SAMPLING TECHNIQUES FOR MEASURING AND FORECASTING CROP YIELDS

ECONOMICS, STATISTICS, AND
COOPERATIVES SERVICE

U.S. DEPARTMENT
OF AGRICULTURE

ESCS
No. 09

SAMPLING TECHNIQUES FOR
MEASURING AND FORECASTING CROP YIELDS

By

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

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 current-year 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.

CONTENTS

Page

CHAPTER 1 - A REVIEW OF YIELD MEASUREMENT TECHNIQUES

CHAPTER 2 - MODELS FOR FORECASTING YIELDS BASED ON PLANT RESPONSE

Page

CHAPTER 3 - DATA COLLECTION CONCEPTS USED IN FORECASTS FOR SPECIFIC CROPS

CHAPTER 1 - A REVIEW OF YIELD MEASUREMENT TECHNIQUES

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) crop-cutting 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, $100 \leq n \leq 400$), even though counter examples have been cited based on sampling from inappropriate but convenient populations by reporters or officials usually using non-probability 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 1--Grower Concepts Involved in Yield Measurements
(Column concepts are not necessarily related horizontally)

Area	Production or field-reporting units	Harvested form of crop	Use
Planted	Standard volume container	Husked heads (or bean, berry)	Hauled from field to farm
Harvested	Weight basis	Unhusked heads (or bean, berry)	Delivered to market
Contracted	Number of bunches	Threshed grain	Sales
Gov't. allotment	Number of heads (or fruit)	Whole leaves	Consumed as feed or food
Number of trees	Sized fruit or head	Brushed roots or tubers	Destroyed or "dumped"
Interplanted area		Whole fruit	Processed
Equivalent solo planted (or harvested)		Stalks or whole plant	Seed

EXHIBIT 1 - EXAMPLE OF DATA COLLECTED FROM GROWERS ON GRAIN CROPS

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.	Acres	Total production harvested and to be harvested
FIELD CROPS		
1. Corn planted for all purposes..................................		▨
2. Corn harvested and to be harvested for grain....................		Bu.
3. Corn cut for silage...		Tons
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................		Bu.
8. Soybeans 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.......................................		Bu.
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......................................		Bu.
14. Barley used for hay, silage, pasture only, plowed under or abandoned..		▨

EXHIBIT 2 - COMMON USES REPORTED FOR SOYBEAN CROP

SOYBEAN INQUIRY

REPORT FOR THE FARM YOU OPERATE	Answer here ↓

1973 CROP PRODUCTION AND PURCHASES

1.	Soybeans HARVESTED for beans on this farm, last year's crop......... Bushels	
2.	Soybeans BOUGHT FOR SEED 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.	Soybeans SOLD AND TO BE SOLD between Sept. 1, 1973 and Sept. 1, 1974......... Bushels	
5.	Soybeans USED FOR SEED on this farm for planting the 1974 crop.............. Bushels	
6.	Soybeans FED AND TO BE FED to livestock on this farm (beans fed whole or ground) between Sept. 1, 1973 and Sept. 1, 1974......... Bushels	
7.	Old-crop soybeans expected to be on hand Sept. 1 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: Sold in 1973 September........................... 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 grower-reported 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)

Crop	Data source	Units reported	Frequency
Cotton	Gins	No. bales, gross or net weight	Monthly
Soybeans	Crushers	Oil, cake or meal	
Rice	Mills	Milled	For season
Coffee	Exporters	Roasted	
Oranges	Gov't. inspection and grading	Juice, fresh fruit	Seasonal
Grapes	Wineries	Tons crushed for wine	
Cherries	Private processors to trade assoc.	Containers packed	Semimonthly
Tobacco	Private auctions	"Hands"	
Wheat	Flour mills	Milled	
Sugar beets	Sugar factories	Tons of brushed roots, or sugar	

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.

EXHIBIT 3: BALE WEIGHT REPORT OF COTTON GINNED PRIOR TO OCTOBER 1

Crop of 1976

1. a. Total number of bales of cotton ginned from this crop prior to October 1 →	Total bales
b. Total weight of the bales reported in item 1a above →	Total weight lbs.

c. The weight reported above is: ☐ NET (Excludes bagging and ties)

☐ GROSS (Includes bagging and ties)

2. Enter the AVERAGE weight of bagging and ties used per bale here →	Average weight of bagging and ties used per bale lbs.

3. 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 more 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).

The bale weights listed below are: ☐ NET ☐ GROSS

Line No.	Bale number (a)	Bale weight (Pounds) (b)	Bale number (a)	Bale weight (Pounds) (b)	Bale number (a)	Bale weight (Pounds) (b)	Bale number (a)	Bale weight (Pounds) (b)	Bale number (a)	Bale weight (Pounds) (b)
1										
2										
3										
4										
5										
6										
7										

EXHIBIT 4: SOYBEANS MONTHLY REPORT OF PRIMARY PROCESSORS

OILSEEDS, BEANS, AND NUTS

Report period - *Mark with an "X" the box which best describes each reporting period* →					January	February	March	April
					☐ 1 Cal.Mo.	☐ 1 Cal.Mo.	☐ 1 Cal.Mo.	☐ 1 Cal.Mo.
					☐ 4 Weeks	☐ 4 Weeks	☐ 4 Weeks	☐ 4 Weeks
Product code	Item description		Unit of measure	Item code	☐ 5 Weeks	☐ 5 Weeks	☐ 5 Weeks	☐ 5 Weeks
	SOYBEAN							
0011611	Beans, crushed		S. tons	0100				
2075111	Crude oil produced (degummed weight)		M. lbs.	0105				
2075113	Cake and meal produced for _	Animal feed	S. tons	0111				
2075115		Edible protein products	S. tons	0112				
2075142	Lecithin produced		S. tons	0114				
2075261	Millfeed produced		S. tons	0115				
0011611	Stocks	beans	S. tons	0120				
2075111		crude oil	M. lbs.	0125				
2075211		cake and meal	S. tons	0130				
2075261		millfeed	S. tons	0135				

EXHIBIT 5: FLORIDA WEEKLY REPORTS FOR CITRUS

PRELIMINARY WEEKLY PROGRESSIVE REPORT OF FRUIT RECEIVED AT PROCESSING PLANTS

Week Ending March 27

In units of 1-3/5 bu.)

	Grapefruit	Early-Mid Oranges	Late Oranges	Navel 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 PROGRESSIVE REPORT OF FRUIT RECEIVED AT PACKING HOUSES

In units of 4/5 bu.)

	Grapefruit	Early-Mid Oranges	Late Oranges	Navel 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 yield is to be estimated
2. Random selection of fields	Number of fields or area of each to determine probabilities of selection for individual fields
3. Subdivision of field into plots	Dimensions of field or number of crop rows in field, used to determine plots of a given size and shape
4. Random selection of plots	Identification of randomly selected fixed-size plots to be measured or marked off by a preconstructed frame
5. Selection of certain plant parts for measurement	An enumeration of all the plant parts (normally the basic yield components)
6. Selection of some plant parts for cutting	Weight or other measure of heads or other plant parts
7. Selection of grain to be forwarded for laboratory analysis	Determination of grain fraction, moisture and, in some cases, quality factors
8. Selection of plants and area to be gleaned after commercial or normal harvest procedure	Number of heads and weight of grain attached to heads as well as loose grains on ground missed or lost from harvesting equipment

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 plantings with tree crop
2. Random selection of blocks of trees	Number of trees, age, variety for all blocks for deriving probabilities for selecting individual blocks
3. Random selection of trees	Rows of trees and trees per row are used to determine selection probabilities
4. A random selection of a small portion of tree is to be made since complete harvesting is costly	The main trunk and primary-limb sizes and number, as basis for selection probabilities
5. Terminal limb selection (and possibly paths to limbs)	Identification of terminal limbs from which to count fruit
6. Random selection of fruit or clusters to be removed from tree	Weight and/or size of fruit removed
7. Random selection of fruit "berry" or nut "meat" at special field stations	Ratio of fruit "berry" or nut "meat" weight to total fruit weight
8. Random selection of trees and ground area to be gleaned after commercial harvest	Number of fruit and weight of berries on trees and ground missed or lost in harvesting

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 16½ ft (1/80 acre) (50 x 25 (links))	Rectangular
16½ ft x 16½ ft (1/160 acre)	Rectangular
33 x 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 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 in. x 26.136 in. (1/10,000 acre)	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.

CHAPTER 2 - MODELS FOR FORECASTING YIELDS BASED ON PLANT RESPONSE

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 a 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.

2.2 Mathematical Models

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 within-year 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 x_t = past data for the t^{th} period (usually years) for a yield characteristic $x_1, x_2, \ldots, x_t$

ε_t = the normal random error for time t

μ = a constant mean

and we wish to forecast the characteristic for time t+k.

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

$$\bar{x}_{t+k} = \frac{1}{t}(x_1 + x_2 + \ldots + x_t)$$

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 over-years 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):

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

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

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

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:

$$x_t = \frac{\alpha}{1 + \beta\rho^t} + \varepsilon_t$$

where x_t = given data value for time t in a sequence of times during crop season for a yield characteristic

α, β, ρ, = constants or model parameters

ε_t = random error for time t

and we wish to forecast the characteristic for time t+k.

Some of the other models commonly encountered are as follows:

Linear trend: $x_t = \alpha + \beta t$ for all t (i.e., the time variable).

Linear regression: $x_t = \alpha + \beta z_t$ where z_t is another variable.

Autoregression: $x_t = \alpha + \beta x_{t-1}$ 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: $x_t = \varepsilon_t - \Theta\varepsilon_{t-1}$ where Θ is a constant between -1 and 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 t+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.

FIGURE 1 - CORN HARVESTED YIELD VS. GROWER PROBABLE YIELD - AUGUST 1 1959 - 1973

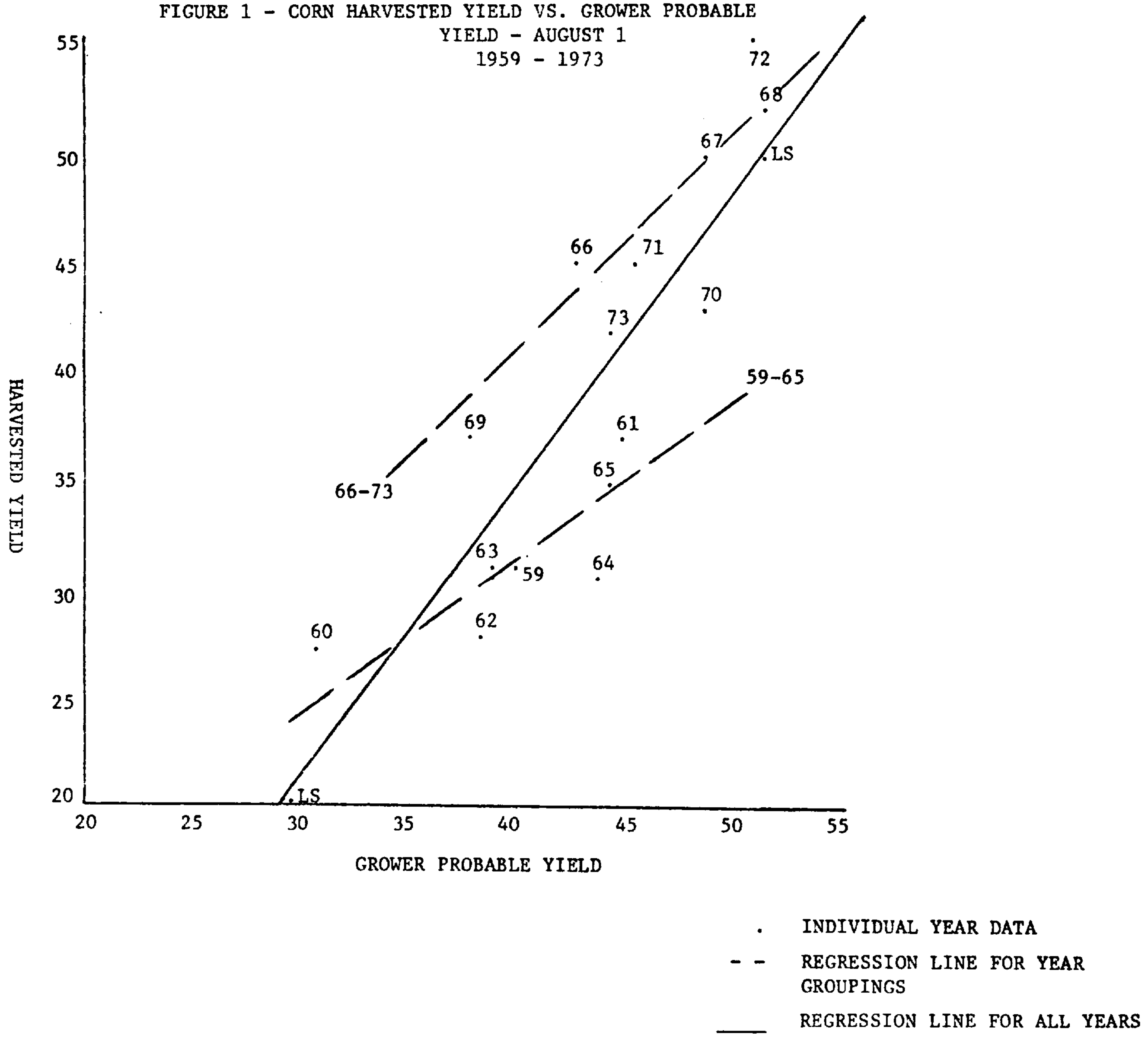


Table 6--Correlation Coefficient and Regression of Farm Operators' Reported Yield (Y) after Harvest on Farm Operators' Projection of Yield (X) at the Beginning of Fruit Setting (for probability samples)

(a)

1972

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

(b)

1973

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

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.

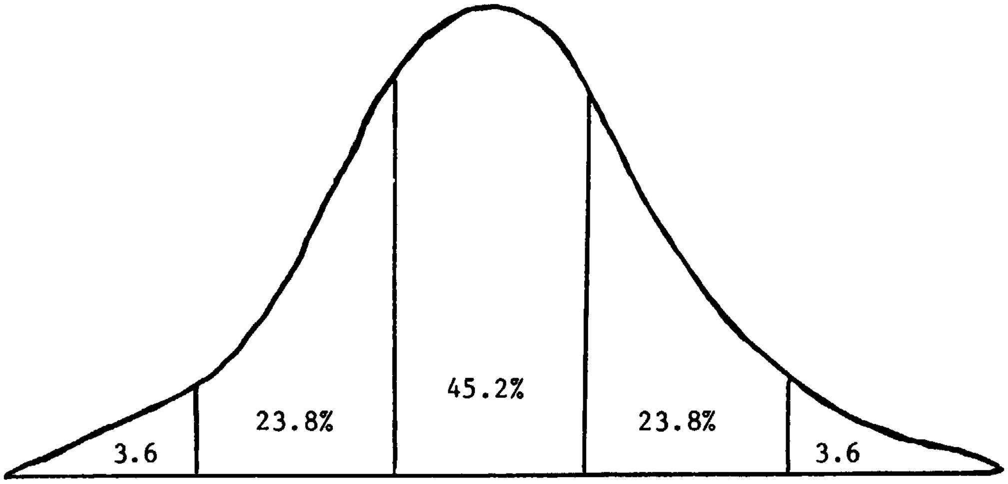
The descriptive ratings provided by the growers are assumed to be distributed normally, as in the grading system commonly used by teachers when a large number of students are to be graded. Thus, the name "grower-graded yield appraisal" is given to the method, since the grower, in effect, "grades" his 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 description corresponding to grades	Crop failure	Below average	Average	Above average	Much above average
	Very poor crop	Poor crop	Normal crop	Good crop	Very good crop
	No harvest				Outstanding Excellent
Standardized (uniform) yield scale	0 .4	.8	1.2	1.6	2.0
Midpoint of interval X_i	.2	.6	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).

$$E(X) = \sum_{1}^{5} p_i X_i = 1.00 \text{ (average yield)}$$

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 probability in interval Z_i	.08	.32	1.00	1.68	1.92

2.4 Crop-Weather Relations for Predicting Yields

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 $|\Theta| << 1$ (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-year 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 <u>no more than half</u> 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:

$$Y = a + bT + cR + d(TR) \qquad (1)$$

or

$$Y = a + bT + cR + d(TR) + gT^2 + hR^2 \qquad (2)$$

where T = average monthly temperature

R = monthly rainfall

and $a, b, c, d, g,$ and h are regression parameters.

The individual monthly 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).

$$Y_t = \bar{Y} + \frac{b}{2}(T_t - \bar{T}) + \frac{c}{2}(R_t - \bar{R}) + \frac{d}{2}(TR - \overline{TR}) \qquad (3)$$

where $\bar{Y}$ = is the normal yield based on trend (or base-period average yield if no trend is present)

$\bar{T}$, $\bar{R}$, and $\overline{TR}$ are the averages for the base period

T_t, R_t, and TR 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 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 - TEN-YEAR SIMPLE MOVING AVERAGES OF YIELD PER ACRE

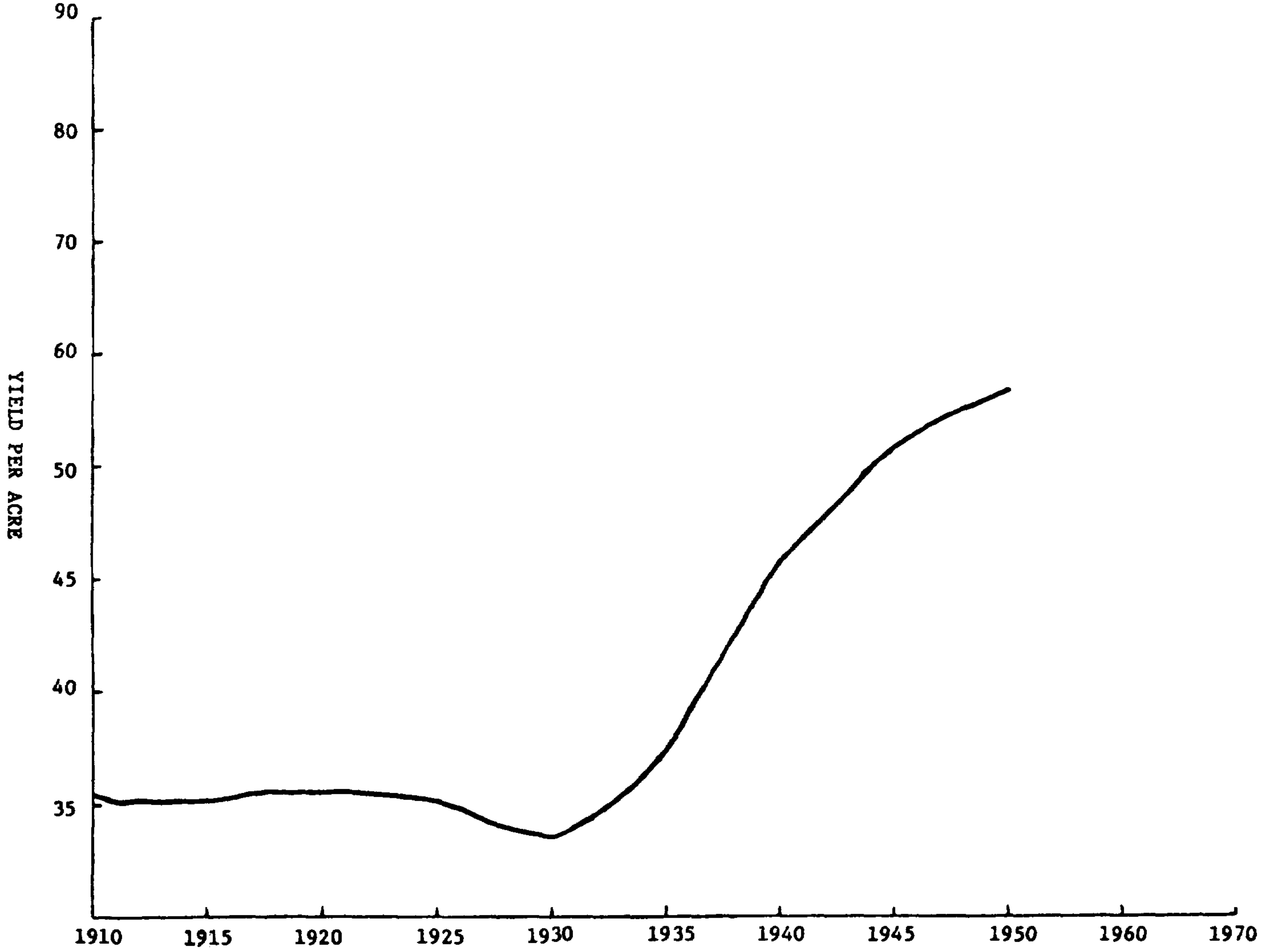


CHART 3 - YIELD DEPARTURE ISOGRAMS BASED ON JUNE RAINFALL AND TEMPERATURE

REGRESSION EQUATION: $Y' = 173.801-43.275R-2.475T+0.6208RT$

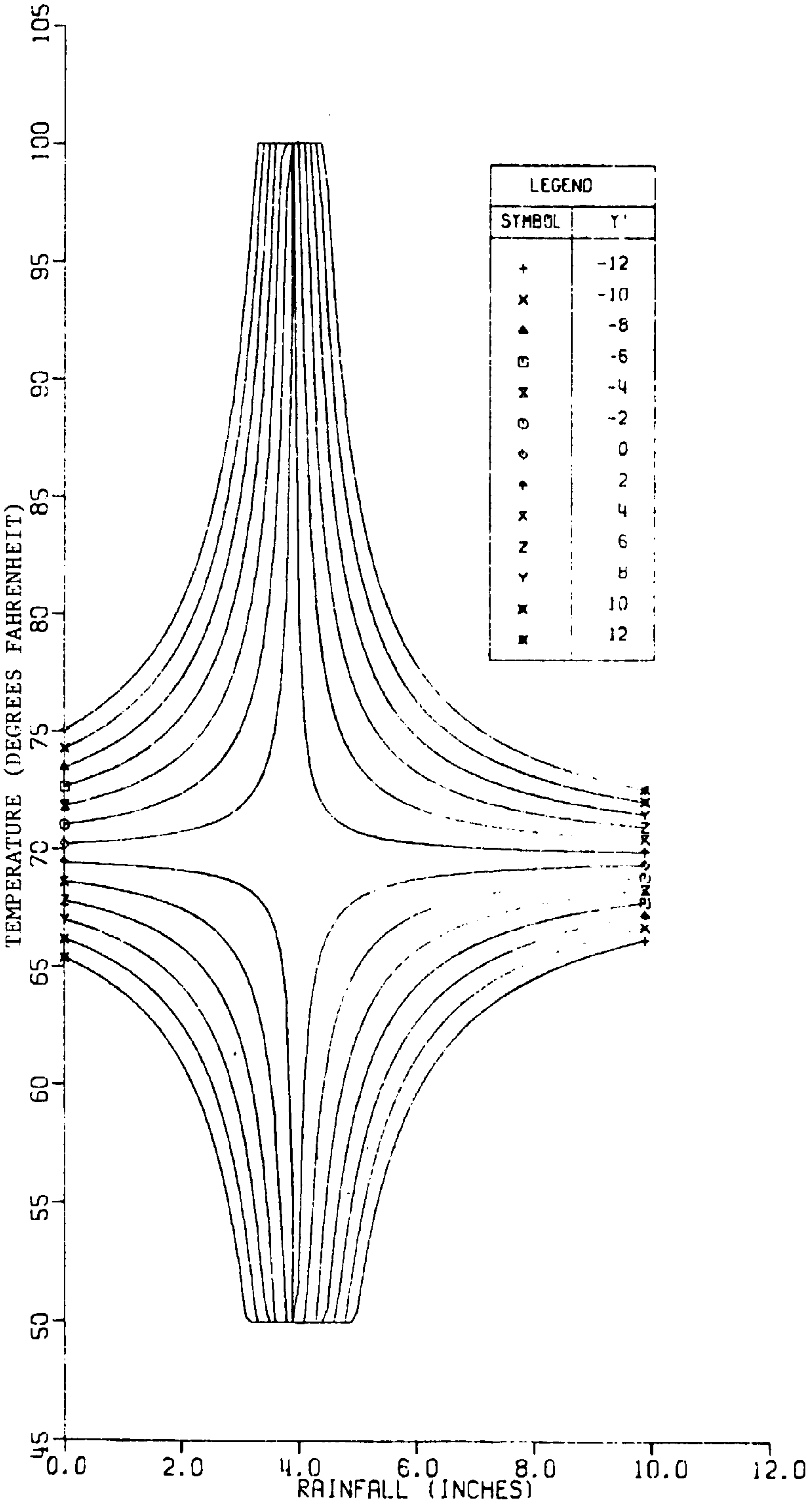


CHART 4 - YIELD DEPARTURE ISOGRAMS BASED ON JULY RAINFALL AND TEMPERATURE

REGRESSION EQUATION: $Y' = 89.939-23.66R-1.263T+0.3397RT$

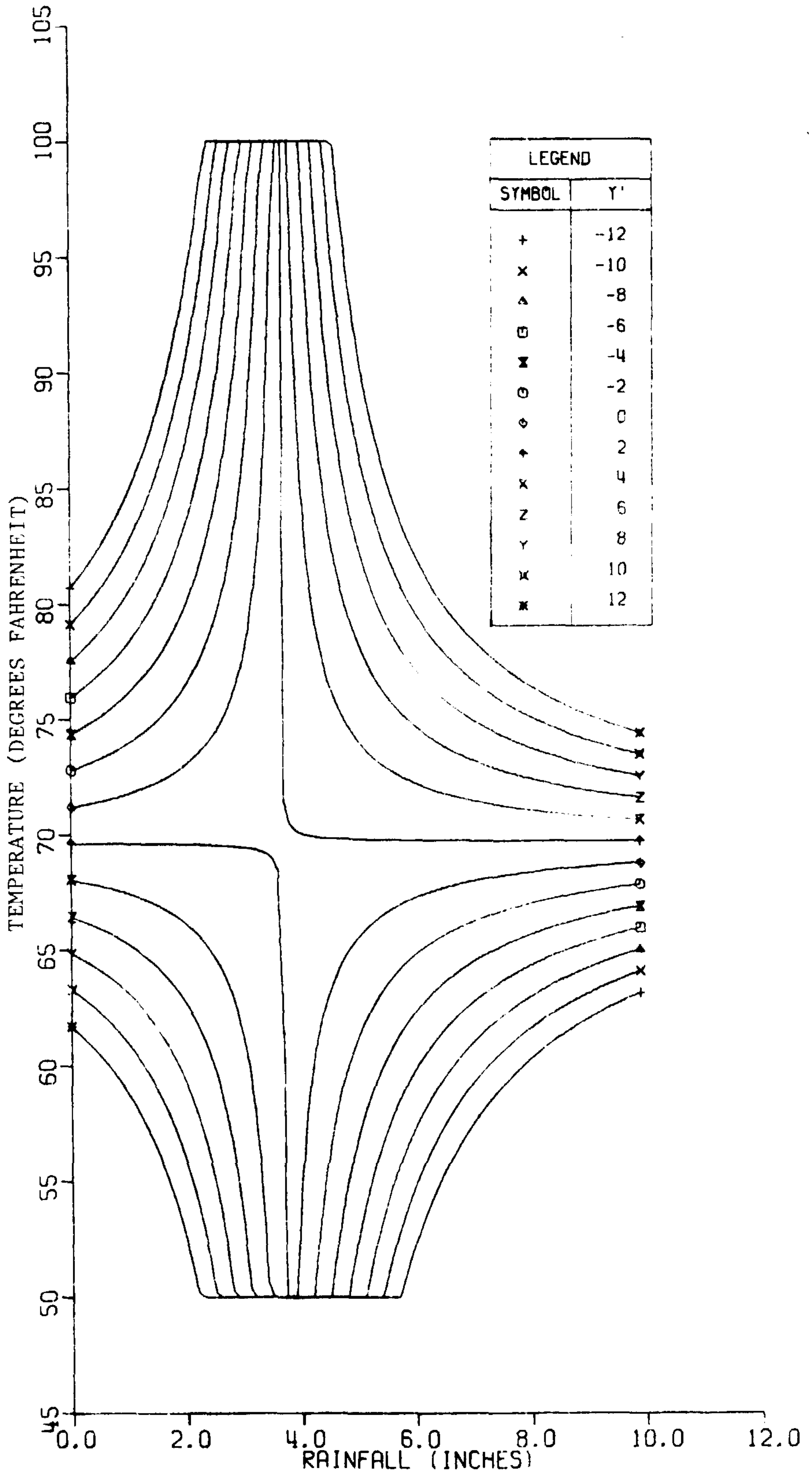
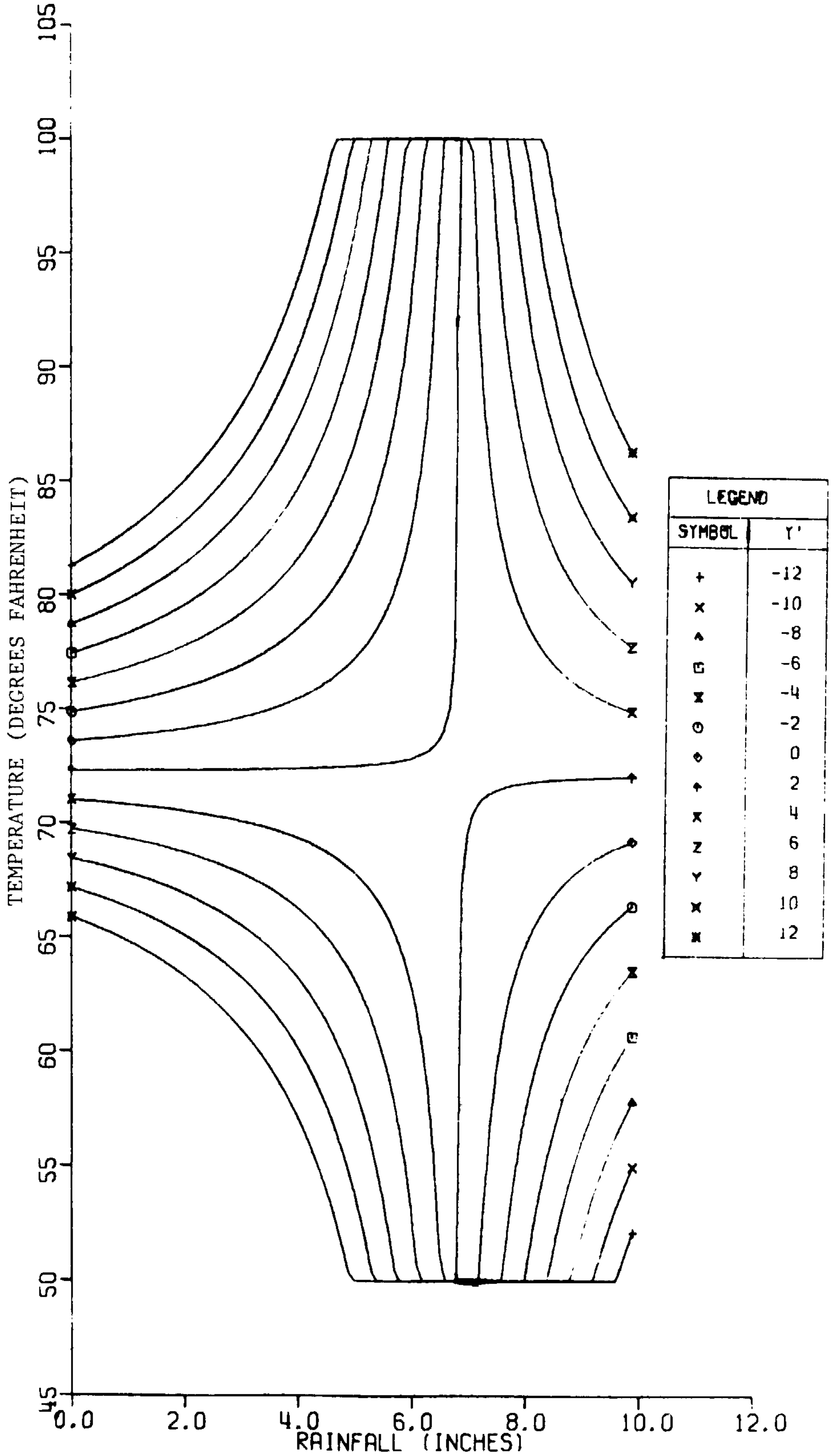


CHART 5 - YIELD DEPARTURE ISOGRAMS BASED ON AUGUST RAINFALL AND TEMPERATURE

REGRESSION EQUATION: $Y' = 114.710-16.328R-1.559T+0.2261RT$



2.4.3 Agrometeorological Forecasting of Crop Yields

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 one-meter 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 black-earth 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:

$$Y = 0.22X + 199.0 \qquad r = 0.75$$

And for the winter wheat of Bezostaya 1 kind

$$Y = 0.24X + 241.2 \qquad r = 0.79$$

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 high-quality 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:

$$Y_L = 0.24X - 16.0$$

(b) highest yields (Y_h) under the most favorable weather conditions of April, May, and June:

$$Y_h = 0.24X - 4.4$$

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

$$Y = 0.24X - 10.2$$

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 (Y_L) under unfavorable conditions of weather of April, May, and June are as follows:

$$Y_L = 0.2X - 15.0$$

(b) the highest yields (Y_h) under the most favorable weather conditions of April, May, and June have an outlook:

$$Y_h = 0.2X - 7.2$$

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

$$Y = 0.2X - 11.1$$

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, evaporation-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 R^2 values between the observed yield and predicted yields was as follows:

$$Y = A \prod_{n=1}^{3} (\Sigma(T/E_o)_n)^{\lambda_n}$$

where Y = bushels winter wheat per acre

n_1 = period from emergence to jointing

n_2 = period from jointing to heading

n_3 = period from heading to soft dough

λ_n = growth-stage weighting factor

T = actual transpiration (daily)

E_o = potential evapotranspiration from a wet soil (daily)

A = multiplier constant

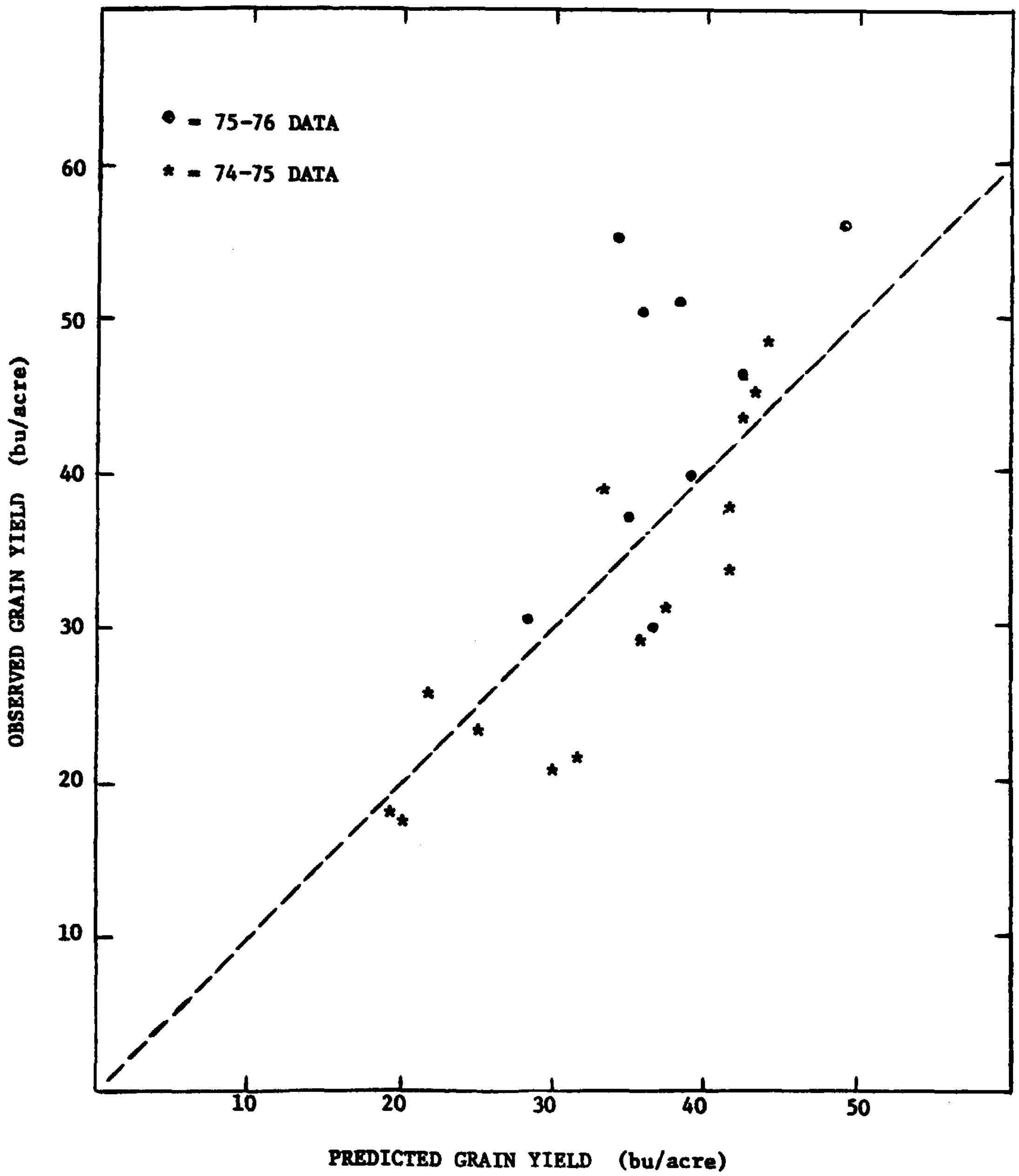
The fitted model is as follows:

$$Y = 2.856\ (\Sigma(T/E_o))_1^{.172} \cdot (\Sigma(T/E_o))_2^{.104} \cdot (\Sigma(T/E_o))_3^{.646}$$

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 $I = 12P/(T + 10)$, where P represents precipitation in millimeters and T represents temperature in degrees Celsius. For example, the January 1970 temperature of 4.2° 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° 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:

$$Y = 9.18 + 0.00098F - 0.0148JF + 0.0706MJ$$

and $R^2 = 0.70$ $SD = 1.074$

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, or 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 dry-matter 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 green-up 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 of sample fields	Approximate size of plot in acres 1/	Approximate size of population		Standard error of estimated yield per acre
			Acres in millions	Percent of U.S. total	
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 of fields	Harvested acres (measured)	Preharvest Survey							Commercial harvest of grain with 15.5% H_2O
			Pairs of plots per field	Ears per plot (60' row)	Grain weight per ear with 15.5% H_2O	H_2O content of grain in field	Gross yield of grain with 15.5% H_2O	Net yield of grain with 15.5% H_2O	Standard error net yield	
					(lb)	(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:

Y = 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 individual-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 stalk count	Mature ears produced	August 1 stalk count	Mature ears produced
10	10	45	45
15	16	50	50
20	21	55	55
25	26	60	59
30	31	65	64
40	41		

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 with "ears" August 1	Ratio of August 1 "ear" count to mature ears produced	Stalks with "ears" August 1	Ratio of August 1 "ear" count to mature ears produced
(pct)		(pct)	
5	.10	60	.87
10	.23	70	1.00
20	.36	80	1.14
30	.49	90	1.27
40	.62	100	1.40
50	.74		

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:

$$Y = 0.0854X - 0.304 \qquad (4)$$

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:

$$Y = 0.0886X - 0.310 \qquad (5)$$

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 11 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.)	(lb)	(lb)
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	

2.5.3.2 Relations for a September 1 Yield Forecast

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:

$$Y = 0.0890X - 0.215 \qquad (6)$$

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 Wet-Kernel Weight and Fraction of Total Dry Matter Laid Down

Average ratio of dry-kernel weight to wet-kernel weight	Average fraction of total dry matter laid down	Average ratio of dry-kernel weight to wet-kernel weight	Average fraction of total dry matter laid down
(pct)	(pct)	(pct)	(pct)
10	5	50	70
20	15	60	85
30	30	70	95
40	50	80	100

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		
				Number of kernels	Dry-kernel weight	Wet-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 Dry-Ear Weight by Days After Silking

Days after silking	Maximum circum. (after husk removed)	Kernel surface area 1/ (husk removed)	Total ear weight (wet)	Cob length over husk	Surface area 2/ (over 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 <u>estimate</u> of grain weight at any given time after silk emergence. The <u>forecast</u> 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 <u>survival</u> (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:

$$\hat{Y}_t = \frac{1}{\hat{\alpha} + \hat{\beta}(\hat{\rho})^t} \quad .$$

An alternative but equivalent form is:

$$\hat{Y}_t = \frac{\hat{\alpha}'}{1 + \hat{\beta}'(\hat{\rho}')^t}$$

This is a nonlinear model, where t is the independent time variable, Y_t is the dependent variable, and α, β and ρ are the parameters which can be estimated from data sets of the form:

$$\begin{array}{cc} t_1 & Y_{t_1} \\ t_2 & Y_{t_2} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{array}$$

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 kernel) 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 $\hat{Y}_m = \lim_{t\to\infty} \hat{Y}_t = \frac{1}{\hat{\alpha}}$. That is, for large values of t, $\hat{Y}_t$ depends upon $\hat{\alpha}$. Therefore, the parameter α 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, $\hat{\alpha}'$, 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 $\hat{Y}_0$ to $\hat{Y}_m$ as t increases. For $0<\rho<1$ the model is indeed a growth model and ρ 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 non-linear 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 = dry weight per kernel at time t', the initial or starting values used were $\hat{\alpha} = 3.8$, $\hat{\beta} = 130$, 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 (IV)	Six (VI)	Seven (VII)	Eight (VIII)	Nine (IX)
α	.0059597	.0061557	.0062934	.0063149	.0063487
β	.12777	.12930	.14616	.14869	.15380
ρ	.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 $\hat{Y}_{t'}$ = estimated dry weight (gm) per kernel at t' days after silks emerged, based upon data from visits I-VII, the model is

$$\hat{Y}_{t'} = \frac{1}{3.8654 + 333.95(.87113)^{t'}} .$$

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 silks started to dry (t)	Estimated dry-grain weight per ear ($\hat{Y}_t$)	Time after silks emerged (t')	Estimated dry-grain weight per kernel ($\hat{Y}_{t'}$)
(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	120	.2587
∞	158.90	∞	.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, α.

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 $(\frac{\hat{\sigma}_{\alpha}}{\hat{\alpha}})$.

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	Independent time variable	Data from visits	Departure of forecast from actual mean dry weight	Relative standard deviation of estimate of primary parameter ($\hat{\alpha}$)
			(pct)	(pct)
Dry weight of grain per ear (Y_t)	Days after silk starting to dry (t)	I(only)	(No convergence to model)	
		I & II	+22.0	34.46
		I - III	+0.7	6.96
		I - IV	+7.8	4.32
		I - V	+8.5	2.74
		I - VI	+4.4	1.74
		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 weight of grain per kernel ($Y_{t'}$)	Days after silk emerged (t′)	I(only)	-89.6	16.07
		I & II	-71.4	12.09
		I - III	-37.3	10.10
		I - IV	+4.6	6.61
		I - V	-6.0	2.24
		I - VI	-1.8	1.59
		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	Age of Tree (years)
1	4- 9
2	10-14
3	15-24
4	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)				Indicated[2/] sample size	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, (x_i/p_i), and derivation of the probability, (p_i), 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 "limb 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 ≐ 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 of citrus	Diameters of fruit size classifications		
	"Regular" bloom	"First late" bloom	"Second late" bloom
	(in.)	(in.)	(in.)
Grapefruit	over 1 1/14	13/16 - 1 1/4	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-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 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.

2.5.5.4 Forecasting Average Harvest 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 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:

$X = 4.34 + .964X_1 - .159X_2 - .002X_3$ (for early-mid oranges on the October 1 forecast date)

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 $S = a + bX + \frac{c}{X}$, 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:

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

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

Y = yield per tree in boxes of fruit

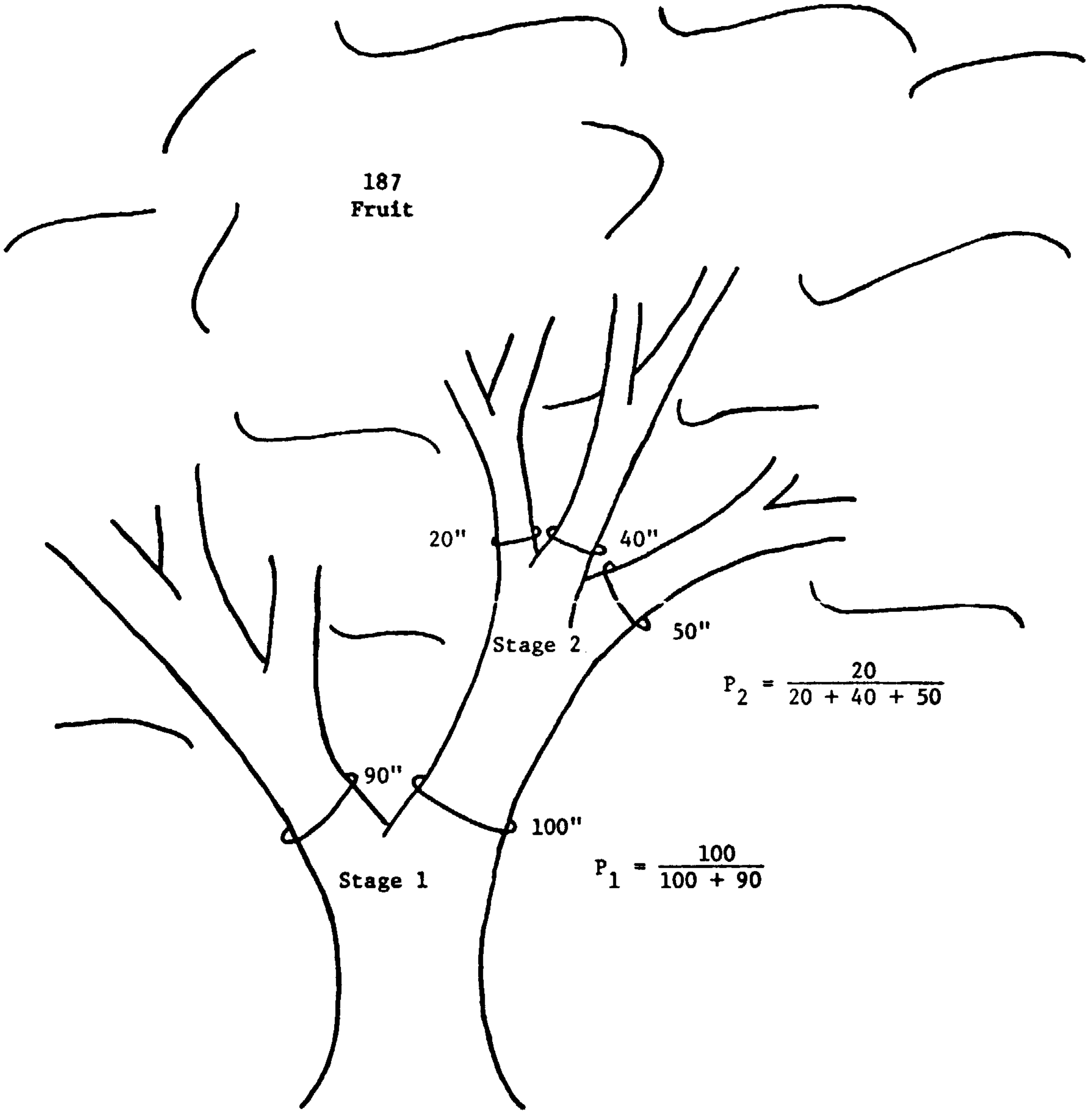
F = number fruit per tree at time of limb-count survey

H = proportion of fruit to be harvested

S = harvest size of fruit expressed in fruit per box

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

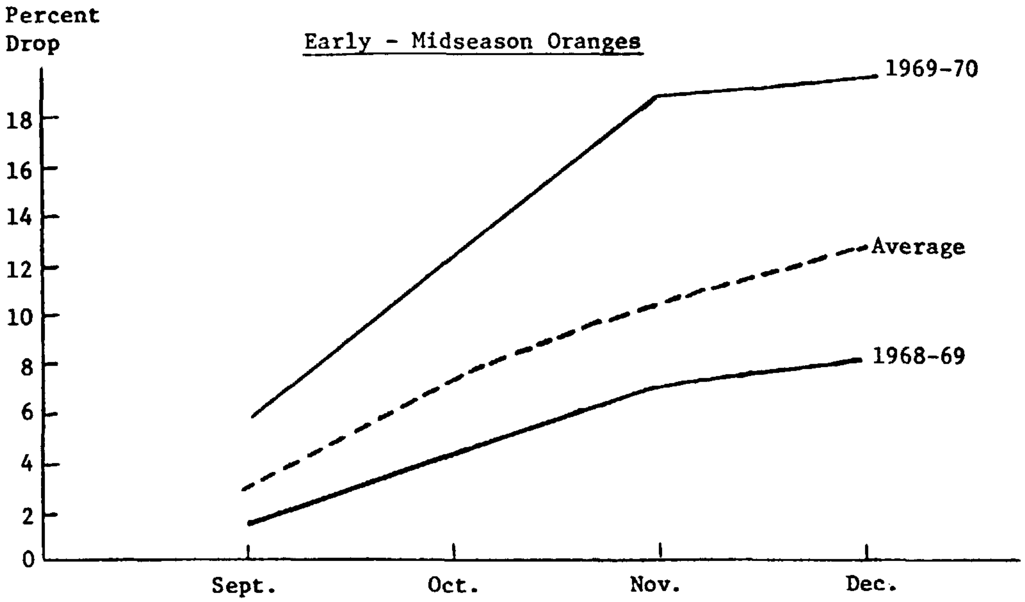
Figure 2: Random Limb Selection With Probability Proportional to Cross-Sectional Area



Estimated Fruit per Tree

$$\text{Fruit Count} \times \frac{1}{P_1} \times \frac{1}{P_2} = 187 \times \frac{100 + 90}{100} \times \frac{20 + 40 + 50}{20} \doteq 1954$$

Figure 3: Fruit Drop Curves
Extreme Years and Average of 1963 - 1969 Seasons



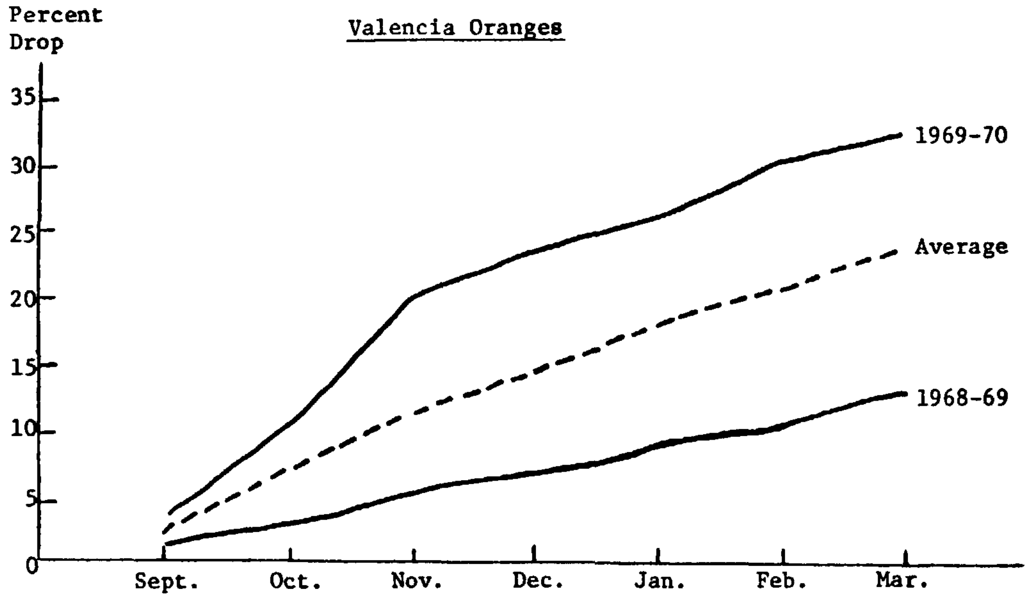
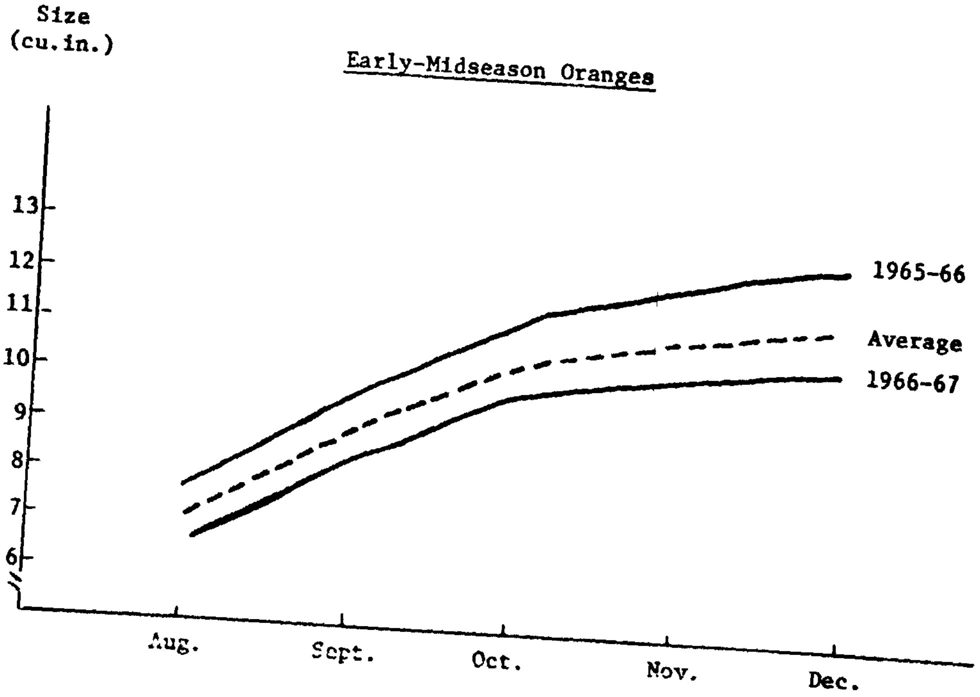


Figure 4: Fruit Growth Curves
Extreme Years and Average of 1963-1969 Seasons



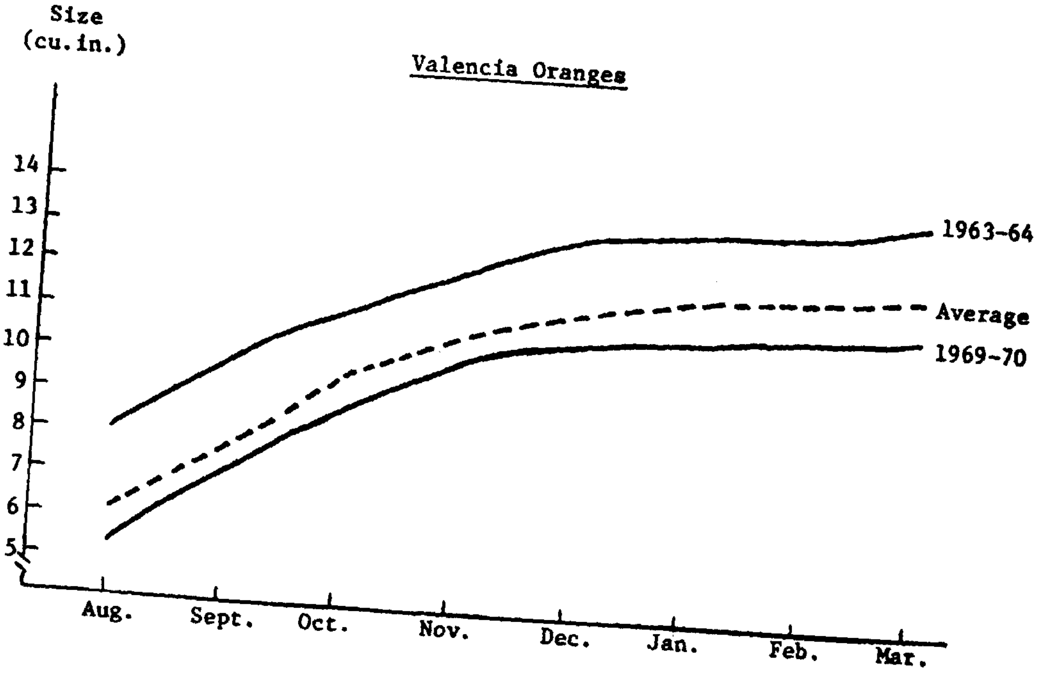


Figure 5: Converting Volume to Fruit per Box, Seedless Grapefruit

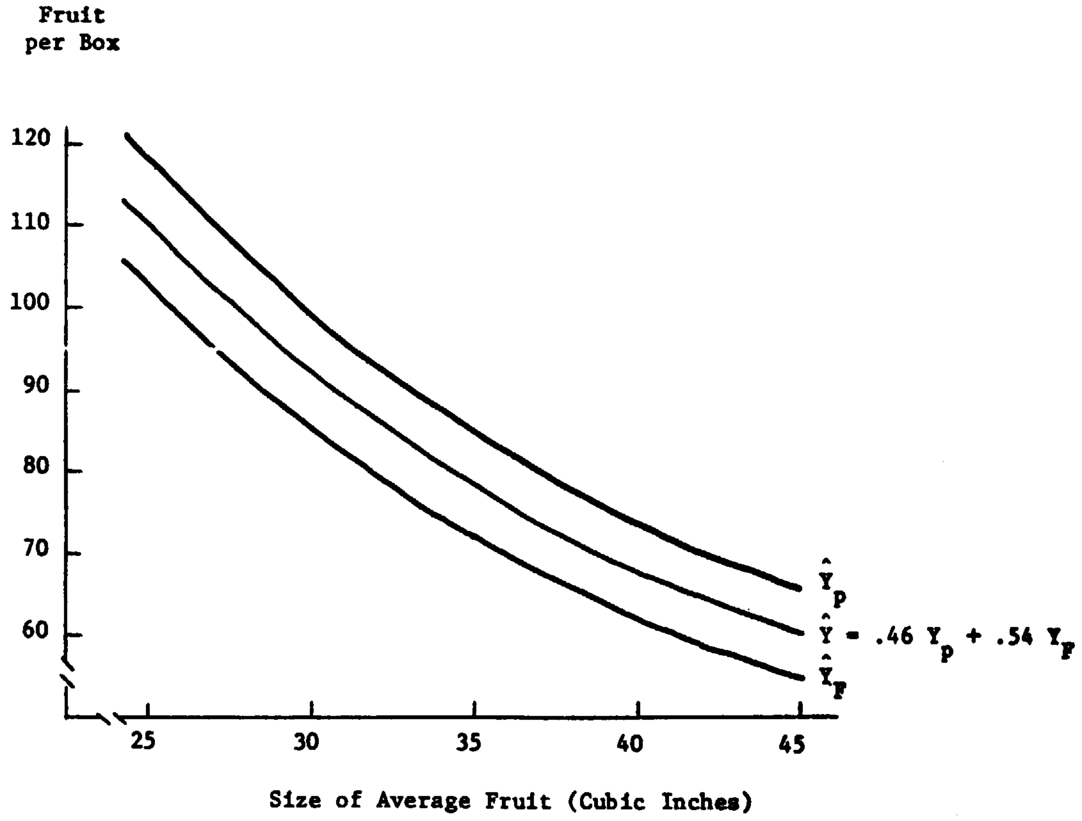
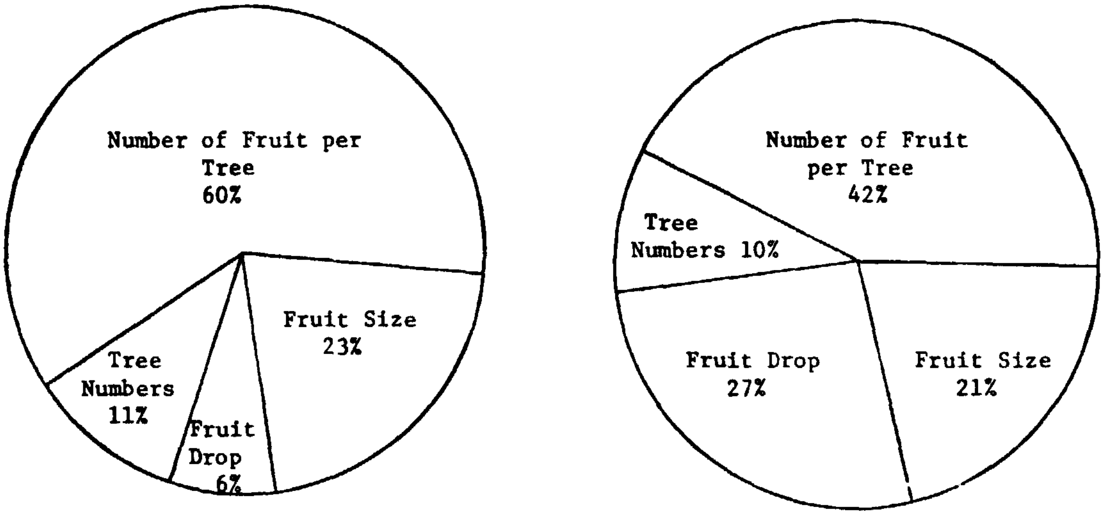
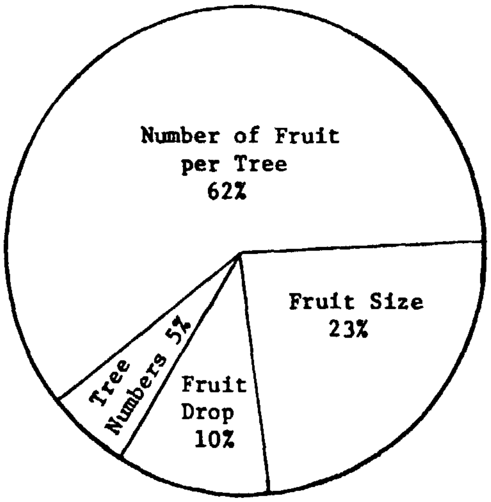


Figure 6: Relative Importance of Factors Affecting Average Annual Change in Florida Citrus Production 1960-61 to 1967-68



Early and Midseason Oranges

Valencia Oranges



Seedless Grapefruit

2.6 Simulation Models Based on Plant Physiology

2.6.1 Introduction

Crop-growth simulation models which consider the soil-plant-atmosphere 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 Model

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

Figure 7: Flow Chart for Sorghum Simulation Model

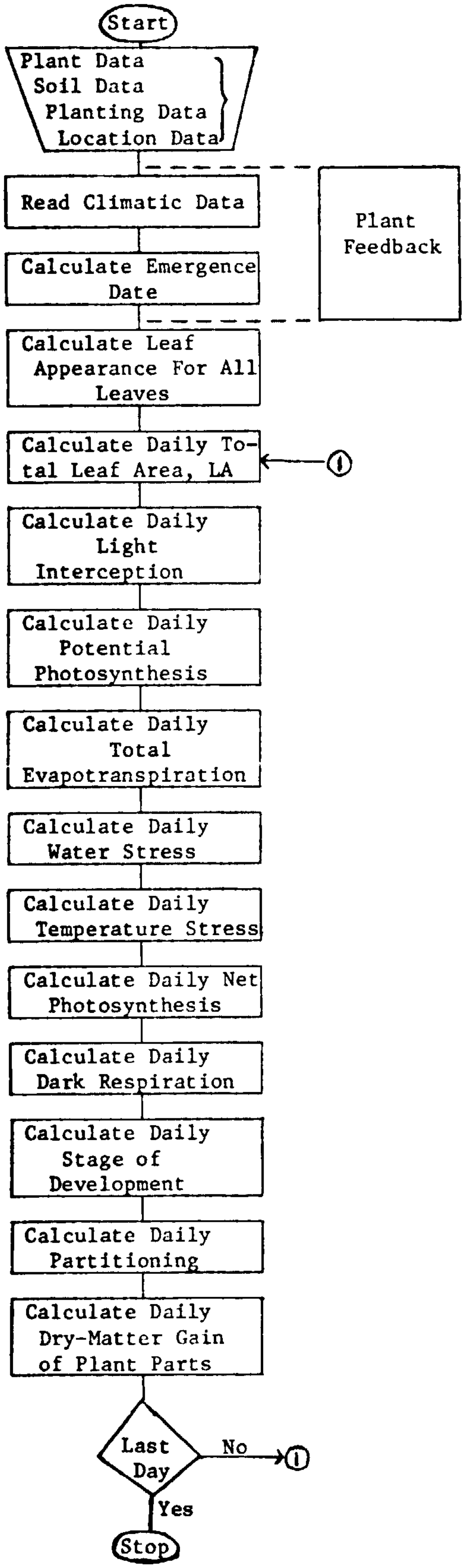


Table 20--Input Data Required for Sorghum Simulation Model

A. <u>For Model Without Feedback Data</u>

Plant Data

Leaf number: total number of leaves produced
Leaf area: maximum area of each individual leaf, 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, °C
Minimum temperature, °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. <u>For Model With Plant Feedback Data - for Specific Date(s) During Growing Season</u>

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°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 μ 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°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 CO_2 fixed during the daylight hours on a ground area basis for nonlimiting water and temperature conditions, 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