A DISCRETE STATE MODEL FOR SCHEDULING ONCE-OVER HARVEST OF PICKLING CUCUMBERS

Thesis for the Degree of Ph. D.
MICHIGAN STATE UNIVERSITY
ASHOK KUMAR PATEL
1973

This is to certify that the

thesis entitled

A DISCRETE STATE MODEL FOR SCHEDULING ONCE-OVER HARVEST OF PICKLING CUCUMBERS

presented by

ASHOK KUMAR PATEL

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in Dept. Agri. Engg.

John B. Holtman
Major professor

Date Feb. 22, 1973

O-7639

ABSTRACT

A DISCRETE STATE MODEL FOR SCHEDULING ONCE-OVER HARVEST OF PICKLING CUCUMBERS

By

Ashok Kumar Patel

The general problem of scheduling formed the basis of the thesis. In particular a cucumber fruit development model was developed to assist the manager in making a harvest time decision for once-over mechanical harvest. The application of the model for the grower is in making reliable forecasts of the harvest data of a field 2-3 days in advance to optimize his expected gains. The application of the model for the processor includes scheduling to reduce field losses due to unforseen shortages of equipment and/or labor and regulation of the flow of desired size distribution mix into the plant.

The model proposed is a class of general linear dynamic models of discrete time and constant parameter form. The model can be viewed in light of discrete state space theory or population dynamics. Both views are helpful. The states of the model are the number of fruit of various size grades at time T. The output from the model is the weight of fruit of various size grades at time T. The input to the model is a unit step that corresponds to a constant number of new fruit of smallest size grade entering into the system each day. The parameters in the model were estimated using the technique of multiple regression with

least squares criterion.

Model verification was attempted using both the available statistical tests as well as analyses of the time behavior from the known historical record. Model verification along with sensitivity analysis has proved the general adequacy of the model in making harvest time decisions for different locations and seasons.

The model was implemented for on-line applications via a teletype-telephone-time share computer hook-up.

The following general conclusions can be made:

(1) The model proposed for cucumber fruit development is valid under the assumptions made in its development.

(2) The model shows promise of being a valuable tool in accurately scheduling 2-3 days in advance the optimal harvest date.

(3) The model implementation is feasible economically and operationally.

Approved [signature]
Major Professor

Approved D. R. Heldman
Department Chairman

A DISCRETE STATE MODEL FOR SCHEDULING ONCE-OVER HARVEST OF PICKLING CUCUMBERS

By

Ashok Kumar Patel

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Agricultural Engineering

1973

ACKNOWLEDGMENTS

The author wishes to express his appreciation to: Dr. J. B. Holtman for his guidance, encouragement and support throughout the course of the thesis.

Dr. J. B. Kreer, Dr. T. J. Manetsch and Dr. Habib Sakhi for their suggestions, constructive criticisms, and review of the thesis.

The National Pickle Grower's Association and Pickle Packer's International for providing financial support.

Daley Pickles, Saginaw, Michigan for providing the data utilized in the study.

The staff of the Department of Agricultural Economics for their help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Michigan is the leading state in the production of cucumbers for pickling, producing 25 to 85 per cent of the nation's total. The cucumber is the most important processing vegetable grown in Michigan, representing 60 per cent of the acreage and 50 per cent of the value of processing vegetable crops (1). Pickling cucumbers also account for a significant part of the vegetable processing industry in Wisconsin, California and North Carolina.

1.1 Purpose of the Study

The trend toward growing more cucumbers strictly for processing, the rising cost of farm labor and its relatively uncertain availability in the past few years, coupled with a need to increase grower income and worker productivity all contribute to the need for the mechanical harvest of this crop. As a result, in just a few years mechanical harvest of pickling cucumbers has increased from zero to a substantial majority of all acreage. In 1968 mechanical harvesters harvested approximately 20 per cent of the Michigan acreage. Estimates place the 1972 acreage mechanically harvested at 90% (2).

Initial attempts to develop a multi-pick cucumber harvester failed primarily because of their poor and inconsistent performance (3). This led to the development of machines for destructive or once-over harvest. In 1964 Stout et al. (3) described a harvester prototype using the constriction principle for fruit removal combined with a mechanism to convey plants to the rollers. The fruit were conveyed to containers and the vines discharged. Commercial production of this machine, with modifications, began in 1964.

The use of once-over machines solved some of the earlier problems of high labor costs and decreasing productivity. However, it gave birth to entirely new problems of decreasing quantity as well as quality of product in relation to the hand harvested product. Research conducted by Cargill and associates in 1969 (2) indicated that hand harvested production averages approximately 200 bushels per acre; whereas once over machine harvested production averages approximately 100 bushels per acre--a 50% reduction. The same research also revealed that both the quantity as well as the quality of the product delivered by present mechanical systems of harvesting, handling and unloading is dependent upon many factors among which physical characteristics of the equipment and the time of harvest are very critical. The latter is the subject matter of this thesis.

1.2 Scope of the Study

Time of harvest influences not only the grower's income and the processor's raw product size distribution, but it has also been found to be related to such complicating factors as the presence of the number of stems and the extent of damage in mechanically harvested cucumbers. The study conducted by Cargill and associates (2) showed that the cucumbers harvested in the afternoon had more stems and less mechanical injury than cucumbers harvested in the morning.

Table 1 shows two examples of the variation in cash value of fields of cucumbers over time, utilizing a common price schedule. Other price schedules produce similar levels of variability. These trajectories are not atypical. In 24 hours the value of a field typically changes 5 to 10 per cent and sometimes as much as 20 per cent or more. Daily increase in weight ranges from 10 to 30 per cent but under favorable conditions of growth, it may be as high as 40 per cent or more. It has been observed (4) that the maturity change is rapid with pickling cucumbers and a field may go from optimum maturity to overmaturity or culls, in just a day or two. In hot weather with adequate soil moisture, the period for optimum harvest may be as brief as 12 hours. As suggested by Table 1 the maximum yield and the maximum value of the crop does not occur at the same time. Packers often differ in their requirements for cucumber size grades.

Table 1. Two typical fruit development patterns.

Date	N_{1A}	N_{1B}	N_2	N_{3A}	N_{3B}	N_4	W_{1A}	W_{1B}	W_2	W_{3A}	W_{3B}	W_4	Cash* Value/ Acre
						Field A							
7/17/71	35	48	50	16	3	2	0.4	1.8	6.7	4.0	1.0	1.0	$179
7/18/71	30	40	54	31	5	3	0.3	1.3	6.9	7.2	1.7	1.4	$220
7/19/71	20	31	31	37	25	4	0.2	1.0	4.4	9.3	9.5	2.5	$320
7/20/71	14	27	39	32	23	13	0.2	0.9	4.8	7.3	8.2	6.2	$285
						Field B							
8/14/71	27	31	32	25	18	5	0.4	1.3	4.0	6.0	6.4	2.3	$332
8/15/71	25	31	30	23	24	11	0.4	1.3	3.8	5.6	8.1	5.3	$251

*Estimated based upon 70% by weight machine recovery.

Therefore, the grower and the processor usually agree before the growing season on the system of payment.

The harvest scheduling problem is further complicated by the need for a forecast of the optimum harvest date one to two days in advance of the optimum date. This information is required for accurate planning of field, transport and plant processing operations. Sims and Zahara (5) suggest that the entire harvesting operation should be considered as one coordinated operation, involving the fruit to be harvested, personnel for sorting, trucking equipment to receive and to transport the fruit, fork lifts, containers for the fruit, and the handling of the fruit at the packing plant. Delay in any of these operations contributes to yield and/or quality losses.

Recommendations as to the best time of harvest for once-over harvested systems have been made (5, 6, 7, 8). Some of these were purely intuitive while others were based on conventional analyses given the existing data. Nearly all of them recommend a time of harvest based on number and/or color of oversized fruits (larger than 2 inches in diameter) in the field. For example, Putnam (6) proposed that the stage of growth for maximum yield in once-over harvesting of cucumbers usually occurred about 2 days following the appearance of grade 3 fruit (1-1/2 to 2 inches in diameter). Morrison and Ries (7) proposed that the most effective harvest index to obtain the highest dollar yield per acre was the development of a yellow color on the most mature fruit (2 to 2-1/2 inches in diameter). Miller and Hughes (8) indicated that the maximum return per acre occurred when the proportion of fruit greater than 2 inches in diameter ranged from 14 to 31 per cent. The same source also suggested that the planting should be harvested as soon as fruit larger than 2 inches in diameter is found in the field. The workers at the Asgrow seed company (5) suggested that the maximum return in dollars will occur when approximately 5 cucumbers per 15 feet of row (in twin-row beds) have just begun to turn yellow at the blossom end.

Most of these harvest indices have the following shortcomings:

(1) They are highly subjective in nature, therefore they are hard to implement without the help of an experienced person.

(2) The measure of their effectiveness is not clear since, as indicated earlier, there is usually a price contract between the processor and grower which varies with the season and the location.

(3) They generally do not predict in advance of the harvest date.

(4) They do not predict product size composition, i.e., the distribution of number and weight of cucumbers in the various size classes.

1.3 Objective

As an initial step toward eliminating these shortcomings, the following objectives were selected for the present thesis:

(1) Development of a prediction model which can be used to forecast 2-3 days in advance, the number and the weight distribution of cucumbers in a field just prior to the normal time of harvest.

(2) Evaluation of the above model for different locations in two growing seasons and for different fruit price schedules.

(3) Devise possible methods to implement the above model.

2. REVIEW OF LITERATURE

The factors that can affect the outcome of a cucumber crop are numerous--type of soil (5), fertilizers (1, 9); irrigation (9), seed bed preparation, plant spacing (5); plant population (9), planting schedules (5, 9); weed control, insect control, bee activity, diseases variety (5); climate (1, 9, 11); and other environmental factors (10), etc. Of major interest in this study are the factors that affect the rate of fruit enlargement at or near the time of harvest. The literature concerning this aspect, particularly in the case of cucumbers, is rather scarce. It is known, however, that daily fluctuation in enlargement of fruits, in general, is a combination of several factors, among which the intensity of the incident solar radiation, the evaporating power of the air, and night temperature play an important role (12).

2.1 Factors Affecting Rate of Fruit Growth

In the apple, enlargement of developing fruit has been shown to occur mainly during the night (12, 13). The rate of enlargement was a function largely of night-temperatures. Both under field and greenhouse conditions, fruit enlarged only slightly during the daylight hours.

Further studies by Tukey (14) showed that the amount of daily enlargement was influenced mainly by night temperatures. A raising of night-temperature increased the amount of apple fruit enlargement during the night, but reduced the amount of daytime fruit enlargement. Went (15) found that plants grown under a temperature cycle (diurnal thermoperiodicity) of 78.8°F during the day-period and 64.4°F during the night-period grew vigorously and had more fruiting compared to plants grown under a constant temperature. Aldrich and Work (16) reported that the rate of pear fruit enlargement was influenced by soil moisture differences within the available soil moisture range. Harley and Masure (13) felt that with sufficient soil moisture, high evaporating power of the atmosphere had a greater retarding effect on fruit growth than high temperature alone. Nevertheless, these two factors were found to be closely associated.

2.2 Factors Affecting Rate of Growth of Cucumbers

The exact nature of the variables and the specific mechanism with which they affect the rate of fruit enlargement for cucumbers is largely beyond scientific understanding. Some workers believe, however, that the factors which affect the rate of fruit enlargement for fruits in general are also the factors affecting cucumber fruit enlargement. The recent work of Lewin (11) is a case in point. Other workers have found that yield and cash value per acre are more responsive to some factors than to others. Among the most widely studied factors in the literature, under both categories, are: temperature, soil moisture, plant population and response to nitrogen.

2.2.1 Temperature

Several workers have investigated the minimum and optimum temperatures for growth of cucumbers. The work of Lewin (11) is of particular interest. He studied the effect of several controlled phototemperature and nyctotemperature conditions on cucumber fruit enlargement (expressed as fruit diameter enlargement) during a period of 10 days after the fruit set, when grown under constant irrigation and under a moisture stress condition. Some of his findings were:

(1) The most favorable conditions for cucumber fruit enlargement were a phototemperature of 85°F and a nyctotemperature of 65°F under constant irrigation, or nyctotemperatures in the range of 55 to 75°F under intermittent irrigation.

(2) Maximum daily enlargement generally occurred with a phototemperature 20°F higher than the nyctotemperature.

(3) Cucumber fruit enlargement appeared to be limited mainly by phototemperature, than by available soil moisture and finally by nyctotemperature.

According to Bailey (17) the most suitable temperatures for "rapid growth" were between 60 and 65°F at night and up to 100°F in bright sunshine with an ample supply of soil moisture. Miller (1) reported that night temperatures of 60°F produced fruit with a greater length to diameter ratio than those grown at 70°F. Studies by Morrison (9) indicated that 50°F was approximately the critical minimum temperature for development of cucumber plants and fruit.

2.2.2 Soil Moisture

Under controlled growth chamber conditions Lewin et al. (11) observed that the rate of cucumber fruit enlargement appeared to be greater and more uniform when the plants were not under moisture stress conditions for approximately the first five days of enlargement after the fruit set. During the second five day period, the moisture conditions did not appear to influence fruit enlargement.

Morrison (9) observed that inadequate soil moisture resulted in poor seed germination, variable plant growth and maturity. Workers at the Asgrow seed company (5) have observed that the most critical need for soil moisture for cucumbers comes during the fruiting period, when lack of moisture can seriously reduce the yield of marketable fruit.

2.2.3 Plant Population

Concurrent with the development of a machine to harvest cucumbers in a once-over manner has been research to study the most suitable plant populations. Putnam (6) evaluated populations ranging from 22,000 to 87,000 plants per acre. Populations exceeding 43,560 plants per acre did not produce higher yields.

Results from research conducted in North Carolina indicated that the value of the crop increased with increasing plant populations. However, at the highest rates (224,000 plants per acre) the fruit, particularly the larger sizes, tended to be pointed on the blossom end (18).

In studies with a once-over harvesting machine at Michigan State University, the highest value in both bushels and dollars per acre was obtained from plant populations of 31,500 in comparison to lower populations (19).

2.2.4 Fertilizers

Dearborn (20) found that high nitrogen increased vegetative growth, fruit size number, and the percentage of female flowers produced in cucumbers. He concluded that a relatively high nitrogen level is essential for the production of a maximum yield of highly-colored, straight, well-shaped fruit. Vaile (21) reported that the set and number of marketable cucumber fruit were greater under high than under low nutrient conditions. Rodinkov (22) reported that nitrogen was the most limiting nutrient during growth and flowering, while potassium became the dominant element during fruit formaticn.

Miller (1) found that 60 and 90 pounds per acre of nitrogen reduced yields below those produced from 30 pounds per acre. Morrison (9) observed that nitrogen rates of approximately 60 pounds per acre were adequate for production of high once-over harvested yields. A recent publication by McCollum and Miller (23) indicated that although the well-fertilized plants made more vigorous growth than those with some nutritional stress, the increase in the number of fruit set was only about one per plant.

2.3 Rate of Growth Curves for Cucurbit Fruits

Gustafson (24) constructed enlargement curves for muskmelon, cucumber, scalloped summer squash and tomato fruits. When the increase in fruit volume was plotted against time, all fruits had the same general enlargement curve: a slow increase in the beginning, followed by a very rapid increase, and then a gradual decrease and cessation of enlargement. He compared the enlargement of fruits to that found for animals and vegetative plant structures which grow according to the well-known "s" pattern.

Sinnot (25) divided the enlargement period of cucurbit fruit into three parts: (i) from origin to flowering, (ii) from flowering to the break in curve (the end of exponential enlargements), and (iii) from the break in the curve to final size. When the logarithms of fruit volume were plotted against time, measured points fell along an essentially straight line which flattened out at the end of the enlargement period. His main conclusions were as follows:

(1) Enlargement consists of two phases, an initial one at a constant exponential rate, followed by a phase at a continually decreasing rate until enlargement ceases.

(2) There is no relationship between enlargement rate, either during the exponential period or later, and final size.

(3) There are differences in enlargement rates between various genetic lines.

(4) Environmental factors, insofar as these are reflected in seasonal differences, markedly affect the enlargement rate.

(5) There is an inverse relationship between enlargement rate and enlargement duration in various seasons. When the rate is high, final size is attained more rapidly than when the rate is lower.

3. MODEL DEVELOPMENT

3.1 Existing Approaches

The harvest schedule of a crop grown for processing should allow for the optimum use of all resources, including equipment, labor and land. This necessitates a uniform and continuous harvest sequence. To achieve this, a crop must be planted in an orderly manner so that the quantity to be harvested during any given period does not exceed the harvesting or processing capacity (7).

Two methods have been suggested for timing successive plantings for mechanical harvesting. One method, developed by workers in Michigan and California (6), is based on a stage-of-growth system, under which the second planting is made when the first true leaves begin to emerge between the cotyledons in the seedling of the earlier planting. In areas where crops are seeded in midsummer for late summer and fall harvest, this system of timing may lead to gaps and bunching at harvest time (6).

The other system, more widespread in use (26, 27) is based on the accumulation of heat units (H.U.). This system employs the use of a base temperature, below which it is assumed no growth takes place. The mean of daily maximum and minimum temperature is determined, the base temperature is subtracted from this mean, and the resultant value is referred to as heat units. For cucumbers, the base temperature has been found to be 55°F and investigations (6) have indicated that an accumulation of 75-100 H.U. is required for germination and a total of 850 to 1,000 H.U. is required for the period from planting to harvest. In planning a planting schedule according to the heat unit system, it is important to consider such factors as the normal daily mean temperature during the expected time of harvest and the daily acreage capacity of the harvesting equipment (4).

The Michigan pickling cucumber industry utilizes various heat unit systems to schedule crop planting such that the harvest volumes are distributed over the harvesting season--approximately July 20 to September 15. However, these systems are not sufficiently accurate for harvest date scheduling. As pointed out by Arnold (28), heat units, at least in their present state of development, are not reliable because the relationship between temperature and the rate of plant development is assumed to be linear when it is undoubtedly curvilinear. He also suggested that factors in addition to temperature affected the rate of development, and that the temperature measured at a single location is used as a basis for the prediction in the varied micro-climate of many fields. Furthermore, a harvest date scheduling system based on planting date, weather over the growing season, soil type, etc. does not seem practical at this time. The modeling work and the associated weather monitoring system required would be very formidable obstacles to such an endeavor (29).

3.2 An Alternative Approach

Some elementary notions from dynamical systems theory suggested an alternative approach. In light of this theory, we can consider our problem to be the description of the dynamics of fruit size distribution for a five-day period just before harvest. We then define the universe as everything outside this fruit size distribution. Interaction between these sets are abundant, such as the influence of soil chemicals, moisture, light exposure, temperature, etc. on plant growth and hence on fruit growth.

Under controlled growth chamber conditions Lewin et al. (11) observed that maximum rate of growth of cucumber fruit occurred when the maximum air temperature was 85°F and the diurnal cycle was 20°F. Physiological considerations suggest that fruit development might be functionally dependent upon both photosynthetic rate and moisture availability. Incident radiation was hypothesized to be indicative of photosynthetic rate. Moisture availability is critically dependent upon root development, soil type and rainfall. Root development is in turn dependent upon soil moisture conditions early in the growing season.

To include all of the interactions such as described above in a forecast model would necessitate a very complex mathematical model. In general, the mathematical model should be formulated to yield a reasonably accurate description or prediction of the behavior of a given system while minimizing computational and programming cost and time (43). Thus, boundaries were imposed on the scope of the defined problem. Direct measurement of factors such as soil chemicals, nutrient levels, etc. were not included in this study because they are difficult to quantify. The effect of these factors, however, was studied indirectly by estimating parameters from the observed data for different locations (See Chapter 7). Measurement of soil moisture is not only expensive and time consuming, but this particular variable can not be predicted accurately because accurate rainfall forecasts are not available. Although this factor is important, it was not possible to include it in the prediction model. It was thought, however, that some measure of its effect can be obtained by estimating parameters from the observed data for different seasons. Factors such as temperature, radiation and relative humidity were included in the model because, not only is it inexpensive to record them, but also their forecasts are readily available.

Intuitively, it was then reasoned that if the state of a field of pickling cucumbers could be measured on day i, then knowledge of input to the field (number of new fruits and temperature were thought to be of major consequence) might permit the prediction of the state of the field on day i+1. Given this capability i can be set equal to i+1 and the process can be repeated. To minimize the data required for the parameterization of the prediction model, only the time period coincidental with the 5 days just prior to the optimum date of harvest was considered. This limitation seems reasonable because a quick visual inspection is sufficient to determine whether or not a field has entered this period. Furthermore, these inspections are a part of the current commercial mode of operation. Thus, the result of these inspections can be utilized to initiate the measurement of the system state.

The next step is the selection of the state variables. The possible candidates for state variables in this problem are the number and the weights of each fruit size grade. In a later section we will establish that there are reliable weight to number relations for fruits of various size grades; therefore we need only consider fruit weight or number as the state. It is helpful to consider here that the output of the model in any case should be the weight of cucumbers in various size grades. This is so because the cucumbers are valued on the basis of the weight. This suggests that, if a mathematical framework can be worked out based on classification of weight alone, we could select weights as the state variables, which would then also be the output of the model.

This possibility was, therefore, given consideration. It was found early in the investigation, however, that the selection of state variables based on weight, though presenting no mathematical framework problems, does present problems with regard to estimating parameters associated with the growth of fruit. This notion is clearly demonstrated by the following example. Let:

W_c = weight of fruit of a given class c at time T

Z_c = increase in weight of that fruit which remains in class c during the time period T to T + ΔT

Y_{c+1} = the weight of that fruit (measured at time T) which was in class c at time T but in class c+1 at time T+ΔT

Y_c = the weight of that fruit (measured at time T+ΔT) which was in class c-1 at time T but in class c at time T+ΔT.

Assuming no loss of fruit from the system an input-output relationship can be written as follows:

$$W_c(T+\Delta T) = W_c + Z_c + Y_c - Y_{c+1}$$

The terms Z_c, Y_c and Y_{c+1}, however, can not be measured from the observed sample values and, hence, can not be estimated. On the other hand, no such problem arises if we consider the number rather than the weight of fruit of a given class c.

Let:

$N_c(T)$ = Number of the fruit of a given class c at time T

$X_{c+1}(T+\Delta T)$= the number of those fruit which were in class c at time T but in class c+1 at time T+ΔT.

$X_c(T+\Delta T)$= the number of fruit which were in class c-1 at time T but in class c at time T+ΔT.

Based on the assumption that there is no loss of fruit and that there are two measurable states at time T and T+ΔT, the following input-output relationship can be written for the number of the fruit of a given size grade c.

$$N_c(T+\Delta T) = N_c(T) + X_c(T+\Delta T) - X_{c+1}(T+\Delta T)$$

Notice here, the absence of the variable Z_c. This particular elimination has been possible because a single fruit during the period ΔT can not change its number (it remains one); however, in the same period its weight does change. Thus, the simplicity of the mathematical framework dictated the use of numbers as the state variables.

The next problem considered was to relate the output (weight) to states (number). For this, it is helpful to consider that cucumbers are divided into various size grades based on diameter. Since diameters only help to identify various classes, no single breakdown based on diameter is unique. Any given breakdown, however, does suggest the degree of aggregation used in the model. This implies that we can determine a classification, such that meaningful functional characteristics exist not only for state variables, but also between the state and the output variables. These functional considerations suggested a classification of original number of fruit into six size grades (see following section). Thus, while a classification based on four size grades (sometimes used by the industry) could result in both savings in sampling cost and time, functional considerations challenged this approach. The homogenity was somewhat better on the classification based on six size grades. This classification resulted in nearly constant weight to number relationships for the six size grades.

3.3 Data Collection

In cooperation with two Michigan pickle processors some 200 once-over mechanical harvest fields were sampled on successive days just prior to harvest in the 1971 and 1972 seasons. Moving diagonally across a field (typically 10-20 acres), 10 row lengths approximately equally spaced on the diagonal were randomly selected. The row lengths were selected such that a total of 100 square feet of area was sampled (row lengths of 51-1/2 inches in 28 inch rows). The plants in the selected row lengths were pulled, and all of the fruit were removed, graded, counted and weighed. Continuous recordings of temperature, radiation and relative humidity were obtained in the area of the fields throughout the harvest seasons. The data were collected from two major pickle producing areas of Michigan near Saginaw. These areas were classified as West (Breckenridge and vicinity) and North (Pinconning and vicinity). The exact location of a field in each of the two areas was coded by a three digit number. The number of successive daily samplings per field was usually two to four. The last day of the sequence was usually the day of harvest. Table 1 illustrates examples of samples from two fields collected in 1971. Columns two through seven show the number of fruit in each of the six size grades (1A, 1B, 2, 3A, 3B and 4) as sampled in the field on the corresponding date; columns eight through thirteen show the corresponding weights. The size grades based on minimum diameter measurements (D_{min}) are:

Size 1A	3/8"	$< D_{min} \leq$ 3/4"
Size 1B	3/4"	$< D_{min} \leq$ 1-1/16"
Size 2*	1-1/16"	$< D_{min} \leq$ 1-1/2"
Size 3A	1-1/2"	$< D_{min} \leq$ 1-3/4"
Size 3B	1-3/4"	$< D_{min} \leq$ 2"
Size 4	2"	$< D_{min}$

The other size grade classification occasionally used by the processors to reduce the sampling labor and the associated costs consists of dividing the sample into four size grades rather than six size grades as shown above. In this classification size 1A and 1B are lumped together to form size 1. Similarly, size 3A and 3B are lumped together to form size 3, the other sizes remaining unchanged.

The fruit size distribution shifts to increasing numbers of fruit in the larger sizes as time progresses. The cash value of the field often increases initially. For most contract price systems, the value of size 4 (known as

---

*National Grader divides size 2 into 2A (1-1/16" $< D_{min} \leq$ 1-5/16") and 2B (1-5/16" $< D_{min} \leq$ 1-1/2").

oversize or culls) fruit is zero. Typically, when a substantial number of fruit enters this size grade the value of the field drops sharply. This can be seen clearly for both fields illustrated in Table 1.

3.4 Model Formulation

The objectives of this study are stated in general terms at the end of Chapter 1. Formulation of a mathematical model to meet these objectives involves consideration of both the type of outputs desired and the type of inputs available. These are as follows:

3.4.1 Available Inputs

(i) Number and weight distribution of fruit in a field as sampled on a given day.

(ii) U. S. Weather Bureau forecasts for the next three days in the area of the field (temperature, etc.).

3.4.2 Desired Outputs

Number and weight distribution of fruit in the same field for each of the next three days.

3.4.3 Model Structure

The structure upon which the proposed model characterizes the dynamics of fruit number distribution is shown schematically in Figure 1. As indicated in the schematic,

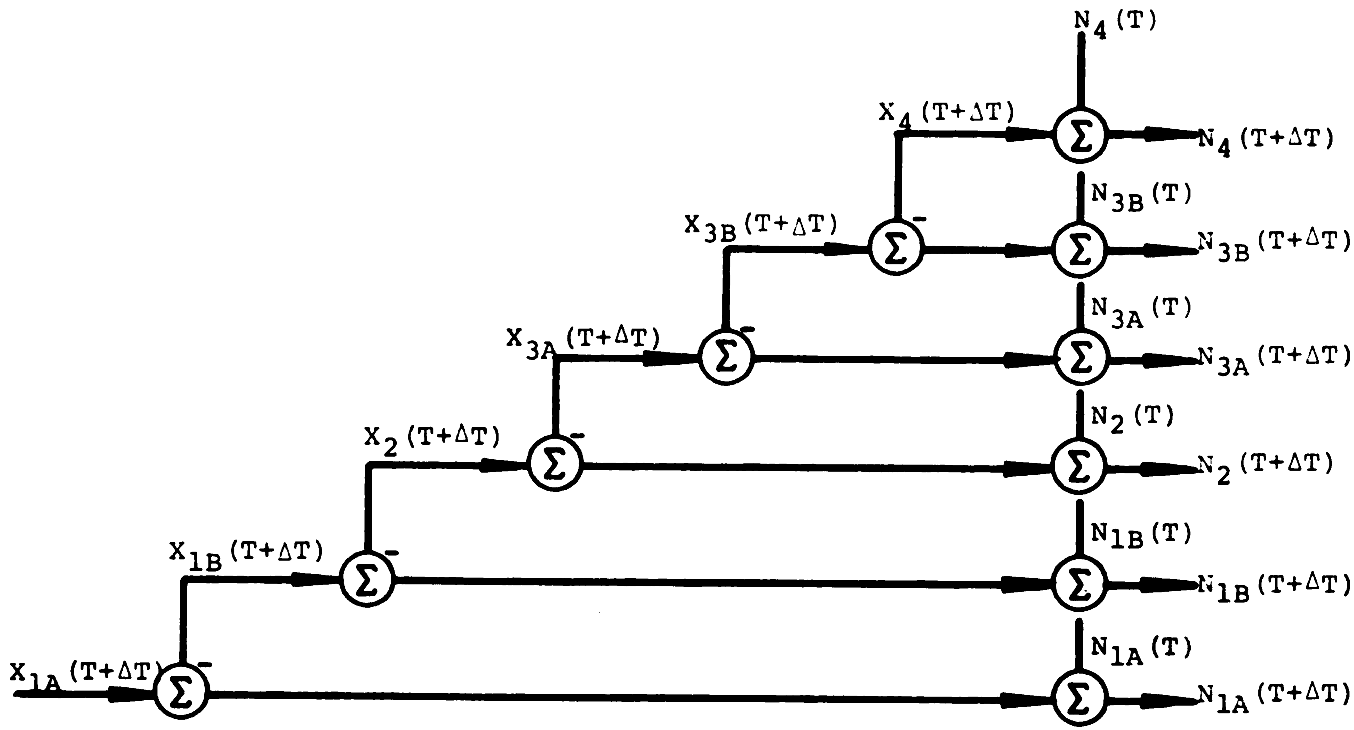

Figure 1. A schematic diagram of the cucumber number-class structure

the basic components of this structure are:

(1) number of fruit in the six size grades

$(N_4, N_{3B}, \ldots, N_{1A})$ at time T

(2) number of new fruit in the six size grades

$(X_4, X_{3B}, \ldots, X_{1A})$ at time $T+\Delta T$

(3) number of fruit in the six size grades at time $T+\Delta T$.

The size (number) distribution of a particular class of fruit at time $T+\Delta T$ is determined by their initial number at time T and the number of new fruit at time $T+\Delta T$. The number of new fruit of a particular size grade at time $T+\Delta T$ is a function of initial distribution and the weather variables, of which temperature is considered to be of major importance. It is to be emphasized that the validity of the resulting model as an instrument of simulation and prediction depends critically on the degree to which time behaviour of number of new fruit can be characterized from the available information and on the assumptions made in the model structure. Basic to the structure proposed here is the assumption that no fruit in any size grade is lost as a result of windfall.

3.4.4 Time Period

Mathematical models can be formulated in either the continuous time form (described by differential equations) or the discrete time form (described by difference equations). The choice of a continuous or a discrete time model depends upon: (1) the level of detail necessary to answer relevant questions; (2) the frequency of events or the flow rate of objects relative to the minimum time interval of interest; and (3) the cost of programming and operating the models (30). The outputs from our model are intended to provide information on a daily basis. The input information is available on a daily basis only. The consideration of these factors along with simplicity and ease of computer implementation led to the selection of a discrete-time (stepping or recursive) model with a time increment of one day. Thus, the ΔT in the ensuing discussion refers to a time interval of one day.

3.5 Functional Relationships

The flow of fruit (Figure 1) through various size grades is characterized by the matrix $\underline{S}$ in the equation:

$$\vec{N}_c(T+\Delta T) = \underline{I}\ \vec{N}_c(T) + \underline{S}\ \vec{X}_c(T+\Delta T) \qquad 3.5.1$$

where:

$$\vec{N}_c = \begin{bmatrix} N_4 \\ N_{3B} \\ N_{3A} \\ N_2 \\ N_{1B} \\ N_{1A} \end{bmatrix} \quad ; \quad \vec{X}_c = \begin{bmatrix} X_4 \\ X_{3B} \\ X_{3A} \\ X_2 \\ X_{1B} \\ X_{1A} \end{bmatrix}$$

$\underline{I}$ is an identity matrix and $\underline{S}$ is a (6 x 6) matrix as shown below:

$$\underline{S} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$\vec{N}_c(T+\Delta T)$ denotes the vector of the number of fruit at time $T+\Delta T$ (e.g., $N_c(T+\Delta T)$ represents the number of the fruit of size grade c at time $T+\Delta T$). $N_c(T)$ denotes the same vector evaluated at time T. $\vec{X}_c(T+\Delta T)$ is a vector representing the number of new fruit by size grade at time $T+\Delta T$.

The notation $\vec{X}_c(T+\Delta T)$ is used rather than $\vec{X}_c(\Delta T)$ to emphasize the fact that the observable states of the system are at time T and $T+\Delta T$. In effect, the number of new fruit in each size grade are in transition, (Figure 1), i.e., the new fruit are in a size grade c-1 at time T but they are in the size grade c at time $T+\Delta T$.

The matrices $\underline{I}$ and $\underline{S}$ remain fixed from one period to the next. The vector $\vec{N}_c(T)$ represents the initial conditions. Thus, knowledge of $\vec{X}_c(T+\Delta T)$ in 3.5.1 will permit the prediction of $\vec{N}_c(T+\Delta T)$.

Since the sequences of field samples of count and weight by size grade, together with the continuous recordings of temperature, radiation and relative humidity were the major sources of data upon which a forecasting model of new fruit could be based, the fruit development model for new fruit was hypothesized to take the form:

$$\vec{X}_c(T+\Delta T) = F[\vec{N}_c(T), \vec{W}(\Delta T)] \qquad 3.5.2$$

where $\vec{W}(\Delta T)$ is a vector of variables summarizing the weather occurring between sampling time T and T+ΔT.

Substituting 3.5.2 along with the initial condition $(\vec{N}_c(T))$ in 3.5.1 will permit the computation of the state at time $T+\Delta T(\vec{N}_c(T+\Delta T))$.

The next problem is to infer the relationships between the output (weight distribution of fruit of each size grade) and the states. This problem will have a relatively straightforward solution if it could be ascertained that there is a definite relationship between the number and the weight of fruit of each size grade c at a given time T. Ideally, this would be the case if the product under consideration is relatively homogeneous (a fixed length to diameter ratio of fruit) and the number of classifications of fruit size grade is infinite. This implies,

$$W_c(T) = e\ N_c(T) \qquad 3.5.3$$

for each size grade c as the number of size grades become infinite. ($W_c(T)$ is the weight of fruit size grade c.)

Let:

$$\vec{W}_c = \begin{bmatrix} W_4 \\ W_{3B} \\ W_{3A} \\ W_2 \\ W_{1B} \\ W_{1A} \end{bmatrix}$$

Where:

$\vec{W}_c(T)$ = a vector representing the weight of fruit at time T by size grade. Thus, $W_c(T)$ is the weight of fruit of size grade c at time T.

$\underline{E}$ = a (6 x 6) matrix of constants having non-zero values on the diagonal and zero elsewhere.

The results of a preliminary regression analysis with the six size grades over the period under consideration (2-5 days before harvest) indicated that there are fairly precise relations between the weight and the number of fruit of each size grade. The results of this analysis are presented later (Section 3.7). Thus:

$$\vec{W}_c(T) \simeq \underline{E}\ \vec{N}_c(T) \qquad 3.5.4$$

and

$$\vec{W}_c(T+\Delta T) \simeq \underline{E}\ \vec{N}_c(T+\Delta T) \qquad 3.5.5$$

That is, the distribution of the weight of fruit (output) at time T+ΔT can be approximately determined from the knowledge of the distribution of the number of fruit

(states) at time T+ΔT.

From 3.5.1 we have:

$$\vec{N}_c(T+\Delta T) = \underline{I}\,\vec{N}_c(T) + \underline{S}\,\vec{X}_c(T+\Delta T)$$

Premultiplying both sides of 3.5.1 by $\underline{E}$ we have:

$$\underline{E}\,\vec{N}_c(T+\Delta T) = \underline{E} * \underline{I}\,\vec{N}_c(T) + \underline{E} * \underline{S}\,X_c(T+\Delta T)$$

or:

$$\vec{W}_c(T+\Delta T) \simeq \underline{E} * \underline{I}\,\vec{N}_c(T) + \underline{E} * \underline{S}\,X_c(T+\Delta T) \qquad 3.5.6$$

Let:

$$\underline{E} * \underline{I} = \underline{P} \qquad \text{and} \qquad \underline{E} * \underline{S} = \underline{P}' \qquad 3.5.7$$

then:

$$\vec{W}_c(T+\Delta T) \simeq \underline{P}\,\vec{N}_c(T) + \underline{P}'\,\vec{X}_c(T+\Delta T) \qquad 3.5.8$$

Thus, the distribution of the weight of fruit at time T+ΔT can also be determined from the knowledge of the distribution of the number of the fruit at time T and the number of the new fruit at time T+ΔT.

The matrices $\underline{P}$ and $\underline{P}'$ can be determined from 3.5.7 or alternatively, they can be estimated directly from the observed sample values. Let the estimated matrices corresponding to $\underline{P}$ and $\underline{P}'$ be $\underline{Q}$ and $\underline{Q}'$ respectively. Then,

$$\vec{W}_c(T+\Delta T) \simeq \underline{Q}\,\vec{N}_c(T) + \underline{Q}'\,\vec{X}_c(T+\Delta T) \qquad 3.5.9$$

and:

$\underline{P} = \underline{Q}$; $\underline{P}' = \underline{Q}'$ if $\vec{W}_C(T) = \underline{E}\,\vec{N}_C(T)$.

That is, any improvement (in the sense of explaining mean square deviation of the dependent variable) made by the relation 3.5.9 over that of 3.5.8 will decrease our confidence in the validity of our assumption in 3.5.4.

It is obvious that the parameters in 3.5.8 can be estimated by specifying the independent variables in several combinations, of which 3.5.9 itself is a special case. Another form which is of considerable interest is as follows:

Let $\vec{O}_C(T)$ represent the vector of number of old fruit at time T by size grade, i.e., the number of fruit excluding the number of new fruit in each size grade at time T. Specifically:

$$\begin{aligned}
O_4(T) &= N_4(T) \\
O_{3B}(T) &= N_{3B}(T) - X_4(T+\Delta T) \\
O_{3A}(T) &= N_{3A}(T) - X_{3B}(T+\Delta T) \\
O_2(T) &= N_2(T) - X_{3A}(T+\Delta T) \\
O_{1B}(T) &= N_{1B}(T) - X_2(T+\Delta T) \\
O_{1A}(T) &= N_{1A}(T) - X_{1B}(T+\Delta T)
\end{aligned} \qquad 3.5.10$$

Substituting 3.5.10 in 3.5.1 we have:

$$\vec{N}_C(T+\Delta T) = \underline{I}\,\vec{O}_C(T) + \underline{I}\,\vec{X}_C(T+\Delta T) \qquad 3.5.11$$

Premultiplying 3.5.11 on both sides by $\underline{E}$ we have:

$$\underline{E}\,\vec{N}_C(T+\Delta T) = \underline{E}\,\vec{O}_C(T) + \underline{E}\,\vec{X}_C(T+\Delta T)$$

then from 3.5.5 we have:

$$\vec{W}_c(T+\Delta T) \simeq \underline{E}\ \vec{O}_c(T) + \underline{E}\ \vec{X}_c(T+\Delta T) \qquad 3.5.12$$

If instead, we estimate 3.5.12 directly from the observed sample values we will have:

$$\vec{W}_c(T+\Delta T) \simeq \underline{H}\ \vec{O}_c(T) + \underline{H}'\ \vec{X}_c(T+\Delta T) \qquad 3.5.13$$

$\underline{H}$ and $\underline{H}'$ should be estimated to check the validity of our assumption in 3.5.4 or 3.5.5. They can be employed to determine the value of $W_c(T+\Delta T)/O_c(T)$ and $W_c(T+\Delta T)/X_c(T+\Delta T)$ if our assumption is not valid.

3.6 Parameter Estimation Procedure

The parameters in 3.5.2, 3.5.3, 3.5.9 and 3.5.13 were estimated by multiple regression utilizing the method of least squares (31, 32, 33). The general linear hypothesis for R explanatory variables and N observations is:

$$\begin{array}{l}
Y_1 = b_0 + b_1X_{11} + b_2X_{21} + \dots + b_iX_{i1} + \dots + b_RX_{R1} + U_1 \\
\vdots \\
Y_t = b_0 + b_1X_{1t} + b_2X_{2t} + \dots + b_iX_{it} + \dots + b_RX_{Rt} + U_t \\
\vdots \\
Y_N = b_0 + b_1X_{1N} + b_2X_{2N} + \dots + b_iX_{iN} + \dots + b_RX_{RN} + U_N
\end{array}$$

Where:

Y_t = observation t of the dependent variable Y

X_{it} = observation t of the explanatory variable X_i

U_t = stochastic disturbance associated with observation t and constants: $b_0, b_1, b_2, \ldots b_i \ldots b_R$.

The method of least squares consists of determining estimates $(\hat{b}_o, \hat{b}_1, \hat{b}_2 \ldots \hat{b}_i \ldots \hat{b}_R)$ of the constants b_0, $b_1, b_2 \ldots b_i \ldots b_R$, such that the sum of the squared residuals is a minimum, i.e.:

$$\sum_{t=1}^{N} \hat{U}_t^2$$

is a minimum.

Where:

$$\hat{U}_t = Y_t - (\hat{b}_0 + \hat{b}_1 X_{1t} + \hat{b}_2 X_{2t} + \ldots \hat{b}_i X_{it} + \ldots + \hat{b}_R X_{Rt})$$

A null hypothesis that the individual b_i's equal zero is established and tested to obtain the regression equation.

An important measure of how much of the variation in the dependent variable may be accounted for by the group of explanatory variables is the coefficient of multiple determination (R^2). R^2 is the proportion of the sum of the squared deviation from the mean of the dependent variable, accounted for by the explanatory variables (32). The positive square root of R^2 is the so-called coefficient of multiple correlation.

3.6.1 Stepwise Addition of Variables

In stepwise addition (34), a candidate (explanatory variable) for entering into a least squares equation is

selected from among the independent variables not presently in the equation. This candidate is the independent variable which will reduce the unexplained sum of squared errors of the dependent variable the most (equivalently, the candidate is the independent variable which will raise R^2 the most). If the candidate does not meet one of the stopping criteria (a preset significance probability*), it is added to the least squares equation, a new candidate is selected, and the procedure is continued. If the candidate meets all preset stopping criteria, the candidate is not entered into the equation and the procedure is terminated.

There is the inherent danger in this procedure that a group of variables which individually account for little of the variation in the dependent variable, but as a group explain much of this variation, may never be entered into the equation (34). Therefore, a relatively high preset significance probability level of 0.05 was used.

3.6.2 Stepwise Deletion of Variables

In stepwise deletion (35), an initial least squares equation is obtained using all of the independent variables. One variable is then deleted from the equation and a new least squares equation estimated. A second variable is deleted and the least squares equation is recalculated. The

---

*Significance probability is the maximum probability of rejecting the hypothesis: $b_i = 0$, when $b_i = 0$ (i.e.) the probability of committing a type I error.

procedure continues until a variable selected as a candidate for deletion meets one or more stopping criteria (a preset significance probability). Since the selection of a candidate variable for deletion is closely tied to the stopping criterion, a preset significance probability level of 0.005 was used.

In general, the stepwise deletion involves moving from a very general hypothesis (with many independent variables in the equation) to a progressively more restricted hypothesis (with fewer independent variables in the equation). The stepwise addition moves from a very restricted hypothesis to a progressively more general hypothesis.

3.7 Estimated Relationships

The estimation process combined statistical estimation procedures and the author's intuitive understanding of the cause-effect relationships governing the fruit development process. All relationships were first estimated using stepwise addition of variables (Section 3.6.1) to allow for large numbers of variables and to avoid the possibility of singularity problems. The method of stepwise deletion of variables (Section 3.6.2) was also employed to permit all variable combinations to account for the variation in the dependent variable.

The resulting regression relationships were then scrutinized. Particular attention was given to the sign and the magnitude of the coefficient of each explanatory variable, their standard errors of estimate as well as the magnitude of the coefficient of determination (R^2) and the overall standard error of estimate for the relationship. Least squares equations were then estimated (36) utilizing the remaining explanatory variables. All relationships were estimated from field samples of a 100-foot square area. The notation used throughout is the same notation formalized previously, with the following addition: R^2, R, and S.E. are the multiple coefficient of determination, multiple coefficient of correlation and standard error of estimate, respectively. The values in parentheses directly below the estimated coefficients will be the standard errors of estimate of each of the coefficients.

The estimated relationships corresponding to 3.5.2 are as follows:

$$X_4(T+1) = \underset{(0.06)}{0.23}\ N_{3B}(T) + \underset{(0.02)}{0.08}\ N_{3A}(T)$$

$$R^2 = 0.50 \qquad R = 0.72 \qquad S.E. = 2$$

$$X_{3B}(T+1) = \underset{(0.03)}{0.45}\ N_{3A}(T)$$

$$R^2 = 0.48 \qquad R = 0.69 \qquad S.E. = 5$$

$$X_{3A}(T+1) = \underset{(0.01)}{0.30}\ N_2(T)$$

$$R^2 = 0.32 \qquad R = 0.56 \qquad S.E. = 6$$

$$X_2(T+1) = \underset{(0.03)}{0.25}\ N_{1B}(T)$$

$$R^2 = 0.20 \qquad R = 0.38 \qquad S.E. = 9$$

$$X_{1B}(T+1) = \underset{(0.07)}{0.30}N_{1A}(T) \simeq 6$$

$R^2 = 0.07$ $R = 0.25$ $S.E. = 11$

$$X_{1A}(T+1) = 6$$

$S.E. = 15$

In estimating the above relationships, an initial hypothesis was made that fruit could not skip a size grade from one day to the next. For example, fruit could not be in size 2 today and size 3B tomorrow. The result of field investigations on two experimental plots indicated that this hypothesis is not true for the case of size 3A cucumbers. A few size 3A cucumbers today may be size 4 cucumbers tomorrow. Therefore, as indicated in the estimated relationship, this possibility was permitted in the statistical estimation process.

The constants in each of the first four estimated relationships were omitted because of the following reasons. (1) It is obvious in these relationships that the dependent variables must be zero if the independent variables are zero. (2) A two-sided test of hypothesis (H_0: $c = 0$ versus H_1: $c \neq 0$) established that the constants were not statistically significant from zero at the 5 per cent significance level.

The results of field investigations as well as the analysis of the data indicated that the rate of formation of new cucumbers per day for the size grade 1A and 1B can not be explained by their initial distribution.

The climatic influence of fruit development has not been statistically verified. Extensive analyses of the fruit development and the climatic data suggest that certain climatic variables may be statistically significant. However, the variability in fruit development explained by them was very minor. The results of one such analysis, based on the studies of Lewin (11), are presented in Appendix A. The exclusion of measures of soil moisture stress is believed to be a major factor in these findings. It is entirely plausible that in some of the sequences the fruit development was photosynthesis limited and in others it was available moisture limited. If this is true the influence of temperatures and radiation would be expected to be inconclusive.

The estimated relationships corresponding to 3.5.4 are as follows (all weights are in units of pounds):

$$W_4(T) = \underset{(0.005)}{0.485}\, N_4(T)$$

$$R^2 = 0.982 \qquad R = 0.991 \qquad S.E. = 0.251$$

$$W_{3B}(T) = \underset{(0.003)}{0.345}\, N_{3B}(T)$$

$$R^2 = 0.982 \qquad R = 0.991 \qquad S.E. = 0.364$$

$$W_{3A}(T) = \underset{(0.001)}{0.238}\, N_{3A}(T)$$

$$R^2 = 0.960 \qquad R = 0.980 \qquad S.E. = 0.406$$

$$W_2(T) = \underset{(0.001)}{0.127}\, N_2(T)$$

$$R^2 = 0.954 \qquad R = 0.977 \qquad S.E. = 0.375$$

$$W_{1B}(T) = \underset{(0.0008)}{0.043}\, N_{1B}(T)$$

$$R^2 = 0.840 \qquad R = 0.917 \qquad S.E. = 0.249$$

$$W_{1A}(T) = \underset{(0.0004)}{0.016}\, N_{1A}(T)$$

$$R^2 = 0.891 \qquad R = 0.944 \qquad S.E. = 0.056$$

The high values of the multiple coefficient of determination (R^2) and low values for the standard error of estimate (S.E.) for the above six equations suggest that there is a definite relationship between the weights and numbers of a particular size grade.

It should be noted here that, in general, we do not expect the relationship between weight and count of 4 to be constant with respect to time because the size grade 4 is an absorbing state. However, the data employed in the estimation were sequences that cover a relatively short period of time (usually two to five days before harvest). It is quite plausible that during this period the relationship between the weight and the number of size four fruit remains approximately constant.

The last two equations have slightly lower R^2 values compared to the remaining four equations because of (1) their greater variability in size, and (2) their g

r

ar

greater range of diameter measurement. The estimated relationships corresponding to 3.5.9 are as follows:

$$W_4(T+1) = \underset{(0.010)}{0.482}\, N_4(T) + \underset{(0.021)}{0.490}\, X_4(T+1)$$

$$R^2 = 0.982 \qquad R = 0.990 \qquad S.E. = 0.253$$

$$W_{3B}(T+1) = \underset{(0.009)}{0.338}\, N_{3B}(T) + \underset{(0.008)}{0.359}\, X_{3B}(T+1) - \underset{(0.025)}{0.360}\, X_4(T+1)$$

$$R^2 = 0.983 \qquad R = 0.991 \qquad S.E. = 0.356$$

$$W_{3A}(T+1) = \underset{(0.005)}{0.244}\, N_{3A}(T) + \underset{(0.006)}{0.228}\, X_{3A}(T+1) - \underset{(0.012)}{0.244}\, X_{3B}(T+1)$$

$$R^2 = 0.969 \qquad R = 0.984 \qquad S.E. = 0.392$$

$$W_2(T+1) = \underset{(0.002)}{0.128}\, N_2(T) + \underset{(0.004)}{0.120}\, X_2(T+1) - \underset{(0.007)}{0.126}\, X_{3A}(T+1)$$

$$R^2 = 0.956 \qquad R = 0.977 \qquad S.E. = 0.375$$

$$W_{1B}(T+1) = \underset{(0.003)}{0.041}\, N_{1B}(T) + \underset{(0.003)}{0.048}\, X_{1B}(T+1) - \underset{(0.004)}{0.040}\, X_2(T+1)$$

$$R^2 = 0.856 \qquad R = 0.925 \qquad S.E. = 0.241$$

$$W_{1A}(T+1) = \underset{(0.0004)}{0.016}\, N_{1A}(T) + \underset{(0.0008)}{0.016}\, X_{1A}(T+1) - \underset{(0.0012)}{0.016}\, X_{1B}(T+1)$$

$$R^2 = 0.892 \qquad R = 0.944 \qquad S.E. = 0.057$$

The estimated relationships corresponding to 3.5.13 are as follows:

$$W_4(T+1) = \underset{(0.010)}{0.481}\, X_4(T+1) + \underset{(0.021)}{0.491}\, O_4(T)$$

$$R^2 = 0.981 \qquad R = 0.990 \qquad S.E. = 0.253$$

The

$\underline{H}$ a

shi

$$W_{3B}(T+1) = \underset{(0.005)}{0.351}\, X_{3B}(T+1) + \underset{(0.009)}{0.333}\, O_{3B}(T)$$

$R^2 = 0.982$ $\quad R = 0.991 \quad$ S.E. $= 0.361$

$$W_{3A}(T+1) = \underset{(0.004)}{0.230}\, X_{3A}(T+1) + \underset{(0.005)}{0.246}\, O_{3A}(T)$$

$R^2 = 0.967$ $\quad R = 0.983 \quad$ S.E. $= 0.399$

$$W_{2}(T+1) = \underset{(0.004)}{0.121}\, X_{2}(T+1) + \underset{(0.001)}{0.128}\, O_{2}(T)$$

$R^2 = 0.956$ $\quad R = 0.977 \quad$ S.E. $= 0.372$

$$W_{1B}(T+1) = \underset{(0.002)}{0.049}\, X_{1B}(T+1) + \underset{(0.0009)}{0.042}\, O_{1B}(T)$$

$R^2 = 0.855$ $\quad R = 0.925 \quad$ S.E. $= 0.239$

$$W_{1A}(T+1) = \underset{(0.0005)}{0.016}\, X_{1A}(T+1) + \underset{(0.0004)}{0.016}\, O_{1A}(T)$$

$R^2 = 0.892$ $\quad R = 0.944 \quad$ S.E. $= 0.057$

The estimated matrices corresponding to $\underline{E}$, $\underline{P}$, $\underline{P}'$, $\underline{Q}$, $\underline{Q}'$, $\underline{H}$ and $\underline{H}'$ referred to in the section on functional relationship are as follows:

$$\underline{E} = \begin{bmatrix} 0.485 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.345 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.238 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.127 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.043 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.016 \end{bmatrix}$$

$$\underline{P} = \underline{E}, \text{ and } \underline{P}' = \begin{bmatrix} 0.485 & 0 & 0 & 0 & 0 & 0 \\ -0.345 & 0.345 & 0 & 0 & 0 & 0 \\ 0 & -0.238 & 0.238 & 0 & 0 & 0 \\ 0 & 0 & -0.127 & 0.127 & 0 & 0 \\ 0 & 0 & 0 & -0.043 & 0.043 & 0 \\ 0 & 0 & 0 & 0 & -0.016 & 0.016 \end{bmatrix}$$

$$\underline{Q} = \begin{bmatrix} 0.482 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.338 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.244 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.128 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.041 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.016 \end{bmatrix}$$

$$\underline{Q}' = \begin{bmatrix} 0.480 & 0 & 0 & 0 & 0 & 0 \\ -0.360 & 0.359 & 0 & 0 & 0 & 0 \\ 0 & -0.244 & 0.228 & 0 & 0 & 0 \\ 0 & 0 & -0.126 & 0.120 & 0 & 0 \\ 0 & 0 & 0 & -0.040 & 0.048 & 0 \\ 0 & 0 & 0 & 0 & -0.016 & 0.016 \end{bmatrix}$$

$$\underline{H} = \begin{bmatrix} 0.491 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.04 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.016 \end{bmatrix}$$

and

$$\underline{H}' = \begin{bmatrix} 0.481 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.351 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.230 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.121 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.049 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.016 \end{bmatrix}$$

On comparing the above matrices, we find that $\underline{P} \simeq \underline{Q}$, $\underline{P}' \simeq \underline{Q}'$ (see 3.5.9), and also $\underline{H} \simeq \underline{H}' \simeq \underline{E}$ (see 3.5.13). Furthermore, we note that the standard errors in the estimated relationships of 3.5.4, 3.5.9 and 3.5.13 are nearly the same. These results support the initial hypothesis of 3.5.4, i.e., there is a definite relationship between the number and the weight of fruits of a given size grade.

3.8 Estimated Models

The estimated relationships in 3.5.2 provide a model to compute the distribution of new fruits at time (T+1) as a function of the distribution of the fruits at time T:

S

where

$$\vec{X}_c(T+1) = \underline{C}\ \vec{N}_c(T) + \underline{B}'\ \vec{U}(T) \qquad 3.8.1$$

Where:

$$\underline{C} = \begin{bmatrix} 0 & 0.23 & 0.08 & 0 & 0 & 0 \\ 0 & 0 & 0.45 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.30 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.30 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underline{B}' = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$U(T) = 6$ for $T \geqslant 0$.

It should be recognized that 3.8.1 is not a mathematical relationship as we have shown above, but rather a statistical relationship. To be precise 3.8.1 should be written as follows:

$$X_c(T+1) = \underline{C}\ \vec{N}_c(T) + \underline{B}'\ U(T) + \bar{\varepsilon}$$

where $\bar{\varepsilon}$ is a vector of observed or estimated errors

corresponding to the assumption of the error term in ordinary least squares. We will, however, continue to disregard it explicitly, but make note of it at appropriate times in the discussion and interpretation of results.

Substituting 3.8.1 in 3.5.1 yields:

$$\vec{N}_c(T+1) = (\underline{I} + \underline{SC})\ \vec{N}_c(T) + \underline{S}\cdot\underline{B}'\vec{U}(T) \qquad 3.8.2$$

The above relationship indicates that the state of the field at time T+1 can be determined from the knowledge of the state of the field and the input at time T.

The weight of fruit in various size grades then, can be determined either from 3.5.4 or 3.5.9 as follows:

$$\vec{W}_c(T) = \underline{E}\ \vec{N}_c(T) \qquad 3.5.4$$

or:

$$\vec{W}_c(T+1) = \underline{Q}\ \vec{N}_c(T) + \underline{Q}'\ \vec{X}_c(T+1) \qquad 3.5.9$$

Where: $\underline{E}$, $\underline{Q}$ and $\underline{Q}'$ are defined as above.

Equations 3.5.4 and 3.5.9 are strictly statistical relations and not mathematical relations.

3.9 A Simulation Model

On the basis of the analysis carried out up to the last section, it is easy to conclude that there is no one unique way to present the model proposed; rather there are several alternate, but equivalent ways in which the model can be presented. One of these forms which is particularly convenient because it has received wide attention in the literature (37) is presented below.

Let:

N_{1A}, N_{1B}, N_2, N_{3A}, N_{3B} and N_4 be the number of fruits of size grade 1A, 1B, 2, 3A, 3B and 4 respectively in a field sample of area 100 ft.^{2}.

W_{1A}, W_{1B}, W_2, W_{3A}, W_{3B} and W_4 be the corresponding weight of fruits in size grade 1A, 1B, 2, 3A, 3B, and 4 respectively.

T be the time in days, T = 1, 2, 3, ...

M(T) be the maturity of a sample at time T, measured as a fraction of number 2's in the sample, i.e.,

$$M(T) = \frac{N_2(T)}{N_{1A}(T) + N_{1B}(T) + N_2(T) + N_{3A}(T) + N_{3B}(T) + N_4(T)}$$

M_i = Initial maturity of a sample as expressed above.

M_f = final maturity of a sample as expressed above.

$(M_i - M_f)$ represent the range of maturity of the sample over which the model below is valid.

Then, for $M_i \leq M(T) \leq M_f$

$$\vec{N}_c(T+1) = \underline{A}\,\vec{N}_c(T) + \underline{B}\,U(T) \qquad 3.9.1$$

$$\vec{W}_c(T) = \underline{E}\,\vec{N}_c(T) \qquad 3.9.2$$

Where:

$$\vec{N}_c(T) = \begin{bmatrix} N_4(T) \\ N_{3B}(T) \\ N_{3A}(T) \\ N_2(T) \\ N_{1B}(T) \\ N_{1A}(T) \end{bmatrix}$$

$$\underline{A} = \begin{bmatrix} 1 & 0.23 & 0.08 & 0 & 0 & 0 \\ 0 & 0.77 & 0.37 & 0 & \cdot & \cdot \\ \cdot & 0 & 0.55 & 0.30 & 0 & \cdot \\ \cdot & \cdot & 0 & 0.70 & 0.25 & 0 \\ \cdot & \cdot & \cdot & 0 & 0.75 & 0.30 \\ 0 & 0 & 0 & 0 & 0 & 0.70 \end{bmatrix}$$

$$\underline{B} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$U(T) = 6$$

$$M_i \lesssim M(T) \lesssim M_f$$

$$M_i = 0.50$$

$$M_f = 0.25$$

$$\vec{W}_c(T) = \begin{bmatrix} W_4(T) \\ W_{3B}(T) \\ W_{3A}(T) \\ W_2(T) \\ W_{1B}(T) \\ W_{1A}(T) \end{bmatrix}$$

$$\underline{E} = \begin{bmatrix} 0.485 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0.345 & 0 & \cdot & \cdot & \cdot \\ \cdot & 0 & 0.238 & 0 & \cdot & \cdot \\ \cdot & \cdot & 0 & 0.127 & 0 & \cdot \\ \cdot & \cdot & \cdot & 0 & 0.043 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.016 \end{bmatrix}$$

Note that we have used maturity $M(T)$ as a dummy variable for time. Our model is valid in a period two to four days before harvest. However, at time $T=0$, when the first sample $\vec{N}_c(0)$ is taken, the day of harvest is not known. Therefore, any reference to harvest time will not be meaningful.

It should also be noted here that the exact relationship between the time and maturity of a sample such as described above is not known. However, the limits 50 and 25 per cent of number of 2's correspond approximately to the time period of our interest. They also represent the limiting sample values from which the model was estimated.

Appendix B contains the required matrices for a simulation model based on size grades 1, 2, 3 and 4. It is obvious that such a model is less desirable because of the homogenity considerations discussed earlier in Section 3.2.

4. MODEL VERIFICATION

Verification of the model was attempted using both statistical tests as well as an examination of the characteristic behavior of the system itself.

4.1 Statistical Tests

4.1.1 Model Structure

The model structure in 3.5.1 was based on (1) availability of data (measurable states of the system), (2) theoretical considerations--fruits in a given size grade during a given period can either enlarge to the next size grade or remain in the same size grade, and (3) the hypothesis that the rate of fruit development depends upon the initial size distribution of the fruit at time T and the influence of weather during the period ΔT. However, the influence of weather was not positively determined and the model in a relatively simple form such as the one given in 3.9.1 was proposed. Recall that the parameters in 3.9.1 were arrived at, first by estimating 3.5.2 and then by appropriate substitution of 3.5.2 in 3.5.1. At first one might be tempted to estimate 3.9.1 by ordinary least squares directly from the observed sample values.

The above reasoning can be made more precise as follows:

Since,

$$\vec{N}_c(T+1) = \underline{I}\,\vec{N}_c(T) + \underline{S}\,\vec{X}_c(T+1) \qquad 3.5.1$$

and neglecting the effect of weather variables in 3.5.2 we have:

$$\begin{aligned}
X_4(T+1) &= a\,N_{3B}(T) + b\,N_{3A}(T)\\
X_{3B}(T+1) &= c\,N_{3A}(T)\\
X_{3A}(T+1) &= d\,N_2(T)\\
X_2(T+1) &= e\,N_{1B}(T)\\
X_{1B}(T+1) &= f\,N_{1A}(T)\\
X_{1A}(T+1) &= g
\end{aligned} \qquad 4.1.1.1$$

Substituting 4.1.1.1 in 3.5.1 we have:

$$\begin{aligned}
N_4(T+1) &= N_4(T) + a\,N_{3B}(T) + b\,N_{3A}(T)\\
N_{3B}(T+1) &= (1-a)\,N_{3B}(T) + (c-b)\,N_{3A}(T)\\
N_{3A}(T+1) &= (1-c)\,N_{3A}(T) + d\,N_2(T)\\
N_2(T+1) &= (1-d)\,N_2(T) + e\,N_{1B}(T)\\
N_{1B}(T+1) &= (1-e)\,N_{1B}(T) + f\,N_{1A}(T)\\
N_{1A}(T+1) &= (1-f)\,N_{1A}(T) + g
\end{aligned} \qquad 4.1.1.2$$

Then, estimating 3.9.1 directly corresponds to estimating 4.1.1.2 from the observed sample values. However, if the parameters a, b, c, ... , g are estimated using the ordinary least squares, the estimated parameters $\hat{a}$, $\hat{b}$, $\hat{c}$, ... , $\hat{g}$ will be biased and inconsistent. The reason for this is that 4.1.1.2 contains not a set of six independent equations, but a set of six simultaneous equations (See 3.5.1). As a result, the errors and the regressors (independent variables) are correlated. This, however, violates the basic assumption of ordinary least squares (32) that the error and the regressor are not correlated. When this assumption is violated the ordinary least squares results in parameter estimates that are biased and inconsistent (32, 38).

Methods are available for finding consistent and unbiased estimates of parameters in simultaneous equation models such as indirect least squares (32), instrumental variable technique (32, 38), etc. It should be remarked here that finding parameter estimates in 4.1.1.1 (as we have done) rather than 4.1.1.2 using ordinary least squares is a special case of the instrumental variable technique.

4.1.2 Test for Autocorrelation

Since the parameters in 3.9.1 and 3.9.2 were estimated using time series data and also because non-autocorrelation is generally assumed in ordinary least squares, the residuals (estimated error terms) in 3.9.1 and 3.9.2 were examined for autocorrelation.

Autocorrelation is indicated, whenever the error e_T at time T is correlated with one or more of its previous values (e_{T-1}, e_{T-2}, etc.). In ordinary least squares estimation, the presence of autocorrelation signals possible inadequacy of the regression model formulation. Generally, autocorrelation does not destroy unbiasedness and consistency of the estimates of the coefficients, but rather of their variances (32). When positive autocorrelation is present, the variances of the coefficients are generally underestimated leading to more frequent rejection of the null hypothesis of b_i equals zero.

A well-known test for the existence of autocorrelation is the Durbin-Watson test (39). A test statistic d for the null hypothesis (H_0: $r = 0$ versus H_1 $r > 0$; r = autocorrelation) of residual independence is computed. This statistic is also called the von Neuman ratio. It is the sum of squares of the first differences of the least squares estimated disturbances, divided by the sum of squares of the estimated disturbances, i.e.:

$$d = \frac{\sum_{t=2}^{N} (\hat{U}_t - \hat{U}_{t-1})^2}{\sum_{t=1}^{N} \hat{U}_t^2}$$

Where:

d = Durbin-Watson test statistic

$\hat{U}_t$ = the least squares estimator of the disturbance for observation t.

If there is no autocorrelation, d is approximately equal to two. Lower values of d indicate positive correlation, while higher d values indicate negative correlation. The regions of acceptance and rejection of the null hypothesis are tabulated for comparison with the computed d values (38).

Table 2 gives the computed values of d in 3.9.1 and 3.9.2.

Table 2. Durbin-Watson test statistic (d).

Relationship (class)	Estimated d
3.5.2 (4)	1.94
3.5.2 (3B)	2.06
3.5.2 (3A)	1.92
3.5.2 (2)	2.20
3.5.2 (1B)	2.34
3.5.4 (4)	1.92
3.5.4 (3B)	1.67
3.5.4 (3A)	1.40
3.5.4 (2)	1.47
3.5.4 (1B)	1.85
3.5.4 (1A)	1.56

The critical values of d at $\alpha = .01$ level of significance for testing the hypothesis H_0: $r = 0$ versus H_1: $r > 0$ for a sample size of 60 and $K = 1$ is 1.45 and for $K = 2$ is 1.48; where K = number of regressors. On the basis of this test we cannot reject the null hypothesis (H_0: $r = 0$). Therefore, no autocorrelation is indicated.

The results of the above tests in addition to the one indicated before in the estimated 3.5.2 and 3.5.4 lead us to believe that the estimated parameters in 3.9.1 and 3.9.2 are the best, linear, unbiased estimators of the true population parameters.

4.2 Time Behavior

The response of the model to a given set of initial conditions such as one obtained by the standard sampling procedure (See Section 3.3) is given in Figures 2 to 8. These figures also show the observed dynamic response of the system for the corresponding time period. The Figures 2 to 7 are drawn on the same scale to give an indication of the relative magnitude of the different variables involved. All inputs and outputs are based on an area of 100 ft^2. In each figure the ordinate for the number of the fruit is given by the left hand Y-axis and the ordinate for the weight of the fruit is given by the right hand Y-axis. The abscissa in each figure represents the number of days after the first sampling. All responses are shown for a limited duration corresponding to the time period over which the model is valid. As indicated in an earlier section this period corresponds approximately to the range of 50 to 25 per cent by weight of number 2 fruit in the sample.

An examination of the Figures 2 to 8 indicate that the time behavior exhibited by the model is in close

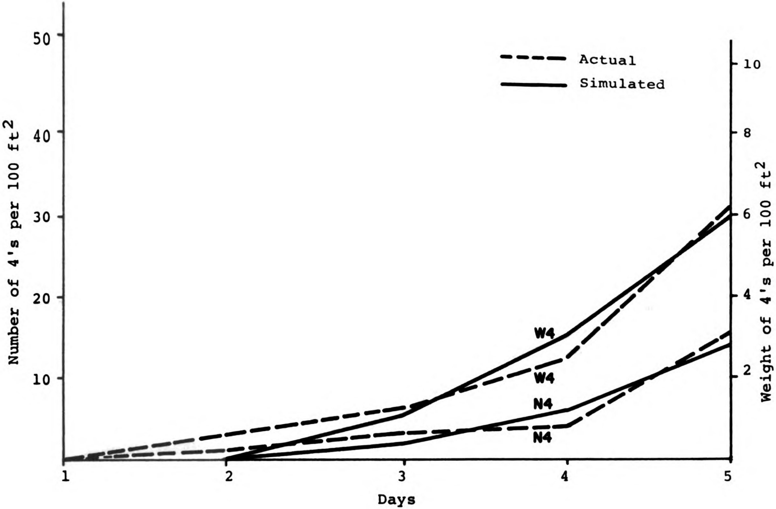

Figure 2. Actual and simulated values for size grade 4.

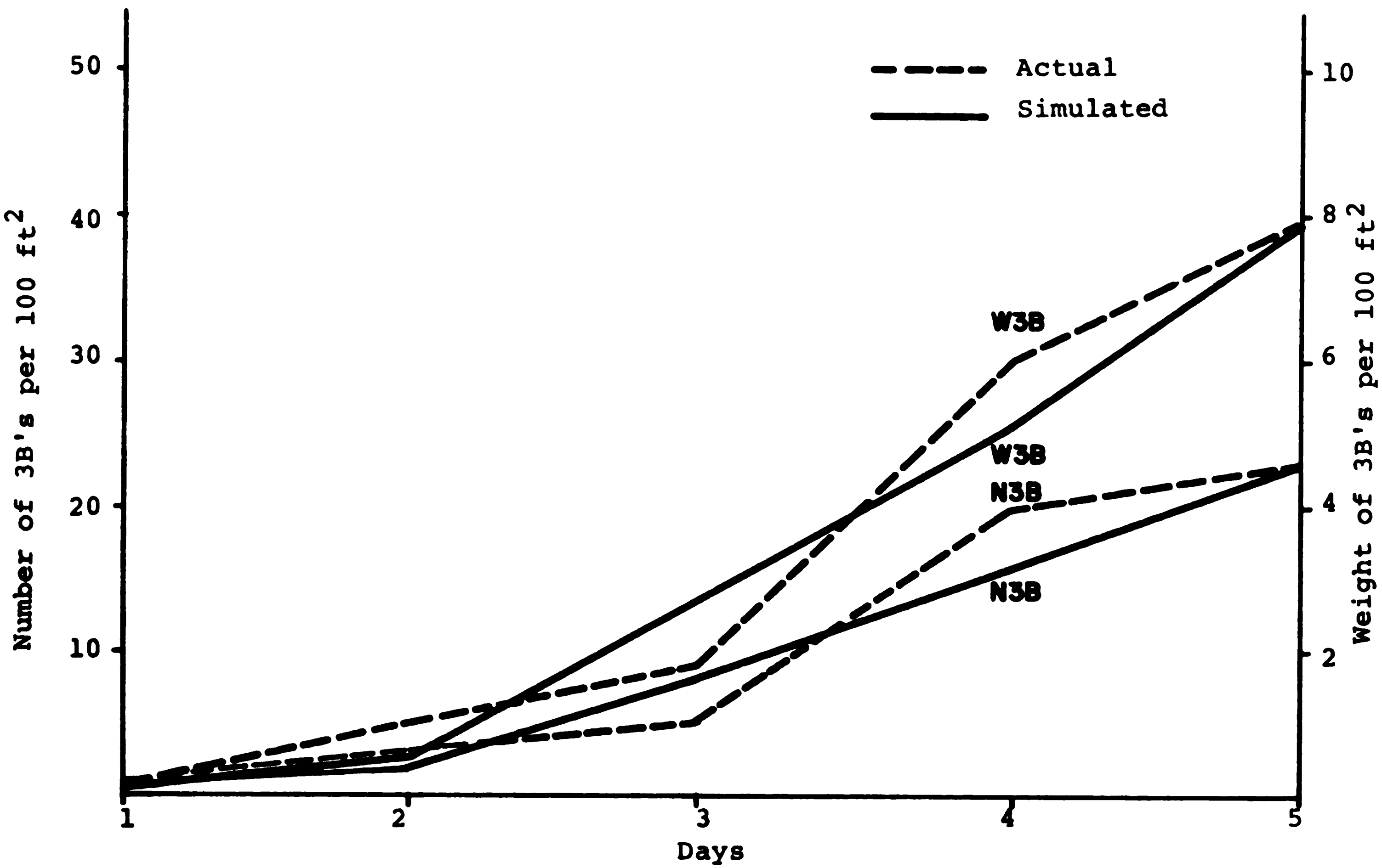

Figure 3. Actual and simulated values for size grade 3B.

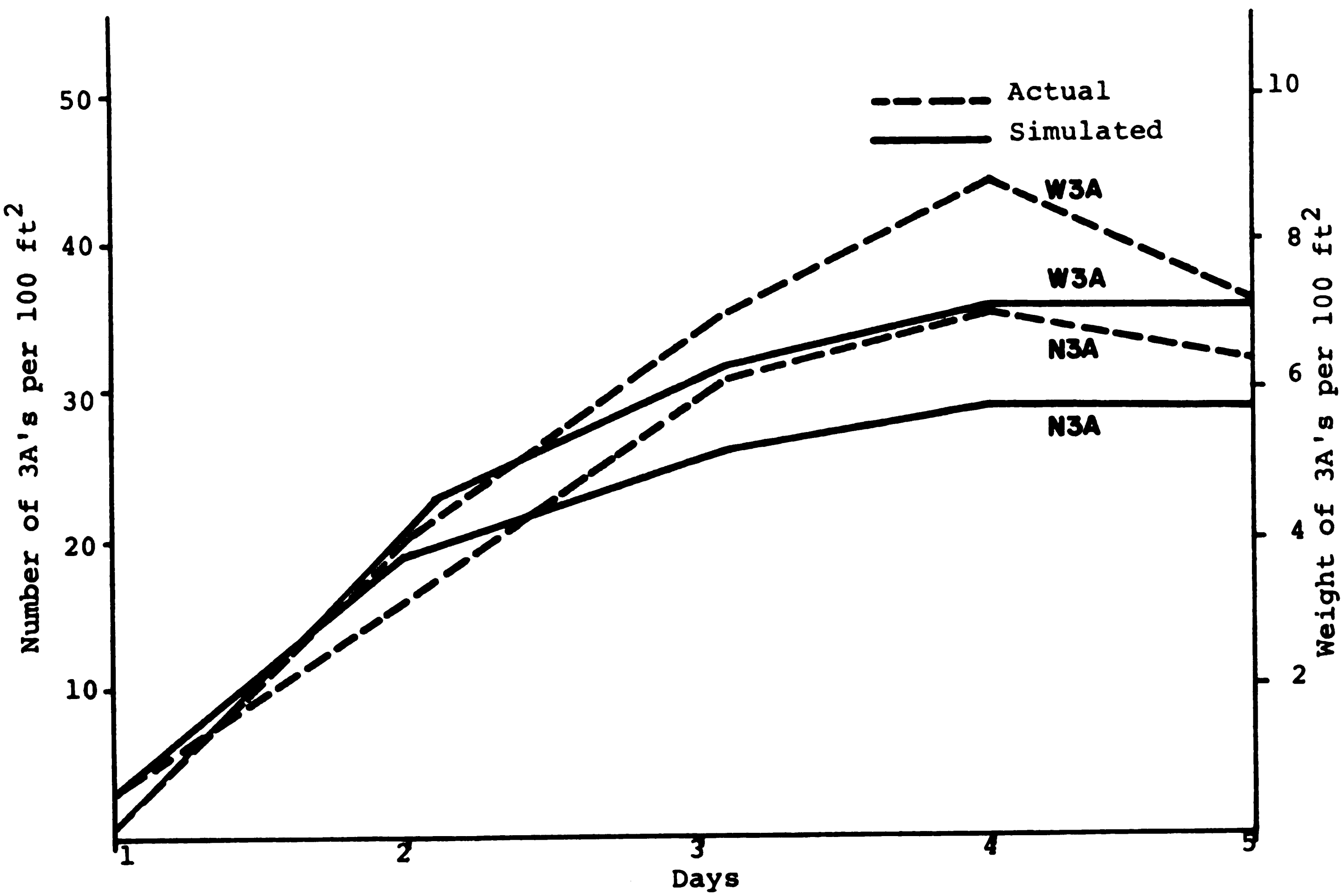

Figure 4. Actual and simulated values for size grade 3A.

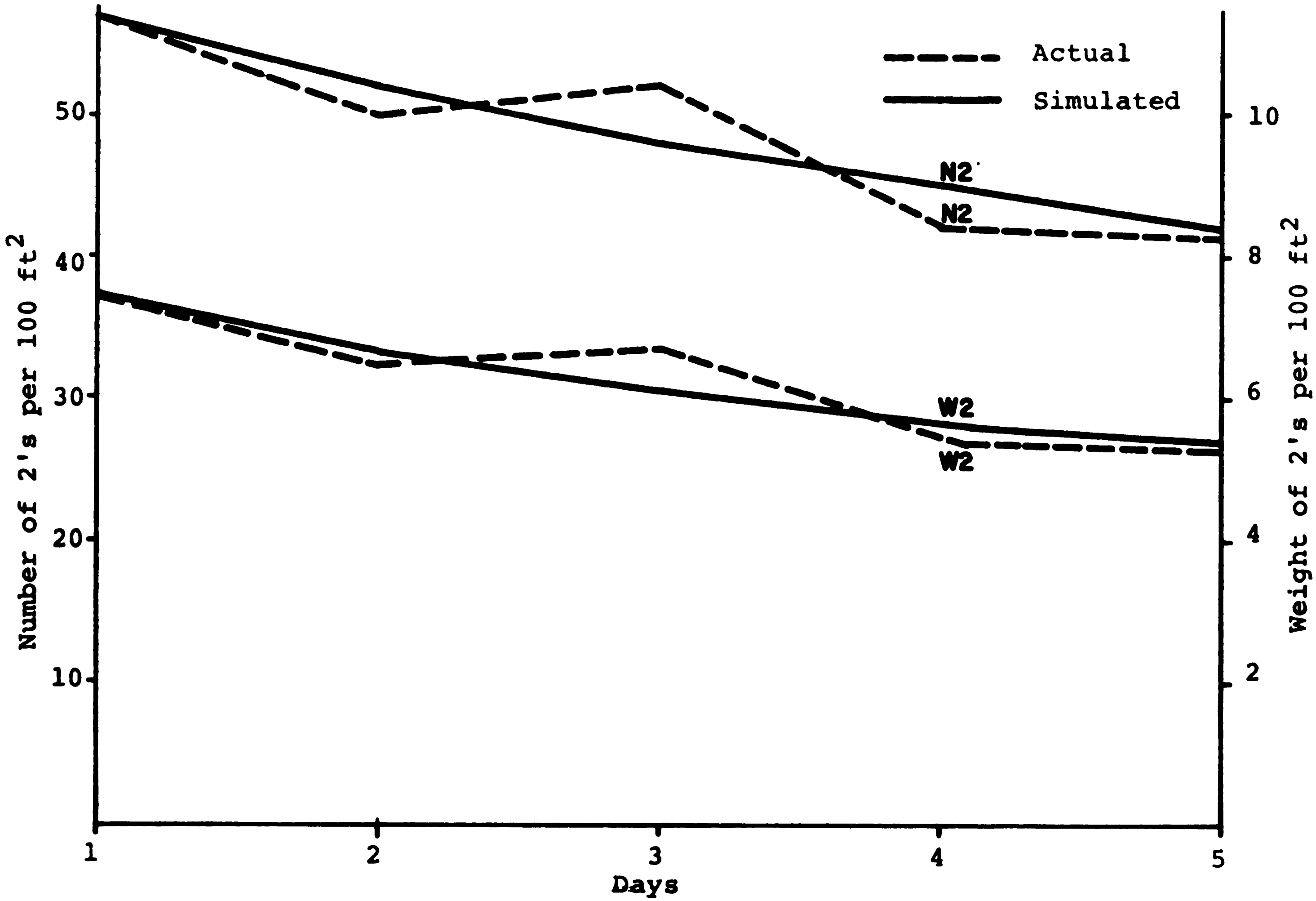

Figure 5. Actual and simulated value for size grade 2.

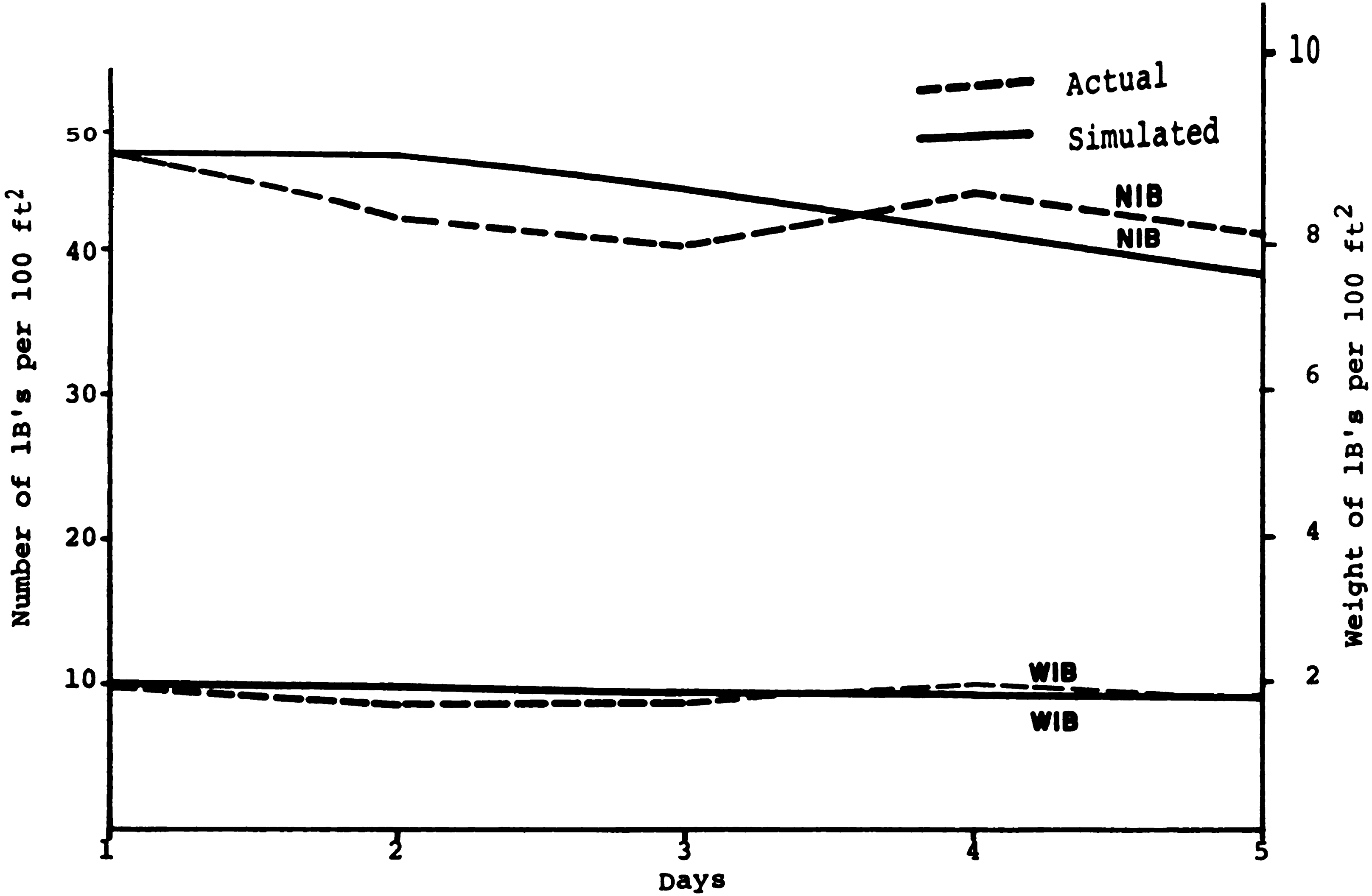

Figure 6. Actual and simulated values for size grade 1B.

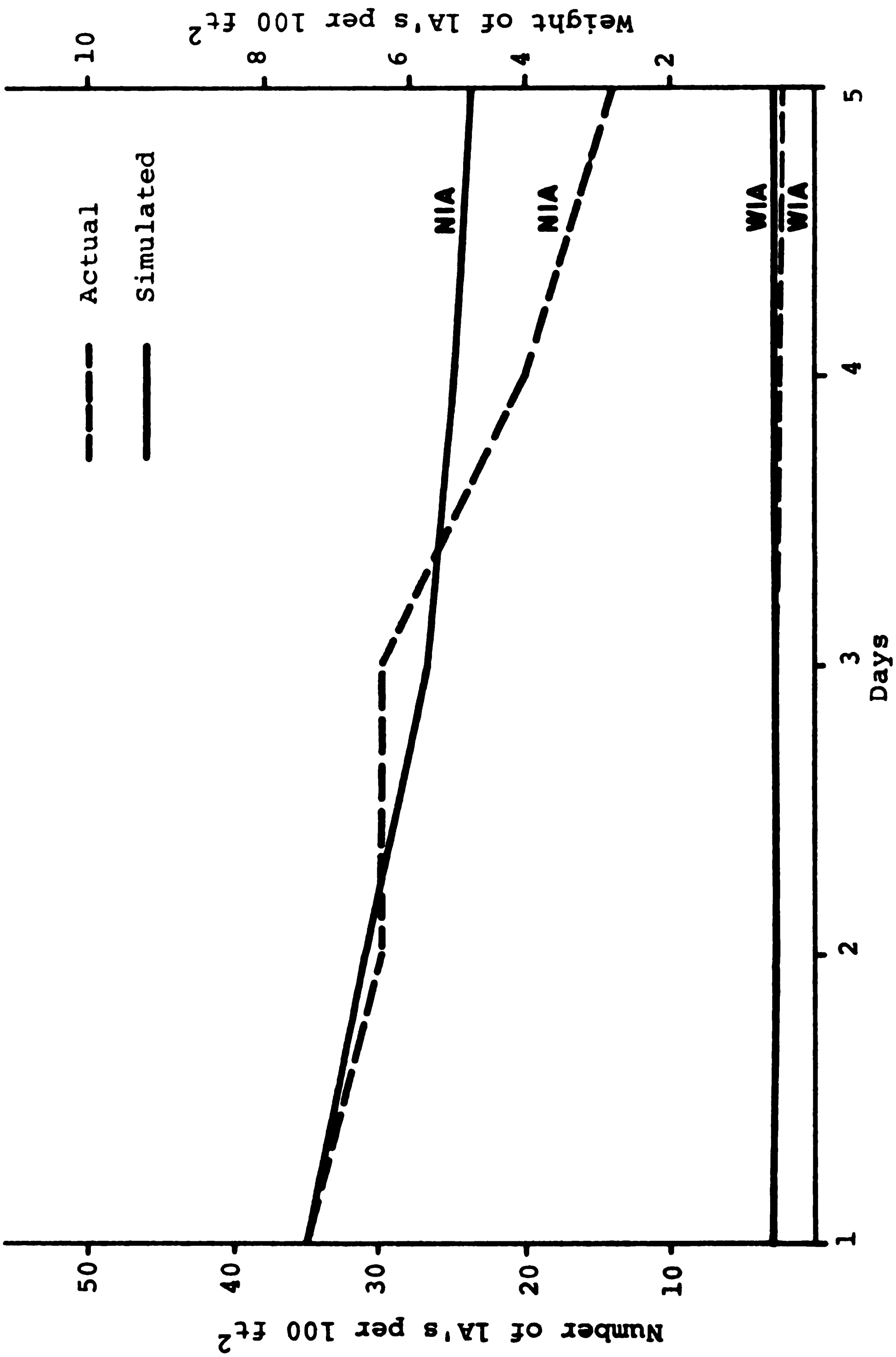

Figure 7. Actual and simulated values for size grade 1A.

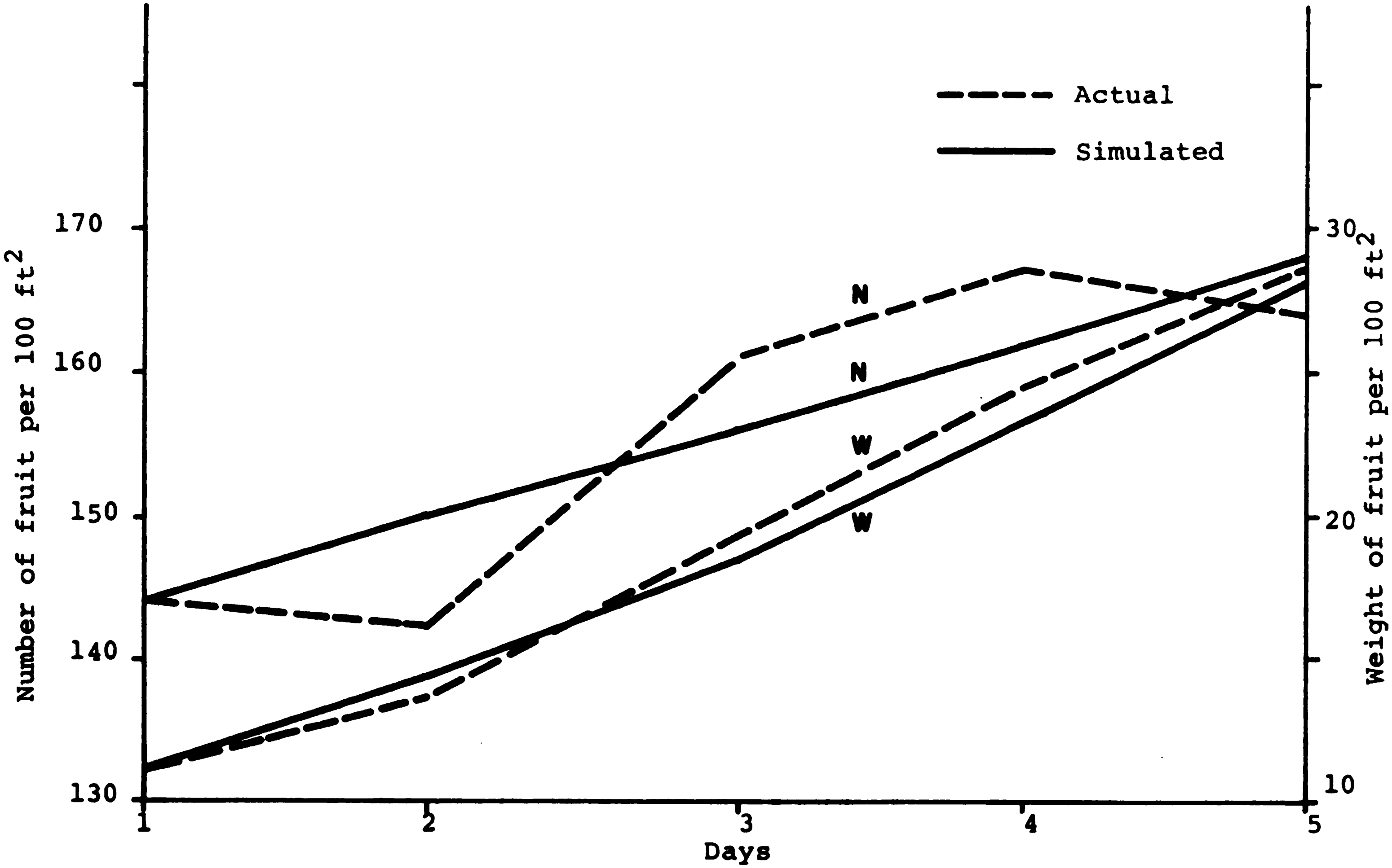

Figure 8. Actual and simulated values for total number and total weight

agreement with the behavior of the system itself. The difference between the actual and the simulated number of fruits of a given size grade at a given time consists of the following components: (1) sampling error, (2) measurement error, and (3) stochastic error. Of these errors, only the first two are controllable. Complete control over the third type of error is not possible because of the inherent irreproducibility of the biological phenomena. This indicates that we do not expect the observed and the simulated behavior to be exactly alike. However, we do expect that the magnitude of the error relative to the magnitude of the variable will be small. Figures 2 to 8 indicate that this is true for all the variables under consideration. Another important point which must be noted here is that the errors in output (weight of fruit) relative to the states (number of fruit) are deamplified rather than amplified. This is not quite apparent from the figures, because we have employed different scales for number and weight. However, this can be seen to be true by examining the elements of the matrix $\underline{E}$ in 3.9.2. All elements in the matrix $\underline{E}$ are found to be strictly less than one. The Figures 2 to 8 also suggest that the errors are not cumulative over successive time periods.

It should be recognized that any inferences based on only one sequence of observations can be misleading. This is so because our model is not deterministic, but stochastic, because of the nature of the relationships in

3.8.1 which are statistical. Therefore, a more reliable index of the quality of the model can be obtained by simulating the model over all the sequences not used in statistical estimation. The results of such a simulation are presented in Table 3.

Table 3 indicates the mean and the standard deviation of the percentage error of prediction of a given variable over the remaining sequences (approximately 50%) from West-1971. The definition of the mean percentage error and its standard deviation for the variables in Table 3 are as follows:

Let:

$AN_c(t)$ = observed number of fruit of size grade c on day t

$SN_c(t)$ = simulated number of fruit of size grade c on day t

$EN_c(t)$ = percentage error in number of fruit of size grade c on day t

MEN_c = mean percentage error in number of fruit of size grade c

SEN_c = standard deviation of percentage error in number of fruit of size grade c

$AN(t)$ = observed number of fruit on day t

$SN(t)$ = simulated number of fruit on day t

$EN(t)$ = percentage error in number of fruit on day t

MEN = mean percentage error in number of fruit

SEN = standard deviation of percentage error in number of fruit

c = size grades, c = 1A, 1B, 2, 3A, 3B, 4

Table 3. Means and standard deviations of percentage error over the range of all sequences.

	Percentage errors in number							Percentage errors in weight						
	N_{1A}	N_{1B}	N_2	N_{3A}	N_{3B}	N_4	N	W_{1A}	W_{1B}	W_2	W_{3A}	W_{3B}	W_4	W
Mean	-1	0	2	1	0	-1	1	0.0	0.4	2.4	0.4	-0.7	-2.0	0.4
Standard Deiva-tion	6	7	8	4	3	2	12	0.7	2.4	7.0	7.5	7.7	4.7	10.7

t = number of days in a given sequence, t = 1, 2, 3, ...

n = total number of sequences used in simulation.

Then:

$$EN_c(t) = \frac{AN_c(t) - SN_c(t)}{AN(t)} \times 100$$

$$MEN_c = \sum_n \sum_t \frac{EN_c(t)}{(t+n)}$$

$$SEN_c = \sqrt{\sum_n \sum_t \frac{EN_c(t)^2}{(t+n-1)} - \frac{(t+n)MEN_c^2}{(t+n-1)}}$$

$$AN(t) = AN_{1A}(t) + AN_{1B}(t) + AN_2(t) + AN_{3A}(t) + AN_{3B}(t) + AN_4(t)$$

$$SN(t) = SN_{1A}(t) + SN_{1B}(t) + SN_2(t) + SN_{3A}(t) + SN_{3B}(t) + SN_4(t)$$

$$EN(t) = \frac{AN(t) - SN(t)}{AN(t)} \times 100$$

$$MEN = \sum_n \sum_t \frac{EN(t)}{(t+n)}$$

and,

$$SEN = \sqrt{\sum_n \sum_t \frac{EN(t)^2}{(t+n-1)} - \frac{(t+n)MEN^2}{(t+n-1)}}$$

The definition of the variables for weight is similar to the one shown for number above and can be obtained simply by replacing W for N at the appropriate places. Table 3 further confirms our earlier observations of the nature of the error term. As indicated by the table, simulation errors in reproducing the past behavior of the system are small.

The analysis in this section led us to believe that our model is an adequate representation of the real system. We note here that our model is valid only over a limited time period corresponding to the range of 50 to 25 per cent by count of number 2 fruit in the sample. Fortunately, this is also the time period in which a decision maker wishes to derive conclusions about the system under consideration.

5. SENSITIVITY ANALYSIS

The purpose of sensitivity analysis is to provide: (1) greater insight about the inner workings of the simulation model, (2) an identification of the critical and less critical parameters, (3) an indication whether some of the constraints should be loosened or tightened, and (4) a more quantitative idea about the expected overall performance of the system being modeled (40).

The parameters that can significantly influence the performance of the model proposed are the parameters estimated in 3.5.2. These parameters correspond to the coefficients of the matrix $\underline{C}$ in 3.8.1. Therefore, these parameters were subjected to a sensitivity analysis (in a very restrictive sense). These restrictions, however, are due to the nature of the problem and not because of the computational cost. Any sensitivity analysis based on variation in parameter value over the one given by the coefficients of matrix $\underline{C}$ is not meaningful here. This is due to the fact that we know that the parameters in $\underline{C}$ are the best, linear, unbiased estimates of the true parameters. This was shown in the model validation section. It was also shown in the same section that these parameters are consistent estimates of the true parameter values.

A more meaningful index of the sensitivity of these parameters can, however, be obtained by estimating the matrix $\underline{C}$ from the data collected from different locations and in different seasons. This presumably will give us an indication of the sensitivity of the parameters in $\underline{C}$ with respect to both the location as well as expected changes in growing season. The results of such an analysis for two locations and two seasons are presented below. The locations in the present discussion will be referred to as West and North and the seasons as 1971 and 1972 (See Section 3.3). Thus, with the given information, it is possible to estimate a set of four matrices, each corresponding to a different location and season. Let these matrices be $\underline{C}_1$, $\underline{C}_2$, $\underline{C}_3$ and $\underline{C}_4$. Here,

$\underline{C}_1$ = estimated matrix from the data set West - 1971 = $\underline{C}$

$\underline{C}_2$ = estimated matrix from the data set North - 1971

$\underline{C}_3$ = estimated matrix from the data set West - 1972

$\underline{C}_4$ = estimated matrix from the data set North - 1972.

As indicated above, the matrix $\underline{C} = \underline{C}_1$ in 3.8.1 was estimated from the data set West - 1971. The matrices $\underline{C}_2$, $\underline{C}_3$ and $\underline{C}_4$ were found as follows. The matrix $\underline{C} = \underline{C}_1$ has been repeated for the sake of comparison.

West - 1971

$$\underline{C}_1 = \begin{bmatrix} 0 & .23 & .08 & 0 & 0 & 0 \\ 0 & 0 & .45 & 0 & 0 & 0 \\ 0 & 0 & 0 & .30 & 0 & 0 \\ 0 & 0 & 0 & 0 & .25 & 0 \\ 0 & 0 & 0 & 0 & 0 & .30 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

North - 1971

$$\underline{C}_2 = \begin{bmatrix} 0 & .23 & .04 & 0 & 0 & 0 \\ 0 & 0 & .46 & 0 & 0 & 0 \\ 0 & 0 & 0 & .32 & 0 & 0 \\ 0 & 0 & 0 & 0 & .26 & 0 \\ 0 & 0 & 0 & 0 & 0 & .29 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

West - 1972

$$\underline{C}_3 = \begin{bmatrix} 0 & .32 & 0 & 0 & 0 \\ 0 & 0 & .48 & 0 & 0 \\ 0 & 0 & 0 & .34 & 0 \\ 0 & 0 & 0 & 0 & .22 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

North - 1972

$$\underline{C}_4 = \begin{bmatrix} 0 & .15 & .10 & 0 & 0 \\ 0 & 0 & .44 & 0 & 0 \\ 0 & 0 & 0 & .30 & 0 \\ 0 & 0 & 0 & 0 & .25 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$\underline{C}_3$ and $\underline{C}_4$ are of dimension (5 x 5) because the 1972 data did not include the decomposition of size 1 fruit into sizes 1A and 1B. When this decomposition is not done, samplers tended not to include the fruit of size 1A into their samples because the fruit of size grade 1A are very small in diameter and length. Therefore, the size grade 1 in $\underline{C}_3$ and $\underline{C}_4$ should be interpreted not as 1 but rather 1B.

Comparing $\underline{C}_1$ with $\underline{C}_2$, $\underline{C}_3$ and $\underline{C}_4$ leads to the following general conclusions:

(1) The distribution of new fruits $X_c(T+1)$ as a function of the distribution of the old fruits $N_c(T)$ in a given season does not appreciably change with the location.

(2) For a given location, the seasons do not appreciably affect the distribution of new fruits $X_c(T+1)$ as a function of the distribution of the old fruit $N_c(T)$.

It should be noted here that the above generalizations c

i

c

t

c

Ta

gr

do not apply to fruits of size grade four. Also, inherent is the assumption that in making these generalizations the cultural practices remain the same within different locations and seasons.

Further investigations into the behavior of fruits of size grade four led to the following conclusions:

(1) The reason that the estimated parameters for the new fruits of size grade four $X_4(T+1)$ as a function of the old fruits of size grade 3B and 3A [$N_{3B}(T)$ and $N_{3A}(T)$] vary among locations and seasons, is that the populations from which the samples were drawn were at different average maturity levels* (See Table 4).

Table 4. Mean and standard deviation of fruit of size grade four for different locations and seasons.

Location	Season	Mean	Standard Deviation
West	1971	4	3
North	1971	4	3
West	1972	6	5
North	1972	3	2

---

*Measured as average number of fruit of size grade 4 in the data set of a given location and season.

(2) A higher average maturity level was found to be associated with an increase in the magnitude of the parameter C_{12} (coefficient in the first row and second column of the matrix $\underline{C}$) and a decrease in the magnitude of the parameter C_{13}.

(3) A lower average maturity level was associated with a decrease in the magnitude of the parameter C_{12} and an increase in the magnitude of the parameter C_{13}.

The above findings are consistent both in theory as well as in actual field experimentations. It can be argued that at relatively low maturity levels where the growth of fruit is not limited by the available nutrient and moisture supplies, it is possible that comparatively more fruits which are size grade 3A today grow to size grade 4 tomorrow. However, at a higher maturity level, where the growth is limited both by the available food supply and the competition among fruits of various size grades, we would expect the growth of the fruit of size grade 3B to size grade 4 to be much higher than the growth of fruit of size grade 3A.

The results of actual field experimentations provided further support to the above theory. In these experiments actual observations were made on a given number of fruits of various size grades over a relatively long duration of time. All fruits in a selected area of 50 ft^2 were labeled. Diameter measurements on these fruits were made daily for 9 days, the harvest date being approximately in the middle of this nine-day period. From these diameter measurements the number and the weights of fruits of various size grades were calculated for each day. The results of an earlier study (41) indicated that the weight is related to the diameter by the following relationship:

$$W = 3.8\, e^{1.96D}$$

Where:

W = weight of fruit in grams

and:

D = diameter of fruit in inches.

The results of one such diameter study for a limited duration are given in Table 5. An analysis of transition behavior of fruit in various size grades from these diameter studies suggested that the conclusions numbered one through three are reasonable. Further experimentation of a controlled nature is needed to get more positive results.

The discussion so far has addressed itself to a broad aspect of the problem, identifying the sensitive parameters (coefficients C_{12} and C_{13} in the matrix $\underline{C}$) and the relative significance of these parameters with respect to time in the model. To use the above information for making decisions, however, one requires reliable estimates of the numerical values of these parameters.

This latter aspect of the problem for the given model can be answered best by means of a parametric simulation. For our purpose, we will define parametric simulation as a process in which the only change from one simulation run to the next is in the level of a given parameter or parameters. The initial conditions and the input remains the same in all simulation runs.

The results of one such parametric simulation at four levels of C_{12} are presented in Table 5. These levels are meaningful intuitively, but otherwise are arbitrary. The parameter C_{13} was retained at its original level and was not subjected to such a study because its contribution in explaining the behavior of fruit of size grade four is very small. Table 5 also shows the actual observed dynamic response of the system for the given time period as determined by the diameter study.

The results of Table 5 indicate that at low maturity levels (first three days), the simulation results remain unaffected for any parameter value within the range of 0.23 to 0.55. However, the effect begins to become apparent on the fourth day. At high maturity levels the growth of fruit of size grade four increases dramatically with the increase in magnitude of parameter C_{12}. This implies that any parameter C_{12} such as one found in matrices $\underline{C}_1$, $\underline{C}_2$, $\underline{C}_3$ and $\underline{C}_4$ will be adequate if the object of the simulation is to predict the distribution of the fruit of various size grade at low maturity levels (approximately 50 to 25 per cent by count of number 2 fruit in the sample) only.

Table 5. The results of parmaetric simulation.

Day	Observed/ Simulated	Level of Parameter C_{12}	Number of					
			N_{1A}	N_{1B}	N_2	N_{3A}	N_{3B}	N_4
1	Observed	---	32	38	68	2	0	0
2	Observed	---	30	36	62	25	0	0
2	Simulated	.23	28	38	57	22	1	0
2	Simulated	.35	28	38	57	22	1	0
2	Simulated	.45	28	38	57	22	1	0
2	Simulated	.55	28	38	57	22	1	0
3	Observed	---	28	36	52	30	6	1
3	Simulated	.23	26	37	49	29	9	2
3	Simulated	.35	26	37	49	29	8	2
3	Simulated	.45	26	37	49	29	8	2
3	Simulated	.55	26	37	49	29	8	2
4	Observed	---	28	36	40	32	14	4
4	Simulated	.23	24	36	44	31	17	6
4	Simulated	.35	24	36	44	31	16	7
4	Simulated	.45	24	36	44	31	15	8
4	Simulated	.55	24	36	44	31	14	9

Table 5. Continued.

Day	Observed/ Simulated	Level of Parameter C_{12}	Number of					
			N_{1A}	N_{1B}	N_2	N_{3A}	N_{3B}	N_4
5	Observed	---	30	36	32	28	20	18
5	Simulated	.23	23	34	40	30	25	13
5	Simulated	.35	23	34	40	30	22	16
5	Simulated	.45	23	34	40	30	20	18
5	Simulated	.55	23	34	40	30	18	20
6	Observed	---	24	30	26	22	14	32
6	Simulated	.23	22	32	36	28	30	21
6	Simulated	.35	22	32	36	28	25	26
6	Simulated	.45	22	32	36	28	22	29
6	Simulated	.55	22	32	36	28	19	32

However, the usefulness of the model might be extended to include high maturity levels (50 to 15 per cent of number 2) simply by using a relatively high magnitude of the parameter C_{12} such as 0.45 to 0.55.

6. LINEAR DYNAMIC MODELS

The model presented in 3.9.1 and 3.9.2 is a class of models known as linear dynamic models. These are described (37) as follows:

Let:

$\vec{U}(T)$ = input or stimulus vector which includes the m inputs

$\vec{Q}(T)$ = state vector that includes the n "internal" variables which relate the input to the output and which account for past inputs and states

and

$\vec{Y}(T)$ = output vector that includes the q outputs of the system model.

Usually $m \leqslant n$ and $q \leqslant n$. Otherwise, some of the inputs and/or outputs are redundant. Using the above notation the structure of a linear constant-parameter discrete time state model is given as follows:

$$\vec{Q}(T+1) = \underline{A}\,\vec{Q}(T) + \underline{B}\,\vec{U}(T) \qquad 6.1$$

$$\vec{Y}(T) = \underline{E}\,\vec{Q}(T) \qquad 6.2$$

Comparing 3.9.1 and 3.9.2 with 6.1 and 6.2 we have:

$$\vec{N}_c(T) = \vec{Q}(T)$$

and:

$$\vec{W}_c(T) = \vec{Y}(T)$$

Thus, it is seen that our model is a linear difference equation model. The theory for such models is well developed (37) and they can be shown very conveniently in the form of a block diagram such as the one shown in Figure 9. The overall system behavior of such models can be determined either by a direct simulation of the system from the block diagram or from the formal solution of the aggregated model. Of course, either solution yields the same input-output behavior.

6.1 Analytic Solution

Given the initial condition $\vec{N}_c(0)$ and the input $\vec{U}(T)$, T = 0, 1, 2, ... t-1, the general solution $\vec{N}_c(T)$ for 3.9.1 can simply be found by iterating for T = 0, 1, 2 ... , etc. Thus, at T = 1

$$\vec{N}_c(1) = \underline{A}\ \vec{N}_c(0) + \underline{B}\ \vec{U}(0)$$

at T = 2

$$\vec{N}_c(2) = \underline{A}^2\ \vec{N}_c(0) + \underline{A}\ \underline{B}\ \vec{U}(0) + \underline{B}\ \vec{U}(1)$$

at T = 3

$$\vec{N}_c(3) = \underline{A}^3\ \vec{N}_c(0) + \underline{A}^2\ \underline{B}\ \vec{U}(0) + \underline{A}\ \underline{B}\ \vec{U}(1) + \underline{B}\ \vec{U}(2)$$

$$\vdots$$

and at T = t

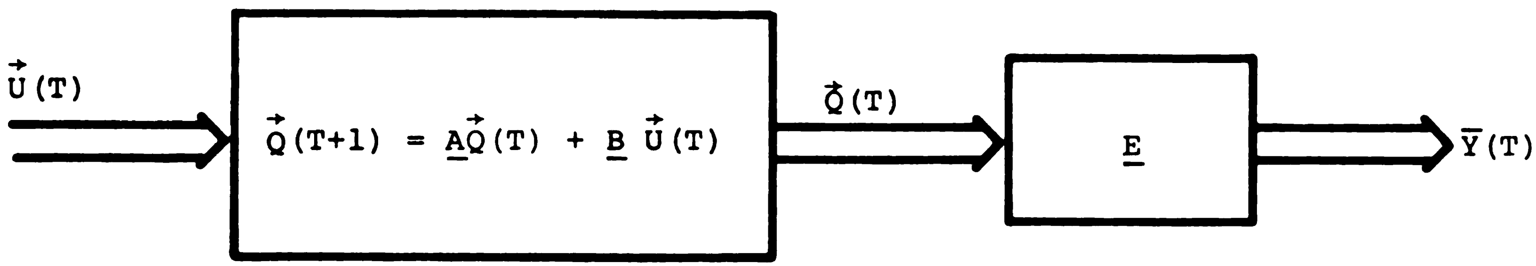

Figure 9. Linear constant-parameter dynamic model.

$$\vec{N}_c(t) = \underline{A}^t \vec{N}_c(0) + \underline{A}^{t-1} \underline{B} \vec{U}(0) + \underline{A}^{t-2} \underline{B} \vec{U}(1) +$$

$$\ldots + I \underline{B} \vec{U}(t-1)$$

or:

$$\vec{N}_c(t) = \underline{A}^t \vec{N}_c(0) + [\underline{A}^{t-1}, \underline{A}^{t-2}, \underline{A}^{t-3}, \ldots, \underline{I}] \begin{bmatrix} \underline{B} \vec{U}(0) \\ \underline{B} \vec{U}(1) \\ \underline{B} \vec{U}(2) \\ \vdots \\ \underline{B} \vec{U}(t-1) \end{bmatrix}$$

and in summation form:

$$\vec{N}_c(t) = \underline{A}^t \vec{N}_c(0) + \sum_{j=0}^{t-1} \underline{A}^{(t-j-1)} \underline{B} \vec{U}(j)$$

The stability behavior of the states is determined by $\underline{A}^t$ which will grow or decay with increasing t depending upon the eigen-values of $\underline{A}$. If they all are less than unity, the system is asymptotically stable (37). Based on this criteria our model in 3.9.1 is found to be unstable. However, we note here that the model in 3.9.1 is developed only to study the transient rather than the steady-state characteristics of the system. Due to the instability, steady-state conditions do not exist. This is so because we are simulating a complex process via a simplified model. It is for this reason that the model was claimed to be valid only over a limited range of field maturity.

7. CASH VALUE PREDICTION

The objectives of this study were stated in general terms at the end of Chapter 1. One of these objectives was to measure the performance of the model in predicting the cash value of fruit from a given production unit. In this chapter the quality of forecast of cash value by the model is evaluated.

The value of a cucumber crop is assessed on a weight basis. A survey of the industry indicated that both aggregate as well as differential pricing systems are in vogue. In either case, the value of the product of size grade four at any time is zero. The following examples give an indication of the two pricing structures.

7.1 Aggregate Price Structure

In aggregate pricing the value of the product is assessed on the useable weight basis. Useable weight [UW(T)] at any time T is defined as the total weight [W(T)] of the product minus the weight of the product of size grade four [$W_4(T)$]. Thus:

$$UW(T) = W(T) - W_4(T)$$

Then, the value of the crop per acre at a given time T is given as follows:

$$A(T) = a \cdot b \cdot [W_{1A}(T) + W_{1B}(T) + W_2(T) + W_{3A}(T) + W_{3B}(T)]$$

or:

$$A(T) = a \cdot b \cdot UW(T)$$

Where:

A = value of the crop, \$/acre

a = a constant depending upon the area used in the sampling procedure. For the standard sampling procedure described in this study $a = 435.0$. This implies that 100 ft^2 area is covered in obtaining the sample.

b = a constant with units of dollars/pound. Its particular value is agreed upon in a contract between the producer and the processor.

W_c = refers to the weight of fruit of size grade c, c = 1A, 1B, 2, 3A and 3B in a given sample.

7.2 Differential Price Structure

In differential pricing the value of the produce is assessed on the basis of the individual weights of fruit of various size grades in the sample.

$$D(T) = a \cdot [e_{1A} \cdot W_{1A}(T) + e_{1B} \cdot W_{1B}(T) + e_2 \cdot W_2(T) +$$

$$e_{3A} \cdot W_{3A}(T) = e_{3B} \cdot W_{3B}(T) + e_4 \cdot W_4(T)]$$

Where:

D = value of the crop, dollars/acre

$a = 435.0$

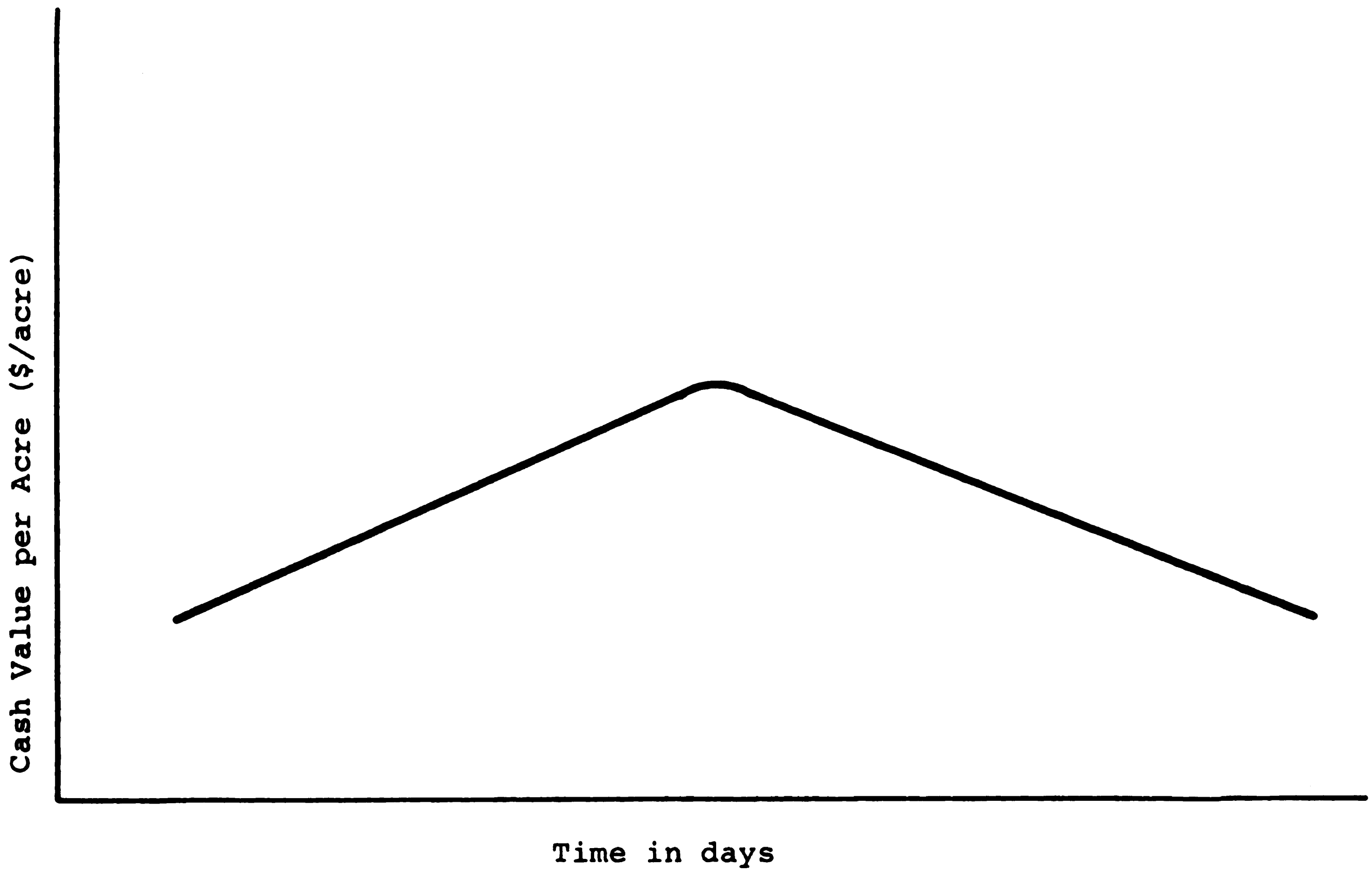

Figure 10. Expected cash value behavior.

e_c = value of the product of size grade c, c = 1A, 1B, ... , 4 in dollars/pound. Usually $e_{1A} \geq e_{1B} > e_2 > e_{3A} > e_{3B} > e_4$. Typically $e_4 = 0$.

The two pricing structures discussed above may have advantages and disadvantages over one another. This, however, is not the basic issue here. The issues to be addressed here are:

(1) Given a price schedule, how can it be used in making decisions regarding scheduling of once-over harvest?

(2) To obtain some quantitative measure of the performance of the model in predicting cash value of the fruit based on the above pricing schedules.

The answer to the first question can be given by the theory and an understanding of the system's behavior. Thus, if cash value per acre (obtained from either schedule) is plotted against time, its expected behavior will be as shown in Figure 10. The figure shows a steady increase in cash value reaching a peak and then a steady decline as time progresses. This particular behavior in aggregate pricing is expected because of the zero value of fruit of size grade four and in differential pricing this feature combined with the relative weight factors ($/lb) of fruit of different size grades. Thus, it is obvious that any of the price schedules can be used for locating an optimum day of harvest.

Figure 11 shows the behavior of useable weight

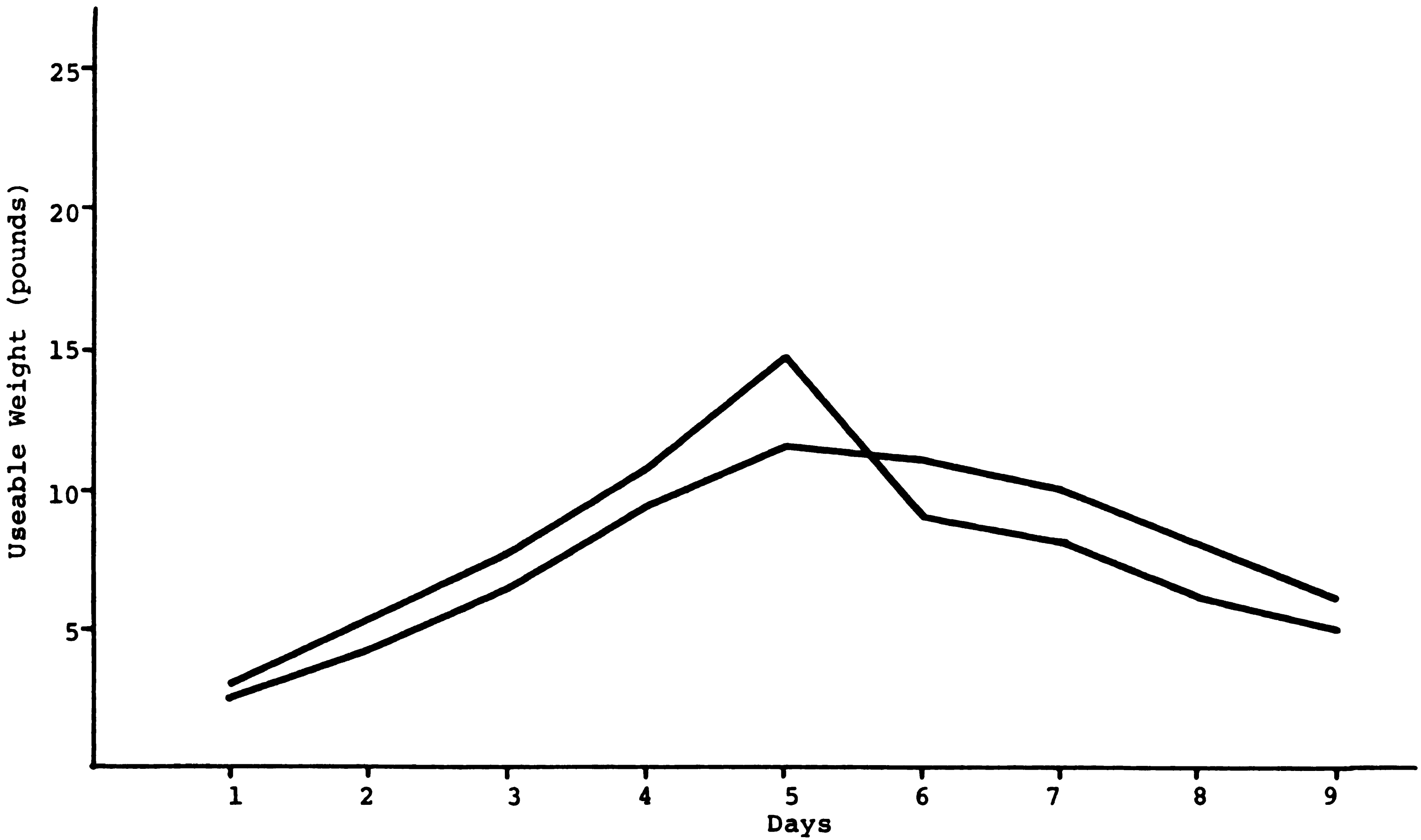

Figure 11. Useable weight behavior.

over a relatively long time period (9 days) determined from two diameter field studies referred to in Chapter 5. Figure 11 underscores the fact that the useable weight and hence, the cash value over time for some fields, may have a more pronounced peak than for other fields. In such situations a decision as to the time of harvest should be based not only on the consideration of optimal cash value, but also the product size composition.

To deal with the second issue a number of simulations were conducted on data-sets from all locations and all seasons. The simulation results relate to not only the narrow problem of cash value prediction, but also gives an indication of the overall performance of the model. Simulations were conducted to determine the model performance with respect to (1) one-day forecasts, (2) two-day forecasts, and (3) up to five-day forecasts. The third gives an indication of the expected errors in forecasts of greater than two days. The results of these simulations are presented in Tables 6 through 10. The notation used in these tables is the same as given in the section on model validation. In all cases considered the standard deviation of error in predicting cash value is approximately 15 per cent.

Table 6. Means and standard deviations of percentage error for one day forecast for 1971 season.

	Numbers							Weights								Cash Value per Acre	
	N_{1A}	N_{1B}	N_2	N_{3A}	N_{3B}	N_4	N	W_{1A}	W_{1B}	W_2	W_{3A}	W_{3B}	W_4	W	Useable Weight	Aggregate Pricing	Differential Pricing
								West - 1971									
Mean	-2	0	1	0	0	0	-1	-0.1	0.3	1.1	-0.5	-0.6	-1.1	-0.8	0.3	-0.04	0.73
Standard Deviation	5	6	7	4	3	1	11	0.7	2.3	6.5	7.3	7.5	4.4	9.6	9.2	10.47	10.86
								North - 1971									
Mean	-3	-2	0	1	1	-1	-3	0.3	-0.6	1.5	1.5	1.9	-1.8	2.8	2.2	2.71	1.84
Standard Deviation	9	11	10	5	4	2	16	2.8	3.6	9.2	9.3	8.9	5.6	12.2	11.7	12.55	15.41

Table 7. Means and standard deviations of percentage error for one day forecast for 1972 season.

	Numbers						Weights						Useable Weight	Cash Value per Acre	
	N_1	N_2	N_{3A}	N_{3B}	N_4	N	W_1	W_2	W_{3A}	W_{3B}	W_4	W		Aggregate Pricing	Differential Pricing
							West - 1972								
Mean	1	-5	3	2	0	1	0.9	-2.0	2.4	5.4	2.5	1.0	1.5	2.67	1.45
Standard Deviation	11	9	8	5	3	13	2.3	5.4	9.1	9.0	9.2	11.0	15.0	16.71	12.34
							North - 1972								
Mean	-4	-2	2	0	-1	-5	-0.3	-1.9	1.7	0.4	-0.3	-0.4	-0.1	-0.32	-3.12
Standard Deviation	12	8	6	4	3	11	3.4	5.7	8.0	7.6	8.0	16.2	12.9	15.17	14.85

Table 8. Means and standard deviations of percentage error for two day forecasts.

	Numbers							Weights							Use-able Weight	Cash Value per Acre	
	N_{1A}	N_{1B}	N_2	N_{3A}	N_{3B}	N_4	N	W_{1A}	W_{1B}	W_2	W_{3A}	W_{3B}	W_4	W		Aggre-gate Pricing	Differ-ential Pricing
								1971 Season									
Mean	-2	-2	1	1	0	-1	-2	0.3	-0.5	2.4	0.9	0.6	-1.9	1.7	2.5	2.61	2.20
Standard Devia-tion	8	8	9	5	4	2	13	2.4	3.0	8.0	8.5	8.2	5.2	11.6	10.8	11.48	11.89
								1972 Season									
Mean	-2		-1	2	0	-2	-3	0.2		-1.4	0.8	0.9	-2.4	-1.9	0.5	0.40	-0.92
Standard Devia-tion	12		8	7	4	3	9	3.4		5.9	8.9	8.6	8.7	13.9	11.2	15.30	11.82

Table 9. Means and standard deviations of percentage error for forecast up to 5 days for 1971 season.

	Numbers							Weights								Cash Value per Acre	
	N_{1A}	N_{1B}	N_2	N_{3A}	N_{3B}	N_4	N	W_{1A}	W_{1B}	W_2	W_{3A}	W_{3B}	W_4	W	Use-able Weight	Aggre-gate Pricing	Differ-ential Pricing
								West - 1971									
Mean	-1	0	2	1	0	-1	1	0.0	0.4	2.4	0.4	-0.7	-2.0	0.4	2.4	2.41	3.28
Standard Devia-tion	6	7	8	4	3	2	12	0.7	2.4	7.0	7.5	7.7	4.7	10.7	10.4	11.80	11.56
								North - 1971									
Mean	-3	-3	0	2	2	-1	-3	0.3	-1.0	1.9	2.5	2.7	-2.4	4.0	3.4	3.76	2.57
Standard Devia-tion	9	10	9	6	4	2	15	2.8	3.5	8.7	10.1	8.3	5.8	13.6	12.5	13.41	14.38

Table 10. Means and standard deviations of percentage error for forecast up to 5 days for 1972 season.

	Numbers						Weights							Cash Value per Acre	
	N_1	N_2	N_{3A}	N_{3B}	N_4	N	W_1	W_2	W_{3A}	W_{3B}	W_4	W	Useable Weight	Aggregate Pricing	Differential Pricing
							West - 1972								
Mean	1	-5	3	2	0	1	1.0	-2.4	2.8	5.8	1.9	3.1	2.1	3.37	2.77
Standard Deviation	11	8	8	4	3	13	2.2	4.8	8.5	8.5	9.3	17.2	14.6	16.29	12.19
							North - 1972								
Mean	-4	-2	3	1	-1	-4	-0.4	-2.0	2.3	1.2	0.0	1.1	1.1	1.25	-2.42
Standard Deviation	11	8	5	4	3	12	3.3	5.9	7.6	7.0	8.3	14.2	11.9	13.63	14.54

8. USE OF THE MODEL

Whenever a complex system is modeled, the behavior of the system predicted by the model must be reviewed critically. All of the assumptions used in developing the model must be considered; all of the possible ways in which the model might differ from the real system must be examined. The behavior predicted for the system can only be as accurate as the model used to represent the system. Therefore, the modeling of a complex system should generally be regarded as an aid to decision making rather than as a decision making process itself (42).

In specifying the way in which the model results relate to and can be used in actual management practice, several items must be considered. The fruit distribution or the cash value suggested by the model for a given field on a given day represents not the distribution or the cash value that one could verify with one sample by the procedure described in Section 3.3, but is rather the <u>expected</u> distribution or the cash value that one would find through the repeated application of the sampling procedure.

The model has been verified only for those varieties and cultural practices common to Michigan commercial pickling cucumber production. Extremes of weather (e.g., cold shock) are not adequately considered in the model either.

The assumption was made in developing the fruit number dynamics structure that no fruit is lost as a result of wind or other natural causes in transition from one size grade to the next. This assumption should be further verified. It appears from the author's own observations of field plots that this assumption is not critical in making a short range (e.g., 3 days) forecast, but its accumulated effect in long range (e.g., 9 days) forecasts needs further investigation.

It should be apparent from the above discussion that although the results from the model presented can be used as a guide in determining fruit size distributions and/or cash values, the role of the manager in the actual decision-making process is of critical importance in achieving a successful transition between the model and reality.

9. IMPLEMENTATION

Two methods of implementation of the fruit development model for forecasting the optimum day of harvest were considered. One possibility is to summarize the major features of the model in tables or charts to be used in the field immediately after a field sample is taken. (Two different forms have been developed and subjected to limited field tests.) The implementation costs (time) and dollars) are very low but the information obtained may not be adequate.

Most experience to date, however, is in the utilization of a teletype-telephone hook-up to a commercial time share computing system. Such a system is surprisingly inexpensive when utilized for all fields of cucumbers contracted by a processor. Total fixed hardware costs (teletype rental and supplies) are less than $200 per harvest season. Individual field forecasts are then obtained at a cost on the order of $0.50 to $1.00 (telephone and computer charges). In addition to the forecast itself an accounting capability is readily available which could be used to chronologically order all of the expected harvests for a processor. Thus, a total daily fruit volume flow into the plant forecast would also be available.

Dr. J. B. Holtman, major professor of the author, in collaboration with Mr. Robert Milligan of the Agricultural Economics Department of Michigan State University, has developed a program for the implementation of the model for a small commercial time-sharing system. This program stores the incoming field samples and returns a forecast of fruit size distribution and cash value for each of the next three days. Utilizing the teletype-telephone hook-up to the time shared system, two Michigan processors evaluated the system during the 1971 season. It was found that this system was a highly reliable and efficient method of implementing the forecasting system. After only four hours of training, a plant clerk (who previously had no related experience) was utilizing the system with no difficulty.

10. SUMMARY AND CONCLUSIONS

The general problem of scheduling formed the basis of the thesis. In particular a cucumber fruit development model was developed to assist the manager in making a harvest time decision for once-over mechanical harvest. The application of the model for the grower is in making reliable forecasts of the harvest date of a field 2-3 days in advance to optimize his expected gains. The application of the model for the processor includes scheduling to reduce field losses due to unforseen shortages of equipment and/or labor and regulation of the flow of desired size distribution mix into the plant.

The model proposed is a class of general linear dynamic models of discrete time and constant parameter form. The model can be viewed in light of discrete state space theory or population dynamics. Both views are helpful. The states of the model are the number of fruit of various size grades at time T. The output from the model is the weight of fruit of various size grades at time T. The input to the model is a unit step that corresponds to a constant number of new fruit of smallest size grade entering into the system each day. The parameters in the model were estimated using the technique of multiple regression with least squares criterion.

Model verification was attempted using both the available statistical tests as well as analyses of the time behavior from the known historical record. Model verification along with sensitivity analysis has proved the general adequacy of the model in making harvest time decisions for different locations and seasons.

The model was implemented for on-line applications via a teletype-telephone-time share computer hook-up.

The following general conclusions can be made:

(1) The model proposed for cucumber fruit development is valid under the assumptions made in its development.

(2) The model shows promise of being a valuable tool in accurately scheduling 2-3 days in advance the optimal harvest date.

(3) The model implementation is feasible economically and operationally.

11. RECOMMENDATIONS

(1) Further refinement in the model can be accomplished by studying the effect of environmental variables on the system considered. Basic studies are needed to relate the effect of such variables as temperature and moisture to the rate of development of new fruits.

(2) The exact behavior of the fruit of the smallest size grade (input) needs to be further explored.

(3) The assumption that the loss of fruits as a result of wind is negligible, needs experimental verification.

(4) The broad framework presented in this thesis for fruit population dynamics is applicable and useful for other types of population studies where interactions depend upon population member size as well as time.

(5) A detailed study of variations among varieties should be considered.

REFERENCES

REFERENCES

1. Miller, C. H. 1957. Studies on the nutrition and physiology of pickling cucumbers. Ph.D. Thesis, Michigan State University.

2. Research project report on post-harvest handling of cucumbers. 1970-72. Department of Agricultural Engineering, Michigan State University.

3. Stout, B. A., M. M. DeLong, D. H. Pettengil, and S. K. Ries. 1964. A once-over mechanical harvester for pickling cucumbers. Mich. Agr. Exp. Sta. Quart. Bul. 46(3):420-430.

4. Sims, W. L., M. B. Zahara. 1968. Growing pickling cucumbers for mechanical harvesting. Agri. Ext. Ser. University of California, AXT-270.

5. Growing cucumbers for processing. 1968. Asgrow Seed Company, Orange, Connecticut.

6. Putnam, Alan R. 1963. Horticultural aspects concerned with the production of pickling cucumbers for once-over harvest. Thesis for the degree of M.S., Michigan State University.

7. Morrison, F. D. and S. K. Ries. 1967. Cultural requirements for once-over mechanical harvest of cucumbers for pickling. Amer. Soc. Hort. Sci. 91:339-346.

8. Miller, C. H. and G. R. Hughes. 1969. Harvest indices for pickling cucumbers in once-over harvested systems. Amer. Soc. Hort. Sci. 94:485-487.

9. Morrison, F. D. 1966. Cultural and environmental parameters for mechanically harvested cucumbers. Ph.D. Thesis, Michigan State University.

10. Hopen, H. J. 1962. Environmental factors affecting the growth of Cucumis Sativus L. with special reference to carbon dioxide. Ph.D. Thesis, Michigan State University.

11. Lewin, I. J. 1970. Effect of photoperiod and nyctoperiod temperatures, and moisture stress on fruit enlargement in the cucumber, Cucumis Sativus L., CV. Mini-Cuke. Ph.D. Thesis, The Pennsylvania State University.

12. Tukey, L. D. 1962. Factors affecting rhythmic diurnal enlargement and contraction in fruits of the apple (Malus domestica Bork.).

13. Harley, C. P. and M. P. Masure. 1938. Relation of atmospheric conditions to enlargement rate and periodicity of Winesap apples. J. Agri. Res. 57:109-124.

14. Tukey, L. D. 1960. Some effects of night temperature on the growth of McIntosh apples. II. Proc. Amer. Soc. Hort. Sci. 75:39-46.

15. Went, F. W. 1944. Plant growth under controlled conditions, II. Thermoperiodicity in growth and fruiting of the tomato. Amer. J. Bot. 31:135-150.

16. Aldrich, W. W. and A. Work. 1934. Evaporating power of the air and top-root ratio in relation to rate of pear fruit enlargement. Proc. Amer. Soc. Hort. Sci. 32:115-123.

17. Bailey, L. H. 1947. The Standard Encyclopedia of Horticulture. 905-907. The MacMillan Company, New York.

18. Bailey, L. H. 1963. Unpublished data.

19. Bailey, L. H. and A. R. Putnam. 1963. The feasibility of a once-over mechanical harvester for pickling cucumbers. Mich. Agr. Exp. Sta. Quart. Bul. 42:2-23.

20. Dearborn, R. B. 1936. Nitrogen nutrition and chemical composition in relation to growth and fruiting of the cucumber plant. Cornell Univ. Agri. Exp. Sta. Memoir 192.

21. Vaile, J. E. 1942. Ark. Expt. Sta. 53rd Ann. Rpt. Bul. 417:28.

22. Rodnikov, N. I. 1944. The rate of maturity and the yield of cucumbers grown under glass and their relationship to mineral nutrition. Proc. Sci. Conf. Timirjazev Agri. Acad. 3-10 June, No. 1, pp. 45-46 (Hort. Abst. 16:39).

23. McCollum, R. E. and C. H. Miller. 1971. Yield, nutrient uptake, and nutrient removal by pickling cucumbers. J. Amer. Soc. Hort. Sci. 96(1): 42-45.

24. Gustafson, F. G. 1926. Growth studies on fruits. Plant Phys. 1:265-272.

25. Sinnot, W. E. 1945. The relation of growth to size in cucurbit fruits. Amer. J. Bot. 32:439-446.

26. Bomalski, H. H. 1948. Growing degree days. Food Packer, 29:51-59.

27. Reath, A. N. and S. H. Wittwer. 1952. The effects of temperature and photoperiod on the development of pea varieties. Proc. Amer. Soc. Hort. Sci. 60:301-310.

28. Arnold, C. Y. 1959. The determination and significance of the base temperature in a linear heat unit system. Proc. Amer. Soc. Hort. Sci. 74: 430-445.

29. Holtman, J. B., A. Patel, F. Panol and B. F. Cargill. 1972. A mathematical model to schedule cucumber harvest. Amer. Soc. Agri. Eng. 72-149.

30. Manetsch, T. J. 1971. Unpublished class notes prepared for systems science 812. Sys. Sci. Dept., Michigan State University.

31. Kmenta, J. 1971. Elements of Econometrics. MacMillan, New York, 750 pp.

32. Kane, E. J. 1968. Economic Statistics and Econometrics. Harper & Row Publishers, New York, 437 pp.

33. Draper, N. R. and H. Smith. 1966. Applied Regression Analysis. John Wiley & Sons, Inc., New York, 407 pp.

34. Rafter, M. E. and W. L. Ruble. 1969. Stepwise addition of variables. Agr. Expt. Sta. Stat. Series Description No. 9. Michigan State Univ.

35. Rafter, M. E. and W. L. Ruble. 1969. Stepwise Deletion of Variables. Agr. Expt. Sta. Stat. Series Description No. 8. Michigan State University.

36. Rafter, M. E. and W. L. Ruble. 1969. Calculation of Least Squares Problems. Agri. Expt. Sta. Stat. Series Description No. 7. Michigan State Univ.

37. Manetsch, T. J. and G. L. Park. 1972. System Analysis and Simulation with Applications to Economic and Social Systems. Preliminary Edition. Department of Electrical Engineering and Systems Science. Michigan State University.

38. Wonnacott, R. J. and T. H. Wonnacott. 1970. Econometrics. John Wiley & Sons, Inc., New York.

39. Durbin, J. 1960. Estimation of parameters in time series regression models. Journal of the Royal Statistical Society, Vol. 22, Series B, pp. 139-153.

40. Asimow, Morris. 1962. Introduction to design. Prentice-Hall Inc., Englewood Cliffs, N. J., 135pp.

41. Bingley, G. W. 1959. Construction, evaluation, and efficiency studies of a mechanical cucumber harvester. M.S. Thesis, Michigan State University.

42. Patel, Ashok K. 1970. Development of a forage-feed planning model for the Rutgers dairy herd: An activity analysis approach. M.S. Thesis, Rutgers University, N. J.

43. Naylor, Thomas H., Joseph L. Balintfly, Donald S. Burdick and Kong Chu. 1968. Computer Simulation Techniques. John Wiley and Sons, Inc., New York, 322 pp.

APPENDICES

APPENDIX A

Based on the studies of Lewin (11) the development of the new fruits was hypothesized to take the following form:

$$X_c(T+1) = \left(a + b\,\frac{N_{c-1}(T)}{W_{c-1}(T)} + c \cdot TX + d \cdot TN + e \cdot DT\right) \cdot N_{c-1}(T)$$

$$TX = (TMAX - 85)^2$$

$$TN = (TMIN - 65)^2$$

$$DT = (TMAX - TMIN - 20)^2$$

Where:

X_c = number of new fruit of size grade c

N_c = number of old* fruit of size grade c

W_c = weight of old fruit of size grade c

TMAX = maximum temperature between the sampling times, T and T+1

TMIN = minimum temperature between the sampling times, T and T+1.

The constants 85 and 65 in TX and TN are respectively the most favorable phototemperature, and nyctotemperature for cucumber fruit enlargement. The constant 20 in DT is the estimate of diurnal thermoperiodicity which produced maximum growth rate as found by Lewin (11).

---

*"Old" as defined in 3.5.10.

The parameters a, b, ... , e in the above equation were estimated by multiple regression utilizing the method of stepwise deletion of variables. The final results of such a run were presented in Section 3.7.

The independent variables TX, TN, DT and N_{c-1}/W_{c-1} were deleted from the final model for the following reasons:

(1) The sign of their coefficients were inconsistent with results of previous investigations.

(2) Each of them had a significance probability level above 0.01.

(3) The variability in fruit development explained by them was very small.

Changing parameters in TX, TN or DT made little difference in result as compared to the one above.

APPENDIX B

The matrices $\underline{A}$, $\underline{B}$ and $\underline{E}$ in 3.9.1 and 3.9.2 for a simulation model based on size grades 1, 2, 3 and 4 were estimated to be as follows:

$$\underline{A} = \begin{bmatrix} 1 & 0.13 & 0 & 0 \\ 0 & 0.87 & 0.30 & 0 \\ 0 & 0 & 0.70 & 0.20 \\ 0 & 0 & 0 & 0.80 \end{bmatrix}$$

$$\underline{B} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$U(T) = 6$$

$$M_i \leqslant M(T) \leqslant M_f$$

$$M_i = 0.50$$

$$M_f = 0.25$$

$$\underline{E} = \begin{bmatrix} 0.485 & 0 & 0 & 0 \\ 0 & 0.285 & 0 & 0 \\ 0 & 0 & 0.127 & 0 \\ 0 & 0 & 0 & 0.035 \end{bmatrix}$$

APPENDIX C

If the parameters in $\underline{A}$ in 3.9.1 are estimated directly from the observed sample values they will be biased and inconsistent. The estimated parameters that result from ordinary least squares are as follows:

$$\underline{A} = \begin{bmatrix} 1.02 & 0.23 & 0.08 & 0 & 0 & 0 \\ 0 & 0.44 & 0.49 & 0 & 0 & 0 \\ 0 & 0 & 0.70 & 0.24 & 0 & 0 \\ 0 & 0 & 0 & 0.64 & 0.36 & 0 \\ 0 & 0 & 0 & 0 & 0.82 & 0.13 \\ 0 & 0 & 0 & 0 & 0 & 0.47 \end{bmatrix}$$

Notice that the columns do not sum to one. This implies that the model structure in 3.5.1 is not satisfied.