more sophisticated model based on the ideas of category (6) is required.

About three years are needed to develop and implement an operating program of yield estimates for a crop employing preharvest observations. In fact, if the goal is to have a successful preharvest crop-cutting survey on an operational basis during the third year, a well-planned, intensive effort by experienced crop specialists and mathematical statisticians in yield work is needed.

Typically, the first year's effort would be limited to a small number of fields to obtain preliminary measures of variability for establishing size of plots and other aspects of samplè design, and to develop operational definitions and instructions for the concepts to be used for a pilot survey the next year. Alternative techniques of measuring the yield on small plots would be tried. This would include consideration of various means of locating sample plots objectively and ascertaining the advantages of alternative instruments, equipment or concepts. Potential sources of error or bias would be identified and means of control considered. In addition, a means of estimating harvesting losses based on either sample plots being gleaned after harvest or obtaining production records for check fields is quite helpful. Thus, the goal of the first year's effort is to develop, as fully as possible for trial the following year, a set of sound, detailed operating specifications, including training plans and a well-designed plan for measuring the quality of the work done.

The second year's effort could be regarded as an intensive and extensive pilot operation using a sample that might be one-fifth or one-fourth the size anticipated for a fully operational program. From the second year's experience much better information should become available on variance components and time requirements for various parts of the job, so that the sample design can be optimized. Quality checks on the fieldwork should provide a basis for improvement of field procedures, which must be rigorous and tightly controlled.

The third year would be regarded as the first attempt to implement a program on an operational level. The matters of sampling design would be reviewed and discussed at length; being stressed, however, is the importance of a balanced effort giving rigorous, tightly controlled procedures regarding all important sources of error. Experience has indicated that inherent biases can be eliminated or controlled effectively by intensive training of the field staff, close supervision, quality checks, and providing clear, concise, well-defined field procedures; but astute observation is essential for the identification and control of factors affecting the quality of results. This type of experience must either be found or developed in the early years of a program.

Estimates derived from preharvest sampling are available earlier than estimates from farmers' postharvest reports. Prior to harvest, a farmer can report only his appraisal of the crop prospects. On the other hand, estimates based on preharvest sampling must rely on previous years' harvesting losses or be delayed until such time as harvesting losses can be determined from gleaning sample plots after harvest or commercially harvesting ears and recovering the grain from known numbers of ears.

For tree crops there is frequently a major interest in forecasts several weeks prior to crop maturity. These surveys are substitutes for preharvest sampling or crop cutting when growers' reports on amounts sold for processing (especially when the total crop is harvested within a short period) will be available. The harvested quantities are complicated by the fact that the amount of some crops left unpicked as a result of selective harvesting for tree crops may vary considerably from year to year.

In addition to the advantage of objectivity, preharvest sampling provides a means of getting much valuable information that cannot otherwise be easily obtained. By means of laboratory analysis of samples taken from fields, information on various attributes of crop quality can be made
available. Crop quality, components of yield, and harvesting losses can be related to varieties, cultural practices, weather, harvesting equipment or methods used, and other factors. Also, if deemed worthwhile, information on some types of insect damage, such as the number of ears of corn damaged by corn earworms, can be readily obtained.

The forecasting of the yield of a crop at periodic intervals during a growing season is more difficult than estimating yield at time of harvest. It is necessary to discover plant characteristics or variables which may be used to predict components of yield. Forecast formulas should be based upon observable plant characteristics and a comprehensive knowledge of the fruiting behavior of the crop. The formulas must translate plant characteristics observed on any date into accurate forecasts. These techniques are illustrated for corn, and tree crops in the next chapter. In contrast to the development of a program for preharvest sampling, any time-schedule for developing and perfecting forecasting procedures is much more tenuous. A major reason for this is the necessity of having "between-years' experience" for the formulation and testing of models. In fact, one may continue to use more than one model for a particular crop after a forecasting program becomes operational, in order to give the most promising alternatives a longer test.

Research work on early-season forecasting from plant measurements has been less extensive than for late season. For some tree crops the duration of "late season" is quite long, and "early season" forecasts have not been attempted. Cotton, wheat, corn, soybeans, citrus and nut crops have received the most attention in the development of early-season forecast models.

Growth patterns among different plant species are so varied that not much can be said about a general approach for finding a forecasting model. The nature of the problem obviously changes rapidly with and related to the stage of development. An important aid in developing good hypotheses might be examining existing detailed research or experimental farm data on fruiting and plant characteristics, starting in advance of the first
forecast date and continuing at intervals up to harvest. Such data, however, usually come from isolated and controlled studies and therefore should be regarded as unreliable for purposes of establishing model parameters.

Some crops set fruit over a relatively long period and may have many fruit on a plant with a wide range of maturity. Cotton and lemons are good examples. A forecast of number of fruit, when only part of the fruit is set, requires modification of the fruit component in the model so that a term for additional fruit expected at harvest from fruit not set can be included. For an early-season forecast of cotton, the relationship between "the number of cotton bolls at harvest from fruit not set" and a maturity index has been tried. To establish this relation, fruit set at time of observation must be tagged so that bolls at harvest from fruit not set can be counted. Another model for early-season cotton forecasts, called "the rate of fruiting" model, has been developed. This type of model is more complex and will not be discussed here, but a sigmoid type of growth curve frequently will give satisfactory results for bolls set.

For winter wheat, a May forecast of number of heads is made from stalk counts using a relation established from historical data. Weight of grain per head is related in a somewhat imprecise manner to plant density. Hence, head weight can be adjusted for plant density rather than merely assuming the average for several years, or standard varieties.

It appears that a historical average weight per head or fruit may be a satisfactory basis for a forecast when cultural practices are fairly static; such practices as irrigation and the thinning of fruits are controlled, so the density varies only moderately from year to year. A historical average weight may also be satisfactory if the forecast is for a large area, say several States, so that the average environment and crop practices for the whole area are about the same from year to year even though there may be differences or trends for individual 10calities which vary considerably from year to year.

Knowing the probable quantities of the crop by weight may not supply sufficient data for all needs. For example, in some countries a large portion of the citrus crop is exportable. It is obvious that knowing only the total weight of the expected yield may not be enough. A marketing organization may need information about the quantities qualified for export and the reasons why some of the fruits fail to meet export standard requirements. Once this extra information is available, better planning of the exporting strategy is possible and remedies might be applied in order to increase the quantities qualified for export. Another important by-product of crop forecasting is projections of the average harvest fruit sizes by variety for use in marketing the crop.

These by-products of crop forecasting provide more accurate data about the weight and size of the expected export-qualified crop. At the same time the causes of disqualifications can be pinpointed, classified, and analyzed by type, time of the year, variety and the region so that scientists can attempt (based on feasibility studies) to limit the impact of these damages that can be controlled. Table 22 is a good example of information on damage for several recent Israeli citrus crops that has been found useful in marketing the crop.

Table 22--Distribution of the Different Fruit Damages by Variety and Season (by percentage)

| Type of fruit damages | Shamouti oranges |  |  | Late oranges |  |  | Grapefruit |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $\begin{array}{\|r\|} \hline 1972- \\ 73 \\ \hline \end{array}$ | $\begin{gathered} 1973- \\ 74 \\ \hline \end{gathered}$ | $\begin{gathered} 1974- \\ 75 \\ \hline \end{gathered}$ | $\begin{array}{\|r\|} \hline 1972- \\ 73 \\ \hline \end{array}$ | $\begin{gathered} 1973 \\ 74 \\ \hline \end{gathered}$ | $\begin{gathered} 1974- \\ 75 \\ \hline \end{gathered}$ | $\begin{gathered} 1972- \\ 73 \\ \hline \end{gathered}$ | $\begin{gathered} 1973- \\ 74 \\ \hline \end{gathered}$ | $\begin{gathered} 1974- \\ 75 \\ \hline \end{gathered}$ |
| Natural causes* | 25 | 34 | 20 | 16 | 34 | 25 | 19 | 29 | 25 |
| Shape of fruit | 18 | 28 | 29 | 15 | 15 | 18 | 18 | 27 | 27 |
| Physiological damages | 23 | 15 | 18 | 36 | 26 | 29 | 23 | 16 | 13 |
| Insects | 15 | 10 | 14 | 21 | 11 | 14 | 28 | 13 | 20 |
| Green fruit | 11 | 8 | 7 | 4 | 9 | 6 | 1 | 4 | 3 |
| Picking damages | 7 | 4 | 11 | 7 | 4 | 7 | 6 | 6 | 8 |
| Diseases | 1 | 1 | 1 | 1 | 1 | 1 | 5 | 5 | 4 |

[^2]
### 3.1 Introduction

The data-collection needs for yield measurement and forecasting can be considerable and exacting if objective data is to be provided on a uniform basis by different workers and over years. Information collected in this manner is also quite valuable both in evaluating the transfer of agricultural technology from the researcher to the farm and in meeting crop production goals by developing countries. For developing yield relationships involving plant and environmental variables, the joint participation of several agencies in the data collection effort can be difficult, due to different objectives as well as timing priorities in collecting and releasing basic data.

The value of crop yield statistics is dependent on being able to collect data in such a manner that the same statistical concepts can be accumulated or made additive over broad areas to represent an entire country or region. Consequently, a careful plan of operation encompassing a definite timetable for planning, training, and all data collection phases is extremely important. The data needs depend on the different demands which the yield modeling imposes.

Following the illustrations of the operational data-collection concepts, an actual model along with the survey statistics is used to calculate specific yield forecasts. In addition, alternative models are postulated and yield forecasts made not only for comparison purposes, but also to suggest that approximately the same forecasts or preharvest estimates are frequently obtained when using the same sample data in different but appropriate models. It should not be inferred. that many different models or forecasts should be used in preference to using a single model which is based on realistic estimates of parameters (i.e., operational concepts can be defined) for which representative sample data can be made available. The emphasis in making the choice af model should be based on the field workers being able to collect the desired data in the prescribed manner and the ability to validate all or most of the model parameters each year.

### 3.2.1 Introduction

The sample design and size must be planned to give estimates for the desired geographic areas at an acceptable level of reliability. Most yield surveys are stratified by major geographic areas or political divisions. Where possible, many surveys are also substratified by major varieties, crop type, age of trees, or irrigated and nonirrigated lands. The total sample size is usually set to control the errors desired for the primary strata or geographic areas for the plant or crop characteristics at harvest. The sample sizes for the other levels of stratification are usually made proportionate to the area planted to the crop. When the allocation is proportionate to area in the crop or number of trees, the sample data on a per acre or per tree basis is selfweighted. This self-weighting feature is desirable for summarizing the data as well as examining the yields by alternative areas other than the initial strata. Where the information on crop area or production is not available by strata and substrata, farm numbers or frame sampling units must be used in the survey design. In most practical cases, several sampling stages and a number of sampling units will be used within strata. If the strata are large political or administrative divisions, a sample of districts within these divisions might be selected at the first stage and a sample of subunits within the districts at the second stage. Villages with identifiable boundaries that account for all the land within their boundaries can serve as suitable units at some stage of sampling. The ultimate unit at the third or lower stage will be the individual holding, field or parcel having the crop planted or for harvest.

### 3.2.2 Selection of Farm Holdings and Fields

The following examples illustrate some procedures that can be used to select farm holdings and fields in the final stages of the sample design.
(1) Farm holdings can be selected from lists, if lists are available or can be constructed. Lists of farm holdings for individual crops would be needed only for the units (villages, subdistricts, etc.) actually selected in the sample at the preceding stage; if necessary, these could be compiled as part of the field operation. The selection of holdings can be made either with equal probabilities or with probability proportionate to size (assuming that information on size is available or can be obtained). The measure of size might be total land, or cultivated area in the holding, but preferably total area planted in the particular crop for which the yield was to be determined.

Similarly, within each selected holding, a list of fields would be compiled and a sample field(s) selected. Again, selection would be made either with equal probabilities or with probability proportionate to size of the area in the crop of interest.
(2) If maps or aerial photographs are available, these can be used to select fields directly without first selecting holdings. One way to do this is to superimpose on the map or photo a grid on which dots have been placed either in a systematic pattern or at random; each field into which a dot falls is then included in the sample, thus giving the fields (and holdings) probabilities of selection proportionate to their sizes. This procedure requires, of course, that the maps or photos be sufficiently detailed so that the point and the corresponding field can be located on the ground. (This procedure is not easily adaptable
to estimating number of holdings, if that is desired, since it involves identifying the holder and determining the total land in the holding so that the proportion of the selected field to the total holding is known.)

Area segments are very useful sampling units for determining which holdings and/or fields are to be included in the sample. The segments may be constructed either with natural boundaries that can be located on the ground or with imaginary boundaries drawn on a photo or map; the choice depends upon the particular situation. Holdings and/or fields may be associated with area segments in any of the following ways:
(a) Area segments with imaginary boundaries could be used as first-stage sampling units and a sample of segments selected; within the sample segments, fields could be selected as second-stage units in the manner described above in (2).
(b) An alternative procedure would be to include in the sample all fields (or holdings) for which a uniquely defined point falls within the segment boundaries. With this procedure, fields (or holdings) would not be selected with probability proportionate to their sizes; the probability of selection would be the same as the probability of selection of the segment into which the point falls. This is known as an open-segment approach. The segments determine which units are included in the sample, but data are tabulated for some fields (or holdings) lying partly outside the segment, and are not tabulated for other fields (or holdings) lying partly inside the segment when the corresponding unique point falls outside the selected segment.

The unique point must be defined with care. Usually a particular corner of the field (holding) would be designated as the unique point. Because fields (holdings)
may not be rectangular, a specific rule for locating this corner would be needed as well. For example, if the northwest corner were the designated unique point, it could be defined either (1) by identifying the boundary points that lie farthest west and then designating the most northern of these points as the northwest corner, or (2) by identifying the boundary points that lie farthest north and then designating the most western of these points as the northwest corner. If the holding were the unit of analysis, the residence of the holder (provided all such residences had a chance of being included in the sample) would generally be preferred as the unique point, since it would be the easiest point to locate. A combination of rules is, perhaps, even more useful. For example, the residence of the holder might be used when the holder lives on the holding, and a particular corner used when he does not live on the holding. In any case, the point must be defined in a way such that it is truly unique (that is, each unit must have one, and only one, chance of being included in the sample); it should also be fairly easy to identify.
(c) If the unit of analysis is the farm holding, the weightedsegment approach will usually be more efficient than the open-segment approach, but this costs more per unit to enumerate. With this procedure, all holdings having any land in the segment are included in the sample and hence must be contacted. In the estimation, the data from each holding are weighted by a factor based on the proportion of the entire holding lying inside the segment. In almost all applications, the weighted-segment approach requires that the segments have natural boundaries that can be identified on the ground.
(d) Still another possibility is to use the so-called closed-segment approach, in which only those fields or parts of fields lying within the segment are included in the sample. One advantage of this procedure is that it avoids the difficulty of having to define the holding. The fields in the crop of interest may be identified by observation, hence it may not be necessary to contact the holder or farm operator. Of course, if yield information is desired on a farm unit basis, the closedsegment approach is not appropriate since some farms or holdings will certainly extend beyond the segment boundaries.

### 3.3.1 Introduction

In the preceding chapters it was assumed that the area standing in the crop could be determined from planted or reported land areas in a manner consistent with the area of the crop harvested. When the appropriate area figure cannot be, or has not been, derived through a questionnaire, then special procedures must be employed to define the area that corresponds to the area occupied by the crop to be harvested, so that production can be obtained by multiplying area times yield. If the gross area planted to the crop was available from a crop survey, this area could be adjusted to obtain the net area standing in the crop. However, if the growers who grew the crop were known but were unable to report the area planted for individual crops, then the area occupied by the crop must be measured. For interplanted or mixed crops, the gross area planted to all crops constitutes the area occupied by the crop of interest.

### 3.3.2 Deriving Net Area From Gross Area Planted

The acres for harvest can be derived in many cases for the sampling unit and individual fields as is shown in Table 23, page 128 of this chapter. In cases where column 4 is greater than column 5, the area which will not be harvested must be eliminated from the area where sample plots (or plants) are located. This is generally relatively straightforward identification for the grower or by inspection for fields planted to a single crop, but is more difficult for interplanted crops. For interplanted crops, the harvested area for the crop of interest in the yield survey would not be reduced unless the gross area planted to the combination of crops is reduced by a similar amount. That is, the gross area standing for harvest for the combination of crops planted should be used as the harvested area for both crops unless
the area is void of all plants of both crops. If the yield surveys are based on a subsample of sampling units, several alternative estimators of the area for harvest would be considered. A ratio or difference estimator would be used to estimate the area for harvest as a percentage of, or reduction in, the planted area estimate. If all the sampling units used to estimate planted area are included in the yield survey, the harvest area figure will be estimated in the same manner as the original planted area.

### 3.3.3 Deriving Net Area When Planted Area Is Not Known

In this case, the growers with the crop of interest have been identified in an earlier agricultural survey or will need to be identified during the first phase of the yield survey. The fields used for the yield survey will be based on selecting a probability subsample of farms or growers (identified during the first phase of the yield survey) with the crop for which the yield plots are being observed. If the selected growers have more than one field or parcel, only one will be selected at random with a known probability. Frequently, the grower may know only the number of fields planted to the crop or possibly only the number of parcels with the crop (a parcel being a cleared or cultivated area planted to one or more crops, which may include grain crops, root crops, and a home garden). For the selected field or parcel the area to be harvested must be determined either by the grower or enumerator by direct measurement of land area. Generally, this means using plane-surveying techniques, including measurement of distances, angles, differences in elevation, and a sketch drawn to a suitable scale of the area under the crop (or the combination of crops in the case of interplanted crops). The area measurements need to be made rather precisely, but these methods usually require only limited training based on techniques involving a measuring tape, standardized cord, Smith's wheel, topofil, rangefinder compass or a sighting device, without fear of introducing any large systematic errors in the area measurements. The net area for harvest is measured and identified on the sketch of the area.

### 3.4 Yields From Crop-Cutting Surveys

Generally, the data-collection needs and problems are easier and fewer for crop-cutting or preharvest surveys than for early-season forecasts. If the yield procedures are to be evaluated or the quality of field workers is to be assessed, the data-collection requirements are somewhat increased. Prudent survey management requires that both of these be undertaken periodically on a subsample basis, but they are generally mandatory whenever a new program is started. Certain additional information will be needed or at least highly desirable from a preharvest survey if forecasting is to be undertaken for the same crop.

If validation, for example, is to be a part of a corn crop-cutting survey, the collection of information on number of ears and the recovery of weight of grain per ear may be necessary. For example, situations may arise wherein it is necessary to determine if (1) harvesting ears by hand from small plots results in a greater number of ears per acre than that obtained by commercial harvesting equipment, or (2) removing grain from ears using a hand sheller results in a greater weight of grain per ear than that obtained by commercial shelling equipment. The specific data needed to resolve such doubts depend on the survey procedures and the commercial harvesting practices. A second set of questions (or check items) may need to be formulated to determine if the survey definitions and procedures are being followed by the field workers.

To insure that the crop cutting (or objective-yield forecasting surveys) can be carried out in a timely and efficient manner, the total program must evolve over a period of months. The following 10 items are the major steps which normally should be spread over a 6 -month period to insure proper execution, but in an emergency these steps might be completed in a 3 -month period by an experienced data-collection staff.
(1) Determine plant and plot characteristics and measurements that will be needed.
(2) Order new or replacement equipment and supplies.
(3) Prepare forms for field-plot and laboratory work.
(4) Prepare training materials.
(5) Obtain results of acreage surveys to prepare acreage estimates and select sample fields.
(6) Conduct training school for collection of plant data:
a. Cover field-work instruction manual.
b. Present slides of important field tasks and discuss data concepts.
c. Demonstrate plot work in the field.
d. Give practical experience to workers using field forms.
(7) Conduct survey - calendar dates (i.e., Oct. 7-21).
(8) Review daily the completed forms (by field supervisors).
(9) Process plant parts in the laboratory.
(10) Transmit or transfer completed forms to data-analysis unit.

The following summary form, Exhibit A, shows the data-collection concepts derived from the crop-cutting survey for one field where validation work is planned, such as reported in Table 8, page 54. The summary form permits a comparison of the individual yield components as well as verifying whether the composite differences in harvesting procedures are accounted for by the postharvest gleaning work. Most of the data-collection techniques employed are illustrated in the next section. In the case of very large fields, it may be desirable to subdivide the field into smaller subfields for sampling purposes and restrict the commercially harvested area so each phase of the field work can be completed in one day.
$\qquad$

| $\begin{gathered} \text { Line } \\ \text { No. } \end{gathered}$ | Item and Computation |  |
| :---: | :---: | :---: |
| 1 | No. samples harvested |  |
| 2 | No. ears husked per 30 ft of row (15-ft unit in 2 adjacent rows) |  |
| 3 | No. ears with grain husked |  |
| 4 | Ears with grain per 30 ft of row (1ine $3 \div 1$ ine 1) |  |
| 5 | Field weight of ears with grain |  |
| 6 | No. reports of moisture content |  |
| 7 | Average shelling fraction |  |
| 8 | Average motsture fraction |  |
| 9 | Average dry-matter fraction (1.000 - line 8) |  |
| 10 | Average field weight per 30 ft (line 5: 1) |  |
| 11 | Average field weight per ear (line $5 \div 3$ ) |  |
| 12 | Average weight of grain per 30 ft at $15.5 \%$ noisture (line $10 \times$ line $7 \times$ line 9) $\div(.8 \div 5)$ |  |
| 13 | Average weight of grain per ear ar $15.5 \%$ moisture (line $11 \times$ line $7 \times$ line 9$) \div(.845)$ |  |
| 14 | Conversion factor to gross yield per acre (25.929 \% average row spacing) |  |
| 15 | Gross yield per acre (line 1 : : line 14) |  |
| 16 | Gross ears per acre <br> (line $3 \times$ line 14) |  |
| 17 | Weight of grain per ear (line $15 \div$ line 16 ) |  |
| B - Pos | arvest Gleaning of Grain |  |
| 18 | Average weight of grain per 30 ft of row |  |
| 19 | Average moisture content |  |
| 20 | Average weight of grain per 30 ft of row at $15.5 \%$ molsture |  |
| 21 | Conversion factor to grain left in field per acre (. 02858 : row width) |  |
| 22 | Total grain per acre left in field (line $20 \times$ line 21 ) |  |
| C - Net | feld from Crop Cutting |  |
| 23 | Preharvest net yleld per acre (line 15 - line 22) |  |
| D - Dat | from Commercial harvest |  |
| 24 | Total ears in equipment bin taken from field |  |
| 25 | Total pounds of ear corn in equipment bin taken from field |  |
| 26 | Total pounds of shelled corn recovered |  |
| 27 | Moisture content of grain |  |
| 28 | Total pounds of ear corn at $15.5 \%$ moisture |  |
| 29 | Total pounds of shelled corn at $15.5 \%$ motsture |  |
| 30 | Net acreage harvested (as measured) |  |
| 31 | $\begin{array}{ll}\text { Pounds of corn per acre } & \text { a. Shelled } \\ \text { (1lne } 24 \text { bine 26) } & \text { b. Har corn }\end{array}$ |  |
| 32 | Bushels per acre a. if shelled (line $31: 56$ ) <br> b. If ear cotn (line $31: 70$ )  |  |
| 33 | $\begin{aligned} & \text { Number ears per acre } \\ & (11 \text { ne } 24: 1 \text { Inc } 30) \end{aligned}$ |  |
| 34 | $\begin{aligned} & \text { Welight of rain per ear } \\ & \text { (line } 32: \text { line } 33 \text { ) } \\ & \hline \end{aligned}$ |  |
| E - Yield difference |  |  |
| 35 | Preharvest net yleld minus comerclal harvest yichd (1tne 23 - 11 ne 32) |  |

### 3.5 Forecasting Corn Yields From Plant Parts

A forecasting model which has been commonly used is based on counts or observations of the individual plant parts, because the datacollection concepts involve known yield components. These components can be identified without additional research to determine what pertinent variables are needed to forecast yields, since agronomists and other agricultural scientists have already identified the basic components. Of course, alternative yield models can be formulated which would require initial research to identify critical factors or the correct time for the scheduling of data-collection activities that could lead to a superior model. However, the choice of this type of model is based on identifying an initial model which can provide useful results with no practical risk of selecting an unworkable model.

The inventory-component type of model may be formulated in several ways involving only minor differences in the components used. For corn a very basic model with several variations would be:

Model (1) Yield per hectare $=$ plants per hectare $x$ weight of grain per plant,
or Model (2) Yield per hectare $=$ ears per hectare $x$ weight of grain per ear,
or Model (3) Yield per hectare $=$ ears with grain per hectare $x$ kernels per ear with grain $x$ weight per kernel.

The components in the above models can be verified at harvest, so the validity of each component can be evaluated.

If three forecasts were to be made prior to harvest, perhaps all three models might be used: model (1) about 90 days prior to harvest, model (2) about 60 days prior to harvest, and model (3) 30 or fewer days prior to harvest. Assuming use of one of the variations in this type of model, the data-collection requirements are given in Exhibit B where corn is planted in rows. The listing and selection of fields for
a sampling unit are given in sections 3.5.1 and 3.5.2. A variation in the procedure for determining number of plants per hectare would be required if plants are planted in an irregular manner.

Exhibit $B$ could be used at any time after emergence, but the form would be materially shortened if only one of the model variations was to be used on a given occasion. For instance, model (3) might be used 60 days prior to physiological maturity by assuming a norm or historical weight per kernel. In this case, the key data items would be 7 and 11 , with item 10 providing an alternative basis for forecasting weight of grain per ear. The weight forecast might be based on developing a linear relation between kernel-row length and harvest weight of grain per ear. It should be clear that similar reductions in the data items to be collected could be made for a specific single-date forecast.

Yield forecasts based on agrometeorological models likewise would use only a very limited amount of the information in Exhibit B, but would require environmental data from another source. However, the verification of the forecasts would require that some data be collected either at physiological maturity or at the time of commercial harvest. Certainly, the field work to collect plant data would be less frequent and greatly reduced if repeated forecasts during the season were not needed.

The information in Exhibit B permits several different ways each model could be used during the season, and the particular variables adopted might be determined either as a result of a pilot study or previous experience of agriculturalists in the area. Table 22 shows the components and how they might be used in different variations of a forecast model.
gaisit bi meld dain foma

|  |  | 4 |  |
| :---: | :---: | :---: | :---: |
| ¢ |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

> and aumber of karnela for eelected rou (b) .........
11. Number of rown with graia (a)
10. Averase leagth of hergel
rove................. Inches is Teethe rove.................... Inches 6 Teothe MRASUREHENTS UITHIN UNIT 22 ROY 1
12. Measure length and circumference of cob for sill eare coimted in Itell?
Do NOT rembve "ear" or pull back hisk. Record to nearsest $1 / 10$ inch.
If more than 2\& eare, use blank space on right.
right.
Len. cir.
--

-

1. Lea. Cir. , Len. Cir. 13 . Len. Cir.
2. -elected row (b)
for Unit 2, Rou 1.

Observe fule 1 and Rule 2 when laying out the sample varite.

$\qquad$



| silked ear shoota............................. |  |  |  |  |  |
| :--- | :--- | :--- | :--- | :--- | :--- |
|  |  |  |  |  |  |





## 


Maturity of 5 ears or
allked ear shoote .....

If total in_Itien a is is or lese, skip Iteme 9 through 13.
13 or nore, continue. (If any ears in Item 8 are code 2, ruplaoe
code 2 ear with a Code 3 -ear und enter in Item 9.1

Doee Iten ì have 3 or mors code? ears? $\square$ YES, if yee oarmplete Itame 13 \& 14 only.
$\square$ No, If no complate Iteme 10 to 12 cmly .

Table 22--Components and Forecast Parameters

| Time of Season | Component/Source | Forecast Parameter |
| :---: | :---: | :---: |
| 90 <br> days <br> prior <br> to <br> harvest | Plants per hectare <br> Item 3 <br> Weight of grain per plant Items 14, 15 at harvest <br> Items 6, 3 <br> Item 3 <br> Items 14, 15 at harvest | Plants per hectare observed <br> (a) Historical norm for area (or variety) <br> (b) Number ears per plant observed $x$ historical norm for grain per ear for area <br> (c) Seasonal prediction based on linear regression of grain per plant Per hectare or per plot (a historical regression equation) |
| 60 <br> days <br> prior <br> to <br> harvest | Plants per hectare <br> Item 3 <br> Ears per hectare <br> Item 6 <br> Ears with grain per hectare Item 7 <br> Weight per plant <br> Weight per ear <br> Item 12 <br> Items 14, 15 at harvest <br> Weight of grain per ear Item 11 (a) <br> Item 11(b) <br> Items 14, 15 at harvest <br> Item 10 <br> Items 14, 15 at harvest | Plants per hectare observed <br> Ears per hectare observed <br> Ears with grain per hectare observed <br> (a), (b), (c) above <br> Ears per hectare observed $x$ seasonal prediction based on linear regression of grain per ear on ear size measured, length x circumference (a historical regression equation) <br> (d) Kernel rows per ear x kernels per row observed $x$ historical norm for grain weight per kernel <br> (e) Seasonal prediction based on linear regression of grain per ear on <br> length of kernel row per ear (a historical regression equation) |
| 30 days or fewer to harvest (physiological maturity) | ```Ears with grain Item 7 Kernels per ear w/grain Item ll Weight per kernal Items 14, 15 Weight grain per ear Items 14, 15``` | Observed <br> Observed <br> Observed and adjusted to standard moisture content <br> Observed and adjusted to standard moisture content |

### 3.5.1 Listing of Crop Fields for Area Sampling Units

A sample of fields is selected from a probability area survey of crop acreages within each region or State based, on the closedsegment concept. The farm tracts and fields with the designated crop are selected with probabilities proportionate to the expanded acreage of the designated crop, hence the sample will be selfweighting if a constant number of plots is selected in each field. The sampling unit is a farm tract with the designated crop and all the fields planted to that crop.

Table 23 is completed for the desired crop only by entries in columns 2 through 5.

Column 2 - The VARIETY planted is recorded in each field. A field should not consist of more than one variety. (Varieties are ignored in this example.)

Column 4 - Acres actually PLANTED are obtained in each field. Exclude acres in roads, ditches, rockpiles and other nonplanted areas.

Column 5 - Acres for HARVEST are obtained. in each field. Exclude acres already abandoned or otherwise not intended for harvest.

Column 5 - HARVESTED acres are accumulated, field by field, to a total for the entire sampling unit.

The accumulation is obtained by adding the acres for harvest in the top line for each field to the previous accumulated entry. Accumulated acres for last field will always equal the total acres for harvest in the entire sampling unit.

Table 23--Sampling Unit Data

| Field no. | Variety | Office use | Acres planted | Acres for harvest Accumulated |
| :---: | :---: | :---: | :---: | :---: |
| (1) | (2) | (3) | (4) | (5) |
| 1 |  |  | 15.0 | Accum. $\frac{15.0}{15.0}$ |
| 2 |  |  | 20.0 | Accum. $\frac{20.0}{35.0}$ |
| 3 |  |  | 10.0 | Accum. $\frac{10.0}{45.0}$ |
| 4 |  |  | 13.0 | Accum. $\frac{13.0}{58.0}$ |
| 5 |  |  | 23.0 | Accum. $\frac{23.0}{81.0}$ |
| 6 |  |  | 15.0 | Accum. $\frac{15.0}{96.0}$ |
| 7 |  |  | 10.0 | Accum. $\frac{10.0}{106.0}$ |
| 8 |  |  | 17.0 | Accum. $\frac{15.0}{121.0}$ |
| 9 |  |  | 7.0 | Accum. $\frac{6.0}{127.0}$ |
| 10 |  |  | 5.0 | Accum. $\begin{array}{r}5.0 \\ 132.0\end{array}$ |
| 11 |  |  | 115.0 | Accum. $\frac{5.0}{247.0}$ |
| (12) |  |  | 65.0 | Accum. $\frac{63.0}{310.0}$ |
| 13 |  |  | 87.0 | Accum. $\frac{87.0}{397.0}$ |
| 14 |  |  | 120.0 | Accum. $\frac{120.0}{517.0}$ |
| 15 |  |  | 150.0 | $\text { Accum. } \frac{145.0}{662.0}$ |
| (16) |  |  | 160.0 | Accum. $\frac{152.0}{714.0}$ |
| 17 |  |  |  | Accum. |
| 18 |  |  |  | Accum. |
| 19 |  |  |  | Accum. |
| 20 |  |  |  | Accum. |

The total acres (last accumulated entry) for harvest on the land in the area unit is...... ACRES
IS THAT RIGHT?
(a) ( ) NO -- Review all fields, correct Table 23, col. 4.
(b) ( ) YES -- Make selection of sample field(s).

If column 2, Table 24 is zero, conclude interview.

A sample field must be selected for each sample number listed in Table 24 on the next page. The sample number and selected acre for each sample have been entered by the statistical office. For each of these samples, observations will be made and ears will be harvested for the two separate units when mature.

The sample number and selected acre will determine in which field(s) the sample(s) will be laid out. Large fields may have more than one sample selected for the field. If only one field is listed in Table 23, that field will automatically become the sample field if a selected acre is listed in Table 24.

## To select the sample field:

a. Select the first field in Table 23 in which the accumulated harvested acres equal or exceed the selected acre for the sample number listed for the sampling unit.
b. Enter selected field number in Table 24.
c. Circle the selected sample field number in Table 23.
d. For the additional sample shown in Table 24 , repeat steps $a, b$, and $c$ above.

The example on the next page shows that two samples will be laid out for the sampling unit. Select the field for sample no. 24 first--this will be the first field listed in Table 23 for which the accumulated acres equal or exceed 295.

Now select the sample field for sample 25. The selected acre is 670 and the first field for which the accumulated acres equal or exceed the selected acre is field No. 16. Enter this number in Table 24 on sample 25. Circle the field number in Table 23 on sample 25.

Table 24--Selection of Sample Fields on Farm

| Sample <br> number | Selected <br> acre | Selected <br> field number |
| :---: | :---: | :---: |
| 24 | 295 | 12 |
| 25 | 670 | 16 |
|  |  |  |

At this point the field enumerator is ready to go to the field to collect the data shown in Exhibit B.

### 3.5.3 Selection of Units Within Field

The enumerator proceeds from the point of interview to the sample field. The work proceeds in stages, starting with the layout of the field units, recording the various counts and observations, and perhaps (destructively) sampling several ears or plants depending on the model. Not all the data may be obtained at each visit, since the stage of development of the plant will determine what information is appropriate or obtainable. The units are located by use of the random-row and pace numbers entered at the top of the form. Figure 9 illustrates some of the key steps in laying out the unit.

DRILLED


STEP NO. 1:
After the last pace into the field, place dowel stick perpendicular to rows. Anchor $50-\mathrm{ft}$ steel tape just beyond the dowel in row 1 . Insert a florist stake by the anchor.

STEP NO. 2:
Insert a starting florist stake EXACTLY 5 feet from the anchor and an ending stake EXACTLY at the 20-foot mark with flat sides at right angles to the row direction.

STEP NO. 3:
Repeat step 1 for row 2, except that no florist stake should be inserted at row 2 anchor.

STEP NO. 4:
Repeat step 2 for row 2.

STEP NO. 5:
Tie a 4 -foot plece of flagging ribbon near the top of the first plant inside the unit in row 1 and across the row middle to the first plant in row 2 of each unit.

RULE 1: If a plant emerges from the ground exactly at the starting stake, INCLUDE that plant in the unit. INCLUDE the entire hill if any plant in a hill is included at the starting stake.

STEP NO. 6:
Tie a 4 -foot piece of flagging ribbon near the top of the last plant inside the unit in row 1 and across the row middle to the last plant in row 2 of each unit.

RULE 2: If a plant emerges from the ground exactly at the ending stake, EXCLUDE that plant from the unit. EXCLUDE the entire hill if any plant in a hill is excluded at the ending stake.
3.5.4 Concepts for Collection of Plot Data

After the plots have been laid out, the pertinent data in Exhibit B is recorded for each plot.

1. Measure distance from stalks in row 1 to stalks in row 2....................... Feet \& tenths

| Unit | Unit 2 |  |
| :--- | :--- | :--- |
|  |  |  |

At the dowel stick, measure the distance across the first row space with the steel tape. Anchor at the center of the stalks in the first row in the unit and measure to the center of the stalks in the second row in the unit. This is the distance across the first middle. Record this distance in feet and tenths of feet.
2. Measure distance from stalks in row 1 to stalks in row 5...................... Feet \& tenths


Measure the distance across 4 corn row spaces (5 adjacent rows) and record in item 2. You should measure at the dowel stick from the center of the stalks in row 1 , to the center of the stalks in row 5. All measurements will be made with the tape in feet and tenths of feet. See example on page 134.

NOTE: Items 1 and 2, (row space measurements) should be made only on the first visit, or if the units are relocated on later visits.

In the event the field is "skip planted" so that there are several rows of corn and then several rows of a second crop, record the planting pattern in the margin. For example, if the planting pattern is 2 rows corn, then 2 rows soybeans, the measurement recorded in item 2 is the sum of the distances between two rows of corn in four different strips. Apply the same principle if corn is planted in strips of three or four
rows. If corn is planted with one narrow middle and one wide middle, (example: a 7 -inch middle followed by a 40 -inch middle) the measure recorded in item 2 is the sum of 2 narrow middles plus 2 wide middles.

In all cases of unusual row spacing (very narrow or wide row spacing), write an explanatory note in the margin of the form.

MEASURE DISTANCE FROM STALKS IN ROW 1 TO STALKS IN ROW 2: At the dowel stick, anchor the tape at the center of the stalks in row 1 of the unit and measure to the center of the stalks in row 2 of the unit. Record in feet and tenths of feet.

MEASURE DISTANCE FROM STALKS IN ROW 1 TO STALKS IN ROW 5: At the dowel stick, anchor the tape at the center of the stalks in row 1 of the unit and measure to the center of the stalks in row 5. Record in feet and tenths of feet.

> Row measurements must be made with a tape calibrated in feet and tenths of feet.

Row-Space Measurement

3. Number of stalks...............

| Unit 1 |  | Unit 2 |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Row 1 | Row 2 | Row 1 | Row 2 |  |
|  |  |  |  |  |

Count all stalks in each 15 -foot row inside the unit, regardless of size or condition. Do not count tillers (suckers) as stalks. An important identifying characteristic of a tiller or sucker is that it emerges from the ground close to the main stalk, often at a slight slant. Other features are the generally smaller size of the tiller as compared with the main stalk, and usually the lack of brace roots on the tiller. A main stalk and its tillers come from the same seed (see illustration, page 141).

If you continue to be uncertain as to whether it is a tiller, go outside the unit and find a similar plant. Dig it up to determine whether it is a stalk or a tiller.

Any volunteer stalks growing in the row space between row 1 and row 2 are to be included in the count for row 1. Likewise, stalks between row 2 and row 3 should be included in the count for row 2.

Late in the growing season, after the seeded crop has matured, mature seed may fall to the ground and germinate. Any volunteer plants which come from the current year's crop should be excluded from the plant.
4. Number of stalks with ears or silked-ear shoots........

| Unit 1 |  | Unit 2 |  |  |
| :--- | :--- | :--- | :--- | :--- |
| Row 1 | Row 2 | Row 1 | Row 2 |  |
|  |  |  |  |  |

Count the number of stalks in item 3 having ears or silked-ear shoots on the main stalk, or if none on the main stalk, on a tiller from the main stalk. A silked-ear shoot is the early formation of an ear on stalk with some silk protruding beyond the husk. Item 4 cannot be greater than item 3, "total stalks."
5. Number of stalks with ears showing evidence of kermel formation......

| Unit 1 |  | Unit 2 |  |
| :--- | :--- | :--- | :--- |
| Row 1 | Row 2 | Row 1 | Row 2 |
|  |  |  |  |

Count the number of stalks in item 4 having ears in which the kernels have definitely begun to form. To have evidence of kernel formation, ears must be in BLISTER or later stages of maturity. Item 5 cannot be greater than item 4 (stalks with ears or silked-ear shoots). Make item 5 counts in row 1 of each unit.

Do not remove or pull back the husks of ears within the unit to inspect for kernels. Outlines of the kernel rows may be felt through the husks, or kernels may be seen at the top of the cob. See page 141 for a description.
6. Number of ears and silked-ear shoots.......

| Unit 1 |  | Unit 2 |  |
| :--- | :--- | :--- | :--- |
| Row 1 | Row 2 | Row 1 | Row 2 |
|  |  |  |  |

This count will include all ears and all ear shoots on which there is visible evidence that silks have emerged beyond the husks. Only one ear or ear shoot is to be counted from each node. A node is a fruiting position on the stalk. Do not count an ear shoot from a node which has an ear. Ears and silked-ear shoots on tillers (or suckers) are to be included in this count. In cases where a considerable period of time may have elapsed since silking, kernel formation may be taken as evidence of silking, even though silks are no longer visible.
7. Number of ears with evidence of kemel formation...............

| Unit 1 |  | Unit 2 |  |
| :--- | :--- | :--- | :--- |
| Row 1 | Row 2 | Row 1 | Row 2 |
|  |  |  |  |

This is to be a count of all ears in which kernels have definitely begun to form. An ear of com is defined as a cob having at least one kernel. Ears on tillers should be included in the count. To have evidence of kernel formation, ears must be in BLISTER or later stages of maturity. Ears will have
started to enlarge and will have a solid "feel" to them. Most silks protruding from the husks will be turning color or may be brown or dry.

Outlines of the kernel rows may be felt through the husks, or kernels may be seen at the top of the cob. On1y one ear is to be counted from each node.

DO NOT remove or pull back the husks of ears within the unit to inspect for kernels. In doubtful cases, go outside the unit and inspect similar ears or ear shoots for the presence of kernels. After having done this, exclude any questionable ears from item 7.

Ears with kernel formation found loose on the ground in row 1 and row 2 middles are to be included in the count of ears for their respective rows.

Deformities emerging as part of the tassel which resemble a small cob with some kernels are not considered ears and should not be included in the count.

## Next Step:

Husk the first 5 ears or silked-ear shoots beyond row 1 for only the designated unit, then examine for maturity. If ears or silked-ear shoots are not yet present, CHECK () and skip items 8-14. See page 142 for illustrations.

Maturity Stage Code No.
Preblister....... 2
Blister.......... 3
Milk............. 4
Dough............. 5
Dent.............. 6
Mature............ 7
 b

For August 1, husk and inspect the first 5 ears or silkedear shoots beyond unit 2, row 1 for stage of maturity. Enter maturity codes in item 8.

August I - If the total of the maturity codes for the first 5 ears is 12 or less, skip items 9 through 14. If total is 13 or more, continue with item 9 on first 5 ears in maturity code-3 or higher. If any ears in item 8 are code-2, replace these ears with code-3 ears and enter in item 9.

For September 1, husk and examine the first 5 ears or silked-ear shoots beyond unit 1, row 1 for stage of maturity. Enter maturity codes in item 8.

September 1-If the total in item 8 is 12 or Tess, skip items 9 through 14. If total is 13 or more, continue with item 9 for first 5 ears in maturity code-3 or higher. If any ears in item 8 are code-2, replace each code-2 ear with a code-3 ear and enter in item 9.

In case there is more than one ear on a stalk, always count the top ear first for odd-numbered samples. Always count the bottom ear first for even-numbered samples. Pull back the husks without removing the ears from the stalks and classify each ear as to stage of maturity. Enter the proper maturity-stage code number for each ear. The rule is: TOP-ODD; BOTTOM--EVEN.

If the field is in a very early stage of growth and as a result ears or silked-ear shoots are not yet present in the unit or beyond the unit, a check mark should be inserted in the appropriate space in the instruction above item 8; then skip items 8 through 14.

The maturity classification for each ear will be based upon external characteristics of the plant and ear as well as kernel characteristics.

Each maturity stage has several distinct characteristics. All of these characteristics should be considered when assigning the maturity stage.

Maturity code 2 fits definition of "silked ear shoots or cobs without evidence of kernel formation." Maturity codes 3 through 7 refer to "ears with evidence of kernel formation" (item 7).
9. Maturity stage of first 5 ears code 3 or higher

Code No.


Does item 9 have 3 or more code-7 ears?
$\square$ YES. Complete items 12 and 13 only.

$\square$
NO. Complete items 10 and 11 only.

August 1 and September 1 - If the sum of the maturity codes for the 5 ears in item 8 totals 13 or more, copy the maturity code for each ear classified as code 3 or higher directly below to item 9. Whenever the total of the 5 ears is 13 or more and any code-2 ears are listed in item 8, you will select the next ear beyond the unit which is maturity code 3. List its maturity code in item 9.

NOTE: There should not be any ears in maturity code 2 listed in any boxes in item 9. All ears must be code 3 or higher. Code 2 ears should have been replaced with code-3 ears.

For October 1, husk and examine the first 5 ears with evidence of kernel formation beyond unit 1, row 2 for stage of maturity. Enter maturity codes in item 9.

For November 1, husk and inspect the first 5 ears with evidence of kernel formation beyond unit 2, row 2 for stage of maturity. Enter maturity codes in item 9.

Al2 Months - Before breaking an ear to determine the difference between maturity code 6 and code 7, measure and record the average length of kermel row in item 10.

Does item 9 have 3 or more code- 7 ears?
If YES, complete items 12 and 13 only.
If NO, complete items 10 and 11 only.
12. Measuring length of ear

In determining the length of the ear, the zero point of the tape is held at the butt of the cob with one hand. With the other hand, the tape is drawn taut along the length of the ear. When the tip of the cob is felt between the thumb and forefinger, the point on the tape is marked by the thumbnail and the length of the ear read to the nearest one-tenth inch. Any husks projecting beyond the top of the cob should not be included in determining the length. (See page .)

Enter measurements in decimal fractions: as 6.4 not 6 4/10, etc. Do not confuse this cob length measurement with the average length of kernel-row measurement in item 10. On the same ear, the cob length is usually from $1 / 2$ inch to 2 inches longer than the average length of the kernel row.


A tiller or sucker may emerge from the ground close to the main stalk, often at a slight angle. (This tiller is to the left of the stalk). Do not count tillers as stalks.


Code 2 - Preblister
Silk still has green tint and has not begun to turn brown. Only the cob and/ or hard spikelets can be felt through the husk.


Code 3 - Blister
Silk is beginning to turn color and the ear is filling out. Kernels, rather than just a hard cob, can be felt through the husk.


Code 3 - Blister
Most spikelets have partially formed kernels well enlarged and full of watery, clear liquid. Much of the silk has curned color and feels somewhat dzy.


Code 4 - Milk
Most kernels, although not fully grown, are full of a milklike substance and have little or no denting.


Code 5 - Dough
Ear is beginning to lean away from the stalk and shucks are taking on a light rust-colored appearance. Visible silk is completely brown and dry.


Code 5 - Dough
Kernels are fully grown with milk or doughlike substance in all of them. About one-half of the kernels are dented. In this example, the maturity line is noticeable but has not moved halfway to the cob on a majority of the kernels.


To measure the cob, hold the zero point of the tape at the butt of the cob, draw the tape up the ear until the tip of the cob is felt, and mark that point on the tape with your thumbnail.

### 3.5.5 Plant Growth Models

These models rely on detailed plant data collected more frequently during the season, as well as on environmental data. The additional plant data needed is primarily to provide information on the vegetative growth and the stage of development of certain plant parts. These two additional data needs are summarized in Exhibit $C$ to typify the kind of information which might be needed for corn. Meteorological and environmental indices would probably be obtained from an alternative data collection system, but due to the more frequent visits to the fields it may be feasible to also collect the environmental data with automatic recording instruments, using the same field workers.

Part A－Growth Model for Weight of Grain per Plant

Plant no．Unit＿ | （Form may have room for $20-50$ plants per |
| :---: |
| unit） |

## Field Data

1．Has plant tasseled？YES（ ）NO（ ）
0 or 1 $\square$
2．Has plant silked？YES（ ）NO（ ）
0 or 1 $\square$
If＂yes，＂enter silking date（day no．，Jan． 1 ＝1） $\square$
For silked ears：

3．Primary ear on plant

4．Secondary ear on plant Length
Circumference
Evidence of kernel formation YES（ ）NO（ ）
Length
Circumference $\qquad$
Evidence of kernel formation YES（ ）NO（ ）

Harvest ears on plants if 3 or 4 shows evidence of kernels and random number entered equals plant number $\mathrm{RN}=$ $\qquad$ ．

6．Number ears harvested
Identify each ear as from 3，4，or 5，and forward to office or field laboratory．

Lab Data
7．Wet weight of ears（grams）by type
$\qquad$ Total weight $\square$
8．Number kernel rows

$$
\text { 非3. } \quad \text { 非. }
$$

9．Number kernels on random row $\qquad$ ， $\qquad$ ， $\qquad$ \＃3． $\qquad$非 4. $\qquad$ \＃5． $\qquad$
Extract kernels from selected row and dry for 36 hours
10．Wet weight of kernels extracted from 3，4， 5 （grams） $\square$
11．Dried weight of kernels extracted from 3，4， 5 （grams） $\square$
12．Dry－matter percentage $\quad$ line $11 \div$ line 10

Part B - Vegetative Growth of Plant Parts

1. Date of planting (day no., Jan. $1=1$ )
2. Date of emergence (day no., Jan. $1=1$ )
3. Variety $\qquad$ Fertilizer applied $\qquad$
4. Soil moisture immediately after emergence at: .5 meter . 10 meter
5. Row direction $\qquad$
6. Height of plant
7. Number of leaves
8. Size of leaves
a.
L. W.
$\qquad$
$\qquad$
L. W.
$\qquad$
$\qquad$
b. i. $\qquad$
c. $\qquad$ j. $\qquad$
$\qquad$ p. $\qquad$
$\qquad$
d. $\qquad$
k. $\qquad$
$\qquad$ q. $\qquad$
$\qquad$
e. $\qquad$
9. $\qquad$
$\qquad$
r. $\qquad$
$\qquad$
f. $\qquad$ m.
$\qquad$ s. $\qquad$
$\qquad$
g. $\qquad$
$\qquad$ n. $\qquad$ t.

Plant leaf area $\square$
9. Stage of development: (circle one): a b c d e f g h
10. Leaf area index

LAI
Percent $\qquad$

Cut plant at ground level, if selected for laboratory sample.
12. Wet weight of plant parts

## Stem

 ___ gramsLeaves ___ grams

Head $\qquad$
Culm __ grams
Grain $\qquad$ grams

Dry weight of plant parts ___ grams
$\qquad$
___ grams ___ grams ___ grams ___ grams
13. Number of grains $\qquad$

Figure 10: Mature Crop Samples Sent to Laboratory for Weight and Moisture Determination


### 3.5.6 Corn Yield Forecasts

The yield-forecast technique being illustrated is based on data (actual) collected in 1976 approximately in the dough stage (around September 1) in a selected State in the Midwest region of the U.S. The number of ears are counted and the length of ears measured in plots 30 feet long consisting of two adjacent rows, and the row spacing is measured so the area of the plot could be converted to an acreage. The basic model for yield is: Biological yield = ears per acre $x$ weight of grain per ear. The statistics which must be obtained are as follows:
(1) Average number of silked ears per 60 -ft row plot $=103.9$
(2) Average row spacing $=3.32 \mathrm{ft}$
(3) Acreage conversion factor for one plot $=218.5$
(4) Average number of silked ears per acre $=22,695$
(5) Average length of cob for silked ears measured over husk $=7.92$ in.
(6) Historical regression equation (equation (4), page 60) for converting ear length to weight of grain at $15.5 \%$ moisture
$\bar{W}=(.0854 \times 7.92)-.304=.3724 \mathrm{lb}$ or 168.9 gm
(7) Biological yield per acre $=$ 非 x 非 $=8451 \mathrm{lb}$
(8) Estimated net yield per acre to be taken from field (less field and harvesting losses) $=8451(.90)=7606 \mathrm{lb}$ or 135.8 bu

In this forecast a global regression model for weight of grain per inch of cob length was used in conjunction with the survey averages of the inventoried components. Equation (4), page 60, was derived from probability samples of ears from the early 1960's.

An alternative model for the weight of grain per ear will be derived from the observed numbers of kernels per ear and a his－ torical weight（global mean model）for the weight per kernel．

The calculations for the alternative yield model are：
（5a）Average number of kernels per ear $=543$（average count）
（6a）Historical weight per kernel（Table 13，page 65） $W=.300 \mathrm{gm}$ per kerne1 at $15.5 \%$ moisture $x$ number of kernels per ear converts to weight of grain per ear $=$ 162.9 gm or .3591 lb
（7）Biological yield per acre＝非4 x 非6＝ 8150 lb
（8）Estimated net yield per acre to be taken from field＝ 7335 1b or 131.0 bu（less field and harvesting losses：非 x ．90）。

## 3．5．7 Corn Yield Forecast Based on Within－Year Growth Model

The use of the term＂growth model＂applies more correctly to just the dry－matter accumulation per ear or dry matter per kernel．The number of ears，number of kernels per ear and plants with ears at harvest are forecast based on a＂survival model＂ rather than a growth model．The yield model implies the separate modeling of the individual components．

The field data have been collected from a somewhat different plot configuration．The plot is laid out from a random starting point in each field．The plot consists of two parts：the plants in a $50-\mathrm{ft}$ section of a row from the starting point and the first 100 plants commencing with the starting point．The $50-\mathrm{ft}$ section is a part of（a subset of）the 100 plants．

Figure 11：Plot Layout for Collection of Weekly Data for Plant Growth Study


The 1976 statistics（same State）required for this model are as follows：
（1）Average number of plants per acre $=21,540$
（2）Average row spacing $=3.32 \mathrm{ft}$
（3）Acreage conversion factor $=218.5$
（4）Average number of silked plants per acre $=20,380$
（5）Number of silked plants with grain at harvest per acre $=20,258$
（6）Growth equation fitted to observed grain weight per plant after the fourth weekly visit since silking and evaluated at harvesttime（ $t \geq 80$ ）gives：
$\hat{\alpha}=156.4, \hat{\beta}=105.5$ ，and $\hat{\rho}=.863$ when the computer routine terminates based on change in y（ $\varepsilon$ ），where $\mathrm{y}=\frac{\hat{\alpha}}{1+\hat{\beta} \hat{\rho} t_{i}}$ and arithmetically $\hat{y}=156.4 \mathrm{gm}$ when 180 is substituted for $t_{i}$ ．
（7）Expected weight of grain per silked plant at plant maturity adjusted to $15.5 \%=181.8 \mathrm{gm}$ or .4008 lb
（8）Biological yield per acre＝非 x 非5 $=8119 \mathrm{lb}$ or 145 bu
（9）Estimated net yield per acre to be taken from field＝ 7307 1b or 130.5 bu（less field and harvesting losses： \＃8 x ．90）。

The actual weight of grain (with $1.8 \%$ moisture) per plant at maturity was 166 grams rather than the 156.4 forecast. The relative errors in the primary parameter in the growth model were $9.2 \%$ for weight of grain per plant and $0.3 \%$ for the survival parameter for plants with grain. Thus, the differences in the alternative yield forecasts for this sample of 24 fields are well within the sampling error of the forecasts.

The methods illustrated for corn can be applied to almost any crop. The specific plant characteristics used in the modeling should be quite similar for all the grain crops, cotton, and soybeans, as well as vine and tree crops. The use of additional characteristics in the concept of "fruit size" such as diameter, circumference or volume may be needed to improve the size-weight relations.

### 3.6.1 Introduction

Crop reporters, observers or farm operators are frequently requested to report on either an absolute (i.e., bushels per acre) or on a relative basis. The reporting is usually voluntary. Consequently, the questionnaires are short and restricted to several crops planted at the same time.

The concept of "normal condition" or "full crop" was initiated for forecasts when the crop was in the vegetative stage of development. The evaluation of the crop was based primarily on the stand and vigor of the plants but also reflects the appearance of fruit on crops with short fruiting periods. The number "100" is frequently used to designate a normal condition if there has been no damage from unfavorable weather, insects, pests, etc. on field crops. As crops near maturity, reporters are asked to report the probable yield on their farms, fields or for their locality. In either case, the crop condition or probable yields are translated into harvested yields by means of regression charts or equations over a series of years. Consequently, it is necessary to keep the concepts over years, and the sample of reporters or growers must be representative of the crop planted over each region or country. Most growers report at regular intervals during the growing season, according to the crop appearance. As the crop approaches harvest, the forecasts are based on the fruit appearance. In general, crops with well-defined and visible fruiting habits which are subject to a relatively short "critical period" are more accurately forecast. By comparison, root crops are subject to rather large forecast errors.

Exhibit D shows the basic questions for reporting condition, while Exhibit E gives corresponding question for probable yields. Exhibit F combines the two concepts and is the basis for an example of the graded yield appraisal discussed in chapter 2 , where similar questions are also asked after harvest.

## Report for your locality

I. WINTER WHEAT

1. For irrigated wheat, what is the condition now as compared with normal growth and vitality you would expect at this time if there had been no damage from any source?

LET 100 PERCENT represent a normal crop.
Percent
2. For nonirrigated wheat, what is the condition now as compared with normal growth and vitality you would expect at this time if there had been no damage from any source?

LET 100 PERCENT represent a nomal crop.
Percent $\qquad$
II. CORN
3. For corn for grain, what is the condition now as compared with normal growth and vitality you would expect at this time if there had been no damage from any source?

LET 100 PERCENT represent a normal crop.
Percent $\qquad$
III. PEACHES
4. What is the condition of peaches now as compared with that of a full crop if there had been no damage from any source?

LET 100 PERCENT represent a full crop.
Percent $\qquad$
IV. SWEET CHERRIES
5. What is the condition of sweet cherries now as compared with that of a full crop if there had been no damage from any source?

LET 100 PERCENT represent a full crop.
Percent $\qquad$

## Report for your farm

I. CORN

1. For irrigated corn, what probable yield per acre do you expect this year on your farm in 70-1b ear or 56-1b shelled bushels?
2. For nonirrigated corn, what probable yield per acre do you expect this year on your farm in $70-1 b$ ear or $56-1 b$ shelled bushels?
II. SORGHUM FOR GRAIN
3. For irrigated sorghums, what probable yield per acre do you expect this year on your farm in 56-1b bushels?
4. For nonirrigated sorghums, what probable yield per acre do you expect this year on your farm in 56-1b bushels?
III. SPRING WHEAT
5. For Durum wheat, what yield per acre do you expect this year on your farm in 60-1b bushels?
6. For spring wheat other than Durum, what yield per acre do you expect this year on your farm in $60-1 \mathrm{~b}$ bushels?

## RICE

1. How many tareas are planted on irrigated land alone?
2. How much rice do you expect to harvest from the irrigated tareas?

Quantity $\qquad$ Unit $\qquad$ Dry weight per unit $\qquad$
3. How would you describe the expected harvest?

Very good $\square$ Good $\square$ Average $\square$ Poor $\square$ Very poor $\square$
4. How many tareas are planted on dryland alone (this land will not be irrigated)?
5. How much rice do you expect to harvest from the dryland tareas?

Quantity $\qquad$ Unit $\qquad$ Dry weight per unit $\qquad$
6. How would you describe the expected harvest?
Very good $\square$ Good $\square$ Average $\square$ Poor $\square$ Very poor $\square$

## CACAO

7. How many hectares are planted alone this year?
8. How much cacao do you expect to harvest from these hectares planted alone?

Quantity $\qquad$ Unit $\qquad$ Dry weight per unit $\qquad$
9. How would you describe the expected harvest?
Very good $\square$ Good $\square$ Average $\square$ Poor $\square$ Very poor $\square$
10. How many hectares are interplanted with another crop this year?
11. How much cacao do you expect to harvest from these interplanted hectares?

Quantity $\qquad$ Unit $\qquad$ Dry weight per unit $\qquad$
12. How would you describe the expected harvest?
Very good $\square$ Good $\square$ Average $\square$ Poor $\square$ Very poor $\square$

In general, eye estimates by farmers or field workers show considerably less variation than actual yields. Consequently, the regression or relation between reported condition and yield may not be successful in eliminating bias from condition reports. Market prices may also introduce a cash-crop bias in reports by growers on harvested yields.

### 3.6.2 Dry Bean Yield Based on Growers' Appraisals

Each quarter a forecast is made of the yield of beans, which is then multiplied by the tareas ( $1 / 16$ hectare) to get a forecast of production. All the data are collected as part of a quarterly probability survey. The survey is a stratified area sample in which the sampling units within strata are selected with equal probabilities and the closed-segment concept is used. Yield appraisal data were obtained for all fields in the segment. Consequently, the tareas in each field are additive, but any field characteristics must be weighted by the tareas to insure unbiased estimates for the characteristics. Information on yields is obtained for all fields in each area sampling unit. The grower-graded-yield appraisal technique in chapter 2 , page 33 , is employed. The results for one quarterly survey are summarized in Table 25. The $E(z)$ based on the reported data $=1.20=$ (1.92) (.000) $+(1.68)(.427)+(1.00)(.443)+(.32)(.130)+(.08)(.000)$ for the forecast period; $E(z)=1.0$ for an average crop.

The growers' appraisals indicate a yield 20 percent above average for the coming quarter and approximately 10 percent above their harvested yield (not shown) for the last quarter (or crop). Since the absolute level of the yield ( $1.23 \mathrm{cwt} /$ tarea) indicates a better-than-average crop, it is meaningful to ask if the growers' idea of the average yield is higher or lower than might be expected. The derived average yield $1.23 \div 1.20=1.03 \mathrm{cwt} /$ tarea as compared with an after-harvest derived average yield of $.98 \mathrm{cwt} / \mathrm{tarea}$ for a
year ago. The growers' idea of normal appears to be somewhat higher than that of last year; perhaps this is just the result of sample variability. However, this may also be a result of greater use of fertilizers, or other factors.

Table 25--Calculations for Dry Bean Yield

| Condition <br> of beans | Centroid of probability in interval | Fraction of tareas in interval for normal crop | Fraction of reported tareas in interval for this year's crop |
| :---: | :---: | :---: | :---: |
| Very good crop | 1.92 | . 036 | . 000 |
| Good crop | 1.68 | . 238 | . 427 |
| Average crop | 1.00 | . 452 | . 443 |
| Poor crop | . 32 | . 238 | . 130 |
| Very poor crop | . 08 | . 036 | . 000 |
| Expectation | $E(z)$ | 1.00 | $=1.20$ |
| Growers' expected yield (weighted by tarea) |  |  | $1.23 \mathrm{cwt} /$ tarea |
| Derived average yield based on appraisal of forecasted crop |  |  | $\begin{gathered} 1.23 \div 1.2=1.03 \\ c w t / \text { tarea } \end{gathered}$ |
| Growers' harvested yield for forecasted quarter 1/ |  |  | $1.05 \mathrm{cwt} / \mathrm{tarea}$ |

1/ Obtained from following quarterly survey.

A second method is available which leads to essentially the same information. It can be referred to as the "growers'-average-yield-and-appraisal" method. For each planting of their crop, early in the season the growers are asked for the expected yield and what the growers consider an average yield to be for the crop planted in the same field. The grower's expected yield (or
production) and the average yield for the same acreage are reported for the data user's evaluation. The grower's within-year average yield permits the user to judge whether this figure is consistent with the reported yield of the previous year or years.

An equally important phase of the yield information is to obtain similar information from the same growers or a probability sample after harvest. The growers reported a yield of 1.05 cwt/tarea after harvest, which was very close to the derived average yield. This second survey provides information on annual harvested acreage and crop production as well as a grower's evaluation by five categories of the crop just harvested. That is,-the grower is asked to grade the harvested yield (or production) by the categories given. This information provides a basis for evaluating how good the growers are at forecasting their crop during the seasons and whether they evaluate the harvested crop in a manner consistent with the model. Based on several years experience, there appears to be a tendency for the growers to be somewhat pessimistic early in the season and after harvest to have a brighter evaluation with regard to the past crop season for beans.

### 3.7 Forecasting Walnut Yields

### 3.7.1 Introduction

The forecasting models developed have relied on marketedproduction data and objective measurement variables for a midseason forecast of production. In general, the models tested have employed regression methods requiring a series of data points over years before reliable forecasts can be achieved. The forecast date is September 1 and is based on a single field survey in late July and early August of approximately 600 blocks of walnuts for data collection. The crop is mature and harvest is active by October, but the date varies by districts because of the large number of varieties being grown.

### 3.7.2 Block and Tree Selection

The sample of 600 blocks was distributed in proportion to the bearing acreage in each county. The sample blocks were selected at the beginning of the program and retained in the years following. However, the sample is revised each year for blocks removed plus the addition of new blocks to represent new acreage coming into bearing. The blocks were selected with probabilities proportionate to the variety and year of planting. Within each block, two trees were selected with equal probabilities. Each sample consists of two "units" of one tree each. The orchard map has a small table at the upper left which lists the row number and space number for the location of tree 1 and tree 2. Each sample tree is shown and labeled on the orchard map. Near each of these numbers is a small arrow showing which direction is to be traveled in counting rows and spaces. If a sample tree falls into one of the categories listed below, an alternative tree is selected:
(1) The selected tree space is blank space (no tree).
(2) The selected space is occupied by a young, nonbearing tree.
(3) The selected space is occupied by a dead tree.
(4) The selected tree is obviously of a variety other than that specified in the heading of the orchard map.
(5) The selected tree is not a walnut tree.
(6) The selected tree is being used for experimental purposes by someone else (usually can be told by tags, grafts, or other markings on the tree).
(7) The accessible branches have been pruned or none are available.

In selecting an alternative tree, proceed away from the BIT (Block Identification Tag) in the same row as the original tree until you come to the next tree that meets all of the criteria for selection to be counted. If there are no eligible trees for counting in the same row away from the BIT, then select the next eligible tree in the next row towards the BIT. Be sure the alternative tree selected is the proper variety.

### 3.7.3 Measurement of Tree Spacing

In order to determine the number of trees per acre, the tree spacing is determined for each sample block. The procedure requires measuring the distance between trees within rows and between rows, at each sample tree. Each team has a 50 - or 100 -foot tape for measuring the distance between trees. They measure the perimeter of one triangle of trees for most sample trees. The only time the spacing measurements are not made is when the sample trees are in border or irregular orchards. Border plantings have two or three rows of trees; irregular plantings have variable spacing between rows and are usually contour plantings. Irregular and border plantings are described in comments.

Tree-spacing measurements must be taken with both samplers cooperating. Distances are measured from center to center, as shown:


Sampler \#1
Sampler 非2

Each sampler positions his tape at the middle of the trunk when measuring spacing between two trees.

The sample tree is used as one point of the triangle. The two nearest trees to the sample tree are selected as the other two points of the triangle, and are identified with heavy chalk marks around both trunks. The three measurements do not have to be taken in a specific order. The tape is pulled taut, and each distance is read to the nearest foot.


Illustration of required tree spacing measurements

Record each distance on the Random-Path Schedule under "Spacing."

The limb selection has been limited to "accessible branches"; that is, branches which can be reached from a 12 -foot ladder. The supervisor chooses by the random-path method the accessible branch to be used for nut counts. The c.s.a. should be between 5 and 15 percent of the tree's accumulated c.s.a. of the primary limbs. A completed form A which follows on page 165 shows the procedure for one tree. The count of nuts (i.e., 42) is also shown at the bottom of the form. Sampler uses a CSA (cross-sectional area) tape to measure for the cross-sectional area of the trunk and of each primary scaffold stemming directly from the trunk. The primaries are numbered starting with 1 in the direction of the BIT and going clockwise around the tree and are also measured and recorded in this order. After trunk and primary CSA are recorded, continue along the primary on which the accessible branch is located. If the accessible branch is itself a primary, then measurements will be completed. However, in most cases it will be necessary to measure the secondary splits and record these measurements. If these measurements to this point do not take in the measurement of the sample branch, then continue the procedure along the path of the accessible branch until the measurement of this branch and alternative branches are recorded. Finally, after the measurement of the accessible branch, blacken in the small box which corresponds to the accessible branch and indicate on the schedule the path followed back to the trunk by blackening in the proper box for each stage. The TIT (tree identification tag) is hung at the point where the accessible branch starts.


The figure below illustrates where to measure branches with bulges. To aid in understanding bulges, illustrations of an unpruned branch and a pruned branch are included. Solid lines indicate correct placement of measurements. Figure A shows a branch which is not pruned. One measurement is made below the branch split and two measurements are taken above the split. Figure B shows the same branch which has been pruned. The pruned branch has a CSA greater than 0.5 and must therefore be considered. Only one measurement is taken, and it must be made above the prune scar. That is the only location which will reflect the bearing surface. Figure $C$ illustrates the same branch with a bulge. The bulge is the healed wound caused by pruning. Measure all bulged branches (with healed prune scars) above the bulge. As in $B$, that is the only location which will reflect the bearing surface.


### 3.7.5 Counting Nuts

The nut count is started at the base of the accessible limb to be counted or, on old sample blocks, from the tree identification tag that is hung. The lateral branches are counted as they are encountered, progressing from the base of the branch to the terminal end of the branch. The sampler feels along the main part being counted for lateral branches and at the same time uses the marking crayon to mark along the main branch as he proceeds. Each nut is counted as it is encountered, and marked with a crayon. Every fifth nut is picked off and placed in the picking bag. A tally sheet is provided; the sampler recording the nut count will check off each nut as counted by his partner and tally every fifth nut counted in the tally column as follows:


The odd count is entered in the tally column and the total count computed for each stage, recording this in the tally column and also on the schedule in the box provided. The branch stage is labeled in the left margin and a line drawn across the tally columns before proceeding with the next stage.

This procedure continues with the picking off of every fifth fruit; thereafter, the count is recorded and the nuts placed in the picking bag. If it is necessary to move the ladder before completing the count on the sample branch, a marker tag should be hung just past the last lateral counted, so that the starting place for the next ladder set can be easily seen from the ground. The exact off-count at the end of the branch is the last count entered on the schedule. A sizer nut is not picked except when the count reaches 5 . Occasionally, terminal branches will extend so high that some nuts

| Survey code | Year | County | Variety | Elock number | $\begin{gathered} \text { Tree } \\ \text { number } \end{gathered}$ | Checker code | $\begin{gathered} \text { Update } \\ \text { code } \end{gathered}$ | Form number | Month/ day | $\begin{array}{\|l\|} \hline \text { Block } \\ \text { status } \\ \hline \end{array}$ | Number stages |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 291 | 74 | 10 | 5 | 44 | 1 |  |  | 1 | 7/1 |  | 3 |
| Tage: Block Row TreeFoundReplaced |  |  |  | $\frac{\text { Start elme }}{}$ |  | End time | Minutes | Team | Recorder |  | Measurer |
|  |  |  |  | 0820 |  | 0905 | 45 |  | McPeak |  | D. Okimura |

BLOCK STATUS CODE: 1 Sampled, 2 Wet, 3 Pulled out, 4 Abandoned,
5 Sprayed, 6 Substitute, 7 Not vialted, 8 Refusal, 9 Not found

|  | Branch | Stage |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | $\begin{gathered} \text { Trunk } \\ 0 \end{gathered}$ | $\underset{1}{\text { Primary }}$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | CSA | 66.0 | 41.0 | 4.3 | 4.1 | - | . | . | - | J. |
| 2 | CSA |  | 28.0 | 36.0 | 3.9 |  | . | $]$ | J. | J |
|  | Total |  | $69.0$ | 40.3 | 8.0 | . | . | J | $\square$ | $\square$ |
| 3 | CSA |  |  |  | 32.0 | . | . | J | $\square$ | $\square$. |
|  | Total |  | ] |  | 40.0 | . | . | $\square$ | $\square$ | $\square$ |
| 4 | CSA |  |  | J. | . | - | . | $J$ | $\square$ | $\square$ |
|  | Total |  |  | $J$. |  |  | . | - | $\square$ | Ј . |
| 5 | CSA |  |  |  |  |  | . | ]. |  | $\square$ |
|  | Total |  |  |  | . | - | - | . | - | . |
| 6 | csA |  | $\square$. |  | . | . | . | $\cdots$. | -. | J. |
|  | Total |  | - |  | . | . | - | - | . | . |
| 7 | CSA |  |  | $\begin{aligned} & \square \square \square \\ & \square \square \square \\ & \square \square L \end{aligned}$ |  |  |  |  |  |  |
|  | Total |  |  |  |  |  |  |  |  |  |
| 8 | CSA |  |  |  |  |  |  |  |  |  |
|  | Total |  | . |  |  |  |  |  |  |  |
| 9 | CSA |  | . |  |  |  |  |  |  |  |
|  | Total |  | - |  |  |  |  |  |  |  |
| 10 | CSA |  |  |  |  |  |  |  |  |  |
|  | Total |  | . |  |  |  |  |  |  |  |
| Random number |  |  | . | - | - | - | - | - | - | - |
| Selected branch |  |  | 1 | 2 | 1 |  |  |  |  |  |
| Fruit count |  |  |  |  | 42 |  |  |  |  |  |
| Check |  |  |  |  |  |  |  |  |  |  |

TABLE OF RANDOM NUMBERS

| 316.7 | 66.1 | 64.4 | 94.3 | 16.6 | 91.9 | 3.1 | 6.7 | 3.6 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 041.3 | 54.1 | 55.8 | 34.3 | 63.5 | 35.2 | 3.6 | 1.9 | 0.0 |
| 184.1 | 72.8 | 48.1 | 18.3 | 47.9 | 98.2 | 3.4 | 0.5 | 6.1 |
| 192.4 | 18.6 | 19.1 | 36.7 | 41.8 | 61.1 | 7.4 | 5.2 | 4.7 |
| 159.9 | 74.6 | 27.7 | 37.0 | 75.8 | 31.9 | 0.7 | 1.7 | 2.4 |
| 179.4 | 32.3 | 92.1 | 97.6 | 36.1 | 19.9 | 6.3 | 3.1 | 4.1 |
| 408.7 | 78.4 | 64.2 | 25.0 | 11.8 | 62.7 | 0.1 | 3.7 | 4.2 |
| 089.9 | 62.0 | 95.3 | 67.8 | 70.0 | 44.1 | 8.7 | 1.5 | 9.8 |
| 186.9 | 12.3 | 02.8 | 07.8 | 33.2 | 62.4 | 8.3 | 0.0 | 2.3 |
| 234.8 | 76.3 | 78.4 | 16.0 | 56.5 | 96.1 | 0.5 | 0.8 | 6.7 |
| 340.6 | 05.0 | 41.4 | 98.0 | 72.0 | 28.8 | 0.7 | 4.2 | 8.4 |
| 459.3 | 46.9 | 78.3 | 54.8 | 25.9 | 36.2 | 8.2 | 4.2 | 4.3 |
| 434.7 | 47.6 | 65.6 | 43.8 | 29.9 | 78.2 | 8.2 | 3.0 | 2.9 |

Coments:
will be counted by sight only. Samplers should have with them a stick with a hook at the end to help them sight-count the nuts and to pick off the sample nuts. Include all nuts in the count except those which are totally shrivelled, totally blighted, or dwarfed; generally these will fall off when tapped lightly.

### 3.7.6 Selecting Subsamples of Nuts for Sizing and Weighing

(1) Place all nuts stripped from the terminal branch on the counting board, spreading them in a continuous line, single file.
(2) Count the nuts and enter the total counted on the Random Path Schedule in the box which corresponds to the last stage where the terminal branch is recorded.
(3) Select 20 nuts for a sizing sample as follows:
(a) Divide the total counted by 20 and round to the next largest whole number. This is the "sampling interval."
(b) Use the third line of the table of random numbers (on the Random Path Schedule), ignoring the last digit to the right of the decimal. Choose the first number which is 01 or greater but which is not larger than the interval. The number chosen designates the first nut to choose from the line of nuts described above (see (1)).
(c) Select the second nut by adding the "interval" to the random number.
(d) Select the third nut by adding the "interval" to the number for the second nut.
(e) Select the fourth nut by adding the "interval" to the number for the third nut. Proceed until 20 nuts are selected. If you reach the end of the line of nuts before the 20 th nut is obtained, continue the count at the beginning of the line.
(f) If the total of nuts counted is between 10 and 20 , include all nuts for your sample.
(g) Place the sample just selected in a neoprene bag. Place the sizing sample identification marker in the bag.
(h) Date sized. Enter the calendar date when you size the nuts on the sizing card. Use 2 digits for the month, 2 digits for the day, and the last 2 digits for the year in that order.

### 3.7.7 Nut Measurements

a. Hull characteristics

The first two characteristics described will be recorded for every nut sized.
(1) Width

Place the caliper jaws on the hull at the widest point of the hull, making sure that the caliper jaws are parallel to the longest axis of the nut. Rotate the hull so that the calipers are measuring the widest point of the hull. See Figure 11, page 172.
(2) Grade

Make a visual determination of the grade of the hull. Descriptions of the grades are as follows:
(a) SOUND. No damage is visible except for wind scarring and superficial hull damage.
(b) SUNBURN. The hull will begin to turn yellow first. Gradually, the yellowing increases as the center of discoloration turns yellowish brown, then dark brown. Usually, there is no depression in the hull.

Sunburning can cause a flat side on the hull if it occurs before the shell hardens. After the shell hardens no flat sides develop.

Consider the nut sunburned whenever 10 percent or more of the hull surface is affected. You should cut some nuts to determine if the meat has been damaged. In advanced stages, meat turns black and shrivels. There will not be any wet substance inside the skin. Meat damage will vary by district and orchard.
(c) LARVAL DAMAGE. Look at the upper portion of the nut hull near the stem for larvae of the walnut husk fly. The larvae may be tiny whitish specks or larger mature maggots. Blackened hulls are characteristic of nuts infected by husk flies. Cut into the blackened area. Husk-fly larvae should be clearly visible.
(d) BLIGHT. There will be a depression in the hull and the hull inside the depression turns dark brown to black. Generally, the blight will darken the meat by the time the depression is about $3 / 8^{\prime \prime}$ in diameter. Depressions $3 / 8^{\prime \prime}$ and larger in diameter will be coded as blight damage.
(e) SHRIVEL. The outward appearance of the hull indicates that the nut will not mature. The hull shrivels due to factors other than blight and sunburn.

Use the following codes for grades:

| SOUND | $=1$ |
| :--- | :--- |
| SUNBURNED | $=2$ |
| LARVAL DAMAGE | $=3$ |
| BLIGHT | $=4$ |
| SHRIVEL | $=5$ |

The following measurements will be made only on every 5 th nut beginning with nut number 3 .
(3) Cross width

Place the caliper jaws on the nut so that the long axis of the nut and caliper jaws parallel each other. Rotate the nut 90 degrees from the position in (1) "Width." See Figure 12. Record the measurement of the nearest millimeter under "C. width" on Form B (measurement cards).
(4) Length

Place the calipers on the nut so that one caliper jaw passes through the stem scar on the end of the nut. The other jaw should pass over the point on the other end of the nut. See Figure 13. Record the measurement to the nearest whole millimeter under "Length."
(5) Weight

Place the nut on the Mettler balance. Weigh the nut to the nearest one-tenth gram, and record under "Weight."
b. Shell characteristics

Cut away the hull at the suture line to expose the walnut she11.
(1) Width

Place the caliper jaws on the shell at the widest point of the shell so that the caliper jaws are parallel to the longest axis of the nut. See Figure 14, page 173. Record the measurement to the nearest millimeter under "Width."
(2) Cross width

Place the caliper jaws on the shell so that the suture line and caliper jaws parallel each other. The suture line should be about 90 degrees from each caliper jaw. See Figure 15. Record the measurement to the nearest millimeter under "C. width" - Form B (measurement cards).
(3) Length

Place the caliper jaws on the shell so that they embrace the longest dimension of the nut. Position the shell so that the suture line parallels the calipers. The suture line should be nearest the stationary part of the calipers. See Figure 16. Record the measurement to the nearest whole millimeter under "Length."
(4) Grade

Cut the shell in half at the suture and make a visual determination of the grade. Descriptions of the grades are as follows:
(a) SOUND. No damage is visible.
(b) SUNBURN. The kernel turns black and shrivels.
(c) LARVAL DAMAGE. Husk-fly larvae will be visible.
(d) BLIGHT (black kerne1). The kernel turns very dark.
(e) SHRIVEL. The kernel pulls apart from its original area.
c. Complete appropriate steps for the remaining nuts as required.
d. If less than 20 nuts are in the sample, draw a line across the card through spaces provided for the next nut.
e. After the last nut has been sized, make a visual check of the recorded data, the numbers directly below measurement headings indicate the number of digits to be recorded in a particular field for each nut. Check each column from top to bottom to detect errors in recording measurements before proceeding to the next sample.

Dispose of the nuts in an acceptable manner.


Figure 1l. Measuring Width of Hull of a Walnut


Figure 12. Measuring Cross Width of Hull of a Walnut


Figure 13. Measuring Length of Hull of a Walnut


Figure 14. Measuring Width of Shell of a Walnut


Figure 15. Measuring Cross Width of Shell of a Walnut


Figure 16. Measuring Length of Shell of a Walnut

FORM B: WALNUT OBJECTIVE MEASUREMENT FIELD SURVEY
BLOCK IDENTIFICATION

| SURVEY <br> CODE | BLOCK CODE |  |  | TREE <br> NUNBER | SIZER <br> CODE | DATE <br> SAMPLED | DATE <br> SIZED |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| $1-3$ | $4-6$ | $7-8$ | $9-11$ | 12 | 13 | $14-16$ | $17-18$ |
| 045 | 020 | 08 | 044 | 2 | 7 | 226 | 227 |

Measurement cards

| Nut \# | IN HULL MEASUREMENTS |  |  |  |  | IN SHELI |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Width घm | Grade | C. width m m | $\begin{aligned} & \text { Length } \\ & \text { im } \end{aligned}$ | Weight grams | Width mm | C. width <br> mm | Length <br> mm | Grade |
| 20-21 | 22-23 | 24 | 25-26 | 27-28 | 29-31 | 32-33 | 34-35 | 36-37 | 38 |
| 01 | 44 | 1 |  |  |  |  |  |  |  |
| 02 | 44 | 1 | U |  |  |  |  |  |  |
| 03 | 45 | 1 | 44 | $56$ | 60.1 | $36$ | 36 | 42 | 1 |
| 04 | 42 | 1 |  |  |  |  |  |  |  |
| 05 | 43 | 3 |  |  |  |  |  |  |  |
| 06 | 44 | 1 |  |  |  |  |  |  |  |
| 07 | 46 | 1 |  |  |  |  |  |  |  |
| 08 | 46 | 1 | $46$ | 50 | 59.6 | $36$ | 37 | $41$ | / |
| 09 | 39 | 1 |  |  |  |  |  |  |  |
| 10 | 46 | 2 |  |  | $z^{2}+8$ |  |  | $8$ |  |
| 11 | 43 | 1 |  |  |  |  |  |  |  |
| 12 | 41 | 1 |  |  | craies. | - |  |  |  |
| 13- |  |  |  |  |  |  |  | - - |  |
| 14 |  |  |  |  |  |  |  | $x^{2}-2+1$ |  |
| 15 |  |  |  | की |  |  |  |  |  |
| 16 |  |  | $2+2 x+2$ | $+1$ |  | P $\square^{6}$ |  |  | $80$ |
| 17 |  |  |  |  |  |  |  |  |  |
| 18 |  |  |  |  |  |  |  |  |  |
| 19 |  |  |  |  | $72$ | $5$ |  |  |  |
| 20 |  |  |  |  |  |  |  |  |  |

3.7.8 Forecast Based on Tree Data and Market Production Records

In each orchard, a limb on each of two trees has been sampled. The total set per tree is obtained by using a ratio-type expansion, since limb selection is restricted to an accessible limb. The nuts counted on the sample limb are expanded to a total tree count, based on the ratio of the cross-sectional area of the limb sampled to the total cross-sectional area of limbs for the tree. A systematic subsample of every fifth nut is selected to be sized, weighed and number of "sound" nuts determined. The yield-per-tree models are as follows:

Yield per tree $=$ Sound nuts per tree x harvested weight per nut. The production model is:

```
Production = Bearing acres x trees per acre x yield per tree.
```

The model which performs the best in practice consists of adjusting the gross yield (and production) to a net production harvested, based on industry-reported production. A regression based on the historical series is used for this purpose. A further refinement is introduced into the estimated weight per nut based on the set per tree and the in-hull weight of nut. The regression derived is as follows:

```
P - harvested production
p}=\mathrm{ gross production = B x T < S x W Wh
B = bearing acreage
T = trees per acre
S = sound nuts per tree
Wh
P = ae kp or after taking natural logarithms }\operatorname{Ln}P=\operatorname{Ln}(a)+k
```

where $a$ and $k$ are model parameters to be estimated and $e=2.7183$, the natural logarithm base (Ln). The components in the above model
are developed as follows: The yield per tree is a product of $W_{h}$ and $S$ and converted from grams to tons by the divisors 453.59 (grams per pound) and 2,000 (pounds per ton).

$$
W_{h}=W \times \frac{\text { shell volume }}{\text { total volume }}=W \times \frac{\left(\frac{\text { shell suture }}{2}\right)^{3}}{\left(\frac{\text { hull suture }}{2}\right)^{3}}=W \times \frac{\mathrm{SS}^{3}}{\mathrm{HS}^{3}}
$$

where $W$ = in-hull weight per nut on survey date

$$
\begin{aligned}
& \mathrm{SS}=\text { shell suture } \\
& \mathrm{HS}=\text { hull suture }
\end{aligned}
$$

These are calculated as follows:

$$
\begin{aligned}
& W=\frac{1}{N} \sum_{i=1}^{t} \sum_{j=1}^{\sum_{i}^{i}} W_{i j} \\
& S S=\frac{1}{N} \sum_{i=1}^{t} \sum_{j=1}^{N_{i}^{i}}(S S)_{i j} \\
& H S=\frac{1}{N} \sum_{i=1}^{t} \sum_{j=1}^{N_{i}^{i}}(H S)_{i j}
\end{aligned}
$$

and $W_{i j}$, (SS) ${ }_{i j}$ and (HS) ${ }_{i j}$ are the weights and measurements for an individual nut on the $i^{\text {th }}$ tree where $N_{i}$ is the number of nuts sampled on the tree and $t$ is the number of trees sampled.

$$
N=\sum_{i=1}^{t} N_{i}
$$

The number of sound nuts per tree, $S$, is computed from the nuts sampled as follows:

$$
S=S_{A} \sum_{i=1}^{t} \frac{S_{i}}{S_{A i}}=S_{A} \cdot F_{S}
$$

where $S_{A}=$ average number of nuts set per tree $=\frac{1}{t} \sum_{i=1}^{t} S_{A i}$
$S_{i}=$ sound nut on $i^{\text {th }}$ tree
$S_{A i}=$ all nuts on $i^{\text {th }}$ tree
$F_{S}=$ fraction of nuts not damaged

The total number of trees is estimated from the bearing acreage times trees per acre, or $B \times T$.

The sample averages give the following results:

```
\(W_{h}=44.10 \times \frac{(32.5)^{3}}{(40.5)^{3}}=44.10(.5168)=22.79\)
\(S=1729.8(.9601)=1660.8\)
\(T=29.4\)
\(\mathrm{B}=163,234\)
\(a=1.29, k=.70\)
\(p=(163,234) 1660.8 \times(29.4) 22.79 \div(454 \times 2000)=200,048\) tons
\(\operatorname{Ln} P(1974)=10.3796+.000007984(200,048)=11.9768\)
    or \(\mathrm{P}=158,906\) tons
```

The adjustment of the gross production from 200,048 tons to the 158,906 tons is the result of a number of undetermined factors of which the following play a major role.
(1) Weight loss in nuts due to moisture and hull removal from survey date to maturity
(2) Possible bias in procedure for estimating in shell green weight
(3) Market order thinning or delivery quotas
(4) Harvesting losses.

Models based on several regression relations (over years) with factors which are undetermined, plus unknown factors, generally require frequent modification and reevaluation of the parameters when there are obvious trends in yield components. Likewise, the degrees of freedom used in the modeling result in a relatively small number for error determination, unless a long series of historical data exists.

### 3.7.9 Forecast Based on Objective Tree Data

A variation or alternative yield model might be employed; the same survey data are used to illustrate such an alternative.

Weight per nut (in shell) $=$ in-hull weight x adjustment factor to convert to an in-shell weight. (Based on weight per unit volume) $=44.10 \mathrm{x} .620=27.34$ grams

Harvest weight per nut (in shell) $=27.34 \mathrm{x}$ (1-. 00514 D ) ; adjustment of weight for days to harvest ( $D=55$ ) $=27.34 \mathrm{x} .7173=19.61$ grams

Harvest gross yield per tree $=$ number sound nuts x harvest weight per nut
$=1660.8 \times 19.61=32,568$ grams or 71.80 1b

Net yield per tree $=71.80 \mathrm{x}$ adj. for net losses in nuts due to droppage and harvesting
$=71.80(.93)$ or 66.77 pounds

$$
\begin{aligned}
\text { Production }= & \text { number of trees } x \text { net yield per tree } \\
= & 4,628,910 \times 66.77=309,072,320 \text { pounds or } \\
& 154,536 \text { tons }
\end{aligned}
$$

Any adjustment due to marketing orders would need to be applied on either a tree basis or the fraction of sound nuts to be marketed.

### 3.8 Forecasting Yields from Historical Crop and Weather Data

### 3.8.1 Introduction

The yield forecasting based on historical data of this type assumes that a global regression model is valid. Since the network of data points is usually limited and the variables are available only for large geographic areas, the weather factors should not be expected to explain differences in yields over small geographic areas. Consequently, the variables used represent averages which do not reflect the full range of the variables within any year and predictions frequently experience larger errors for individual years than the conventional error levels calculated from the regression model.

### 3.8.2 Corn Yield Forecast

The technique described in chapter 2 , section 2 will be used for illustration purposes. The daily temperature values were averaged. The precipitation for the month for each weather station in the State is shown in monthly published reports of NOAA*. The individual station values were then averaged by 10 districts within the State. Predetermined weights (approximately equal) were applied to the district averages to obtain the monthly values for the State. The State yield for corn was obtained from the published SRS report giving harvested yield based on a statewide survey. The variables for 1972 are given in the table below. The "normal" yield due to technology for 1972 is derived from the moving average shown in Chart 7, page 181.

[^3]Table 26--1972 Variables for Corn Yield

| Variable | June | July | August |
| :--- | :---: | :---: | :---: |
| Average daily <br> temperature ( ${ }^{\circ}$ ) | 69.6 | 74.4 | 73.7 |
| Monthly precipi- <br> taion (inches) | 3.88 | 6.02 | 5.43 | | Moving-average yield (i.e., technology): |
| :--- |

The predicted yield departures from technology level by months were derived from the following equation:

$$
\begin{aligned}
& \text { June } \quad \Delta y_{1 i}=173.801-43.275 R-2.475 T+0.6208 R T=1.280 \\
& \text { July } \quad \Delta y_{2 i}=89.939-23.666 R-1.263 T+0.3397 R T=5.651 \\
& \text { August } \quad \Delta y_{3 i}=114.710-16.328 R-1.559 T+0.2261 R T=1.635 \\
& \text { June and July } \Delta y_{1 i}+\Delta y_{2 i}=6.931 \\
& \text { June, July\&August } \Delta y_{1 i}+\Delta y_{2 i}+\Delta y_{3 i}=8.566
\end{aligned}
$$

The accumulative yield forecasts are sumarized below:

$$
\begin{array}{ll}
\text { June } & =100+1.3 \doteq 101.3 \\
\text { June and July } & =100+6.9 \doteq 106.9 \\
\text { June, July and August } & =100+8.6 \doteq 108.6
\end{array}
$$



### 3.9.1 Introduction

The forecasting model as first developed was discussed in chapter 2. The basic model has changed very little except that the estimators of the key components are now derived mathematically rather than graphically. The first forecast of the crop season is made in early October by type of citrus.

### 3.9.2 Block and Tree Selection

The survey uses the stratified multistage probability sample design by the major citrus types described in chapter 2. The strata within type are four age groups. All trees of bearing age (4 years or older) and all citrus-producing areas are proportionally sampled within strata.

The sample of blocks is selected from an inventory of all commercial citrus plantings (1/3 acre or more). The inventory is obtained from an aerial photography survey of all citrus-producing areas in the State, combined with ground inspections of any previously unidentified plantings. The aerial surveys are done at two-year intervals.

The selected groves are identified by township, range, section, and block. All groves have aerial photo blueprints (ozalid copies), county maps, and instructions giving the location. If for any reason a sample grove does not conform to the description on the instructions, the crew supervisor notifies the statistical office and an appropriate substitute is made.

Within a grove, the procedure discussed in chapter 2 has been modified from the cluster of four trees, and three sample trees are now selected for all oranges, grapefruit, Temple and tangelo groves. In all groves the crew supervisor must (1) cut a fruit from each sample tree to verify proper fruit type, and (2) verify that the tree is in the proper age group. The three trees are obtained
from the cluster of four trees by eliminating one tree at random. Sample trees are changed every 3-5 years in a gradual rotation pattern around the pivot tree. This gradual rotation maintains a high degree of tree identity in successive years and yet provides for unsurveyed trees to enter the population; it also provides a measure of any sample longevity effects on the trees retained for several years.

### 3.9.3 Limb Selection

The final stage of sampling is the selection of a portion of the tree on which the fruit is to be counted. The portion of the tree is selected by the random-path technique discussed in chapter 2. When this multiple-stage process terminates, the selected portion had a probability of selection proportionate to limb crosssectional area (c.s.a.). The reciprocal of this probability of selection times the fruit count provides an unbiased method of estimating the total fruit on the tree. If the limb selected is not too small, the method is more efficient than equal-probability selection because of the positive correlation between limb c.s.a. and fruit numbers.

After the sample limb is selected, it is divided into smaller units for counting purposes. Two separate fruit counts are made, each by a different member of the survey crew. If the two counts do not agree within 5 percent, additional counts are made. A random selection of one limb in a 10-percent random subsample of groves is made as a quality check.

### 3.9.4 Fruit Drop Surveys

A measure of fruit mortality prior to harvest must be introduced into the computed crop forecasts, because initial estimates of the average number of fruit per tree are established from counts in August and September. Natural loss of fruit, from August until the month in which each type of fruit is considered mature, is
measured by a sequence of monthly surveys. Maturity is considered to be reached in predetermined cutoff months which precede the heaviest harvest period. Cutoff dates are: December for tangelos and tangerines, January for early and mid-season oranges, February for Temples and grapefruit, and April for late-season oranges.

The sample trees for droppage surveys are drawn from a route frame rather than the limb count frame, since the route frame is more readily accessible for monthly observations. This sample frame consists of all bearing commercial groves fronting on a 1,600-mile route which traverses producing areas of the most important counties. This microcosm of the citrus population provides a satisfactory base for sampling drop and other relatively uniform characteristics.

The sample for each variety is stratified into four areas (homogeneous county groupings) and the four age groups previcusly discussed. The sample size within strata is based on productivity in a base year.

A sample limb approximately two percent of the trunk c.s.a. is selected near shoulder height, on a designated side of the tree. This limb is tagged and all fruit beyond the tag are counted during successive surveys. The monthly counts are entered on the pocket-notebook-size field sheets. The differences between the initial survey counts and later survey counts indicate the droppage to the time of the survey. The average drop for each age-area stratum is computed and then combined by production weights into the average drop for the State. The sample counts are weighted, because groves are selected with probability proportionate to historical production and the "two percent limb" sampling method tends to put a disproportionate part of the sample in older, more productive trees.

The monthly droppage is projected to the cutoff month to estimate seasonal drop rate for use in the forecast models.


### 3.9.5 Size of Fruit

The fruit size survey coincides with the drop survey. Moreover, the same subsample of trees in sample groves drawn from the route frame is used for both sets of monthly observations. In the size survey 10 sample fruit per tree are measured from a two-tree cluster per sample grove. Frequency distributions of standard fresh-fruit sizes and the estimated average size are obtained each month.

The fruit to be measured are determined by minimum size categories at a specified point on the tree at about shoulder height. This point on the tree is tagged and, for each survey, horizontal circumferences are measured on the 10 regular-bloom fruit nearest
the tag. The photograph illustrates the position of measurements and the device used to obtain the circumference.

These circumference measurements are entered as a tally on the 240-cell field form. Summarization is done in volume, which is linearly correlated to weight and, therefore, additive.

The growth rates of various citrus types were shown in chapter 2. The dates shown are the month in which surveys were conducted; usually surveys were near the third week of each month. The annual growth curves generally parallel each other, thereby allowing these relationships to be a fairly effective tool in forecasting size at maturity. It should be noted that fruit measured on-tree does not reflect harvest size. Early observations are of immature fruit and measurements for forecasts usually cease prior to volume harvest. The size of fruit at maturity is defined as the average size of fruit in groves in a specific month. These cutoff months are the same as in the drop surveys. Prior to the cutoff month, it is necessary to estimate the average size that fruit will attain in the cutoff month.

Figure 16. Measuring Circumference of Citrus Fruit (on the tree) with Calipers


# Florida Crop and Livestock Reporting Service <br> 1222 Woodward Street <br> Orlando, Florida 32803 

CITRUS GROWTH SURVEY
CIRCUMFERENCE CALIPER MEASUREMENTS


| In. | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | 0579 | 0865 | 1231 | 1689 | 2247 | 2917 | 3710 | 4634 | 5699 | 6917 | 8297 |
| 1 |  |  |  |  |  |  |  |  |  |  |  |
| 16 | 0594 | 0885 | 1257 | 1721 | 2285 | 2964 | 3764 | 4696 | 5771 | 6998 | 8388 |
| 2 |  |  |  |  |  |  |  |  |  |  |  |
| 16 | 0610 | 0906 | 1283 | 1753 | 2324 | 3010 | 3818 | 4759 | 5843 | 7080 | 8481 |
| 3 |  |  |  |  |  |  |  |  |  |  |  |
| 16 | 0626 | 0927 | 1310 | 1786 | 2363 | 3057 | 3875 | 4821 | 5616 | 7163 | 8574 |
| 4 |  |  |  |  |  |  |  |  |  |  |  |
| 16 | 0643 | 0948 | 1337 | 1819 | 2403 | 3104 | 3928 | 4884 | 5989 | 7246 | 8668 |
|  |  |  |  |  |  |  |  |  |  |  |  |

* $\mathrm{W}=$ white, $\mathrm{P}=$ pink

Sdy. = seedy, SS . = seedless

### 3.9.6 Florida Citrus Forecast

The objective estimate of citrus production by type is computed from the results of four surveys or different types of data collection activities:
(1) The total number of commercial trees is determined biennially, but is adjusted in intervening years based on trend and data on tree plantings.
(2) Fruit per tree is determined from the limb count survey in August and September.
(3) A fruit loss or "drop" survey is run monthly to give an indication of the changes, and project fruit remaining at harvest.
(4) A fruit size survey is run monthly to determine growth and project fruit volume at harvest.

The estimated number of fruit per tree for early oranges was $F=696$. The estimated drop from August to September was . 1439 . The drop to harvest was estimated using a multiple-regression equation:

$$
D_{h}=a+b_{1} \sqrt{x_{1}}+b_{2} x_{2}+b_{3} x_{3}
$$

where $x_{1}=.1439$ or fraction of fruit dropped through September 15
$x_{2}=696$ or estimated fruit per tree in September $x_{3}=7.25 \mathrm{cu}$. inches or estimated volume per fruit in September
$D_{h}=-.7050+1.472(.3793)+.00001(696)+.045(7.25)=.1865$

The fraction of the September fruit to be harvested is:

$$
H_{h}=1-D_{h}=.8135
$$

The fruit size or volume in cubic inches at harvest is estimated using a multiple-regression equation as follows:

$$
v_{h}=a+b_{1} x_{1}+b_{2} x_{2}+b_{3} x_{3}
$$

where

$$
\begin{aligned}
\mathrm{x}_{1}= & 7.25 \text { or average September volume per fruit in } \\
& \text { cubic inches } \\
\mathrm{x}_{2}= & 696 \text { or estimated fruit per tree in September } \\
\mathrm{x}_{3}= & 2 \text { or monthly change in volume per fruit from } \\
& \text { August to September } \\
\mathrm{V}_{\mathrm{h}}= & 2.909+.915(7.25)-.0028(696)+1.085(2)=9.764 \mathrm{cu.} \text { in. }
\end{aligned}
$$

The regression estimate of volume per fruit is used to derive the number of fruit per box using a regression equation as follows:

$$
\begin{aligned}
& \hat{\mathrm{S}}=65.87-1.95 \mathrm{~V}_{\mathrm{h}}+1772 \div \mathrm{V}_{\mathrm{h}} \quad \begin{array}{l}
\text { fruit per box at cutoff } \\
\text { month }
\end{array} \\
& \mathrm{S}=65.87-1.95(9.764)+(1772 \div 9.764)=228.313
\end{aligned}
$$

The forecasted yield per tree in boxes of fruit is:

$$
\hat{Y}=\frac{F \cdot H}{S}=\frac{696(.8135)}{228.3}=2.48
$$

The expected production is obtained by multiplying yield per tree times number of trees:

$$
\hat{P}=T \cdot \hat{Y}=14,256,000(2.48)=35,355,000 \text { boxes. }
$$

Table 27--Costs of Objective Yield and Related Surveys, 1967-68

| Survey | $\begin{aligned} & \text { Unit } \\ & \text { of } \\ & \text { cost } \end{aligned}$ | Cost classification |  |  |  |  | Total |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  |  | Field |  |  |  | Office |  |
|  |  | Wages |  | Mileage | Per diem | Supplies, clerical \& ADP |  |
|  |  | Within grove | Between groves |  |  |  |  |
| $\text { Limb } \mathrm{count} \text { I/ }$ | Sample grove | \$ 9.43 | \$ 6.29 | \$ 4.87 | \$ 1.02 | \$ 1.62 | \$ 23.23 |
| $\begin{aligned} & \text { Size } \& 2 / \\ & \text { drop } \end{aligned}$ | Sample grove | . 84 | 1.25 | . 45 | . 27 | . 82 | 3.63 |
| Maturity $3 /$ | Sample grove | . 23 | 1.30 | . 47 | . 10 | . 21 | 2.31 |
| Row count 4/ | Survey | 620.00 | 110.00 | 200.00 | 35.00 | 100.00 | 1065.00 |

1/ Costs are based upon a five-man crew consisting of four fieldmen and a supervisor.

2/ Treated as one survey, as both types of observations are made on the same sample trees. Surveys conducted each month. Information usually collected by a two-man crew.

3/ Survey conducted twice each month.
4/ Cost per month.

A dilemna usually faces the person in charge of the yieldforecasting program: (1) Do I choose the simpler model which I know will be reasonably satisfactory four out of five years and give poor results the fifth year? or (2) Do I choose a more elaborate model that is slightly more satisfactory and may provide clues that an unusual season may be occurring in the fifth year? Traditionally, the former has been chosen because of the cost savings and in the convenience of data collection. Also, there is some evidence to suggest that the complicated models may not necessarily reflect the seasonal influences even if they are based on the growing crop. This evidence is not conclusive nor is it based on complete yield-component modeling. Likewise, there is no conclusive evidence to suggest that combining weather or environmental variables with plant characteristics in a model will be any more successful, and these auxiliary data will inflate data acquisition costs. The problem is not hopeless or insolvable theoretically but it may be so practically, because of costs and the unpredictability and dissimilarity of the one in five years with marked departures from near-average crop seasons.

There are perhaps two approaches which will give better answers than are probably currently in use by those providing public information. (1) Greater seasonal detail on plant characteristics and the interrelation among the lead and lag yield components, and (2) A seasonal discriminant analysis to identify the unusual season before harvest from which a decision can be made to employ an alternative set of model parameters or procedures. The discriminant analysis will involve not only more detailed seasonal information on plant characteristics but also a means of measuring and predicting the nutrient uptake or accumulation by the plant parts.

The first solution or approach is realistic in terms of known yield components used in models. For example, in poor and excellent yield years for corn, the change in grain per ear is only partially reflected
in the average cob or kernel-row length, for the number of kernels and weight per kernel are also factors. The clues are present well before harvest, but the model or procedure used in forecasting must be selected by the analyst so it will discriminate such a season from the more typical conditions under which most of the crop data are collected. Of course, the relation between the lead and lag characteristics must then be employed. However, it will be clear what the direction of the lag component is even if the exact relation may be imprecise, since the effects are usually cumulative.

The second approach is probably considerably more costly and involves agricultural scientists not usually involved in operational data-collection procedures for making inferences or forecasts for large geographic areas. However, there are many benefits or research uses which can be obtained from these more detailed measurements of the development of plant parts besides those concerned with the yield modeling.


[^0]:    * 

    Leaf area index

[^1]:    * "A New Approach to Small Area Crop-Acreage Estimation," Harold F. Huddleston and Robert M. Ray III, Annual Meeting of the American Agricultural Economics Association, Pennsylvania, August 1976.

[^2]:    * Natural causes are sunburn, wounds, abrasions and hail.

[^3]:    * National Oceanic and Atmospheric Administration

