AGRICULTURAL RESPONSE TO HYDROLOGIC DROUGHT

by
V.J. BIDWELL

July 1972

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

53

AGRICULTURAL RESPONSE TO HYDROLOGIC DROUGHT

by

V.J. Bidwell*

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80521

July 1972 No. 53

*Research Associate, Colorado State University, Civil Engineering Department, Fort Collins, Colorado

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The research leading to this paper was conducted as part of the large continental droughts project with principal investigator Dr. V. M. Yevjevich and financially supported by the U.S. National Science Foundation under grant GK-11564.

ABSTRACT

The paper deals with the effects and evaluation of the effects of hydrologic drought on the yield of dryland farming. Some of the physiological principles involved in crop growth are reviewed in the context of a system which relates reduction in crop production to deficits in the water input. A possible set of non-linear functions is developed to describe this system by using the physical properties of crop production as constraints on the mathematics. The method of non-linear least squares regression is described as a suitable technique for solving the system function. Application of the analytical methods to a set of corn yield data is demonstrated with several types of system functions, one of which is suggested as a suitable practical model. the predictability of the models is not high because of restriction imposed by the availability of data. Some improvements are suggested for further development when more suitable data are available.

Chapter 1

THE PURPOSE OF A DROUGHT-CROP MODEL

Of all the detrimental effects ascribed to droughts the total or partial loss of agricultural production is probably the most serious. Within this context is included both crops for direct harvesting and those used as pasture for animal grazing.

1.1 Definition of Drought

The question of what constitutes a drought has been asked many times and the answer can be given only in terms of the phenomenon being described. In this paper the case under consideration is dryland farming, i.e., the crop depends entirely on naturally occurring rainfall for its moisture requirements. Drought then occurs when the rainfall fails to supply sufficient moisture for optimum yields. The irrigated farm is not considered here because, ideally, irrigation should be managed in such a way that optimum yields are obtained. The case of insufficient irrigation water supply is therefore also outside the scope of this discussion.

1.2 The Causes of Drought

In section 1.1 it was stated that agricultural drought occurs whenever the yield is not optimal because of insufficient moisture supply. The term optimum yield in this context is defined only with respect to the moisture requirements of the crop. Even with adequate water the yield may not be the 'best' in an absolute sense because of soil, fertilizer, temperature conditions, biological factors etc. Insufficient moisture for optimal growth can be caused by deficiency in the amount and timing of rainfall and by climatic conditions which cause increased requirements. The latter conditions are generally grouped as evapotranspiration factors.

1.3 The Statistical Properties of Agricultural Drought

The factors which cause agricultural drought such as deficits in rainfall and the climatic variables related to evapotranspiration are themselves random with respect to time and space. Hence the effects on crop yields are also randomly distributed in both time and space because they are functions of random variables. The usual ways of combating random natural disasters such as floods and droughts are by economic methods such as insurance or engineering structures and systems designed for a certain level of damage which has a certain probability of occurrence. In either case the designer requires estimates of the probability distribution of the event in time and space. This information can be obtained in two ways:

(1) By a study of the probability distribution of the event.

(2) By relating the event to other variables whose probability distributions can be found.

The latter method is favored in the study of agricultural drought because of the scarcity of data on crop yield which is not confounded with factors other than hydrologic drought. The availability of rainfall data, for example, enables good estimates of its statistical properties to be made. If a suitable functional relationship can be derived between rainfall and drought effects then the statistical properties of droughts can be obtained. This approach can also be applied to areas where a particular crop is not yet grown in order to determine the economic feasibility of planting the crop.

1.4 The Drought-Crop Yield Relationship

It is the purpose of this study to develop a methodology for deriving the relationship between crop yield and the causative conditions. The case of total crop loss due to drought is not dealt with directly because the devastating conditions required are known and hence their probability of occurrence can be estimated. The case of partial crop damage however is more complex because the same value of less than optimal yield can be caused by a large number of combinations of causative conditions hence the need for a functional relationship. The study is addressed mainly to this problem.

The situation may arise when the joint probability distribution of all causative factors cannot be determined due to lack of data. The functional relationship may then be derived for those factors for which statistical information exists. The remaining factors are incorporated as uncertainty in the model. This is the case in the present study in which only rainfall was included as a causative factor due to lack of suitable data.

Chapter 2

WATER-SOIL-PLANT RELATIONSHIPS

2.1 The Factors Contributing to Yield

The purpose of this chapter is to review some of the basic concepts of the physical relationship between applied water and the yield of a crop. It is generally recognized [Blake, 1967]* that the total biological yield of a crop can be expressed as:

Total yield = (duration of photosynthesis) x (intensity of photosynthesis) - (respiration)

The harvestable or economic yield is then some proportion of the total biological yield. This formulation is affected by genetic and biological factors but is generally true for a particular crop situation. The rate of photosynthesis is affected by the following factors [Blake, 1967]:

(1) Light, which depends on day length, cloud cover and the area and orientation of leaf surfaces.

(2) Temperature - The photosynthetic rate doubles for each increase of $10^{o}C$ over the range-$6^{o}C$ to $30^{o}C$.

(3) Availability of atmospheric carbon dioxide which may be limiting only in thick foliage.

(4) Water status; water stress results in a sharp decline in photosynthesis.

The loss of plant matter by respiration of carbon dioxide during the night is affected by temperature in the same way as photosynthesis and can account for 10 percent to 50 percent of the carbon dioxide fixation during the day.

The above comments of course do not take into account the availability of nutrients and other factors. The main point to be considered is that the water status of the plant which is the main interest in a drought study is only one factor in crop production although a very important one. It is quite possible for the effects of low or very high temperatures, for example, to mask the effect of water deficit. The effects of water stress will now be considered in more detail.

*references are given by authors' names followed by the year of a reference at the end of this paper.

2.2 The Role of Water in Plant Growth

Water has several uses in the growth of plants and the following compilation from Currier [1967] outlines the principal roles:

(1) Water is an important structural constituent of the plant.

(2) Water hydrates enzymes and provides the aqueous environment necessary for metabolic reactions.

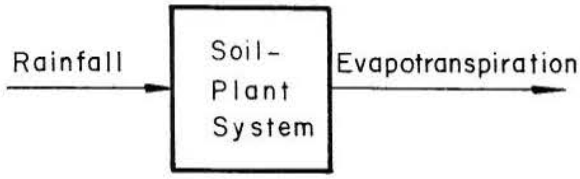
(3) Water is the medium for transport of substances from regions of supply to regions of use; e.g., transport of nutrients from the soil into the plant.

(4) Because of its high heat of vaporization, water is an effective coolant for the plant by transpiration and its high specific heat also helps in protection against low temperature injury.

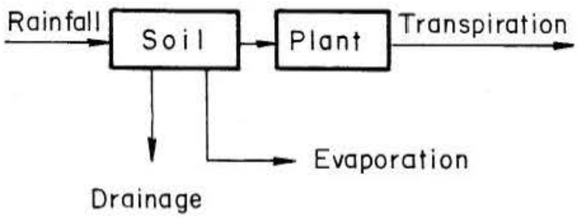
Some schools of thought [Pearson, 1967] maintain that transpiration serves no useful purpose directly but is an indirect effect of photosynthesis. The stomata in the leaf surface must be open to allow atmospheric carbon dioxide to reach the underlying mesophyll tissue where the photosynthetic reactions occur. Since the moist surface of the mesophyll is exposed, transpiration to the atmosphere results. One of the effects of water stress in the plant is that the stomata are closed partially or fully, thus reducing transpiration loss. Whatever the cause and effect relationships are, the end result is that with the stomata closed the growth processes cannot occur. Therefore the effect of mild water stress is to inhibit the growth oriented phases of the plant and more severe stress causes permanent structural damage or kills the plant.

2.3 Water Stress; The Soil-Water-Plant System

In order to determine the causes of undesirable water deficit in the plant, the water-soil-plant relationship can be considered as a system with rainfall input to the soil and transpiration output to the atmosphere.



This system can be further decomposed into two subsystems to account for soil and plant behavior separately.



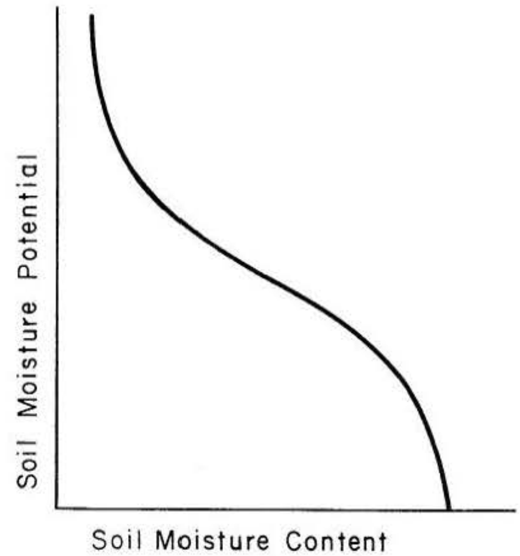


Fig. 1 Relation between soil moisture potential and soil moisture content.

2.4 The Soil System

The soil can be considered as a storage which receives water by infiltrating rainfall and loses it mainly by drainage until a state called field capacity is reached in which the remaining water in the soil is held mostly by surface tension forces. The small amount held by electrochemical forces is not of concern here. Further loss of water from the soil occurs by evaporation from near the soil surface or absorption into the root system of the plant. The evaporation of water from the soil surface is controlled by atmospheric conditions and soil moisture conditions near the surface. For humid climatic conditions and deep soil this loss is usually of a smaller order of magnitude than the water transpired by plants.

The soil water output to the plant is determined by the difference in potential between plant water and soil water and the conductivity of the root system. The soil moisture potential is a measure of the energy status of the soil water with respect to a free surface and can be considered as a function of the soil moisture content shown in Figure 1. This relationship is often complicated by hysteresis, i.e., different curves for increasing and decreasing moisture content.

Water transport can occur in soils in two modes: (1) viscous flow, and (2) diffusion in the liquid and vapor phases. The latter becomes important for quite dry soils although it is favored by large temperature gradients in wetter soils. It is important to realize that the parameters of the soil system behavior depend on the physical structure of the soil and its chemical properties. Thus a functional relationship derived for a particular soil may not apply to another soil or even for the same soil under a different tillage practice.

2.5 The Plant System

The plant receives water from the soil and transpires the bulk of it into the atmosphere. Any excess of transpiration rate over absorption rate from the soil will eventually result in internal water stress and the plant will suffer by reduction of growth rate, adverse structural effects (wilting) or may die. The transpiration rate is dependent on climatic conditions (discussed in a subsequent section) and the diffusion resistance of the leaf surface. This resistance is controlled to some extent by the water status of the plant and becomes limiting only at very high potential transpiration rates or existing water stress in the plant.

The input of soil water to the plant depends on the difference in potential between soil water and internal plant water and the conductivity of the root system. Plant water potential depends on the water status, hence on the transpiration rate, and soil water potential depends upon available soil moisture as mentioned previously. Another source of energy for absorption is the osmotic potential which expresses the solute concentration gradient between the soil water and plant water. If the soil water is too saline the adverse osmotic potential can prevent the absorption of the soil water by the plant.

The parameters of the plant system are highly non-stationary with respect to time because as the plant's foliage and root system develop with growth the absorption and transpiration characteristics change. The growth parameters of the same period in the growing season for an annual crop can change because of damage due to previous water stress or the root system may develop during a period of low soil moisture and make the plant subsequently more drought resistant.

2.6 Factors Controlling Evaporation and Transpiration

Evapotranspiration, as the phenomenon is collectively called, depends on plant or soil factors and climatic conditions. The climatic conditions are usually expressed in the form of the potential evapotranspiration. The actual evapotranspiration is then a proportion of the potential evapotranspiration; the proportion depends on the availability of soil water and the transpiration characteristics of the plant. Some of the many relationships proposed for the ratio of actual to potential evapotranspiration as a function of available soil moisture are shown in Figure 2 [David and Hiler, 1970]. For a crop the total evapotranspiration is also affected by the amount of bare soil and crop cover. For well developed cover and root system the plant population is not a significant factor because generally soil water is limiting.

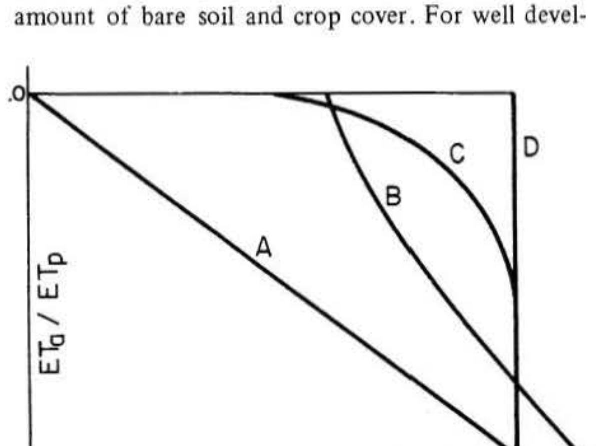


Fig. 2 Soil moisture according to: A) Thornwaite; B) Holmes; C) Pierce; D) Veihmeyer-van Bavel.

The potential evapotranspiration depends on the incident solar energy, relative humidity of the air, wind velocity and turbulence and advection from adjacent areas. There are many empirical formulae available for estimating potential evapotranspiration from solar and climatic data.

The effect of large scale transport of atmospheric moisture into or out of a region is an important factor in the cause of droughts. Without this transport a region could be considered as a closed system in which the soil water transpired into the atmosphere eventually returns to the soil as rainfall excluding the effects of surface streams - and the system would remain stable. However, the advection of moisture laden air from oceanic regions and the accompanying atmospheric circulation is a major factor and fluctuations in this moisture transport due perhaps to shifts in ocean currents may cause a major drought.

2.7 The Yield-Water Stress Relationship

From the previous discussion on plant physiology it is evident that closure of the stomata by water stress reduces photosynthetic activity thus inhibiting growth. This phenomenon has been empirically verified by Denmead and Shaw [1962] who showed that plants ceased to gain dry matter weight when the soil moisture was less than the turgor loss point (a measure of the wilting point). Moore [1961] used a similar concept by proposing an index of potential growth which was a function of percentage moisture depletion. In general then, any internal water stress in the plant causes reduction of growth rate and the total reduction is the product of duration of stress by magnitude of growth loss with respect to the potential growth rate. The concept of potential growth rate is important as there are stages of growth for some crops when the potential green growth is essentially zero and water stress is not harmful. Sometimes water stress can be beneficial in the case of crops grown for seed when the reduction in vegetative growth does not affect the production of the desired harvest. Therefore, a production function derived for one aspect of crop production may not apply to a different product from the same plant. A comprehensive treatment of these factors is given by Richards and Wadleigh [1952].

Chapter 3

PRESENT MODEL THEORIES

There is a large amount of literature dealing with production functions for crops. These are mostly oriented towards irrigated farming and its optimization. In this chapter only those models dealing with the timing of applied water during the growing season are reviewed together with theories directly concerned with agricultural drought. The survey is not intended to be exhaustive but rather to show the range of models and theories in this field.

3.1 Models Directly Related to Droughts Without Regard to Time of Occurrence

Barger and Thom [1949] considered rainfall in n-week totals during a 16-week growing season for corn and correlated yield with the n-week values ($n = 1, 2, ..., 16$) for many years in order to determine a drought base for each n-week total for each station. The yields were then correlated with deficits from the drought base.

Van Bavel [1953] used estimated evapotranspiration (by the Thornwaite method) as the output from a predetermined soil moisture storage. Drought was defined as occurring whenever the storage was empty. The magnitude of the storage was determined from an assumed moisture stress level.

Stewart and Hagan [1969] dealt with water usage over the whole growing season without regard for the timing of water application. They considered water use to be equivalent to the estimated evapotranspiration.

3.2 Models Concerned with Time of Occurrence of Water Deficit

The following theories are concerned with the timing of water application to crops but do not include those models in which feedback is required from soil moisture readings taken during the growing season. Although not all specifically drought oriented the models are of interest because it is now realized that the timing of water deficit is of prime importance in agricultural drought.

Parks and Knetsch [1959] used Van Bavel's [1953] drought day criterion to estimate the drought level during certain periods in the growing season for corn. The corn yield was then correlated with these drought levels with an equation of the form

$$Y = b_o + b_1 A + b_2 B + b_3 C + b_4 D + b_5 C^2 + b_6 CB + b_7 CD + b_8 BD, \qquad (3.1)$$

where A, B, C, and D denote the number of drought days occurring in successive periods of 33, 32, 14, and 31 days, respectively during the growing season. The 14 day C term represents the tasseling period. The Van Bavel's drought day is calculated by subtracting the estimated evapotranspiration (by Thornwaite method) from a known soil moisture until it is depleted and the drought begins. Rainfall refills the storage, and the total storage capacity is determined from an assumed moisture stress level. The authors also correlate yield with a quadratic function of drought index values (determined from the above equation) and nitrogen fertilizer treatment.

Moore [1961] considers the growing season as a number of irrigation periods. During each period the relative growth (compared to potential growth) Gr_i for the ith period is given by

$$Gr_i = \frac{\int_o^{\theta_i} g(x)dx}{\theta_i \times 100} = I_{\theta_i} , \qquad (3.2)$$

where θ_i is the soil moisture stress of the end of the ith irrigation period and $g(x)$ is the functional relationship of relative growth to percent moisture depletion.

The total relative growth for the entire season is

$$G_r = \sum_{i=1}^{n} I_{\theta_i} \frac{t_i}{T} , \qquad (3.3)$$

where T is the length of season and t_i the length of the ith irrigation cycle.

The function $g(x)$ is assumed to be the same at any stage in the crop's growth in the absence of additional information. A critical assumption is that there is no interdependence of relative growth, i.e., stress at any one stage does not affect the response at a later time.

Doll [1967] was concerned with functions which varied with the year t and weather variables $x_t(s)$ at time s within year t, so that both trend and weather effects on a crop were explained. For corn yield data in Missouri for the years 1930-1963 he obtained a function of the form

$$Y_t = b_o + b_1 z_t + b_2 z_t^2 + b_3 t + b_4 t^2 + b_5 t^3 , \quad (3.4)$$

where Y_t is yield in year t ($1930 \rightarrow t = 1$) and $z_t = a_1 x_t(1) + a_2 x_t(2) + \ldots a_8 x_t(8)$ and $x_t(i)$ = rainfall variable for week i in season beginning June 7.

Flinn and Musgrave [1967] also use an additive model similar to that of Moore (above), i.e., the growth at each stage is independent of moisture stress at other stages and the total growth for the whole season is the sum of the relative growth at each stage. The authors recognize that relative growth is different at various stages of growth and base their relative growth calculations on the assumption that growth occurs on those days when actual evapotranspiration equals potential evapotranspiration, i.e., no stress. A dynamic programming solution is also presented for the optimal allocation of irrigation water as the assumed production function is suitable for this technique.

Hall and Butcher [1968] also proposed a dynamic programming algorithm for their model

$$y = a_1(w_1)a_2(w_2) \ldots . a_n(w_n)\, y_{max} \quad (3.5)$$

in which y_{max} is the maximum possible yield when there is no moisture stress during any of the n time periods during the growing season, w_i is the soil moisture during period i, and if all other periods have soil moisture at field capacity w_f then the resulting yield y is given by

$$y = a_i(w_i)\, y_{max} \quad (3.6)$$

where the growth coefficient $a_i(w_i)$ for period i depends on the soil moisture for period i. Thus in general, the proportion of maximum yield obtained is given by the product of the coefficients $a_i(w_i)$. The production function was put into a form suitable for dynamic programming by taking the logarithm of the function. In this model also there is no interdependence of moisture stress effects at different times in the growing season but the effects are multiplicative rather than additive. The authors did not present any empirical proof of the model.

Jensen [1968] suggests two model types, one for what he calls determinate crops, i.e., crops with a definite flowering or heading stage and another for indeterminate crops such as grasses. The model for a determinate crop when soil moisture is limiting is

$$\frac{y}{y_o} = \prod_{i=1}^{n} \left(\frac{w_{et}}{w_{oc}}\right)_i^{\lambda_i} , \quad (3.7)$$

where y/y_o is the relative yield of the crop (y_o represents the yield when soil moisture is not a limiting factor); $(w_{et}/w_{oc})_i$ is the ratio of actual water use to potential water use during period i of the growing season, and λ_i is a parameter to be determined. In the case of indeterminate crops which tend to recover from periods of water stress the suggested relationship is

$$\frac{y}{y_o} = \frac{\sum_{i=1}^{n} \lambda_i (w_{et})_i}{\sum_{i=1}^{n} \lambda_i (w_{oc})_i} , \quad (3.8)$$

which implies that the effects of stress during one period are additive rather than multiplicative in contrast to the determinate crop model which applies to crops which suffer permanent damage due to moisture stress.

Brown and Vanderlip [1969] correlated production of winter wheat at several locations in Kansas with monthly moisture departure values and also with monthly values of the Palmer Drought Index [Palmer, 1964]. This index is based on a soil moisture accounting scheme which introduces some antecedent weather effects. The production functions used are of the form

$$y = b_o + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + b_6 x_2^2 + b_7 x_4^2 + b_8 x_5^2 , \quad (3.9)$$

where y is yield and the x_i are monthly climatic values. Note that although the quadratic terms are included there are no interaction cross-products so again the effects are considered independent and additive.

Yaron [1971] in his review paper of crop production functions suggests a generalized optimal path type procedure for optimizing water usage and other relevant plant factors. The approach is largely conceptual and no empirical evidence is offered in its support.

In conclusion it can be said that the principle differences between models and thus the factors to be considered are:

(1) The length and number of periods during the growing season.

(2) The choice of state variable for each period.

(3) The functional relationship between yield and the state variables at each period.

In the next chapter a general model structure will be considered and some particular examples given.

Chapter 4

MODEL DEVELOPMENT

4.1 The Systems Approach

The response of crop production to agricultural drought can be considered as a system with climatic variables as input and harvestable yields as output. The input variable set could include rainfall and the climatic variables which control evapotranspiration. One way of analyzing a system is to construct a model on the basis of knowledge of the physical behavior of the system and then evaluate the model parameters by using input-output data and a suitable optimization technique. The usual objective function is the minimization of the error mean square of the predicted output. A second approach is to treat the system as a "black box" and derive a mathematical response function from an analysis of only the input and output data. The latter method has been selected in the present research although a certain amount of physical knowledge has been used to assist the analytical procedures.

4.2 A Discussion of Errors

The use of a system model in the simulation of such a complex physical phenomenon as the growth of a crop is at best a workable approximation. It is useful, therefore, to consider the closeness of this approximation or conversely the types and causes of errors. The sources of these errors can be considered in two main groups:

(1) Errors in the data; these may be further divided into:

(a) Measurement errors; this type of error is relatively unimportant in the present state of the art because the accuracy of instrumentation tends to exceed the resolution of the analytical procedures.

(b) Missing data; this could be, for example, monthly rainfall data instead of the desired weekly amounts or complete lack of information on some of the variables related to evapotranspiration.

(2) Model errors; again, these can be grouped as:

(a) Incorrect selection of input variables. The effect is the same as error 1(b) above in which the variables are chosen correctly but no data is available.

(b) Inadequate simulation of the physical behavior of the process because the mathematical formulation is not sufficiently flexible. Sometimes the choice of mathematical functions is restricted by the lack of sufficient data to enable significant estimates to be made of a large number of parameters.

(c) Non-stationary behavior of the process e.g., an annual crop may develop different root structure in different years and thus have a varying reaction to the same degree of water stress. For some crops the root development depends on the moisture stress in the early stages of growth. Sometimes apparent non-stationarity is really due to errors of type 1(b) as for example the behavior of crops growing in different soils. If the soil properties cannot be quantified and used as an additional input variable then the error may well be dismissed as being due to non-stationarity. Finally it should be noted that the varying behavior of a particular crop throughout the growing season is not considered to be non-stationary because the climatic inputs a each period of time are considered to be separate input variables thus removing the time dependence.

4.3 Development of the System Function

Consider the growing season of a crop divided into n time periods, not necessarily of equal length and let x_i be the climatic input variable in the ith period. The variable x_i could be rainfall, irrigation water, drought index or a function of several other variables related to precipitation and evapotranspiration for example. The yield y of the crop at harvest time can be expressed generally as

$$y = f(x_1, x_2, \ldots, x_n) \ . \qquad (4.1)$$

The function y may have different forms for yield expressed as green matter, dry weight, fruit, etc. Provided the function is continuous and all derivatives exist over the range considered it may be expressed as a multidimensional polynomial

$$y = \sum_{i=1}^{n} a_i x_i + \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} a_{ijk} x_i x_j x_k + \dots \quad . \qquad (4.2)$$

In order to reduce the number of terms to be considered in subsequent analyses an economy may be affected by factorizing the polynomial

$$\begin{aligned} y = & \sum_{i=1}^{n} (b_1^{(i)} x_i + b_2^{(i)} x_i^2 + b_3^{(i)} x_i^3 + \dots) \\ & + \sum_{i=1}^{n} \sum_{j=1}^{n} (c_1^{(i)} x_i + c_2^{(i)} x_i^2 + \dots) \\ & \qquad (c_1^{(j)} x_j + c_2^{(j)} x_j^2 + \dots) \\ & + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=j}^{n} (d_1^{(i)} x_i + d_2^{(i)} x_i^2 + \dots) \\ & \qquad (d_1^{(j)} x_j + d_2^{(j)} x_j^2 + \dots) \\ & (d_1^{(k)} x_k + d_2^{(k)} x_k^2 + \dots) + \dots , \qquad (4.3) \end{aligned}$$

and then expressing each polynomial factor as a single function

$$\begin{aligned} y = & f_1(x_1) + f_2(x_2) + \dots f_n(x_n) \\ & + f_{12}(x_1) f_{21}(x_2) + f_{13}(x_1) f_{31}(x_3) + \dots \\ & + f_{123}(x_1) f_{213}(x_2) f_{312}(x_3) + \dots \quad . \qquad (4.4) \end{aligned}$$

The main purposes for expressing the function in this form is that all interactions between periods are considered and yet the $f_{ij}...(x_i)$ can be expressed as single parameter functions thus economizing on parameter estimation or optimization. The choice of function is dependent on the climatic variable x_i used and the physical behavior of the system. In the empirical work carried out for this report the x_i were monthly rainfall values and the functions f_i were of the form, x_i^{α} where $0 \leqslant \alpha \leqslant 1$ so that a "diminishing returns" type of behavior is obtained. However, other functional forms may better suit other variables and models.

4.4 Simplification of the General System Function

Even with the above reduction to the cross-products of single parameter functions there will generally be too many terms for analysis. The main reasons being lack of sufficient data to give significant estimates and the physical limitations of computer facilities. A further reduction in the number of terms to be considered can often be done by considering the boundary conditions and critical properties of the system. For the drought-crop production model the questions to be considered are:

(1) What are the lengths of the time periods for each x_i and does a zero x_i value (e.g., no rain for two weeks) during that period destroy the crop? If so, this period is said to be critical with respect to the input x_i.

(2) Which of the n periods are critical and which are non-critical (i.e., zero x_i does not destroy the crop)?

(3) Is there interaction between periods, i.e., does the crop response to stress at period i depend on what occurred during period j?

(4) Does the model satisfy the condition that no input during the growing season equals zero output?

It should be emphasized at this point that a model developed in terms of inputs other than rainfall or applied water may have a different set of critical and boundary conditions.

As an example of the application of the above boundary conditions, consider the model proposed by Hall and Butcher [1968]

$$y = a_1(w_1)\, a_2(w_2)\, a_3(w_3) \dots a_n(w_n)\, y_{max} \, , \qquad (4.5)$$

where w_i is applied water during period i and $a_i(w_i)$ is the growth coefficient.

The condition of no inputs giving zero output is easily satisfied if any one of the $a_i(w_i)$ has a zero intercept, i.e., any $a_i(0) = 0$.

Any of the i periods which are critical must satisfy $a_i(0)_c = 0$. However, any period not critical must satisfy $a_i(0)_{nc} = k_i$ so that it represents the fact that the crop is not destroyed by zero input in those non-critical periods. Now consider the case when period i is non- critical as is period i+1 but the period of time equivalent to i plus (i+1) is critical. Therefore $a_i(0) = k_i$ $a_{i+1}(0) = k_{i+1}$ and yet zero inputs during periods i and i+1 do not destroy the crop. Obviously there is some interaction between periods and the model does not take this into account. The only situation in which this model will satisfy the boundary conditions is when the periods are chosen to be of such lengths that all periods are critical. This, of course, requires some prior knowledge of the plant physiology.

A superficial knowledge of which periods are likely to be critical during the crop's growing season can greatly assist in reducing the orders of function cross-products to be considered in the analysis. For example, consider a growing season consisting of four monthly periods say, May, June, July and August. These months receive rainfall inputs x_1, x_2, x_3 and x_4, respectively, and it is known that only July is critical, i.e., no rain in July will destroy the crop. Now

$$y = \phi\left\{f_1(x_1), f_2(x_2), f_3(x_3), f_4(x_4)\right\} \quad , \qquad (4.6)$$

where ϕ denotes some order of products of the functions within the brackets.

Since period 3 is critical, $f_3(0) = 0$. Consider the product

$$y = f_1(x_1)\,f_2(x_2)\,f_3(x_3)\,f_4(x_4) \quad , \qquad (4.7)$$

which form does not seem feasible because of the reasons given above. If no rain during May and June would destroy the crop, then the model breaks down as $f_1(0) = k_1$ (non-critical) and $f_2(0) = k_2$ but $f_1(0)\,f_2(0) \neq 0$.

The next order of products to be considered is the third order set

$$f_1(x_1)\,f_2(x_2)\,f_3(x_3) \ , \ f_4(x_2)\,f_5(x_3)\,f_6(x_4) \ ,$$
$$f_7(x_1)\,f_8(x_2)\,f_9(x_4) \ , \ f_{10}(x_1)\,f_{11}(x_3)\,f_{12}(x_4) \ . \qquad (4.8)$$

Note that this is not simply products of the original four $f_i(x_i)$ taken in sets of three but each product contains different functions of the x_i. The purpose of this form is to give the maximum generality to the production function and especially to represent any interaction between periods. If there was no interaction effects then

$$f_1(x_1) = f_7(x_1) = f_{10}(x_1) \quad , \qquad (4.9)$$

but in general the possibility of interaction must be considered.

In order to satisfy the critical nature of x_3 only those products containing a function of x_3 should be considered, i.e.,

$$y = b_1 f_1(x_1)\,f_2(x_2)\,f_3(x_3) + b_2 f_4(x_2)\,f_5(x_3)\,f_6(x_4)$$
$$+\, b_3 f_6(x_1)\,f_{11}(x_3)\,f_{12}(x_4) \ . \qquad (4.10)$$

If no period was critical then all four products would have to be included. In this form all $f_i(0) = 0$ will still satisfy the non-critical behavior as any non-critical period appears in only two of the three terms and hence zero inputs in any one does not give zero output. However, zero inputs in any two periods will give zero y value. These need not be adjacent periods in the above model as $x_1 = x_4 = 0$ will give $y = 0$ for all $f_i(0) = 0$. If $x_1 = x_4 = 0$ gives $y \neq 0$ then the model would have to be modified by considering other product forms such as second order terms combined with the third order ones above. It is often easier to work from high order products to low order in developing the model because most low order terms do not satisfy the boundary conditions. There is the greatest number of terms to be considered in the middle range of orders of products and, in general, these terms should be added to the model only if there is some prior reason for doing so as the model can easily become unwieldy.

4.5 Some Physical Considerations

It is useful at this point to consider the physical phenomena underlying crop production and their relationship to the aforementioned mathematical functions. In Chapter 2 it was deduced that the water soil plant system is basically a storage with rainfall input and evapotranspiration output. Whenever the storage runs low or becomes empty the plant growth is reduced or it may die. Some of the models reviewed in Chapter 3 use accounting procedures, essentially, to determine when stress occurs. This behavior can be included in the systems analysis by means of a function defined as

$$\dim(x,a) = \begin{cases} x-a & x>a \\ 0 & x \leq a \end{cases} \quad . \qquad (4.11)$$

The use of this function is apparent in the response from the ith period in the growing season which becomes:

$$f_i \left\{ \dim(x_1 + x_2 + \ldots + x_i, a_i) \right\} . \qquad (4.12)$$

Note the implicit accounting procedure which tests whether the storage is likely to be empty at period i. The parameter a_i can be estimated in the analysis. This storage concept may explain some of the interaction between growth periods as the "carry over" of soil moisture to give a higher input to the succeeding period. There is still some "true" growth interaction apart from this factor due to water stress damage affecting later growth. Both types of interaction are included in the cross product terms of the general system functions.

There has been some argument in the literature as to whether the production function should be additive or multiplicative. This seems to be largely a question of definition of the yield response for each time period. For the additive model, $f_i(x_i)$ refers to the actual growth increment whereas in the multiplicative model $f_i(x_i)$ is the percentage of the maximum final yield if only x_i is limiting as defined by Hall and Butcher [1968]. Neither model includes any interdependence of response between different periods of the growth cycle.

4.6 Solution of the Production Function and Data Requirements

The principal advantage of the method of analysis, as described in the following text, lies in the efficient use of data. Some models such as that of Hall and Butcher [1968] require controlled inputs to crops in order to determine the production function. The method proposed by Hall and Butcher is to keep inputs to all periods at maximum except for the period under observation if no interaction is presumed to occur. If interactions occur the number of controlled experiments required becomes enormous.

For the least squares estimation to be discussed the inputs can be randomly distributed as would occur from rainfall on a crop and the number of input sets (i.e., number of crop yields) considered need be only great enough to estimate the production function parameters. This point is discussed in detail later.

The remainder of this chapter is concerned with the development of the mathematical theory underlying the algorithm for solving the system function. Considerable use is made of matrix notation in the presentation of the theory and the non-mathematically minded reader may proceed to Chapter 5 without loss of continuity.

Consider the situation in which there are m observations on each of the input variables $x_1, \ldots, x_n$ and the output yield y. Suppose that from the physical considerations outlined previously y can be expressed as a function of the n input variables

$$y = f(x_1, x_2, \ldots, x_n; \theta_1, \theta_2, \theta_3, \ldots, \theta_p) , \qquad (4.13)$$

with the parameters $\theta_1, \theta_2, \ldots, \theta_p$ to be estimated, so that

$$y = \theta_1 x_1^{\theta_2} x_2^{\theta_3} x_3^{\theta_4} + \theta_5 x_1^{\theta_6} x_3^{\theta_7} x_4^{\theta_8} + \ldots \qquad (4.14)$$

The method of obtaining the solution is the non-linear least squares regression procedure given by Jennrich and Sampson [1968]. This is the Gauss-Newton method with modifications due to Hartley [1961] and Jennrich and Sampson in order to ensure convergence and prevent instability. The technique consists essentially of approximating the above production function with a Taylor series expansion in terms of the θ_i and solving the resulting equations by the least squares methods of multiple linear regression. Let $\underset{\sim}{x}$ denote the vector $(x_1, \ldots, x_n)$ and $\underset{\sim}{\theta}$ the vector $(\theta_1, \theta_2, \ldots, \theta_p)$.

For some initial values $\underset{\sim}{\theta}^o$ of the parameters the function $f(\underset{\sim}{x}; \underset{\sim}{\theta}^o)$ can be approximated by the first order terms of the Taylor series

$$f(\underset{\sim}{x}; \underset{\sim}{\theta}^o + \Delta\underset{\sim}{\theta}) \approx f(\underset{\sim}{x}; \underset{\sim}{\theta}^o) + \Delta\theta_1 \frac{\partial}{\partial\theta_1} (f(\underset{\sim}{x}; \underset{\sim}{\theta}^o) +$$

$$\Delta\theta_2 \frac{\partial}{\partial\theta_2} f(\underset{\sim}{x}; \underset{\sim}{\theta}^o) + \Delta\theta p \frac{\partial}{\partial\theta p} f(\underset{\sim}{x}; \underset{\sim}{\theta}^o) \qquad (4.15)$$

and

$$y - f(\underset{\sim}{x};\theta^o) = \Delta\theta_1 \frac{\partial}{\partial\theta_1} f(\underset{\sim}{x};\theta^o) + \ldots \Delta\theta_p \frac{\partial}{\partial\theta_p} f(\underset{\sim}{x};\theta^o) + e \quad , \quad (4.16)$$

where e is the error term. Let y_t, $f_t(\underset{\sim}{x};\underset{\sim}{\theta}^o)$, where $t = 1,2,\ldots,m$, denote the observations on y and $f(\underset{\sim}{x};\underset{\sim}{\theta}^o)$. Then the observations on the above equation can be written in matrix form

$$\begin{vmatrix} \frac{\partial}{\partial\theta_1} f_1(\underset{\sim}{x};\theta^o) & \ldots & \frac{\partial}{\partial\theta_p} f_1(\underset{\sim}{x};\theta^o) \\ \frac{\partial}{\partial\theta_1} f_2(\underset{\sim}{x};\theta^o) & \ldots & \frac{\partial}{\partial\theta_p} f_2(\underset{\sim}{x};\theta^o) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \frac{\partial}{\partial\theta_1} f_m(\underset{\sim}{x};\theta^o). & \ldots & \frac{\partial}{\partial\theta_p} f_m(\underset{\sim}{x};\theta^o) \end{vmatrix} X \begin{vmatrix} \Delta\theta_1 \\ \Delta\theta_2 \\ \cdot \\ \cdot \\ \cdot \\ \Delta\theta_p \end{vmatrix} = \begin{vmatrix} y_1 - f_1(\underset{\sim}{x};\theta^o) \\ y_2 - f_2(\underset{\sim}{x};\theta^o) \\ \cdot \\ \cdot \\ \cdot \\ y_t - f_m(\underset{\sim}{x};\theta^o) \end{vmatrix} + \begin{vmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_m \end{vmatrix} \quad (4.17)$$

or representing the above matrices in the condensed form

$$|f'(\underset{\sim}{x};\theta^o)| \, |\Delta\theta| = |y - f(\underset{\sim}{x};\theta^o)| + |e| \quad . \quad (4.18)$$

This formulation is analogous to the multiple linear regression of a variable y on a set of variables $x_1,\ldots, x_n$, which can be represented as

$$|X| \, |B| = |Y| + |e| \quad , \quad (4.19)$$

where $|X|$ is a m x n matrix of the m observations on the x_i, $|Y|$ is a (m x 1) matrix of observations on y , $|e|$ is the error term matrix, and $|B|$ is the matrix of regression coefficients in the equation

$$b_1x_1 + b_2x_2 + \ldots + b_nx_n = y \quad . \quad (4.20)$$

The least squares estimation of the b_i is obtained by the procedure.

$$|X|^T |X| \, |\hat{B}| = |X|^T |Y| \quad . \quad (4.21)$$

Thus

$$|\hat{B}| = (|X|^T |X|)^{-1} |X|^T |Y| \quad , \quad (4.22)$$

where $|X|^T$ denotes the transpose of $|X|$, and $|\hat{B}|$ is the matrix of estimates of the regression coefficients.

Hence the $\Delta\theta_i$ in the nonlinear regression are given by

$$|\Delta\theta| = (|f'(\underset{\sim}{x};\theta^o)|^T | f'(\underset{\sim}{x};\theta^o)|)^{-1} \; |f'(\underset{\sim}{x};\theta^o)|^T | y - f(\underset{\sim}{x};\theta^o)| \quad . \quad (4.23)$$

New values of $\theta_i = \theta_i^o + \Delta\theta_i$ are calculated and the procedure is repeated until it converges to final values of θ_i. The criterion for convergence is the residual sums of squares

$$\sum_{t=1}^{m} (y_t - f_t(\underset{\sim}{x};\underset{\sim}{\theta}))^2$$

or the error mean square

$$\sum_{t=1}^{m} [y_t - f_t(\underset{\sim}{x};\underset{\sim}{\theta})]^2/(m-p) \; .$$

The procedure is continued until the relative increase in error mean square is less than a chosen value, i.e.,

$$\frac{[|S_{n+1}{}^2 S_n^2|]}{S_{n+1}{}^2} < C$$

$$\text{where } s_n^2 = \sum_{t=1}^{m} [y_t - f_t(x,\theta)]^2 \quad . \qquad (4.24)$$

4.7 Modifications to Assist Convergence

Sometimes the error mean square does not decrease when $\underset{\sim}{\theta}$ is incremented by $\Delta\underset{\sim}{\theta}$ because the linear approximation used for $f(\underset{\sim}{x};\theta)$ fails at that point. The usual modification is to increment $\underset{\sim}{\theta}$ by $\epsilon\Delta\underset{\sim}{\theta}$ where $0<\epsilon<1$ and re-evaluate the error mean square. A practical value to use is $\epsilon = 1/2$ and if the error mean square still increases use $\epsilon = 1/4$ etc., halving the step each time until convergence occurs.

4.8 Singularity

As $\underset{\sim}{\theta}$ changes its values during the iteration procedure it can happen that the partial derivatives $\partial f(\underset{\sim}{x};\theta)/\partial\theta_i$ become linearly dependent or nearly so and thus the matrix $|f'(\underset{\sim}{x};\theta)|$ becomes singular or almost singular. The regression procedure cannot then continue because the inverse cannot be evaluated. This problem of singularity can be handled by carrying out the regression procedure in a stepwise manner. The stepwise regression also provides a method of introducing parameter constraints into the analysis.

4.9 Stepwise Regression

The algorithm given here is essentially that due to Efroymson [1960]. In order to invert a n x n matrix $|A|$ the Gauss-Jordan pivotal method is used so that if a_{ij} are the elements of $|A|$ before pivoting then after pivoting on a non-zero diagonal element a_{kk} the new matrix has elements

$$a'_{ij} = \begin{cases} a_{ij} - a_{ik}a_{kj}/a_{kk} & i \neq k \quad j \neq k \\ -a_{ik}/a_{kk} & i \neq k \quad j = k \\ a_{kj}/a_{kk} & i = k \quad j \neq k \\ 1/a_{kk} & i = k \quad j = k \end{cases} \qquad (4.25)$$

After all diagonal elements have been used as pivots the matrix $|A|$ is completely inverted. If any of the diagonal elements are zero or become zero the inversion cannot be completed. By partioning $|A|$ into $|A_{11}|$ the submatrix containing the diagonal elements which are to be used as pivots and the other three submatrices $|A_{12}|$, $|A_{21}|$, $|A_{22}|$, so that

$$|A| = \begin{vmatrix} |A_{11}| & |A_{12}| \\ |A_{21}| & |A_{22}| \end{vmatrix} , \qquad (4.26)$$

then pivoting on the chosen elements yields the matrix

$$\begin{vmatrix} |A_{11}{}^{-1}| & |A_{11}{}^{-1} A_{12}| \\ |-A_{21}A_{11}{}^{-1}| & |A_{22}-A_{21}A_{11}{}^{-1}A_{12}| \end{vmatrix} \qquad (4.27)$$

This result is important in the application of stepwise inversion to multiple linear regression used in the following section.

Now let $|Z|$ be a (m x n) matrix of observations on a set of n variables expressed as deviations from their mean values, i.e., $z_{ij} - \bar{z}_i$, $i = 1, \ldots, n$, $j = 1, \ldots, m$. One of these variables z_n is the dependent variable in a regression equation on a subset of the other n - 1 variables which are to be tested in a stepwise manner as possible independent variables. The independent variables are added to the regression equation one by one, so that at any stage the matrix $|Z|$ may be considered as two submatrices

$$|Z| = ||X| \ |Y|| \quad , \qquad (4.28)$$

where $|X|$ is the matrix of variables already in the regression equation and $|Y|$ is the matrix of variables not in the regression, including the chosen dependent variable. The covariance matrix of all the variables is given by

$$||X| \ |Y||^T ||X| \ |Y|| = \begin{vmatrix} |X|^T|X| & |X|^T|Y| \\ |Y|^T|X| & |Y|^T|Y| \end{vmatrix} . \qquad (4.29)$$

The stepwise inversion algorithm may be applied to the covariance matrix by pivoting only on the diagonal elements of the submatrix $|X|^T|X|$, i.e., the covariance matrix of the independent variables already in the regression. This partial inversion of the total covariance matrix gives the result according to the stepwise algorithm of Equation 4.25 to give result of Equation 4.27,

$$\begin{vmatrix} (|X|^T|X|)^{-1} & (|X|^T|X|)^{-1}\ X^TY \\ -|Y|^T|X|\,(|X|^T|X|)^{-1} & |Y|^T|Y| - |Y^T||X|(|X|^T|X|)^{-1}|X|^T|Y| \end{vmatrix} \tag{4.30}$$

and using the standard results for the multiple linear regression of Equation 4.22 this matrix can be written as

$$\begin{vmatrix} (|X|^T|X|^{-1} & |B| \\ -|B| & |S| \end{vmatrix} \tag{4.31}$$

where |B| is the matrix of regression coefficients of each of the "y" variables on the variables already in the regression and includes the regression of the chosen dependent variable on the "x" variables. The matrix |S| is the residual covariance matrix and the diagonal elements contain the sums of squares of the residuals of the "y" variables estimated from "x" variables. The essence of stepwise regression lies in moving variables from the "y" set to the "x" set and vice versa in order to minimize the residual sum of squares in the chosen dependent variables y_d. The variable to be next entered into the regression is the y_i with the largest value of b_{id}/s_{ii}, where b_{ij} and s_{ij} are elements of |B| and |S| respectively. In order to avoid singularity in the matrix inversion procedure the magnitude of the diagonal element to be used as the next pivot is checked by comparing it with an arbitrary tolerance value, e.g., 0.001. If the element is smaller than the tolerance value then the corresponding variable is not entered into the regression. Variables may also be removed from the regression if they lose significance. The F-test is commonly used as a significance test for entering or removing variables. The stepwise procedure continues until there is no further reduction in the error variance of the dependent variable or the tolerance of diagonal pivot elements becomes too small.

4.10 Application to Non-linear Least Squares

In the case of the non-linear least squares the independent variables are replaced by the partial derivatives of the regression equation with respect to the parameters of Equation 4.17,

$$|X| = |f'(\underset{\sim}{x};\underset{\sim}{\theta}^o)| \quad , \tag{4.32}$$

and the residual error becomes the dependent variable

$$|Y| = |y - f(\underset{\sim}{x};\underset{\sim}{\theta})| \quad . \tag{4.33}$$

Thus the initial data matrix becomes

$$||X|\ |Y|| = ||f'(\underset{\sim}{x};\underset{\sim}{\theta}^o)|\ \ |y - f(\underset{\sim}{x};\underset{\sim}{\theta}^o)| \quad , \tag{4.34}$$

and the "covariance" matrix is formed in the same manner as that in Equation 4.29. Application of the inversion procedure and recognizing that $|\Delta\theta| \equiv |B|$ gives the result according to Equation 4.31 as

$$\begin{vmatrix} (|f'(\underset{\sim}{x};\underset{\sim}{\theta}^o)|^T|f'(\underset{\sim}{x};\underset{\sim}{\theta}^o)|)^{-1} & |\Delta\theta^o| \\ -|\Delta\theta^o| & |(y - f(\underset{\sim}{x};\underset{\sim}{\theta}^o))^2| \end{vmatrix} \tag{4.35}$$

The use of a stepwise procedure becomes important in that if any of the diagonal elements of

$$|f'(\underset{\sim}{x};\underset{\sim}{\theta}^o)|^T|f'(\underset{\sim}{x};\underset{\sim}{\theta}^o)|$$

have a low tolerance value (and thus likely to cause singularity), they are not used as pivots and the corresponding $\Delta\theta_i$'s are zero. Otherwise, $\Delta\theta_i$ is given in the $|\Delta\theta|$ matrix. This is also a means of handling parameter constraint. When a particular θ_i has reached its boundary value (this is in the input to the procedure) the corresponding diagonal element is no longer used as a pivot. When a set of $\Delta\theta$ have been found from the assumed initial values $\underset{\sim}{\tilde{\theta}}^o$ the new values for the next iteration are:

$$\underset{\sim}{\theta}' = \underset{\sim}{\theta}^o + \Delta\underset{\sim}{\theta} \quad , \tag{4.36}$$

and the procedure continues. If

$$\theta_i^{j+1} = \theta_i^j + \Delta\theta_i$$

exceeds its boundary value the matrix is unpivoted on the corresponding element and an $\alpha\Delta\theta_i$ found for $\alpha \leqslant 1$ so that the boundary condition is satisfied. The above computational procedures are those employed in the empirical section of this report by using the Biomedical Computer Program BMDX85 (1970).

4.11 Three Additional Notes on the Procedure

(1) Throughout the foregoing analysis the usual properties of least squares regression have been assumed. That is, the error term is considered to be symmetrically distributed with constant variance and to be independent. Sometimes, the constant variance assumption may not be satisfied, e.g., large y_i having larger errors or some observations may be unreliable. This situation can be handled by introducing a weighting variable w_i with each observation.

The value of w_i is proportional to the reliability of the ith observation. Usually the weights are normalized so that

$$\sum_{i=1}^{m} w_i = 1 \qquad (4.37) \qquad (4.37)$$

where m is the number of observations. The set of observation equations (4.19) become

$$|w|\ |X|\ |B| = |w||Y| + |w||e| \qquad (4.38)$$

so that the error $w_i e_i$ has constant variance for all observations. The least squares estimate of the regression coefficients is

$$|B| = (|X|^T |w|\ |X|)^{-1}\ |X|^T |w|\ |Y| \qquad (4.39)$$

The values to be chosen for the weights w_i are usually estimated from physical considerations such as instrumentation errors or if it is known that the error is proportional to the magnitude of the dependent variable y then the weights are scaled accordingly.

(2) An interesting property of the non-linear least squares procedure becomes evident when analyzing functions of the type

$$y = \theta_1 x_1^{\theta_2} x_2^{\theta_3} \qquad (4.40)$$

such as occurs in the Hall and Butcher [1968] model. The non-linear least squares procedure gives the least squares estimates of the parameters $\theta_1, \theta_2, \theta_3$, whereas the commonly applied technique of taking logarithms

$$\log y = \log \theta_1 + \theta_2 \log x_1 + \theta_3 \log x_2\ , \qquad (4.41)$$

and employing linear least squares regression leads to biased estimates. The reason for the bias is that the sums of squares of the logarithms of the errors are minimized which is not usually the correct error model. Edwards [1962] gives examples of this effect using the Cobb-Douglas multiplicative production function.

(3) With certain types of non-linear functions and data sets the non-linear regression procedure may not converge if the assumed initial set of parameter values $\underset{\sim}{\theta}^{o}$ is poorly chosen. There are no easily usable rules to determine when such a situation is likely to occur so the possibility should be recognized in any analysis. Often the initial parameter values can be estimated from the physics of the problem. An alternative procedure is to use linear least squares to estimate the linear parameters in a function

$$y = \theta_1 x_1^{\theta_2} + \theta_3 x_2^{\theta_4} \qquad (4.42)$$

the parameters θ_1, θ_3 can be calculated and initial values of $\theta_2 = \theta_4 = 1$ assumed for the non-linear parameters.

Chapter 5

APPLICATION OF THE MODEL

5.1 Description of the Data

Two types of data were used to test the model. Firstly, 210 observations of synthetic data were generated according to a certain non-linear function, in order to test some properties of the estimation techniques. The second data type consisted of corn yield values, corresponding monthly rainfall and soil moisture at planting from 20 experimental plots in Minnesota and South Dakota for the years 1960-1961. The initial soil moisture data was missing from six of the plots for 1961 and therefore only 34 independent observations were available. These data were abstracted from a U.S.D.A. Agricultural Research Service Publication [Holt *et. al.* 1965] and are given in Table 1. There were also data for stand and fertilizer treatment for the plots but inclusion of these additional variables in the model would have required more observations to avoid loss of degrees of freedom. There was no evaporation or evapotranspiration data available for the plots and no attempt was made to estimate this factor and include it in the data input. There were two reasons for this apparent omission. Firstly, the evapotranspiration data would have necessitated further functions and parameters because of the lack of enough information to conduct a soil moisture balance and, therefore, cause a lack of degrees of freedom. Secondly, the model is for the purpose of estimating the frequency of agricultural drought from the statistical properties of the input variables. Although the statistical properties of rainfall can be easily found from the abundant data available this is not true for the joint distribution of rainfall and the climatic factors involved in evapotranspiration. Further study in this area will provide the answers but until they are found and more crop data becomes available the analysis is confined to rainfall input.

5.2 Demonstration of Analytical Capability Using Synthetic Data

In order to demonstrate the capability of the non-linear least squares technique in detecting the correct model among a limited set, the data used were generated without any error term by means of the function

$$\begin{aligned} y = {} & 2.0\,x_1^{0.38} + 3.0\,x_2^{0.38} + 4.0\,x_3^{0.38} \\ & + 5.0\,x_4^{0.38} + 3.0\,x_5^{0.38} + 2.5\,(x_1x_3)^{0.21} \\ & + 3.0\,(x_1x_4)^{0.21} + 4.0\,(x_2x_3)^{0.21} + 4.5\,(x_2x_4)^{0.21} \\ & + 4.0\,(x_3x_4)^{0.21} + 3.0\,(x_3x_5)^{0.21} + 2.0\,(x_4x_5)^{0.21}, \end{aligned} \tag{5.1}$$

where x_1, x_2, x_3, and x_5 were values of monthly rainfall at 5 stations with the statistical properties shown in Table 2. The model proposed for analysis included all the second order cross products whereas three of these were omitted from the generation scheme. Thus the model is

$$\begin{aligned} y = {} & p(1)\,x_1^{p(2)} + p(3)\,x_2^{p(4)} + p(5)\,x_3^{p(6)} \\ & + p(7)\,x_4^{p(8)} + p(8)\,x_5^{p(10)} + p(11)\,(x_1x_2)^{p(12)} \\ & + p(13)\,(x_1x_3)^{p(14)} + p(15)\,(x_1x_4)^{p(16)} \\ & + p(17)\,(x_1x_3)^{p(18)} + p(19)\,(x_2x_3)^{p(20)} \\ & + p(21)\,(x_2x_4)^{p(22)} + p(23)\,(x_2x_5)^{p(24)} \\ & + p(25)\,(x_3x_4)^{p(26)} + p(27)\,(x_3x_5)^{p(28)} \\ & + p(29)\,(x_4x_5)^{p(30)}, \end{aligned} \tag{5.2}$$

with the constraints

$$0 \leqslant p(i) \leqslant 1.00, \quad i = 3, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30,$$

$$p(i) \geqslant 0 \quad , \; i = 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 \;,$$

and initial values

$$p(i) = 1.00 \text{ for all } i \;.$$

The results of the parameter estimation after 50 iterations of the procedure are shown in Table 3.

TABLE 1

Corn Yield Data 1960-1961

Yield, Bu./acre	Initial Soil Moisture	Planting to May 31	Rainfall - inches June	July	August	September 1 - 15
74.0	7.25	0.00	2.64	3.41	0.85	2.00
30.0	7.42	0.00	1.90	1.12	1.17	2.76
96.0	6.81	0.00	3.22	3.65	4.17	2.90
48.0	8.60	0.00	1.14	1.79	2.99	2.42
74.0	8.51	0.65	1.75	2.66	3.34	2.31
66.0	6.53	0.00	2.34	3.30	3.00	2.69
61.0	6.62	0.00	4.70	4.25	3.15	1.15
48.0	8.36	0.35	1.26	2.17	3.55	3.80
24.0	7.28	0.75	0.80	0.87	1.91	2.88
61.0	7.42	0.32	3.06	4.21	0.00	0.70
38.0	7.00	0.33	3.71	3.48	3.69	0.60
24.0	5.12	0.00	2.10	0.40	2.40	1.32
60.0	7.76	0.80	1.78	3.99	2.40	0.69
96.0	6.26	0.47	7.20	1.16	6.86	0.80
84.0	7.23	0.00	3.63	1.17	5.50	1.20
38.0	8.52	0.66	3.05	2.55	6.00	0.90
29.0	3.26	0.52	3.71	1.55	3.25	1.03
67.0	6.46	1.07	5.02	2.18	6.73	0.70
38.0	7.14	0.00	3.99	0.55	6.00	1.60
59.0	9.01	0.00	2.16	0.49	5.73	1.85
50.0	4.61	0.00	4.27	0.39	4.20	2.44
60.0	7.22	0.12	2.98	1.75	6.67	2.65
76.0	7.49	0.66	3.98	1.20	5.96	0.70
64.0	8.51	0.00	4.63	1.36	4.73	1.45
106.0	10.47	0.00	3.01	1.98	4.20	1.13
50.0	7.33	0.54	3.62	1.33	7.82	0.50
64.0	6.04	0.84	6.08	1.74	5.89	0.61
78.0	7.76	3.19	2.32	1.50	5.77	1.30
62.0	10.45	7.35	1.25	0.00	1.50	0.60
102.0	6.82	5.24	2.21	4.31	5.30	3.25
71.0	8.57	6.64	5.09	2.26	2.59	0.84
66.0	7.36	0.00	2.78	3.95	4.31	0.92
44.0	7.37	0.00	2.65	7.09	1.50	1.49
56.0	4.32	0.00	5.14	2.09	3.20	0.82

TABLE 2

Statistical Properties of the Synthetic Data

Variable	x_1	x_2	x_3	x_4	x_5
Mean	1.30	1.56	1.70	1.41	1.42
Standard Deviation	0.33	0.34	0.36	0.42	0.36

Correlation Matrix

Variable	x_1	x_2	x_3	x_4	x_5
x_1	1.000	0.265	0.093	0.080	0.109
x_2		1.000	0.067	0.056	0.255
x_3			1.000	-0.088	0.004
x_4				1.000	0.035
x_5					1.000

TABLE 3

Results of Parameter Estimation for Synthetic Data

Parameter	True Value	Estimated Value
1	2.00	1.98
2	0.38	0.38
3	3.00	3.00
4	0.38	0.38
5	4.00	3.99
6	0.38	0.38
7	5.00	5.00
8	0.38	0.38
9	3.00	3.00
10	0.38	0.38
11	0.00	1.21×10^{-6}
12	0.00	4.44×10^{-16}
13	2.50	2.53
14	0.21	0.21
15	3.00	2.99
16	0.21	0.21
17	0.00	2.33×10^{-3}
18	0.00	7.10×10^{-15}
19	4.00	4.00
20	0.21	0.21
21	4.50	4.50
22	0.21	0.21
23	0.00	1.32×10^{-4}
24	0.00	3.55×10^{-15}
25	4.00	3.99
26	0.21	0.21
27	3.00	2.98
28	0.21	0.21
29	2.00	2.01
30	0.21	0.21

It is interesting to note that the terms actually missing in the true model compared to the proposed model were removed by driving the relevant parameters to insignificantly small values. This type of behavior can be expected only if the functions used as terms in the model vanish at certain parameter values. The large number of iterations used here is not always necessary just to obtain "true" parameter values because the value of the dependent variable is often relatively insensitive to changes in some of the parameters. In this example the variance of y was 73.96 and after 20 iterations the error mean square was 2.60. Since the data are error free this value represents model error of only 3.5 percent unexplained variance. At 50 iterations the error mean square had been reduced to 7.46×10^{-9}. The number of iterations required in any particular problem depends on the number of parameters, the parameter correlation structure, the initial values and the constraints imposed on the parameters.

5.3 Assumptions in the Formulation of Models for Corn-Yield Data

The rainfall data are in monthly totals except for the parts of May and September included in the growing season. The relative coarseness of this time increment means that the critical days or weeks in the growth cycle cannot be identified closely with the respective precipitation inputs and a good model simulation cannot be expected. The data come from plots on four different soil types and this factor has not been taken into account because the reduction in degrees of freedom by introduction of the relevant terms would seriously affect the significance of the results. The evapotranspiration at any one time is considered to be the same for all plots for the two years of data collection although variation during the growing season is implicitly taken into account. Again the relevant terms could be included if more observations are available.

Throughout all the models to be tested the basic building block function is $x_i^{\alpha_i}$, where x_i is the precipitation input during period i except for the soil moisture at planting time. Other functions could be chosen which theoretically may fit the plant behavior better but for the assumptions and data considered x^{α}, has sufficient flexibility. An advantage of x^{α} is the fact that the effect of any particular $x_i^{\alpha_i}$, in product form with other terms can be eliminated by driving α_i to zero in order to remove model error. For example, if the 'true' model is

$$y = x_1^{\alpha_1} x_2^{\alpha_2}$$

then it can be obtained from a proposed model

$$y = x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4}$$

when the analytical procedure puts $\alpha_3 = \alpha_4 = 0$.

For all the models tested the basic variables were

x_1, the soil moisture at planting plus rainfall in May from planting time to the end of the month.
planting time to the end of the month,

x_2, the rainfall in June,

x_3, the rainfall in July,

x_4, the rainfall in August, and

x_5, the rainfall in September up to September 15.

5.4 Formulation and Testing of Some Models

Instead of proposing one particular model on the basis of the boundary conditions as suggested earlier, several types of model are considered in order to demonstrate the effects of model error and the analytical capability of the procedure.

(1) The 5th order product

$$y = ax_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4} x_5^{\alpha_5} \quad . \quad (5.3)$$

This model assumes that all periods are critical. Before using the non-linear least squares in this case a simple least squares regression on the logarithms was carried out, i.e.,

$$\log y = \alpha_1 \log x_1 + \alpha_2 \log x_2 + \alpha_3 \log x_3 + \alpha_4 \log x_4 + \alpha_5 \log x_5 + \log a \quad . \quad (5.4)$$

The least squares regression yielded

$$y = 6.76\, x_1^{0.67}\, x_2^{0.41}\, x_3^{0.17}\, x_4^{0.13}\, x_5^{0.17} \quad . \quad (5.5)$$

The parameter estimates are biased because of minimizing the mean square of logarithm of the assumed multiplicative error term. $R^2 = 0.526$ although this value is uncorrected for degrees of freedom and difficult to interpret in terms of the untransformed data. The parameter values look feasible even though there were no constraints. Corrected error mean square is 581.

The same model was analyzed by the non-linear least squares procedure with the constraints

$$a \geqslant 0 \;,\; 0 \leqslant \alpha_i \leqslant 1, \text{ for } i = 1, 2, 3, 4, 5,$$

with the initial values

$$a = 0.1 \qquad \alpha_i = 1.00 \quad i = 1, \ldots, 5 \;.$$

The set of partial derivatives, $f'(\underset{\sim}{x}, \underset{\sim}{\theta})$ of the assumed function of Equation (5.3), which constitute the independent variables in the non-linear regression, are

$$\frac{\partial y}{\partial a} = x_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4} x_5^{\alpha_5}$$

$$\frac{\partial y}{\partial \alpha_1} = \log_e(\alpha_1)\, ax_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4} x_5^{\alpha_5}$$

$$\frac{\partial y}{\partial \alpha_2} = \log_e(\alpha_2)\, ax_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4} x_5^{\alpha_5} \quad (5.6)$$

. . . .

etc.,

with the "dependent variable"

$$y - ax_1^{\alpha_1} x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4} x_5^{\alpha_5} \quad ,$$

and the "regression" coefficients

$$\Delta a, \Delta\alpha_1, \Delta\alpha_2, \Delta\alpha_3, \Delta\alpha_4, \text{and } \Delta\alpha_5 \quad ,$$

for each iteration in the parameter estimation.

The procedure would not converge thus suggesting that the model is not feasible. The procedure was also run with the initial parameter values equal to those estimated by the log transformation method as given above, but still did not converge.

(2) This model was then extended by considering the possibility of some non zero intercept forms which admit non-critical periods and represent a general model of the type proposed by Hall and Butcher [1968]

$$y = a_o(a_1 + x_1^{\alpha_1})(a_2 + x_2^{\alpha_2})$$

$$(a_3 + x_3^{\alpha_3})(a_4 + x_4^{\alpha_4})(a_5 + x_5^{\alpha_5}) \quad , \quad (5.7)$$

with the constraints

$$a_i \geqslant 0, \quad i = 0, 1, \ldots, 5$$

$$0 \leqslant \alpha_i \leqslant 1 \ , \quad i = 1, 2, \ldots, 5 \ ,$$

and the initial values

$$a_i = .00001 \ , \ i = 1, 2, \ldots, 5 \quad .$$

Again the procedure did not converge. These results suggest that the model is inadequate probably because the interactive effect between periods is more important than was assumed in this model.

(3) The next model considered was the 4th order one including all the five product terms, namely

$$y = a_1 x_2^{\alpha_2} x_3^{\alpha_3} x_4^{\alpha_4} x_5^{\alpha_5}$$

$$+ a_2 x_1^{\beta_1} x_3^{\beta_3} x_4^{\beta_4} x_5^{\beta_5}$$

$$+ a_3 x_1^{\gamma_1} x_2^{\gamma_2} x_4^{\gamma_4} x_5^{\gamma_5} + a_4 x_1^{\delta_1} x_2^{\delta_2} x_3^{\delta_3} x_5^{\delta_5}$$

$$+ a_5 x_1^{\epsilon_1} x_2^{\epsilon_2} x_3^{\epsilon_3} x_4^{\epsilon_4} \ . \quad (5.8)$$

It was expected that not all terms or all parts of a term would be significant. The number of parameters is large (25) but this will be reduced if any of them vanish in the analysis.

This model converged very quickly in 19 iterations to the form

$$y = .0063\, x_1 x_3 x_4 x_5 + 2.59\, x_1 x_4^{0.17}$$

$$+ 9.73\, x_1^{.10} x_2^{0.51} x_3^{0.34} x_5^{0.38}$$

$$+ 0.39\, x_2 x_4 \ . \quad (5.9)$$

The constraints on the parameters were

$$0 \leqslant \alpha_i, \beta_i, \gamma_i, \delta_i, \epsilon_i \leqslant 1.00 \ , \quad \text{for } i = 1, \ldots, 5$$

and

$$a_i \geqslant 0$$

Some of the parameters did take on boundary values and removal of the constraints tended to yield non-physical results. The error mean square corrected for the degrees of freedom for the parameters in the final form of the model was 471 and uncorrected for degrees of freedom was 249. The latter value is included as a measure of the simulation obtained as the sparse data employed in the study does not give an optimistic view of the capabilities of the model. The above value of error mean square corresponds to an explained variance of 46 percent. The derived model also has the interesting boundary property that although no one zero rainfall period will destroy the crop any two successive zero periods will do so. This fact is very much in keeping with the physical situation as the zero values appearing in the data lead to a similar conclusion.

(4) A consideration of all the possible 3rd order products in a model would lead to a large number of parameters, 40 in this case, which exceeds the number of observations. Therefore, some prior knowledge such as the boundary conditions would have to be used. However, any group of five or fewer third order terms are a special case of the five fourth order terms with certain parameters equal to zero and thus a good model of this form would have been detected in the previous analysis.

(5) The first and second order terms were included in one model although each product term had a common exponent

$$y = a_1 x_1^{\alpha_1} + a_2 x_2^{\alpha_2} + a_3 x_3^{\alpha_3} + a_4 x_4^{\alpha_4} + a_5 x_5^{\alpha_5}$$

$$+ a_8 (x_1 x_2)^{\alpha_6} + a_7 (x_1 x_3)^{\alpha_7} + a_8 (x_1 x_4)^{\alpha_8}$$

$$+ a_9 (x_1 x_5)^{\alpha_9} + a_{10} (x_2 x_3)^{\alpha_{10}} + a_{11} (x_2 x_4)^{\alpha_{11}}$$

$$+ a_{12} (x_2 x_5)^{\alpha_{12}} + a_{13} (x_3 x_4)^{\alpha_{13}} + a_{14} (x_3 x_5)^{\alpha_{14}}$$

$$+ a_{15} (x_4 x_5)^{\alpha_{15}} \quad , \quad (5.10)$$

with constraints $a_i \geqslant 0$, and $0 \leqslant \alpha_i \leqslant 1$, which resulted in the model

$$y = 2.79\, x_1 + 2.99\, x_2 + 1.79\, (x_1 x_3)^{0.47} + 0.16\, (x_1 x_4)$$

$$+ 0.42\, (x_2 x_4) + 1.04\, (x_2 x_5) + 1.96\, (x_3 x_5) \quad (5.11)$$

in 26 iterations. Some of the exponents took on the boundary values but when the constraints were removed the procedure did not converge. The error mean square corrected for degrees of freedom was 391 and uncorrected was 230 which corresponds to 50 percent explained variance.

(6) By using the previous model results to show which product terms are important the components of each of these products can be given separate exponents and the model re-analyzed. However, this model should appear as a special case when all the second order trems only are considered as the x_1^{α} and x_2^{β} will occur if their other product component is driven to unity by the exponent going to zero. When the model

$$y = a_1 x_1^{\alpha_1} x_2^{\beta_1} + a_2 x_1^{\alpha_2} x_3^{\beta_2} + a_3 x_1^{\alpha_3} x_4^{\beta_3} + a_4 x_1^{\alpha_4} x_5^{\beta_4} + a_5 x_2^{\alpha_5} x_3^{\beta_5} + a_6 x_2^{\alpha_6} x_4^{\beta_6} + a_7 x_2^{\alpha_7} x_5^{\beta_7} + a_8 x_3^{\alpha_8} x_4^{\beta_8} + a_9 x_3^{\alpha_9} x_5^{\beta_9} + a_{10} x_4^{\alpha_{10}} x_5^{\beta_{10}} \quad (5.12)$$

was analyzed with constraints

$$a_i \geq 0, \ 0 \leq \alpha_i \leq 1, \text{ and } 0 \leq \beta_i \leq 1 \ ,$$

the resulting model obtained in 32 iterations was

$$y = 3.24\, x_1 + 0.68\, x_1 x_3^{0.31} + 0.71\, x_2 x_4 + 2.99\, x_2 x_5^{0.46} + 2.12\, x_3 x_5 \ . \quad (5.13)$$

The error mean square corrected for degrees of freedom was 379 and uncorrected 223 which is a simple explained variance of 51 percent. Note that in the above model x_1 appeared singly as the exponents α_1 and β_1 in the product $x_1^{\alpha_1} x_2^{\beta_1}$ were driven to one and zero, respectively.

(7) The final model in the order series is the first order assembly

$$y = a_1 x_1^{\alpha_1} + a_2 x_2^{\alpha_2} + a_3 x_3^{\alpha_3} + a_4 x_4^{\alpha_4} + a_5 x_5^{\alpha_5} \ , \quad (5.14)$$

with constraints $0 \leq \alpha_1 \leq 1$ and $a_1 \geq 0$, which resulted in

$$y = 2.86\, x_1 + 3.99\, x_2 + 5.64\, x_3^{0.78} + 2.56\, x_4 + 2.96\, x_5 \ , \quad (5.15)$$

with corrected error mean square being 380 and simple explained variance of 42 percent. The relaxation of the constraints resulted in lack of convergence.

The final class of models that were tested were those involving the function

$$\dim(x, a) = \begin{cases} x - a, & x > a \\ 0, & x < a \end{cases} \qquad \frac{\partial \{\dim(x,a)\}}{\partial a} = -\frac{\dim(x,a)}{(x-a)} \quad (5.16)$$

This particular function is a standard library subroutine in the computer softwave package. Two model types were tested, a simple additive model and a multiplicative model. The additive model was

$$y = b_1 \{\dim(x_1, a_1)\}^{\alpha_1} + b_2 \{\dim(x_1 + x_2, a_2)\}^{\alpha_2} + b_3 \{\dim(x_1 + x_2 + x_3, a_3)\}^{\alpha_3} + b_4 \{\dim(x_1 + x_2 + x_3 + x_4, a_4)\}^{\alpha_4} + b_5 \{\dim(x_1 + x_2 + x_3 + x_4 + x_5, a_5)\}^{\alpha_5} \ , \quad (5.17)$$

with constraints $b_i \geq 0$, $0 \leq \alpha_i \leq 1$, and $a_i \geq 0$. The resulting equation was

$$y = 0.67 \{\dim(x_1, 16.29)\} + 2.47 \{\dim(x_1 + x_2, 0)\}^{0.80} + 4.68 \{\dim(x_1 + x_2 + x_3, 1.50)\}^{0.0} + 0.94 \{\dim(x_1 + x_2 + x_3 + x_4, 8.89)\}^{0.69} + + 1.56 \{\dim(x_1 + x_2 + x_3 + x_4 + x_5, 0.31)\} \ . \quad (5.18)$$

The corrected error mean square was 566 which corresponded to a simple explained variance of 31 percent. This result is rather unsatisfactory and some of the parameter values do not appear to be physically feasible. The additive model does not satisfy expected drought behavior as not all terms will vanish for certain zero inputs.

The multiplicative model was of the form

$$y = a_o \left\{\dim(x_1, a_1)\right\}^{\alpha_1} \left\{\dim(x_1 + x_2, a_2)\right\}^{\alpha_2}$$
$$\left\{\dim(x_1 + x_2 + x_3, a_3)\right\}^{\alpha_3} \times$$
$$\times \left\{\dim(x_1 + x_2 + x_3 + x_4, a_4)\right\}^{\alpha_4}$$
$$\left\{\dim(x_1 + x_2 + x_5, a_5)\right\}^{\alpha_5} \qquad (5.19)$$

with constraints

$$a_i \geqslant 0\ ,\ i = 0, 1, \ldots, 5,\ \text{and}\ \alpha_i \geqslant 0\ ,$$

and initial values

$$a_i = 0,\ a_o = 1.00,\ \text{and}\ \alpha_i = 1.00\ .$$

The resulting equation was

$$y = 2.60 \left\{\dim(x_1, 0)\right\}^{0.0} \left\{\dim(x_1 + x_2, 3.44)\right\}^{0.0}$$
$$\left\{\dim(x_1 + x_2 + x_3, 7.53)\right\}^{0.0}$$
$$\times \left\{\dim(x_1 + x_2 + x_3 + x_4, 0)\right\}^{0.0}$$
$$\left\{\dim(x_1 + x_2 + x_3 + x_4 + x_5, 0\right\}^{1.00} \qquad (5.20)$$

with corrected error mean square of 394 and explained variance of 42 percent. This model is the most attractive from the point of view of drought simulation. If any of the terms becomes zero because of insufficient rainfall the yield is zero. Also note that three of the terms have zero (or near zero) exponents, which produce a binary effect in the function. These terms have no direct effect on yield variation but if the term becomes zero the crop is "destroyed."

The values of the a_i parameters could correspond to an average minimum requirement of applied water to sustain the plant during the ith period. The excess over this amount then contributes to plant growth. This still does not take into account the variation of evapotranspiration from year to year. The actual parameter values in this example may not be meaningful because of data limitations but they illustrate the desired action. This model does not take into account interaction between growth periods due to partial damage in one period and other orders of products would need to be tested. However the limited amount of data did not permit an investigation of other functional combinations because of the larger numbers of parameters involved.

At this point it is interesting to compare the above results with those obtained by using simple stepwise linear regression on a quadratic equation involving the initial soil moisture and precipitation data. For this model the variable x_1 was resolved into its two components, initial soil moisture SM and May rainfall p_1. The remaining variables $p_2,\ldots,p_5$ are identical to the $x_2,\ldots,x_5$ used in the previous models. The following terms were considered in the stepwise regression procedure

$$SM, p_1, p_2, p_3, p_4, p_5$$

$$SMp_1, SMp_2, SMp_3, SMp_4, SMp_5$$

$$p_1p_2, p_1p_3, p_1p_4, p_1p_5$$

$$p_2p_3, p_2p_4, p_2p_5$$

$$p_3p_4, p_3p_5$$

$$p_4p_5$$

The Biomedical Computer Program BMD02R [1967] was used with the tolerance level 0.001, F level for inclusion 0.01, F-level for deletion of variables 0.005. The equation with the smallest standard error of estimate corrected for degrees of freedom was

$$y = 54.37SM + 39.93p_2 - 26.76p_3 + 34.70p_4$$
$$- 5.06\ SMp_1 - 301\ SMp_2 - 5.61\ SMp_4 - 8.19\ SMp_5$$
$$+ 9.83\ p_1p_2 - 18.24\ p_1p_3 + 3.36\ p_1p_4 + 21.33\ p_1p_5$$
$$- 0.86\ p_2p_3 - 3.01\ p_2p_4 + 1.48\ p_3p_4 + 14.96\ p_3p_5$$
$$+ 5.11\ p_4p_5\ . \qquad (5.21)$$

The error mean square corrected for degrees of freedom was 249 and the uncorrected explained variance was 84 percent. This relatively high explained variance compared with the previous models immediately suggests the use of the model as a good predictive tool. However, an examination of the signs of some of the regression coefficients and the types of terms included in the regression shows that the model lacks physical feasibility and is unlikely to satisfy critical boundary conditions. It may be useful for prediction within the range of the data used in the analysis but extrapolation is hazardous. The other models have been constrained by a priori physical consideration and although their predictive value on the data used appears to be low they can probably be used more safely in prediction of new variables. This loss of explained variance is to be expected whenever a regression procedure is constrained by physical considerations. Part of the problem lies with the data used for the analyses. Ideally the data should cover a much wider range of events from near optimality to near total drought. The monthly values are a crude approximation, especially during critical periods of the crop growth. However, finer definition may require the use of more variables and thus more observations are required in order to obtain significant results. The scarcity of the data does not even allow use of part of the sample as a test for prediction of new variables.

Finally it should be remembered that other factors affect plant growth such as soil type, temperature, changes in evapotranspiration, use of fertilizers, etc. If data are available these factors can be included in a suitably constructed model, but again as more independent variables are added the data requirements for significant results increase rapidly. An advantage of the type of analysis used here is that the data do not have to be controlled in the sense that only one period is varied at a time but can be random values in all periods provided there is not too much correlation between the rainfall on different plots. Any such correlation will, of course, reduce the effective sample size.

CONCLUSIONS

The principal contribution of this report is the development and demonstration of methods for constructing and analyzing models to explain the relationship between the crop, yields and hydrologic droughts. The examples contained herein are for rainfall input to dryland farming and for a particular class of functional relationships. More useful models could be constructed for inputs which include climatic and soil variables but the lack of sufficient data precludes the inclusion of these relationships in the present study. One of the main purposes of the model is to reduce the uncertainty in estimation of the statistical properties of agricultural drought over large areas. The abundance of rainfall data encourages the development of models which relate it to drought effects. In this report it is shown that the models can explain about 50 percent of the variance in crop yields on the basis of rainfall input. The amount of unexplained variance is due to several factors, namely

(1) The models do not take into account directly the other contributary inputs such as evapotranspiration data.

(2) The data is sparse both in number of observations and distribution throughout the growing season, i.e., monthly rainfall instead of weekly.

(3) The crop yields cover only a small range and do not include cases of severe droughts. Extrapolation of results from such a narrow "band" is hazardous.

It has also been demonstrated that the models do not explain as much variance in the yield data as the use of polynomial function solved with linear regression. The reason for this behavior is the physical constraints imposed in the non-linear procedure. This use of constraints is a way of infusing some a priori knowledge into the analysis and makes the resulting model more suitable for use outside the range of the sample data.

Finally, it is hoped that the presentation of the mathematical tools and examples in this paper will encourage the collection of more suitable data and the construction of models to make the best use of such data.

REFERENCES

Barger, Gerald L. and Thom, H.C.S., 1949, A method for characterizing drought intensity in Iowa: Agronomy Journal, Vol. 41, pp. 13-19.

Biomedical Computer Programs, X-series Supplement, ed. W. J. Dixon, Univ. of Calif. Press 1970, pp. 177-186.

Biomedical Computer Programs, ed. W. J. Dixon, Univ. of Calif. Press 1967, pp. 233-255.

Blake, C.D., ed., 1967, Fundamentals of modern agriculture: Sydney University Press.

Brown, Merle J. and Vanderlip, R.L., 1969, Relationship of Palmer drought variables to yield of winter wheat: Agronomy Journal, Vol. 61, pp. 95-98.

Currier, H.B., 1967, Nature of plant water in "Irrigation of Agricultural Lands": Ed. Hagan, *et al.*, American Soc. of Agron., pp. 307-308.

David, Wilfredo P. and Hiler, Edward A., 1970, Predicting irrigation requirements of crops: Proc. A.S.C.E. Vol. 96, IR3.

Denmead, O.T. and Shaw, R.H., 1962, Availability of soil water to plants as affected by soil moisture content and meteorological conditions: Agronomy Journal, Vol. 54, pp. 385-389.

Doll, John P., 1967, An analytical technique for estimating weather indexes from meteorological measurements: Journal of Farm Economics, Vol. 49, pp. 79-88.

Edwards, Clark, 1962, Nonlinear programming and non-linear regression procedures: Journ. of Farm Economics, Vol. 44, pp. 100-114.

Efroymson, M.A., 1960, Multiple regression analysis in "Mathematical models for digital computers": A. Ralston and H.S. Wilf, Wiley, pp. 191-203.

Flinn, J.C. and Musgrave, W.F., 1967, Development and analysis of input-output relations for irrigation water: The Australian Journal of Farm Economics, Vol. 11, No. 1.

Hall, Warren A. and Butcher, William S., 1968, Optimal timing of irrigation: Proc. A.S.C.E., Vol. 94, No. IR2.

Hartley, H.O., 1961, The modified Gauss-Newton Method for the fitting of nonlinear regression functions by least squares: Technometrics, Vol. 3, No. 2.

Holt, R.F., Timmons, D.R., Allmaras, R.R., Caldwell, A.C., Gates, C.E., and Shubeck, F.E., 1965, Production factors for corn in the northwest corn belt area: U.S.D.A. ARS41-110.

Jennrich, R.I. and Sampson, P.F., 1968, Application of stepwise regression to non-linear estimation: Technometric, Vol. 10, No. 1.

Jensen, M.E., 1968, Water consumption by agricultural plants in "Water Deficits and Plant Growth" Vol. 2, Academic Press, Chapter 1, pp. 1-22.

Moore, Charles V., 1961, A general analytical framework for estimating the production function for crops using irrigation water: Journal of Farm Economics, Vol. 43, pp. 876-888.

Parks, W.L. and Knetsch, J.L., 1959, Corn yields as influenced by Nitrogen level and drought intensity: Agronomy Journal, Vol. 51, pp. 363-364.

Pearson, Lorentz C., 1967, Principles of agronomy: Reinhold.

Richards, L.A., and Wadleigh, C.H., 1952, Soil water and plant growth in "Soil Physical Conditions and Plant Growth" Agronomy Monographs, Vol. II, Ed. Shaw; B. T. Academic Press, pp. 73-251.

Stewart, J. Ian and Hagan, Robert M., 1969, Predicting effects of water shortage on crop yield: Proc. A.S.C.E., Vol. 95, IR1.

Van Bavel, C.H.M., 1953, A drought criterion and its application in evaluating drought incidence and hazard: Agronomy Journal, Vol. 45, pp. 167-172.

Yaron, Dan, 1971, Estimation and use of water production functions in crops: Proc. A.S.C.E., Vol. 97, IR2.

KEY WORDS: Agricultural drought, crop production, non-linear models.

ABSTRACT: The paper deals with the effects and evaluation of the effects of hydrologic drought on the yield of dry-land farming. Some of the physiological principles involved in crop growth are reviewed in the context of a system which relates reeuction in crop production to deficits in the water input. A possible set of non-linear functions is developed to describe this system by using the physical properties of crop production as constraints on the mathmetics. The method of non-linear least squares regression is described as a suitable technique for solving the system function. Application of the analytical methods to a set of corn yield data is demonstrated with several types of system functions, one of which is suggested as a suitable practical model. The predictability of the models is not

KEY WORDS: Agricultural drought, crop production, non-linear models.

ABSTRACT: The paper deals with the effects and evaluation of the effects of hydrologic drought on the yield of dry-land farming. Some of the physiological principles involved in crop growth are reviewed in the context of a system which relates reeuction in crop production to deficits in the water input. A possible set of non-linear functions is developed to describe this system by using the physical properties of crop production as constraints on the mathmetics. The method of non-linear least squares regression is described as a suitable technique for solving the system function. Application of the analytical methods to a set of corn yield data is demonstrated with several types of system functions, one of which is suggested as a suitable practical model. The predictability of the models is not

KEY WORDS: Agricultural drought, crop production, non-linear models.

ABSTRACT: The paper deals with the effects and evaluation of the effects of hydrologic drought on the yield of dry-land farming. Some of the physiological principles involved in crop growth are reviewed in the context of a system which relates reeuction in crop production to deficits in the water input. A possible set of non-linear functions is developed to describe this system by using the physical properties of crop production as constraints on the mathmetics. The method of non-linear least squares regression is described as a suitable technique for solving the system function. Application of the analytical methods to a set of corn yield data is demonstrated with several types of system functions, one of which is suggested as a suitable practical model. The predictability of the models is not

KEY WORDS: Agricultural drought, crop production, non-linear models.

ABSTRACT: The paper deals with the effects and evaluation of the effects of hydrologic drought on the yield of dry-land farming. Some of the physiological principles involved in crop growth are reviewed in the context of a system which relates reeuction in crop production to deficits in the water input. A possible set of non-linear functions is developed to describe this system by using the physical properties of crop production as constraints on the mathmetics. The method of non-linear least squares regression is described as a suitable technique for solving the system function. Application of the analytical methods to a set of corn yield data is demonstrated with several types of system functions, one of which is suggested as a suitable practical model. The predictability of the models is not

high because of restriction imposed by the availability of data. Some improvements are suggested for further development when more suitable data are available.

REFERENCE: Bidwell, V.J., Colorado State University, Hydrology Paper No. 53 (July, 1972) "Agricultural Response to Hydrologic Drought".

high because of restriction imposed by the availability of data. Some improvements are suggested for further development when more suitable data are available.

REFERENCE: Bidwell, V.J., Colorado State University, Hydrology Paper No. 53 (July, 1972) "Agricultural Response to Hydrologic Drought".

high because of restriction imposed by the availability of data. Some improvements are suggested for further development when more suitable data are available.

REFERENCE: Bidwell, V.J., Colorado State University, Hydrology Paper No. 53 (July, 1972) "Agricultural Response to Hydrologic Drought".

high because of restriction imposed by the availability of data. Some improvements are suggested for further development when more suitable data are available.

REFERENCE: Bidwell, V.J., Colorado State University, Hydrology Paper No. 53 (July, 1972) "Agricultural Response to Hydrologic Drought".