OEY, Hong-Sioe, 1937-

ESTIMATION OF THE EFFECT OF LAND USE AND TREATMENT ON THE YIELD OF WATER BY USE OF COMPONENT ANALYSIS AND MULTIPLE REGRESSION TECHNIQUES.

The University of Oklahoma, Ph.D., 1966
Engineering, civil

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

ESTIMATION OF THE EFFECT OF LAND USE AND TREATMENT ON THE YIELD OF WATER BY USE OF COMPONENT ANALYSIS AND MULTIPLE REGRESSION TECHNIQUES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

HONG-SIOE OEY

Norman, Oklahoma

1966

ESTIMATION OF THE EFFECT OF LAND USE AND TREATMENT ON THE YIELD OF WATER BY USE OF COMPONENT ANALYSIS AND MULTIPLE REGRESSION TECHNIQUES

APPROVED BY

John C. Brixey

[illegible]

Edwin H Keeler

Robert J. Nelson

DISSERTATION COMMITTEE

ACKNOWLEDGEMENTS

The author wishes to acknowledge the guidance of Professor George W. Reid in conducting this research and his advice on the preparation of this dissertation. He is also grateful to the other members of his doctoral committee, Dr. J.C. Brixey, Dr. R.Y. Nelson, and Dr. E.H. Klehr for reviewing the manuscript and earlier discussions.

The author wishes to express his appreciation to his colleaques in the Department of Civil Engineering and Environmental Sciences at the University of Oklahoma for their generosity with their time in discussing portions of this work and advancing suggestions for improvement.

His appreciation also goes to Dr. J.R. Assenzo, presently at the Upjohn Co., Kalamazoo, Michigan, who gave him his first interest in the application of Statistics in Engineering.

Thanks are due to the Bureau of Reclamation of the United States Department of the Interior for providing the funds which made this research possible.

And last but not least the author is grateful to Mrs. Linda Hawkins for her valuable assistance in typing the manuscript.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ESTIMATION OF THE EFFECT OF LAND USE AND TREATMENT ON THE YIELD OF WATER BY USE OF COMPONENT ANALYSIS AND MULTIPLE REGRESSION TECHNIQUES

CHAPTER I

INTRODUCTION

Many efforts have been made towards rationalizing the estimation of the effects of land use and treatment on the yield of water in a watershed. A large number of these estimates are still being made by inadequate means. In the last several years water supply has become critical in many areas in the United States. Many regions which have adequate supply of water at the present time must eventually seek ways to increase this supply to keep up with increasing demands. There are two primary sources of water: groundwater and surface supply. Of these, there is already overdevelopment of the groundwater sources in many parts of the country with resulting falling water table and dried up wells.

Water Yield

Precipitation is the source of water yield. Land releases water to streams and underground basins under either of the two conditions.

1. When precipitation is delivered to the soil surface more rapidly than it can be absorbed, the surface water flows downhill on the surface.
2. When more water enters the soil than it can hold the surplus moves downhill beneath the surface, some of it reappearing in springs, streams, and lakes, and some of it flowing into groundwater basins.

The quantity of water released is therefore that portion of the precipitation supply after demands are satisfied, including absorption and storage by the soil, evaporation from land, vegetation, and water surfaces.

Under natural conditions water is rarely yielded in ways that are entirely satisfactory so far as man's activities are concerned. There are many circumstances, some natural and some due to man's entry upon the land, that have made control over the yield of water necessary.

Considered in the most general terms, control over water yield can be exercised by three kinds of activities.

1. Control over the release of water from the atmosphere

which is meteorological in nature.

2. Use of engineering works and structural modifications such as reservoirs, water retaining structures, etc.

3. Make use of the wild or cultivated vegetation that clothes water yielding areas.

Effects of Land Use and Treatment

The sum of the volumes of water used by the vegetative growth of a given area of transpiration and building of plant tissue and that evaporated from adjacent soil, snow or intercepted precipitation on the area in any specified time, divided by the area, is called the "consumptive use". Because these losses represent water which would otherwise be available for human use, any method by which consumptive use can be decreased is to be desired. No other single factor affects runoff in so many ways as does vegetation. The condition of a watershed's vegetation can determine how much of the rainfall will be lost to evaporation, appear as surface runoff, or enter the soil. This study is concerned with the effects of land use and treatment which is parallel to vegetation, or runoff, and where possible, ways of changing the land practices to increase runoff will be suggested.

Objectives

The purpose of this study was to estimate the effect of land use and treatment on the yield of water by a mathematical model. If this could be achieved successfully it would provide a means of predicting water yield for the future by applying the predicted future values of the parameters in the model. The present work should be viewed as a pilot study to test the validity of the approach on readily available data.

Need for the Study

A major goal in water resources planning is to derive a mathematical model which will maximize the net benefit from the operation of a water resources region. To optimize the net benefit it is essential that all the variables, and their interactions, be well defined and that their effect on the operation of the region be estimable.

A very important part of this is the yield of water in watersheds of a water resources region. This paper is concerned with the estimation of the effect of land use and treatment on the yield of water. Methods will be presented for estimating water yield as streamflow by use of component analysis and multiple regression techniques.

In Tables 1a and 1b the percent reduction of water yield in the nation as a whole, and the increase of yield due to improved irrigation is shown.

Historical Review and Work by Others

The practice of watershed management was started in the U.S.A. over 75 years ago. The State of New York in 1868 began an investigation into the adverse effect that destruction of forest cover was having on fish. From such a beginning was launched the present Forest Commission which in 1955 controlled 2.4 million acres of forest preserve for the primary purpose of protecting water supplies (2). This pioneer work has lead to many other complicated works of which water yield estimation is of most importance.

Many methods of evaluating effects of watershed treatment on streamflow have been studied. Federal, State and local agencies involved in the control and use of the nation's water resources are participating in this area of study. The Bureau of Reclamation, Soil Conservation Service, and Agricultural Research Service are conducting the "Cooperative Water Yield Procedures Study" to develop and test methodology for estimating the effect that conservation activities may have on the yield of streamflow. This study group has developed the so-called "Rational Procedure" method which consists of applying logic, and known effects, to the problem.

TABLE Ia

Percent Reduction in Downstream Water Yield
Assuming Current Irrigation Efficiency

Item	1980	2000
Land use and treatment	0.09	0.23
Structures	0.23	0.34
Irrigation	1.18	4.43
TOTAL.	1.50	5.00

TABLE Ib

Percent Reduction in Downstream Water Yield
Assuming Increased Irrigation Efficiency

Item	1980	2000
Land use and treatment	0.09	0.23
Structures	0.23	0.34
Irrigation	-0.68	0.65
TOTAL.	-0.36	1.22

(where the negative values indicate increase in water yield. The difference between 5.00 and 1.22 percent in the year 2000 represents an amount of water of approximately 50,000,000 acre-ft.)

It breaks the problem down to its elements on the basis of climate, evapotranspiration, soils, topography, vegetation, land use and treatment, and streamflow, then treating only those elements subject to effects of conservation use and treatment of land. The rational procedure provides reasonable estimates of average annual effects of watershed treatment on streamflow in the dry-subhumid-to-arid areas such as the Great Plains, Midwest, and Southwest.

Reports (12) indicated that a procedure, based only on statistically significant results secured from studies of river basins and research watersheds, could not be developed. This does not mean that a combination of rational reasoning and statistical results could not be employed. In this paper the author attempted to utilize as much as possible the statistical results, and in very uncertain situations rational judgments are used.

As mentioned before a method of the estimation of the effects of land treatment on streamflow is essential for a better water resources management. The main objective is to reduce the consumptive use. Nearly all of the methods known are based mainly on the estimation of losses due to evapotranspiration and seepage.

In the rational procedure, evapotranspiration is regarded as one of the most important elements. Many efforts have been made to reduce evaporation by applying hexadecanol and octadecanol on water surfaces. The results of pan-evaporation studies are very encouraging; evaporation reductions of 33 to 60 percent have been obtained (1). However, experiments carried out on small ponds by the Oklahoma State University and the Southwest Research Institute of San Antonio, Texas, have not been so encouraging. The Institute reported a reduction of 18 percent in evaporation from a 3-acre pond near San Antonio during the period October 10-31, 1956. Oklahoma State University reported reductions in evaporation of 4 to 5 percent from specially designed 0.3 acre ponds. Although reductions of evaporation could be artificially achieved, it is a question whether the resulted benefit could be higher than the cost.

CHAPTER II

DISCUSSION OF OPERATIONAL TECHNIQUES

Variables Used in the Study

The estimation of the effect of land use and treatment on the yield of water involves no less than 30 variables. Many of these variables are dependent on some others. It is therefore desirable to reduce the number of variables to be used in the model to a reasonably small number of independent variables. The usual method of doing this is by choosing the most significant variable using a step-wise regression analysis, and discard the rest. This method does not guarantee to give variables which are independent of each other. No attempt was made to compare results of the two methods.

A more mathematical way of doing this is by using Component Analysis Techniques, finding the principal components of the original variables, and using these principal components as our variables for the model. It turns out that these principal components are independent of each other.

In this study component analysis is used as the primary step to reduce the number of variables, followed by the use of regression analysis to determine the model.

Mathematical Techniques

This section is meant as a review of some of the basic ideas commonly used to explain the physical meaning of some statistical theories.

Points in n-Dimensional Space

A point in n-dimensions is defined as an ordered set of quantities $(x_1, x_2, \ldots x_n)$. These ordered n-tuples can also be interpreted as vectors in an n-dimensional space. We shall refer to this space as an S_n. Any equation in the variables x determines a sub-space in S_n, which is also called a variety.

Coordinate Transformations

Consider a transformation

$$Y_i = \sum_{j=1}^{n} l_{ij} x_j + a_i \qquad i = 1, 2, \ldots n$$

This can be regarded as a displacement of the origin, represented by a_i, and a "rotation", represented by the coefficients l_{ij}. It is usually convenient to consider this separately. Of particular importance is the so-called "orthogonal" transformation.

$$Y_i = \sum_{j=1}^{n} l_{ij} x_j$$

where

$$\sum_{j=1}^{n} l_{ij} l_{jk} = \delta_{ik}$$

$$= 1, \quad i = k$$

$$= 0, \quad i \neq k$$

Coefficients l can always be chosen as to obey the above condition. Writing L for the matrix (l_{ij}) and L' for its transpose we see that LL' = I where I is the unit matrix. Thus $|L||L'| = 1$, and since the determinants of the matrix and its transpose are equal, we have $|L|^2 = 1$. We shall take the determinant as +1. The negative sign corresponds to the "left-handed" transposition and does not affect the properties with which we are concerned. Also

$$Y = LX$$

and $L'Y = L'LX = X$

Thus the transformation is bi-orthogonal.

Equation of a Flat or Hyperplane Through Given Points

In n-dimensions the general equation of an (n-1)-flat is $a_1x_1 + a_2x_2 + \ldots a_nx_n = k$. If this passes through n points with coordinates $x_{1i}, x_{2i}, \ldots x_{ni}$, $i = 1, 2, \ldots n$, we have n equations typified by

$$a_1x_{1i} + a_2x_{2i} + \ldots a_nx_{ni} = k \qquad i = 1, 2, \ldots n$$

Eliminating the (n+1) constants a_1, a_2, ... a_n, k from the (n+1) equations, we have for the equation of the (n-1) - flat

$$\begin{vmatrix} x_1 & x_2 & \ldots\ldots & x_n & 1 \\ x_{11} & x_{21} & \ldots\ldots & x_{n1} & 1 \\ \cdot & \cdot & \cdot\cdot\cdot\cdot & & \\ \cdot & \cdot & \cdot\cdot\cdot\cdot & & \\ \cdot & \cdot & \cdot\cdot\cdot\cdot & & \\ x_{n1} & x_{n2} & \ldots\ldots & x_{nn} & 1 \end{vmatrix} = 0$$

If the matrix (x_{ij}) is not of rank n the points are not independent and the (n-1) - flat is not uniquely determined. The generalization to n dimensions of results which are familiar in two and three dimensions will offer no difficulty. For example, with n = 3, if the matrix (x_{ij}) is of rank 2 the three points lie on a line and a single infinity of planes will contain them.

Component Analysis

Suppose we have p variates x_1, x_2, ... x_p, each observed on n individuals. We can represent the observations in a matrix form:

$$\begin{bmatrix} x_{11} & x_{12} & \cdots\cdots & x_{1n} \\ x_{21} & x_{22} & \cdots\cdots & x_{2n} \\ \cdot & \cdot & \cdots\cdots & \\ \cdot & \cdot & \cdots\cdots & \\ \cdot & \cdot & \cdots\cdots & \\ x_{p1} & x_{p2} & \cdots\cdots & x_{pn} \end{bmatrix}$$

where x_{ij} represents the j*th* observation on the i*th* variate. The object of component analysis is to reduce the number of variates. To do this we shall introduce linear transformation of type

$$\zeta_i = \sum_{j=1}^{p} a_{ij}\, x_j$$

It is hoped that we can express the data in terms of fewer than p of the ζ's. We thus reduce the dimension of the problem to m<p variates. But this is not always possible, therefore, we shall try to carry out an approximate reduction in this sense.

We shall choose the a_{ij} coefficients so that the first of our new variates ζ_1 has as large a variance as possible; we shall then choose ζ_2 orthogonal to the first and to have as large a variance as possible; and so on. In this way we transform into new uncorrelated variates which account for as much of the variation as possible in descending order.

It may be that the first two or three variates account for nearly the whole of the variation, say 85 or 90 percent, and the contribution of the other p - 2 or p - 3 is small. We can then say that the variation is represented approximately by the first two or three variates and in favorable circumstances may be able to neglect the remainder.

The technique of extracting orthogonal components in order of decreasing variance is called principal component analysis, and the sets of ζ- values are the principal components. It can be shown that the rank of the correlation matrix is equal to the number of transformed variates, ζ_i, required to represent the data. For example, if a four-variate model has a correlation matrix with minors and cofactors such that its rank is two, the implication is that two sets of variates, ζ_1 and ζ_2 are sufficient. This does not imply that any two of the original input variates, such as x_1 and x_4, can be found explicity in terms of the remaining two, x_2 and x_3. Instead, two blends, ζ_1 and ζ_2, are so defined that all the characteristics of the data are maintained. To evaluate the components, ζ_i, it is necessary to solve the following characteristic matrix equation

$$\left| r_{ij} - \lambda I \right| = 0$$

in which r_{ij} represents the correlation matrix; I denotes the unit matrix and λ indicates an underdetermined scalar multiplier.

This can be explained as follows: Let us again examine the set of p variables, each observed on n individuals. This can be considered as a set of p vectors in S_n. They determine an S_p which may be considered as immersed in the S_n. In this S_p we can find linear transformations to new variables $y_1, y_2, \ldots y_p$ which are orthogonal. We can do so in many ways. Let us choose for y_1 the axis for which the corresponding variables $\sum l_{ik} x_k$ has the greatest variance, i.e., such that the sum of projections of the vectors on to it is a maximum. Measuring as usual from a origin at the means and taking the vector to have unit length, we have to minimize

$$(\sum l_i x_i)' \quad (\sum l_i x_i)$$

subject to $\sum l_i^2 = 1$. This leads to finding the unconditioned maximum of $\sum' \; l_i l_j r_{ij} - \lambda \sum l_i^2$ and finally to

$$|r - \lambda I| = 0$$

where r is the correlation matrix.

λ is analogous to a lagrange multiplier. The above equation yields, in general, a p-rooted polynomial in λ, whose roots are the eigen-values of the matrix.

If the matrix of coefficients, r_{ij}, is positive-definite as all correlation matrices must be, it can be shown that

$\sum_{i=1}^{p} \lambda_i = p$, and moreover, if the original variates are derived from a normal population or a population rendered normal by a transform,

$$\frac{\lambda_i}{p} \, 100\% = P_i$$

in which P_i represents the percentage of variation accounted for by the ith component. All the λ_i's are real and non-negative.

Multiple Regression

Suppose y is an observable random variable, and x_1, x_2, ... x_n are observable mathematical variables. Suppose that y depends on p other variates x_1, x_2, ... x_p. These need not be independent, and in fact may all be powers of x, a single variate.

We shall call x_1 ... x_p the predictors and y the predicted variates. The usual problem is to find the best linear predicting equation for y of the form:

$$\hat{Y} = \sum_i b_i x_i \qquad i = 0, 1, 2, \ldots p.$$

To avoid introducing a separate constant term, the first variate x_0 is a dummy which always takes the value 1. The coefficient b_i are called partial regression coefficient.

They are estimators of the true regression coefficient β_i which are supposed to characterize the population, and they are calculated from a set of observations of each of the p+1 variates made on N individuals from the population. Let us denote the observed value of x_i on the individuals numbered α by $x_{i\alpha}$.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots\cdots & x_{1n} \\ x_{21} & x_{22} & \cdots\cdots & x_{2n} \\ \cdot & \cdot & \cdots\cdots & \\ \cdot & \cdot & \cdots\cdots & \\ x_{p1} & x_{p2} & \cdots\cdots & x_{pn} \\ y_1 & y_2 & \cdots\cdots & y_n \end{bmatrix}$$

The true regression equation in the population is

$$\eta = \sum_i \beta_i x_i \qquad i = 0, 1, 2, \ldots p$$

The x_i's are fixed numbers, or at least the errors in x_i are small compared with the error in y.

The b_i will therefore be chosen to minimize the sum of squares of the difference between the observed y_α and the theoretical η_α. This is the method of least squares. We want

$$\sum_\alpha (y_\alpha - \sum \beta_i x_i)^2 = \text{minimum.}$$

Differentiating the above expression with respect to β_i and equating the derivative to 0, we obtain $\hat{\beta}_i$, which is equal to b_i.

$$\sum_{\alpha} x_{i\alpha} \left(Y_{\alpha} - \sum_{j} \beta_j x_j\right) = 0$$

$$i,\ j = 0,\ 1,\ 2,\ \ldots\ p$$

or

$$\sum_{\alpha} \sum_{j=0}^{p} x_{i\alpha} x_{j\alpha} \hat{\beta}_j = \sum_{\alpha} x_{i\alpha} y_{\alpha}$$

$$i = 0,\ 1,\ 2,\ \ldots\ p.$$

This is a system of p+1 linear equations in the p+1 unknowns $\hat{\beta}_j$, with $\hat{\beta}_j = b_j$:

$$\begin{aligned}
&b_0 a_{00} + b_1 a_{01} + \ldots b_p a_{0p} = g_0 \\
&b_0 a_{10} + b_1 a_{11} + \ldots b_p a_{1p} = g_1 \\
&\cdot \qquad \cdot \\
&\cdot \qquad \cdot \\
&b_0 a_{p0} + b_1 a_{p1} + \ldots b_p a_{pp} = g_p
\end{aligned}$$

where:

$$a_{ij} = \sum_{\alpha} x_{i\alpha} x_{j\alpha}$$

$$g_i = \sum_{\alpha} x_{i\alpha} y_{\alpha}$$

This is the so-called "normal equations" of the regression problem. We than compute the b_i's from this system of linear equations, and find the regression equation

$$\hat{Y} = \beta_0 + \beta_1 x_1 + \ldots \beta_p x_p$$

which is the best linear predicting equation for Y.

When one is using the same data to derive several forms of a linear equation it is believed that the criterion to use for selection of the form that "best" fits is to choose the form which has the highest coefficient of determination, R^2. This is the simple correlation coefficient between Y and $\hat{Y}$.

$$R^2 = 1 - \frac{\text{var } (\hat{Y})}{\text{var } (Y)}$$

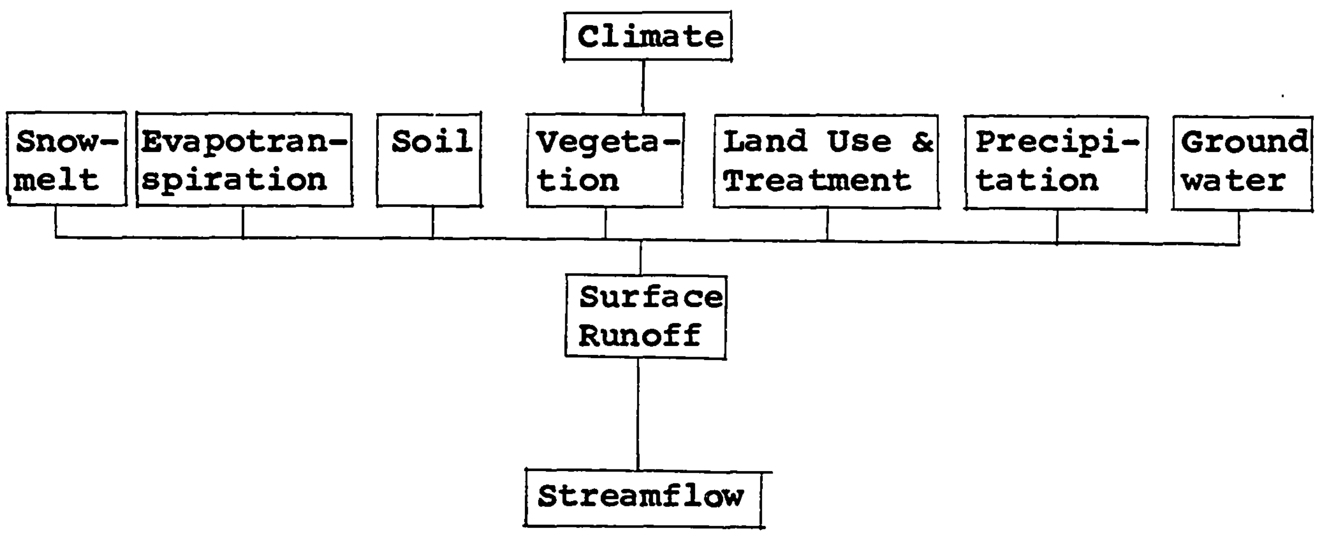
It specified the ratio of the variance explained by the prediction equation to the total variance of the dependent variable. Thus an R^2 of 0.90 signifies that only 10 percent of the variance in the sample is unexplained by that particular regression equation. The square root of the coefficient of determination, R, is called coefficient of multiple correlation.

CHAPTER III

BASIC MODEL FOR OKLAHOMA

Description

The main factors and their inter-relationships can be seen in a simplified scheme shown below.



For the state of Oklahoma the effect of evaporating water surfaces such as farm ponds, reservoirs, etc., is very great on the yield of water due to the long hot summers. Snow-melt gives very little contribution to streamflow. In this model the effects of land practices on the yield of water is measured in terms of streamflow reduction. Thus the model is treated in a global sense instead of breaking it down in small details. Streamflow is a final form of water yield readily available to men. It was decided that annual records would give good results because these periods include a complete annual climatic variation. The effect of groundwater variation is this covered by the effect of land practices and vegetational variation. It was also decided that the four factors, evapotranspiration, soil characteristics, vegetation and land use and treatment could be reasonably grouped together and represented by two new variables that account for most of their variations. It was later determined that this group could be very well represented by the number of farm ponds and the first principal component of the most significant land practices such as contour farming, cover cropping, etc.

Sparkman (3) shows that the state of Oklahoma could be roughly divided into three general rainfall regions. The eastern region is bounded on the west by the 40 inch rainfall line, which roughly follows the line of 96 degree longitude. The central division is in turn bounded on the west by the 30 inch rainfall line, roughly paralleling the 98 degree longitude line; and the western region lies beyond the 30 inch rainfall line. In the eastern division rainfall varies from a low of 40 inches to a high of 51 inches per year. The central region has an annual rainfall varying from 30 to 40 inches. The water situation in the western part of the state is the most critical. A mean annual rainfall of 30 inches in the eastern part of this region decreases to less than 17 inches at the western border. See Figures Ia and Ib.

Analysis and Solution

The scheme shows the main factors which we shall designate by x_1, x_2, x_3, x_4, x_5, x_6 and x_7 respectively from left to right. Using multiple regression analysis an equation of the form $Y = F(x_1, x_2, \ldots x_7)$ can be derived, where

Y = streamflow

F = a linear function for example

Suppose we have n sets of observations. We can consider these as n points in a seven-dimensional space. The problem could not be interpreted physically as finding a flat or hyperplane in this seven-dimensional space which contains the n points. It could be that the main factors have in itself subfactors. Let us consider x_5, land use and treatment which can be controlled by men. There are about 30 practices that are classified as land use and treatment. It is desirable to find one or two, preferably one parameter which represents

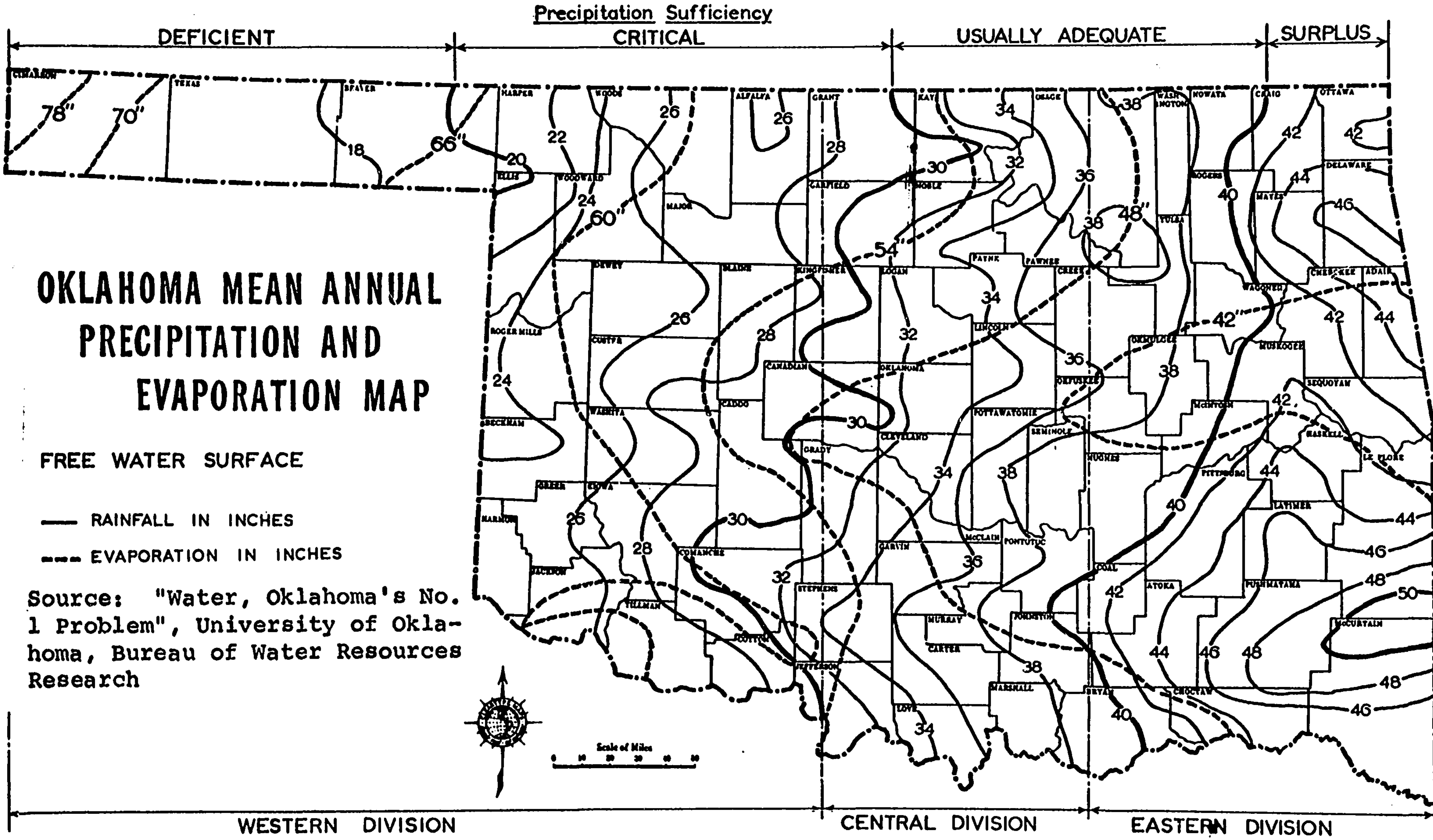
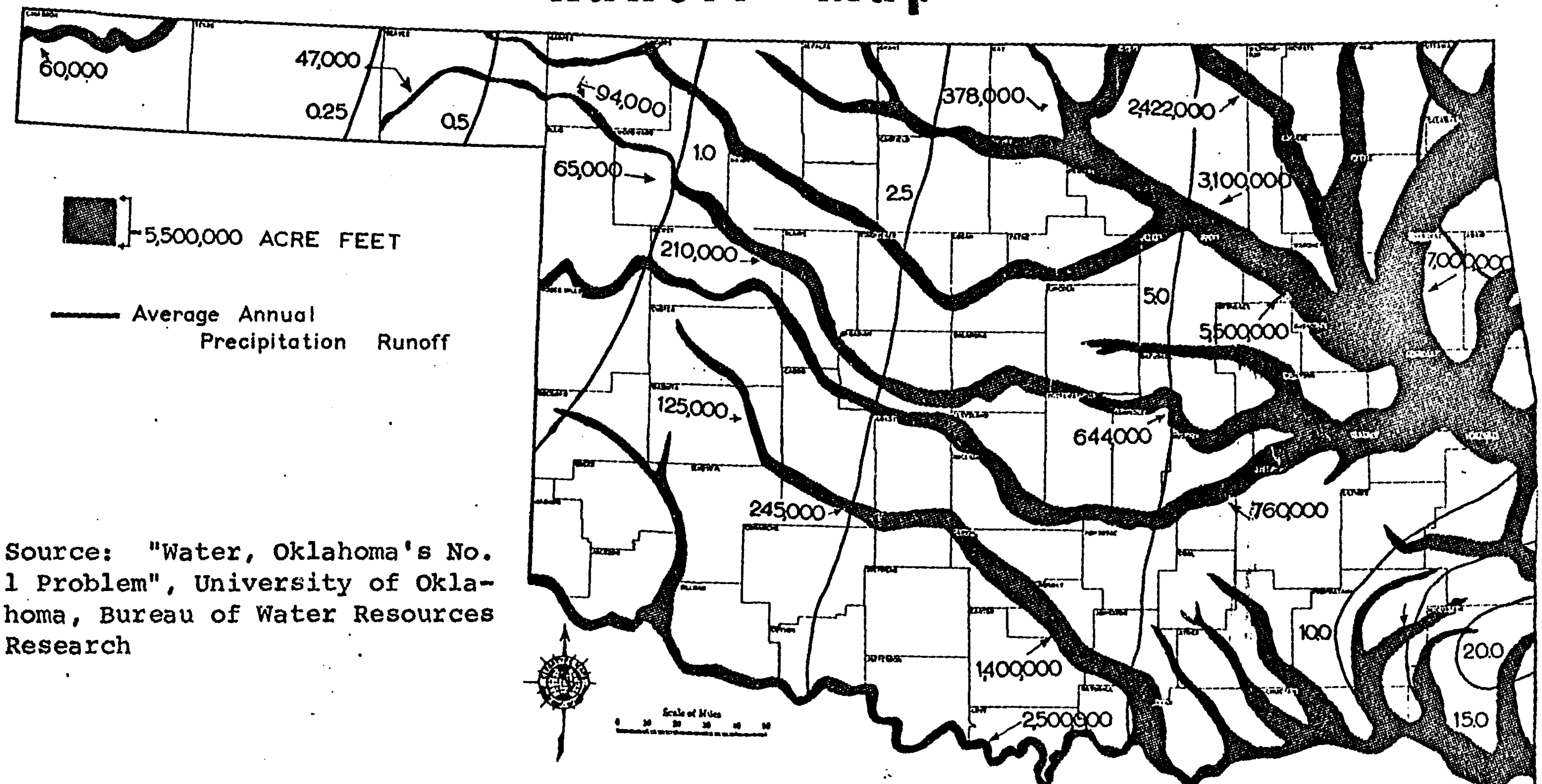

Precipitation Sufficiency
DEFICIENT
CRITICAL
USUALLY ADEQUATE
SURPLUS
OKLAHOMA MEAN ANNUAL PRECIPITATION AND EVAPORATION MAP
FREE WATER SURFACE
RAINFALL IN INCHES
EVAPORATION IN INCHES
Source: "Water, Oklahoma's No. 1 Problem", University of Oklahoma, Bureau of Water Resources Research
WESTERN DIVISION
CENTRAL DIVISION
EASTERN DIVISION

Figure 1a

Runoff Map



Source: "Water, Oklahoma's No. 1 Problem", University of Oklahoma, Bureau of Water Resources Research

Figure 1b

the whole group, if such a parameter exists. To do this we use component analysis as our tool. Suppose we were successful in finding this "principal component" which represents for 90% of the variations of the entire group. This means an important reduction of the dimension of the problem. If as the final result we have only three variables left x_1, x_2, and x_3, we simply try to find the equation

$$Y = F(x_1, x_2, x_3)$$

which is a flat or subspace in four dimensional space. From this equation we can predict values of Y for the future by substituting the future values of the x_i's in the equation.

Records in county offices of the Department of Agriculture in 77 counties in Oklahoma list the land practices over a period of 17 years. The types of records available to this study is shown in the appendix. By thorough examination of the data and consultation with experienced workers in this field it was decided that the diversified land practices can be reduced and well represented by 7 parameters. Several graphs of these parameters versus runoff were plotted to get a general idea of their interactions. This study is meant to show the validity of the method and is therefore far from being perfect. It is hoped that when more complete data is available a better result could be achieved using this same method. In this work a study was made on three watersheds in Oklahoma: Washita River above Clinton in the western part of the state, Black Bear Creek above Pawnee in the northern part, and Clear Boggy Creek above Caney in the south-eastern part of the state. The areas of these watersheds are 1977, 580, and 720 square miles respectively.

CHAPTER IV

MODEL TEST

Procedure

The systematic steps of the procedure are as follows:

1. Using past experiences and logical judgments, the variables are chosen and divided according to their importance.

2. Compute the correlation matrices of the various groups of variables according to need.

3. Compute the principal components by computing the eigen-vectors of the correlation matrix.

4. Determine the best linear regression equation using the most significant parameters, in our case a combination of original input data and first principal components.

5. Predict the future values of the parameters and substitute in the equation to predict future values of the regression.

The computations are so cumbersome as to require the use of a digital computer. A program for use with the IBM digital computer is on file in the School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, Oklahoma.

Results show that the effects of farm ponds are not always as would be expected, namely, reducing the yield of water. In the Washita watershed above Clinton farm ponds improve the yield of water when the number of ponds is small. When the number increases to a certain level then the ponds will start reducing water yield.

If we strictly follow statistical results, regression equations with largest coefficients of multiple correlation should be chosen. However, due to errors in the data and unexplained climatic effects, such is not the case. Instead, the equations which give the best explanations are preferred. Thus statistical results are used as much as possible within the limitations of rational judgments.

In this study the first principal components of land use in all cases account for more than 60% of the group variations. On account of this and in order to damp uncertain fluctuations, only the first principal components are used to represent the groups. The remarkable thing in this study is the finding that the main agricultural practices could be represented by their first principal component with a high degree of accuracy.

In the regression analysis we seek for an equation which best estimates streamflow as a function of precipitation, number of farm ponds, and the first principal component mentioned above. Standardized variables are used in order to eliminate dimension, thus making the variables uniform in the equation.

Discussion of Results

Case I: Washita Watershed above Clinton

Area: 1977 square miles

Mean annual precipitation during the period 1946 to 1960: 24.09 inches

The first three land practices tested for their intercorrelations are contour farming, cover cropping, and range and pasture seeding. Annual data over a period of 15 years, 1946 to 1960, Table 2a, was used. A component analysis is applied and the principal components of these three variables computed.

Results:

Correlation matrix.

could be represented by their first principal component with a high degree of accuracy.

In the regression analysis we seek for an equation which best estimates streamflow as a function of precipitation, number of farm ponds, and the first principal component mentioned above. Standardized variables are used in order to eliminate dimension, thus making the variables uniform in the equation.

Discussion of Results

Case I: Washita Watershed above Clinton

Area: 1977 square miles

Mean annual precipitation during the period 1946 to 1960: 24.09 inches

The first three land practices tested for their intercorrelations are contour farming, cover cropping, and range and pasture seeding. Annual data over a period of 15 years, 1946 to 1960, Table 2a, was used. A component analysis is applied and the principal components of these three variables computed.

Results:

Correlation matrix.

TABLE 2a

Percentage Area of Land Practices in Washita Watershed above Clinton, Oklahoma

	Percent Acreage Contour Farming	Standard-ized	Percent Acreage Cover Cropping	Standard-ized	Percent Acreage Range & Pasture Seeding	Standard-ized	First Principal Component
	x_1	y_1	x_2	y_2	x_3	y_3	z_1
1946	5.48	-2.184	4.00	-1.246	0.56	-1.119	-0.026
1947	7.00	-1.672	5.82	-0.819	1.08	-0.946	-2.010
1948	0.50	-1.167	8.55	-0.178	1.25	-0.889	-0.015
1949	10.00	-0.662	11.20	0.443	1.48	-0.813	-0.914
1950	11.15	-0.275	13.35	0.948	1.60	-0.773	-0.497
1951	12.55	0.196	15.80	1.523	2.26	-0.552	0.114
1952	14.05	0.701	18.45	2.145	2.71	-0.402	0.713
1953	12.55	0.196	3.73	-1.309	3.82	-0.032	-0.162
1954	11.52	-0.150	4.00	-1.246	3.40	-0.172	-0.492
1955	12.60	0.213	8.50	-0.190	3.75	-0.055	0.078
1956	13.75	0.600	10.20	0.209	4.22	0.101	0.548
1957	15.00	1.021	12.40	0.725	4.85	0.312	1.00

TABLE 2a (Continued)

	Percent Acreage Contour Farming	Standard-ized	Percent Acreage Cover Cropping	Standard-ized	Percent Acreage Range & Pasture Seeding	Standard-ized	First Principal Component
	x_1	y_1	x_2	y_2	x_3	y_3	z_1
1958	15.75	1.274	10.33	0.239	7.25	1.112	1.700
1959	15.40	1.156	7.20	-0.495	9.96	2.016	2.050
1960	14.20	0.752	6.12	-0.748	10.55	2.213	1.820
Means	11.97	0	9.31	0	3.92	0	-0.001
Standard Deviation	2.97	1	4.26	1	2.99	1	1.780

TABLE 2b

Runoff, Rainfall, and Number of Ponds in Washita Watershed above Clinton, Oklahoma

	Runoff		Rainfall		Number of Ponds	
	acre-ft per year	Standardized	inches per year	Standardized	Original number	Standardized
1946	62150	-0.5593	23.51	-0.0982	743	-1.6080
1947	144100	0.8597	26.56	0.4217	897	-1.4130
1948	63490	-0.5361	22.79	-0.2209	1041	-1.2310
1949	182000	1.5190	28.56	0.7627	1287	-0.9210
1950	89440	-0.0868	21.90	-0.3726	1425	-0.7466
1951	177900	1.4440	25.39	0.2222	1675	-0.4307
1952	19090	-1.3040	15.95	-1.3860	1850	-0.3596
1953	23550	-1.2270	18.86	-0.8908	2100	0.1062
1954	93280	-0.0203	17.80	-1.0710	2300	0.3589
1955	63650	-0.5333	25.92	0.3126	2326	0.3917
1956	16380	-1.3510	13.96	-1.7260	2545	0.6685
1957	137900	0.7523	30.90	1.1610	2770	0.9528

TABLE 2b (Continued)

	Runoff		Rainfall		Number of Ponds	
	acre-ft per year	Standard-ized	inches per year	Standard-ized	Original number	Standard-ized
1958	34450	-1.0380	22.24	-0.3146	2968	1.2030
1959	168700	1.2850	36.35	2.0900	3083	1.3480
1960	140500	0.7973	30.60	1.1100	3229	1.5320
Means	94452	0	24.09	0	2016	0
Standard Deviations	57750	1	5.87	1	791	1

$$\begin{array}{ll} & \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\ \begin{array}{l} \text{Contour farming} \quad (x_1) \\ \text{Cover Cropping} \quad (x_2) \\ \text{Range and pasture seeding} \quad (x_3) \end{array} & \begin{bmatrix} 1.0000 & 0.3647 & 0.7479 \\ 0.3647 & 1.0000 & -0.1561 \\ 0.7479 & -0.1561 & 1.0000 \end{bmatrix} \end{array}$$

The first principal component is

$$Z_1 = y_1 + 0.2984\ y_2 + 0.8952\ y_3$$

with $\lambda = 1.8$. Hence this accounts for $\frac{1.8}{3}$ X 100% = 60% of the whole variations of x_1, x_2, and x_3.

x_1, x_2, x_3 are percentages of areas used for contour farming, cover cropping, range and pasture seeding respectively.

where y_1, y_2, y_3 are standardized values of x_1, x_2, x_3

λ = characteristic value of the matrix.

The correlation matrix of runoff, precipitation, number of ponds and the first principal component Z_1, of contour farming, cover cropping and range and pasture seeding is:

$$\begin{array}{ll} \begin{array}{l} \text{Runoff} \\ \text{Precipitation} \\ \text{No. of Ponds} \\ Z_1 \end{array} & \begin{bmatrix} 1.0000 & 0.7943 & -0.0111 & -0.1254 \\ 0.7943 & 1.0000 & 0.2227 & 0.2294 \\ -0.0111 & 0.2227 & 1.0000 & 0.8307 \\ -0.1254 & 0.2294 & 0.8307= & 1.0000 \end{bmatrix} \end{array}$$

This result is in accordance to what one would expect. The correlation between runoff and precipitation is 0.7942, positive. This means that increased rainfall results in increased runoff. The correlation between runoff and number of farm ponds is -0.0111, negative. It means that increased number of farm ponds results in less runoff when other conditions remain constant. This is to be expected.

Using precipitation, number of ponds and Z_1 as the independent variables we then seek for the best fit equation for the prediction of runoff as a function of the three independent ariables. Multiple linear regression analysis was performed on the data of Tables 2a and 2b. For computation, the Doolittle procedure was used. Several equations are tried and the results are shown below:

	Equation	R^2
1.	$Y = 0.876 + 0.9551 X_1 + 0.0429 X_2 - 0.3303 X_3 + 0.1166 X_1^2 - 0.2370 X_2^2 + 0.0186 X_3^2$	0.7819
2.	$Y = 0.0937 + 0.9630 X_1 + 0.0777 X_2 - 0.3628 X_3 + 0.1213 X_1^2 - 0.2152 X_2^2$	0.7814
3.	$Y = - 0.1279 + 0.8500 X_1 + 0.1749 X_2 - 0.4198 X_3 + 0.1278 X_1^2$	0.7600

4. $$Y = -\ 0.0002 + 0.8615\ X_1 + 0.2110\ X_2 - 0.3734\ X_3$$ 0.7446

Using R, the correlation of multiple regression as the criteria of testing the goodness of fit, it was found the following equation:

$$Y = 0.0876 + 0.9551\ X_1 + 0.0429\ X_2 - 0.3303\ X_3 + 0.1166\ X_1^2 - 0.2370\ X_2^2 + 0.0186\ X_3^2$$

is the best fit. The equation gives results which is in accordance to what a rational judgment would expect. Thus it is justifiable to use it as our model for this watershed. The closeness of the values predicted from this equation to the observed values is shown in Figure 2d.

We then perform the so-called "Sensitivity Analysis" on the independent variable. All except one of the variables in the equation is allowed to vary through out its full possible range, and the effects on the predicted runoff values are noted. Results are shown in Figures 2a, 2b, and 2c.

Relative Effect of Rainfall
on Runoff

Washita Watershed above
Clinton, Oklahoma

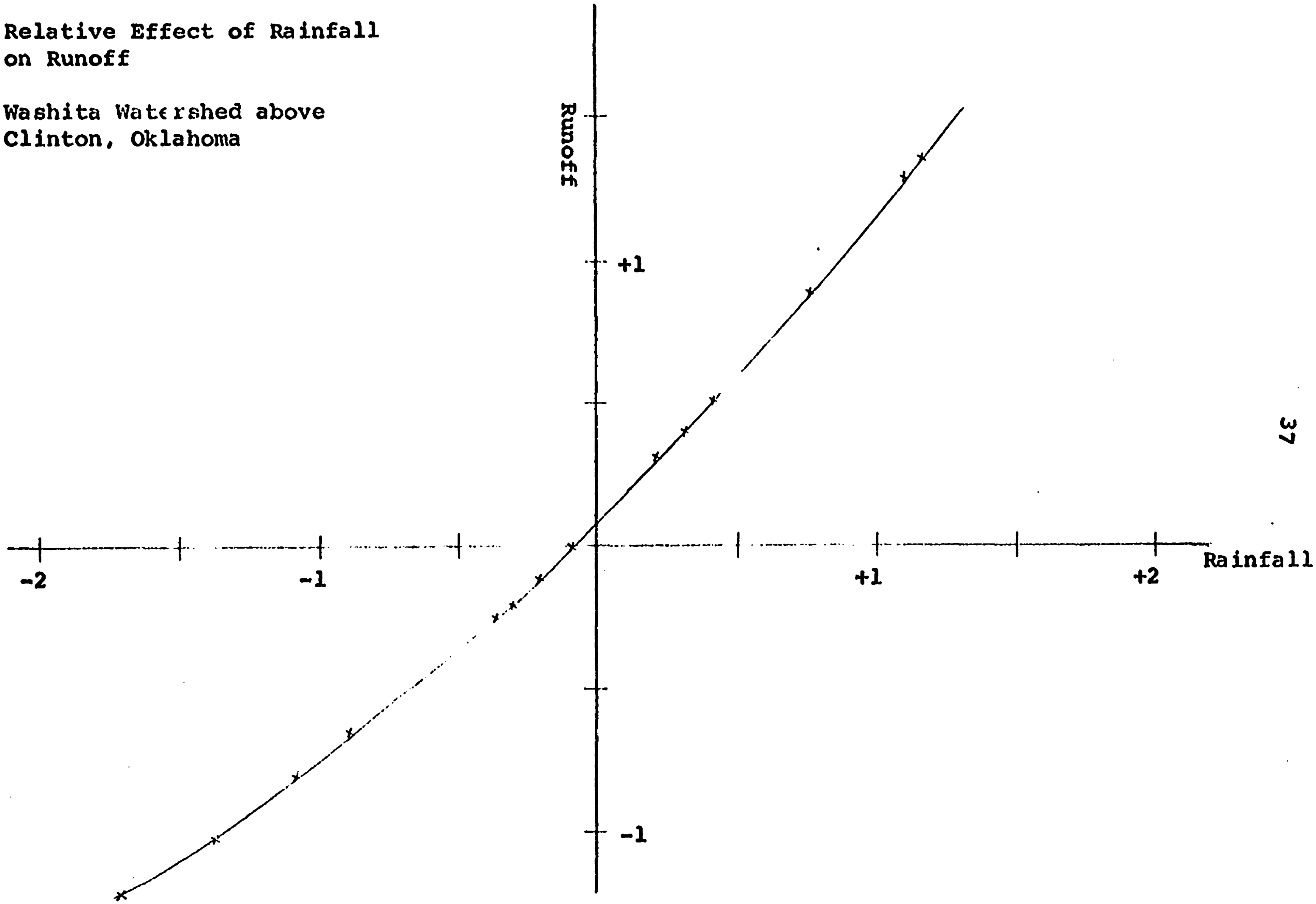
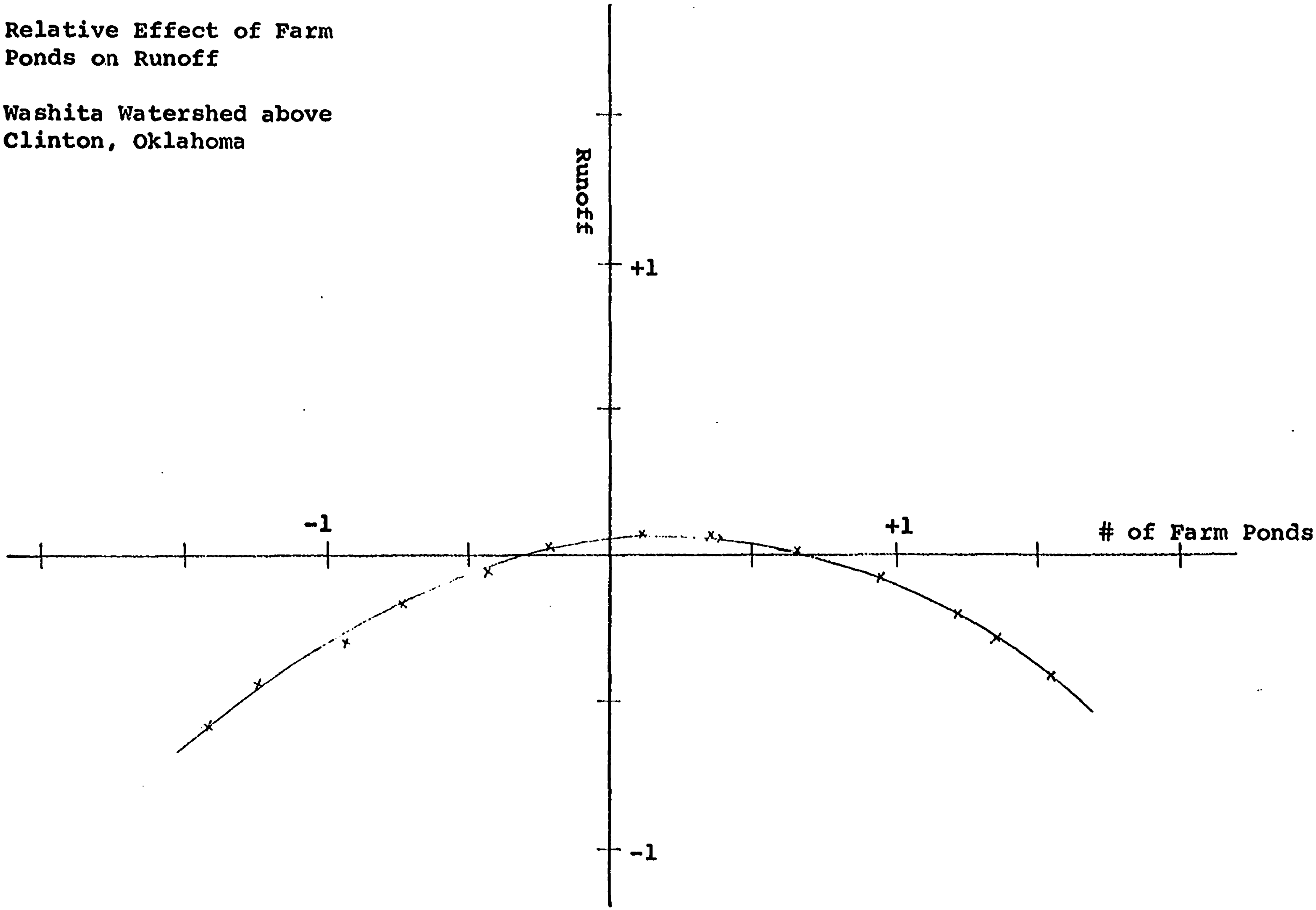


Figure 2a



Relative Effect of Farm Ponds on Runoff
Washita Watershed above Clinton, Oklahoma
Runoff
+1
-1
-1
+1
# of Farm Ponds

Figure 2b

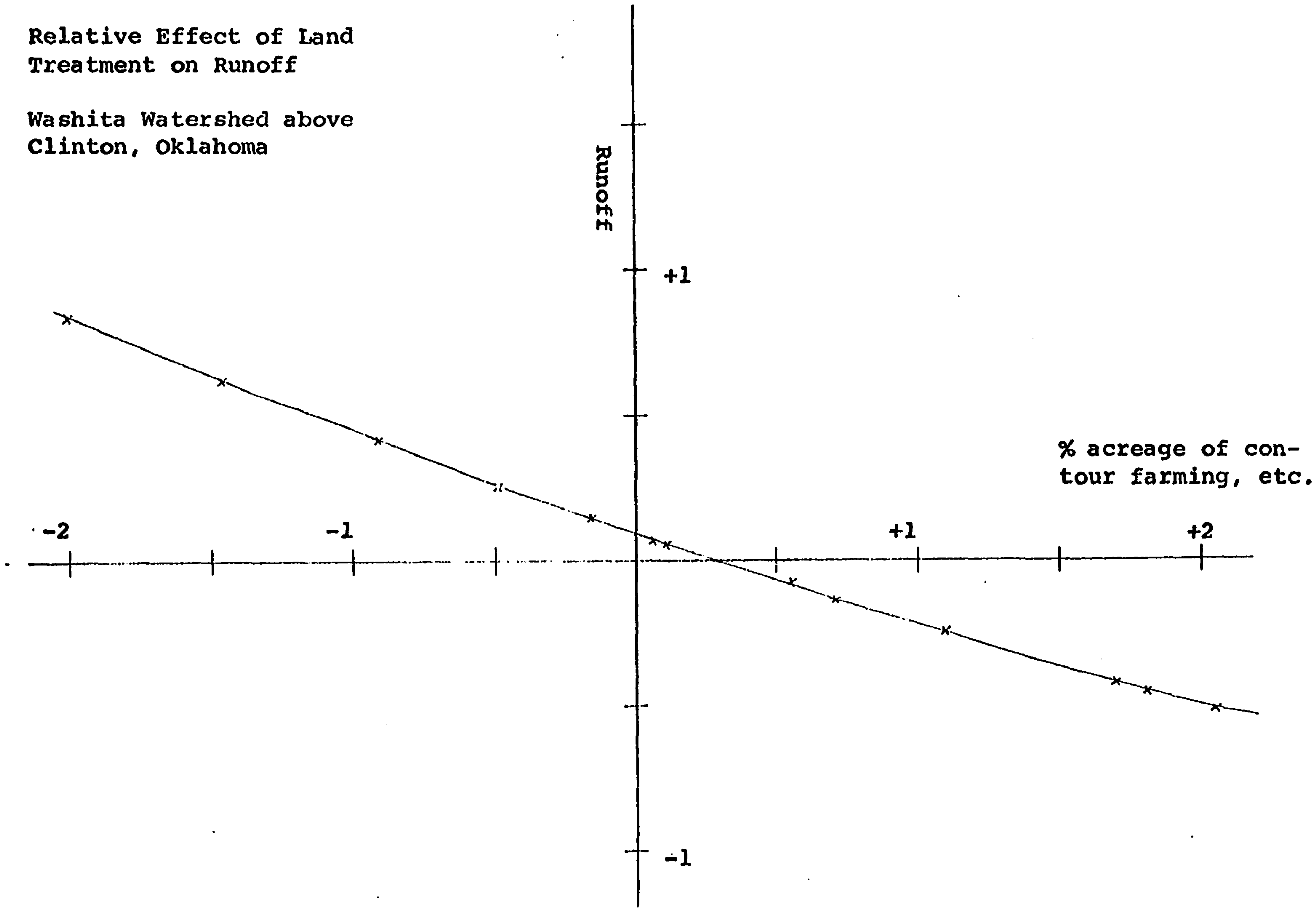

Relative Effect of Land Treatment on Runoff
Washita Watershed above Clinton, Oklahoma
Runoff
+1
-1
% acreage of contour farming, etc.
-2
-1
+1
+2

Figure 2c

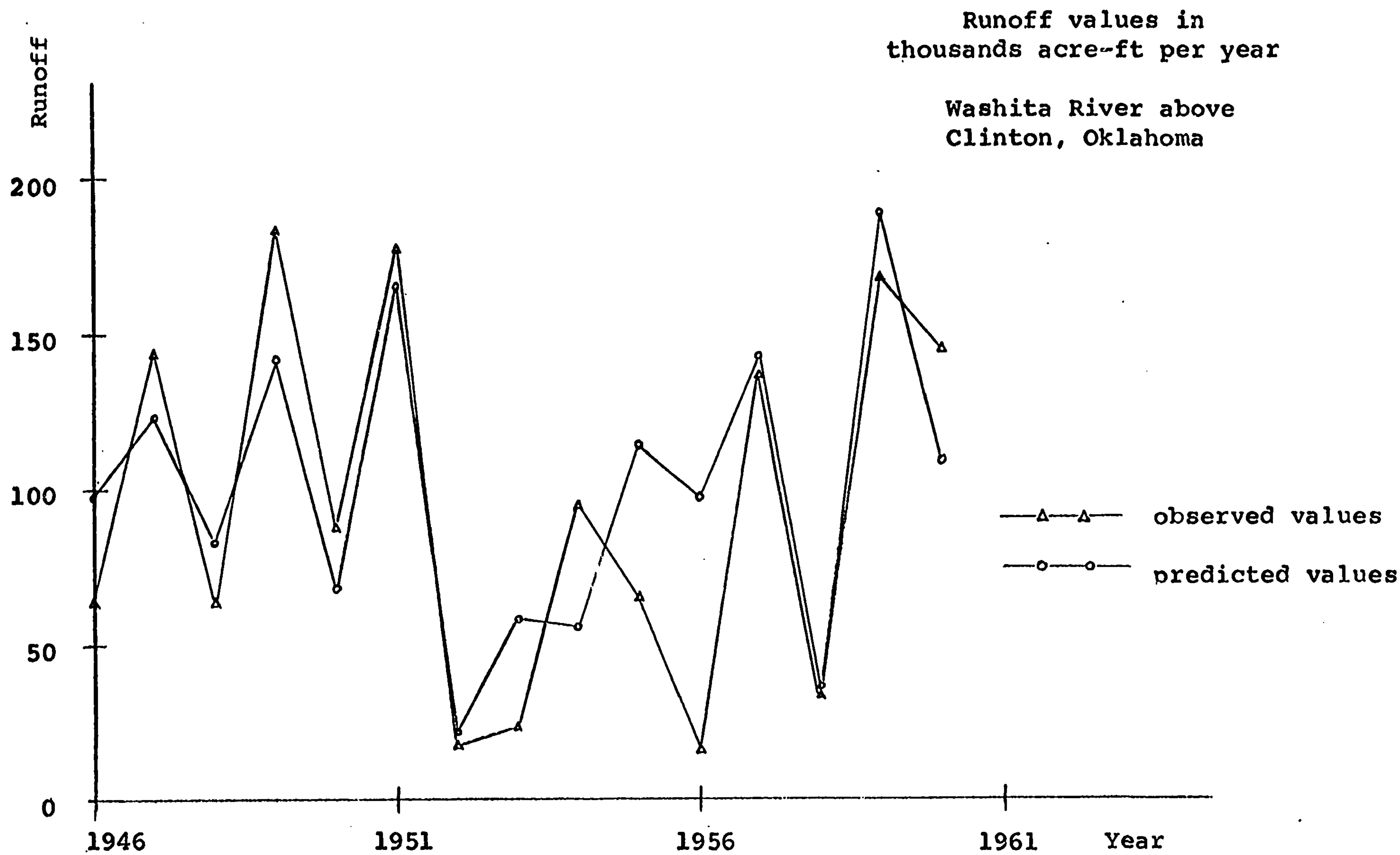

Runoff values in
thousands acre-ft per year
Washita River above
Clinton, Oklahoma
Runoff
200
150
100
50
0
1946
1951
1956
1961
Year
observed values
predicted values

Figure 2d

Case II: Black Bear Creek above Pawnee

Area: 576 square miles

Mean annual precipitation during the period

1946 to 1960: 29.642 inches

The first four land practices tested for their inter-correlations are contour farming, cover cropping, range and pasture seeding, and crop residue management.

As in Case I, data over a period of 15 years, Table 2a, was used, and similar analysis applied.

Results:

Correlation matrix.

$$\begin{array}{ll} & \begin{array}{cccc} x_1 & x_2 & x_3 & x_4 \end{array} \\ \begin{array}{ll} \text{Contour farming} & (x_1) \\ \text{Cover cropping} & (x_2) \\ \text{Range \& pasture seeding} & (x_3) \\ \text{Crop residue management} & (x_4) \end{array} & \begin{bmatrix} 1.0000 & 0.6185 & 0.8851 & 0.7931 \\ 0.6185 & 1.0000 & 0.2480 & 0.1314 \\ 0.8851 & 0.2480 & 1.0000 & 0.9761 \\ 0.7931 & 0.1314 & 0.9761 & 1.0000 \end{bmatrix} \end{array}$$

The first principal component is:

$$Z_1 = Y_1 + 0.5091\ Y_2 + 0.9942\ Y_3 + 0.9425\ Y_4$$

with $\lambda = 2.945$.

Hence this accounts for $\frac{2.945}{4} \times 100\% - 74\%$ of the whole variations of x_1, x_2, x_3, and x_4, where Y_1, Y_2, Y_3, Y_4 are standardized values of x_1, x_2, x_3, x_4.

λ = characteristic value of the matrix.

TABLE 3a

Percentage Area of Land Practices in Black Bear Watershed above Pawnee, Oklahoma

	Percent Acreage Contour Farming	Standard -ized	Percent Acreage Cover Cropping	Standard -ized	Percent Acreage Range & Pasture Seeding	Standard -ized	Percent Acreage Crop Residue Management	Standard -ized	First Principal Component
	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4	z_1
1946	0.75	-1.740	0.42	-1.815	0.11	-1.285	1.29	-1.290	-5.160
1947	1.27	-1.508	0.88	-1.424	0.25	-1.156	2.90	-1.046	-4.370
1948	1.81	-1.263	1.22	-1.127	0.38	-1.047	3.69	-0.927	-3.752
1949	2.48	-0.962	1.77	-0.649	0.55	-0.899	4.87	-0.747	-2.891
1950	3.15	-0.665	2.48	-0.041	0.77	-0.704	6.34	-0.524	-1.880
1951	3.78	-0.381	3.46	0.808	1.02	-0.481	7.27	-0.383	-0.809
1952	4.63	-0.002	4.28	1.520	1.25	-0.282	8.78	-0.155	0.350
1953	4.68	0.025	2.93	0.354	0.90	-0.591	6.59	-0.487	-0.841
1954	5.57	0.422	3.46	0.810	1.74	0.145	8.45	-0.204	0.787
1955	6.23	0.720	3.65	0.972	1.93	0.315	9.52	-0.041	1.490
1956	7.04	1.080	3.73	1.040	2.10	0.461	10.90	0.167	2.227
1957	7.65	1.358	3.76	1.072	2.38	0.705	12.24	0.371	2.956

TABLE 3a (Continued)

	Percent Acreage Contour Farming	Standard -ized	Percent Acreage Cover Cropping	Standard -ized	Percent Acreage Range & Pasture Seeding	Standard -ized	Percent Acreage Crop Residue Manage-ment	Standard -ized	First Principal Component
	x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4	z_1
1958	6.85	0.996	2.43	-0.081	2.91	1.170	17.07	1.104	3.159
1959	6.42	0.806	1.60	-0.802	3.46	1.621	22.71	1.960	3.857
1960	7.10	1.111	1.79	-0.636	3.89	2.030	24.33	2.206	4.887

TABLE 3b

Runoff, Rainfall, and Number of Ponds in Black Bear Watershed above Pawnee, Oklahoma

	Runoff		Rainfall		Number of Ponds	
	acre-ft per year	Standard-ized	inches per year	Standard-ized	Original number	Standard-ized
1946	84490	-0.4381	26.42	-0.3773	17	-1.2210
1947	136800	0.0660	29.54	-0.0119	31	-1.1610
1948	80090	-0.4805	21.30	-0.9768	45	-1.1010
1949	138800	0.0853	32.57	0.3429	62	-1.0290
1950	81520	-0.4667	27.21	-0.2848	85	-0.9313
1951	145200	0.1469	36.88	0.8475	121	-0.7777
1952	66360	-0.6128	21.96	-0.8995	190	-0.4833
1953	38460	-0.8816	27.05	-0.3035	263	-0.1719
1954	12020	-1.1364	17.23	-1.4534	369	0.2802
1955	156900	0.2597	31.01	0.1602	424	0.5149
1956	2900	-1.2243	16.37	-1.5541	492	0.8050
1957	358200	2.1995	49.12	2.2808	539	1.0010
1958	59230	-0.6815	29.51	-0.0155	592	1.2310
1959	288300	1.5259	38.61	1.0501	652	1.4870
1960	300000	1.6386	39.85	1.1953	669	1.5600
Means	129551	0	29.64	0	303	0
Standard Deviations	103774	1	8.54	1	234	1

The correlation matrix of runoff, precipitation, number of ponds and the first principal component Z_1 of contour farming, cover cropping, range and pasture seeding, and crop residue management is:

$$
\begin{array}{l}
\text{Runoff} \\
\text{Precipitation} \\
\text{No. of Ponds} \\
Z_1
\end{array}
\begin{bmatrix}
1.0000 & 0.9220 & 0.4457 & 0.4349 \\
0.9220 & 1.0000 & 0.3579 & 0.3636 \\
0.4457 & 0.3579 & 1.0000 & 0.9616 \\
0.4349 & 0.3636 & 0.9616 & 1.0000
\end{bmatrix}
$$

This result is not in accordance to what one would expect. The correlation between runoff and precipitation is 0.9220, positive, which is good. But the correlation of runoff with the number of ponds and Z_1 is positive, which is not in accordance with rational judgment. Such a case could be explained by the fact that there are errors in the data and that there are masked fluctuations in streamflow.

We try to level off the effect of these inconsistencies by seeking for the regression equations which best describes the relationships. Several equations are tried and the result shown below:

Equation		R^2
1.	$Y = 0.8719\ X_1 + 0.2684\ X_2 - 0.0468\ X_3$	0.867
2.	$Y = -\ 0.0997 + 0.8469\ X_1 + 0.2523\ X_2 - 0.0590\ X_3 + 0.0984\ X_1^2$	0.880

Equation		R^2
3.	$Y = 0.5122 + 0.7217\,X_1 - 0.1511\,X_2 + 0.0558\,X_3 + 0.1481\,X_1^2 + 0.3621\,X_2^2$	0.915
4.	$Y = -\,0.4692 + 0.7323\,X_1 - 0.3812\,X_2 + 0.1581\,X_3 + 0.1274\,X_1^2 + 0.0716\,X_2^2 + 0.0291\,X_3^2$	0.921
5.	$\ln Y = -\,0.2502 + 0.3430\ln X_1 + 0.6801\,X_2 + 0.0529\,X_3$	0.4153

To describe the effects of land practices on runoff, equation 2 gives the best fit. It's R^2-value is 0.89 which is very reasonable. This equation also best describes the relationships of precipitation and streamflow. To describe the effects of farm ponds on streamflow, equation 4 is chosen. The closeness of the predicted values to the observed values is shown in Figure 3d.

Results of the "Sensitivity Analysis" are shown in Figures 3a, 3b, and 3c.

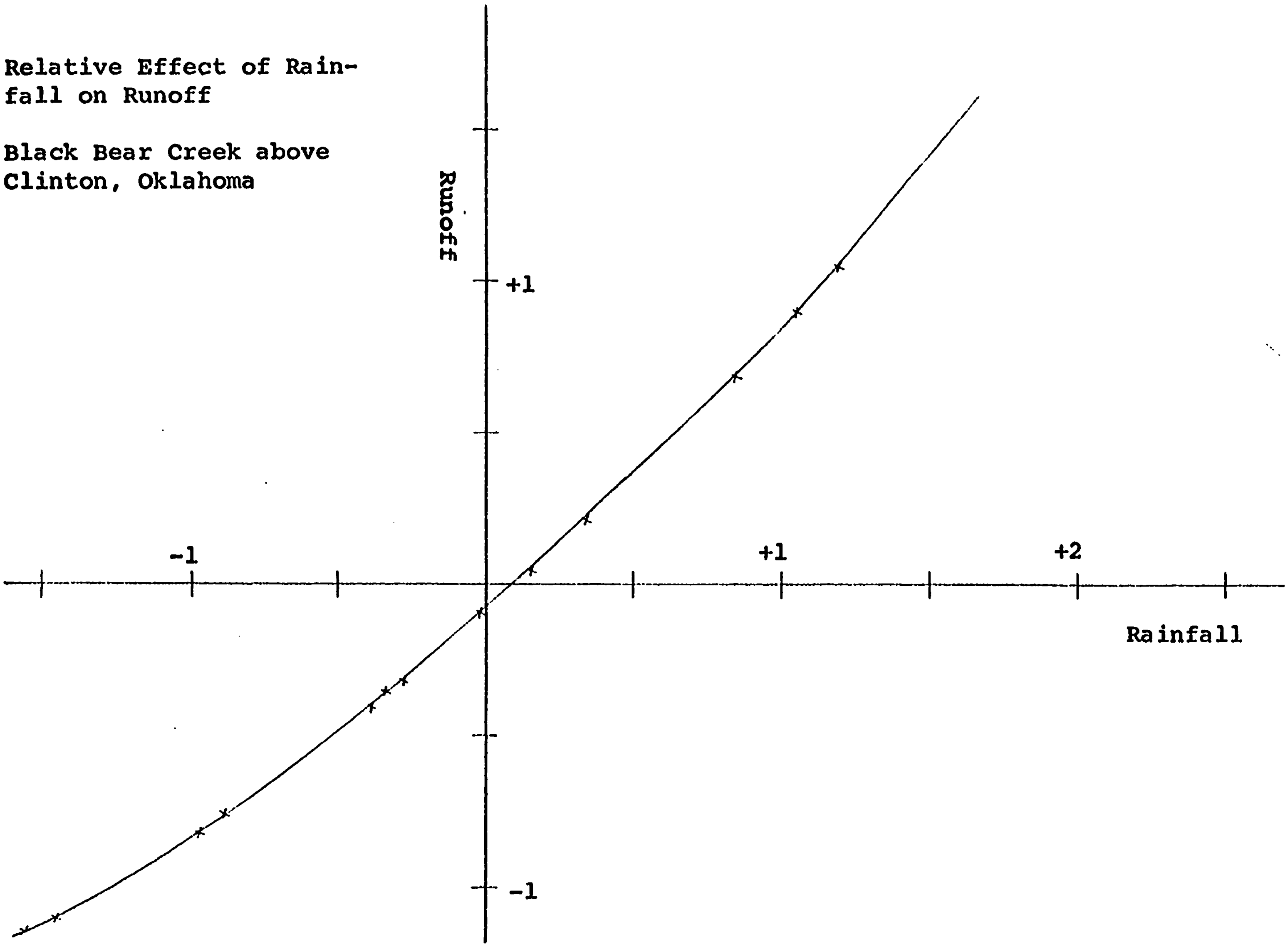

Relative Effect of Rainfall on Runoff
Black Bear Creek above Clinton, Oklahoma
Runoff
+1
-1
-1
+1
+2
Rainfall

Figure 3a

Relative Effect of Farm Ponds on Runoff

Black Bear Creek above Pawnee, Oklahoma

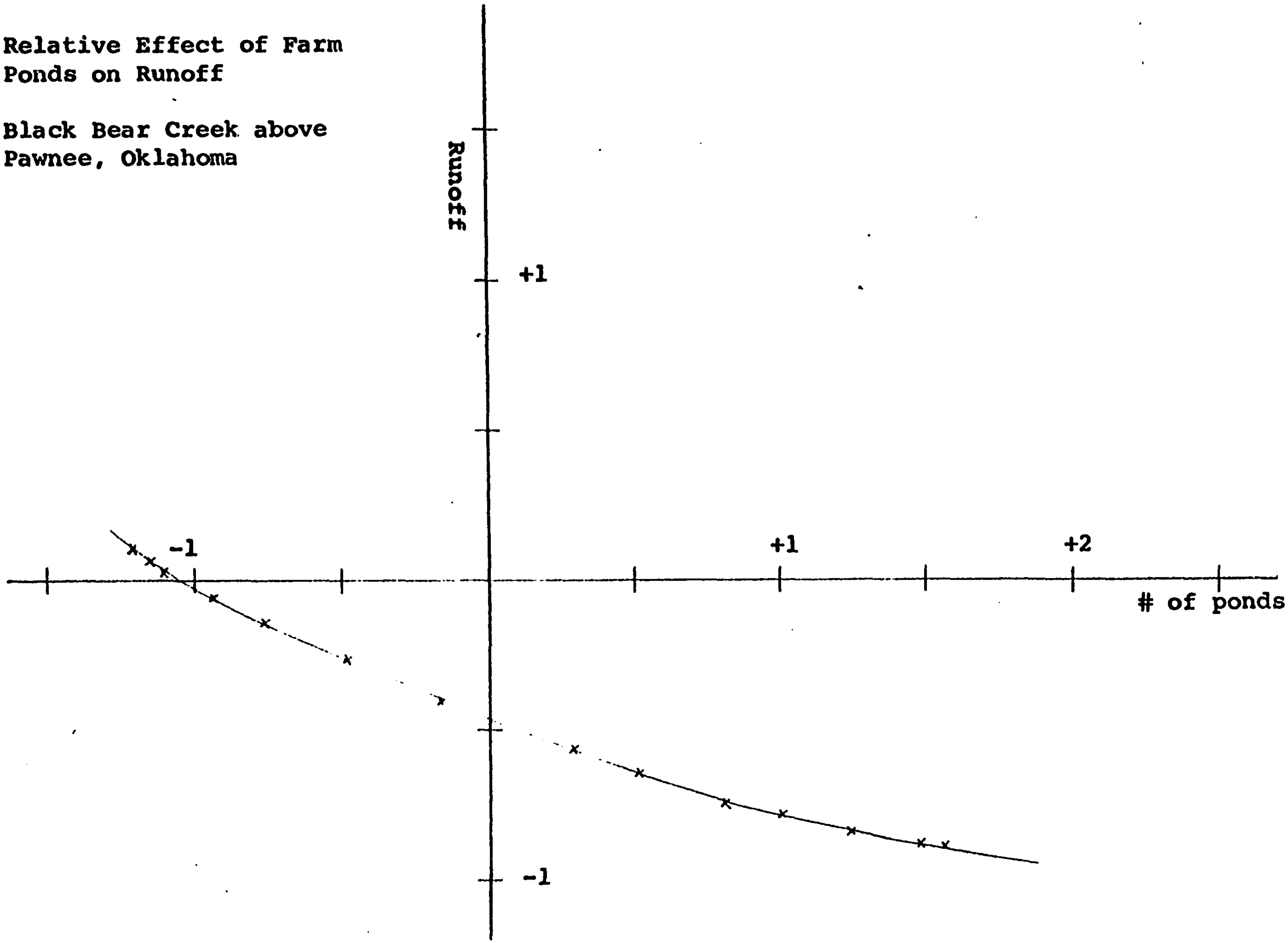


Figure 3b

Relative Effect of Land Treatment on Runoff

Black Bear Creek above Pawnee

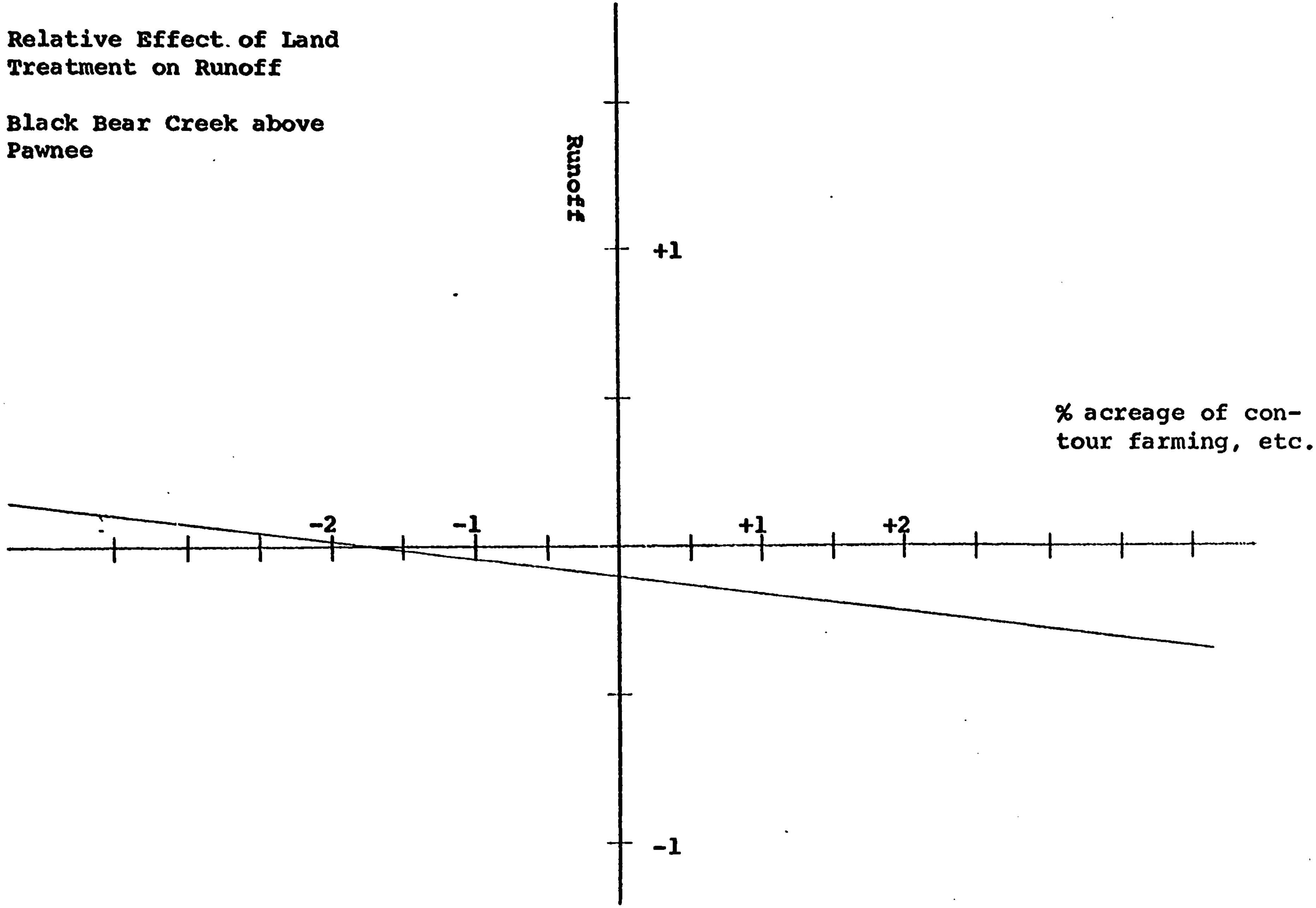


Figure 3c

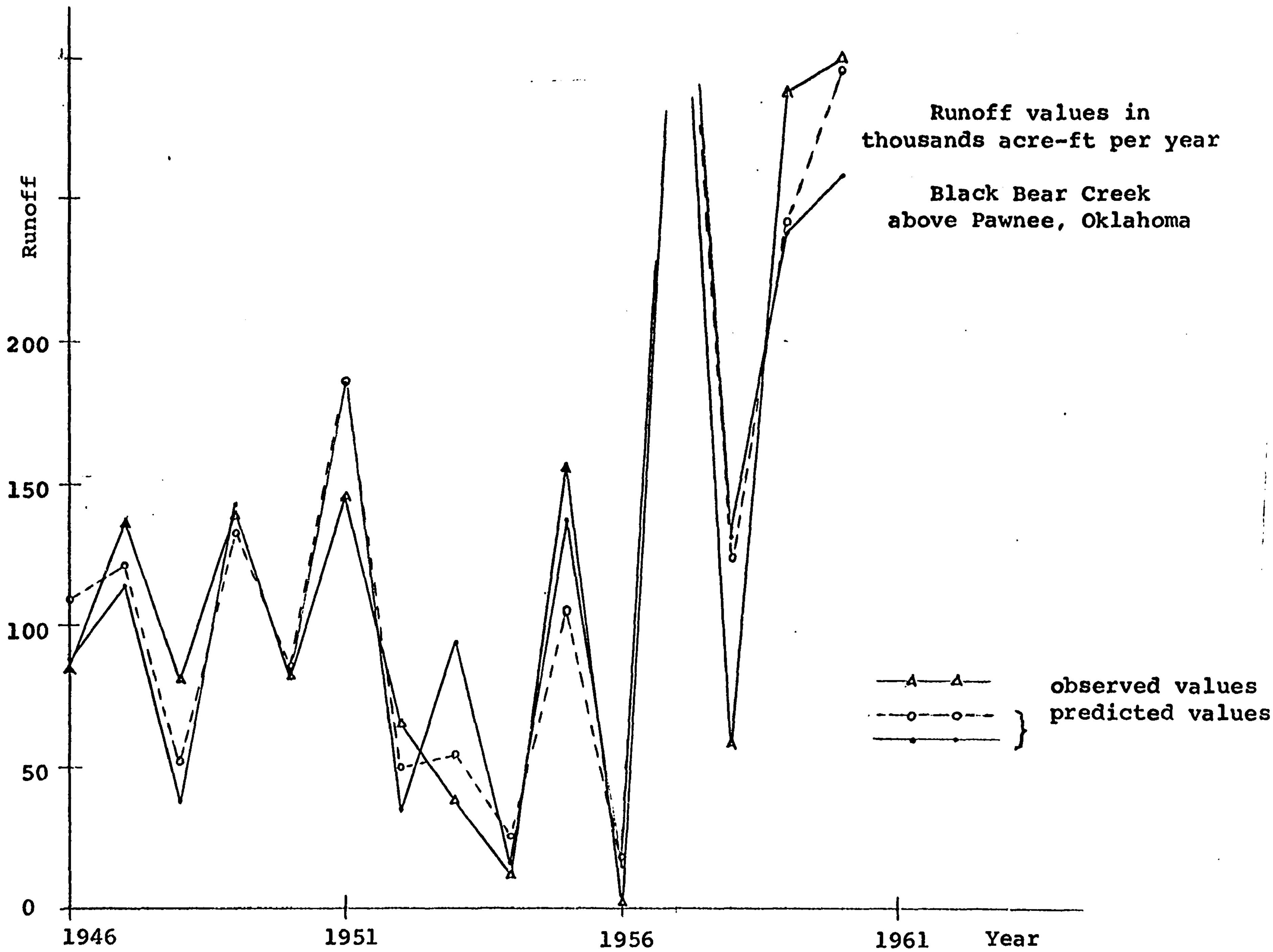

Runoff values in
thousands acre-ft per year
Black Bear Creek
above Pawnee, Oklahoma
Runoff
200
150
100
50
0
1946
1951
1956
1961
Year
observed values
predicted values

Figure 3d

Case III: Clear Boggy Creek above Caney

Area: 720 square miles

Mean annual precipitation during the period

1946 to 1960: 39.66 inches.

The first three land practices tested for their intercorrelations are contour farming, cover cropping, and crop residue management.

As in Case I and II, data over a period of 15 years, Table 4a, was used, and similar analysis applied.

Results:

Correlation matrix.

$$
\begin{array}{lc}
 & \begin{array}{ccc} x_1 & x_2 & x_3 \end{array} \\
\begin{array}{ll}
\text{Contour farming} & (x_1) \\
\text{Cover cropping} & (x_2) \\
\text{Crop residue management} & (x_3)
\end{array} &
\begin{bmatrix}
1.0000 & 0.4844 & 0.7986 \\
0.4844 & 1.0000 & 0.8760 \\
0.7986 & 0.8760 & 1.0000
\end{bmatrix}
\end{array}
$$

The first principal component is

$$Z_1 = 0.8458\ Y_1 + 0.8858\ Y_2 + Y_3$$

with $\lambda = 2.4514$. Hence this accounts for $\frac{2.4514}{3} \times 100\% =$ 82% of the whole variations of x_1, x_2, and x_3 where y_1, y_2, and y_3 are standardized values of x_1, x_2, and x_3.

λ = characteristic value of the matrix.

The correlation matrix of runoff, precipitation, number of ponds and the first principal component Z_1, of contour

TABLE 4a

Percentage Area of Land Practices in Clear Boggy Creek Watershed above Caney, Oklahoma

	Percent Acreage Contour Farming	Standard-ized	Percent Acreage Cover Cropping	Standard-ized	Percent Acreage Crop Residue Manage-ment	Standard-ized	First Principal Component
	x_1	y_1	x_2	y_2	x_3	y_3	z_1
1946	2.34	-1.5255	1.34	-1.9134	2.34	-2.4932	-5.4784
1947	2.79	-0.4490	1.87	-1.0224	3.06	-0.7206	-2.0059
1948	3.09	0.2562	2.29	-0.3264	3.27	-0.2062	-0.2786
1949	3.24	0.6142	2.45	-0.0627	3.24	0.1751	0.6389
1950	3.50	1.2387	2.60	0.1916	3.61	0.6314	1.8488
1951	3.56	1.3723	2.62	0.2189	3.70	0.8551	2.2096
1952	3.72	1.7554	2.82	0.5603	3.95	1.4685	3.4495
1953	2.84	-0.3396	1.77	-1.1904	3.05	-0.7397	-2.0814
1954	2.94	-0.1049	2.12	-0.6037	3.16	-0.4745	-1.0980
1955	3.06	0.1781	2.79	0.5045	3.56	0.5201	1.1176
1956	3.09	0.2690	3.29	1.3326	3.66	0.7651	2.1730
1957	3.16	0.4246	3.84	2.2416	3.94	1.4472	3.7919
1958	2.73	-0.6041	2.95	0.7776	3.42	0.1676	0.3454
1959	2.33	-1.5491	2.14	-0.5714	2.94	-1.0038	-2.8202
1960	2.33	-1.5363	2.40	-0.1361	3.19	-0.3915	-1.8114

TABLE 4b

Runoff, Rainfall, and Number of Ponds in Clear Boggy Creek Watershed above Caney, Oklahoma

	Runoff		Rainfall		Number of Ponds	
	acre-ft per year	Standard-ized	inches per year	Standard-ized	Original number	Standard-ized
1946	604600	1.2698	47.89	0.8839	219	-1.4366
1947	289100	-0.1472	38.62	-0.1117	323	-1.2170
1948	343700	0.0980	34.70	-0.5327	386	-1.0240
1949	295000	-0.1207	43.05	0.3641	451	-0.9467
1950	561700	1.0771	44.48	0.5177	494	-0.8559
1951	207500	-0.5137	32.65	-0.7529	607	-0.6173
1952	128100	-0.8703	27.30	-1.3276	758	-0.2985
1953	302000	-0.0892	43.82	0.4468	847	-0.1106
1954	213500	-0.4867	30.58	-0.9752	966	0.1406
1955	138900	-0.8218	37.17	-0.2674	1090	0.4024
1956	35210	-1.2875	25.10	-1.5639	1241	0.7212
1957	961400	2.8723	62.70	2.4747	1324	0.8965
1958	268600	-0.2393	34.32	-0.5735	1440	1.1414
1959	238900	-0.3727	47.87	0.8818	1615	1.5109
1960	240000	-0.3677	44.67	0.5381	1730	1.7538

farming, cover cropping and crop residue management is:

$$\begin{array}{l} \text{Runoff} \\ \text{Precipitation} \\ \text{Number of Ponds} \\ Z_1 \end{array} \begin{bmatrix} 1.0000 & 0.8354 & -0.1470 & 0.0166 \\ 0.8354 & 1.0000 & 0.1290 & -1.1980 \\ -0.1470 & 0.1290 & 1.0000 & 0.1206 \\ 0.0166 & -0.1980 & 0.1206 & 1.000 \end{bmatrix}$$

As in the other two cases, the correlation between runoff and precipitation is positive, between runoff and number of ponds negative as one would expect, and very small between runoff and Z_1. Results of the regression analysis are shown below:

	Equation	R^2
1.	$Y = -0.3143 + 0.7197 X_1 - 0.3790 X_2 + 0.0187 X_3 + 0.2524 X_1^2 + 0.0701 X_2^2 - 0.0013 X_3^2$	0.8795
2.	$Y = -0.3145 + 0.7205 X_1 - 0.3746 X_2 + 0.0211 X_3 + 0.2467 X_1^2 + 0.0679 X_2^2$	0.8795
3.	$Y = -0.2410 + 0.7399 X_1 - 0.3543 X_2 + 0.0096 X_3 + 0.2410 X_2$	0.8778
4.	$Y = -0.0001 + 0.9199 X_1 - 0.2940 X_2 + 0.946 X_3$	0.8444

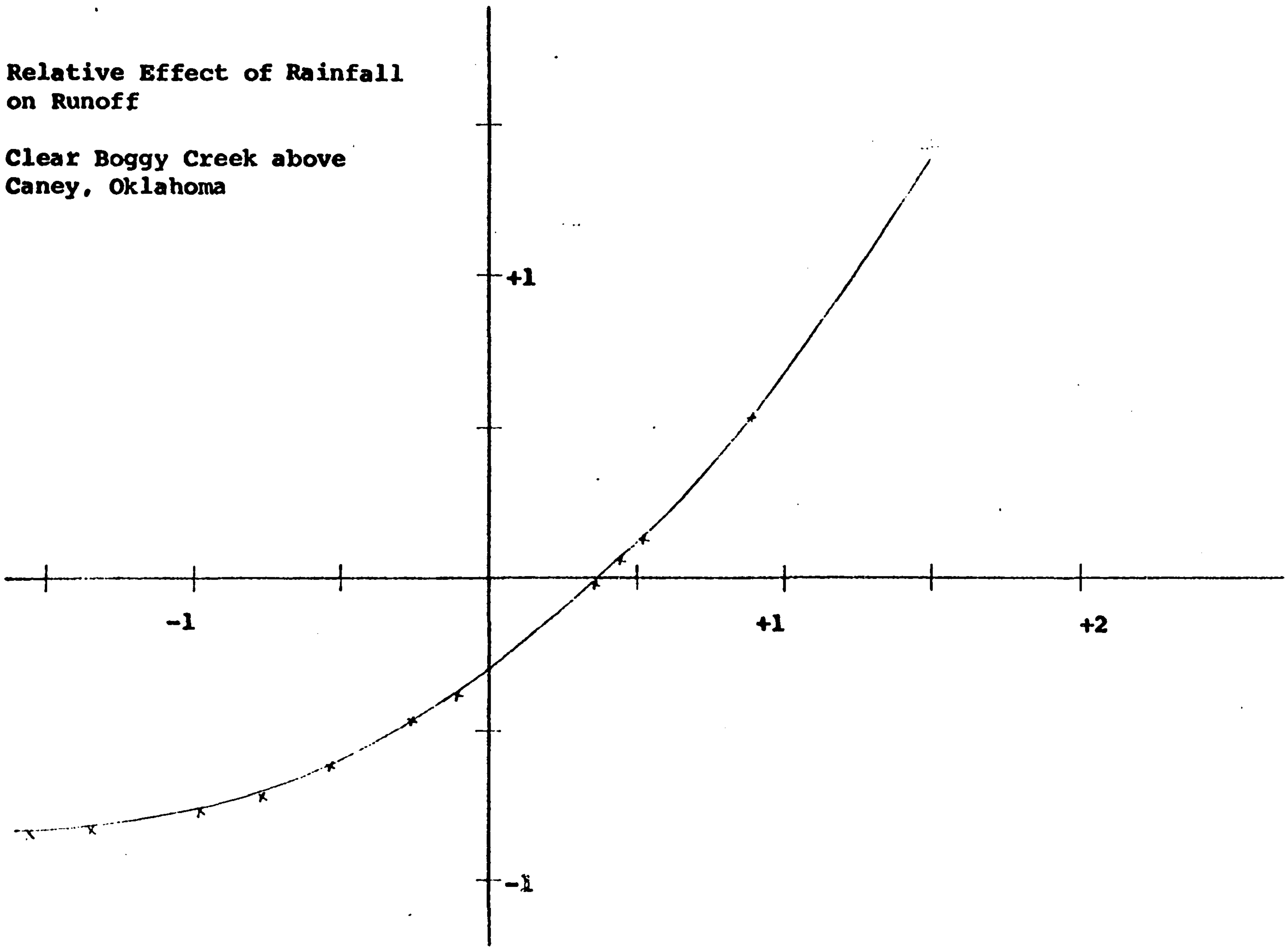

Relative Effect of Rainfall
on Runoff
Clear Boggy Creek above
Caney, Oklahoma
+1
-1
-1
+1
+2

Figure 4a

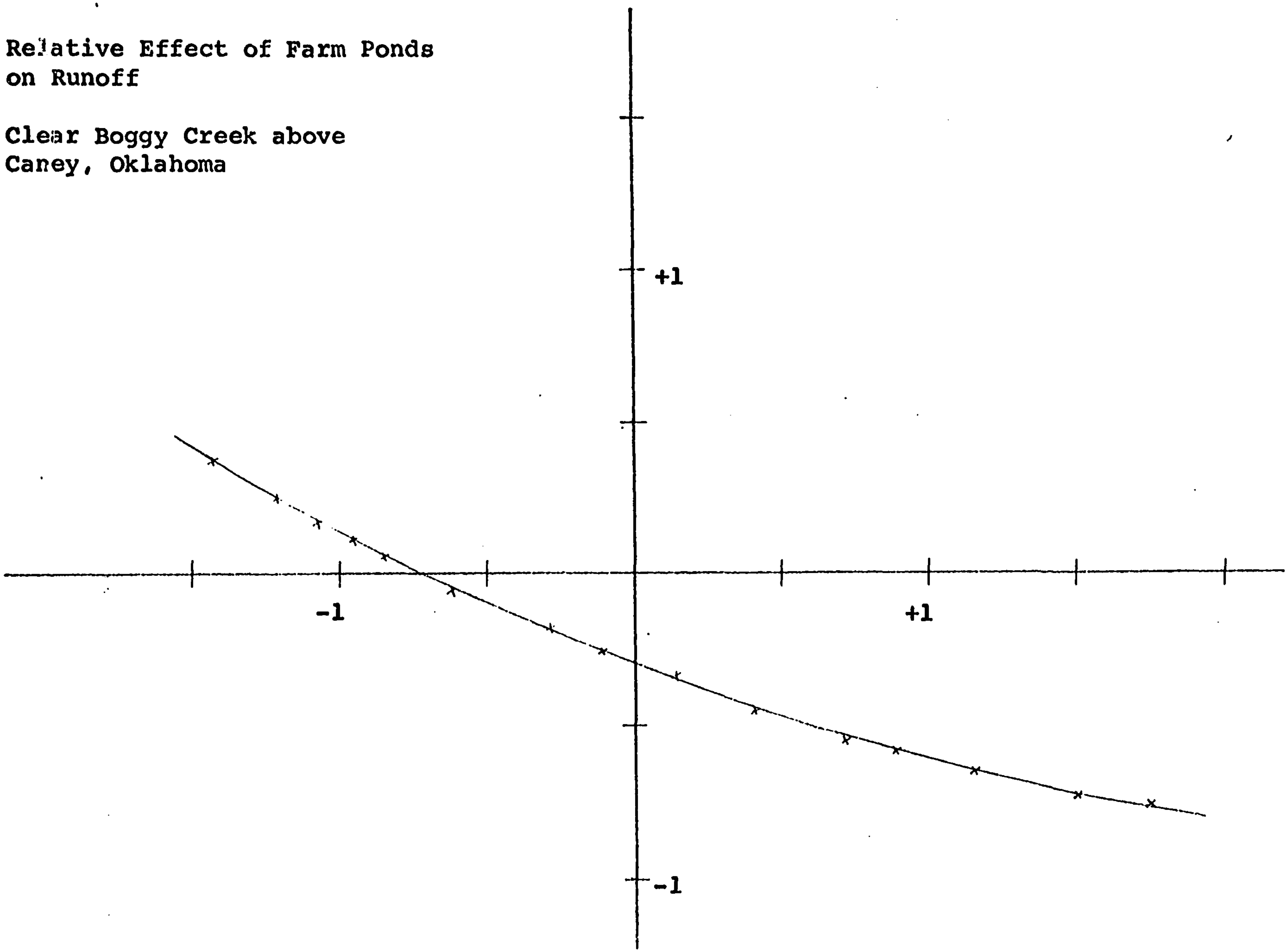

Relative Effect of Farm Ponds
on Runoff
Clear Boggy Creek above
Caney, Oklahoma
+1
-1
-1
+1

Figure 4b

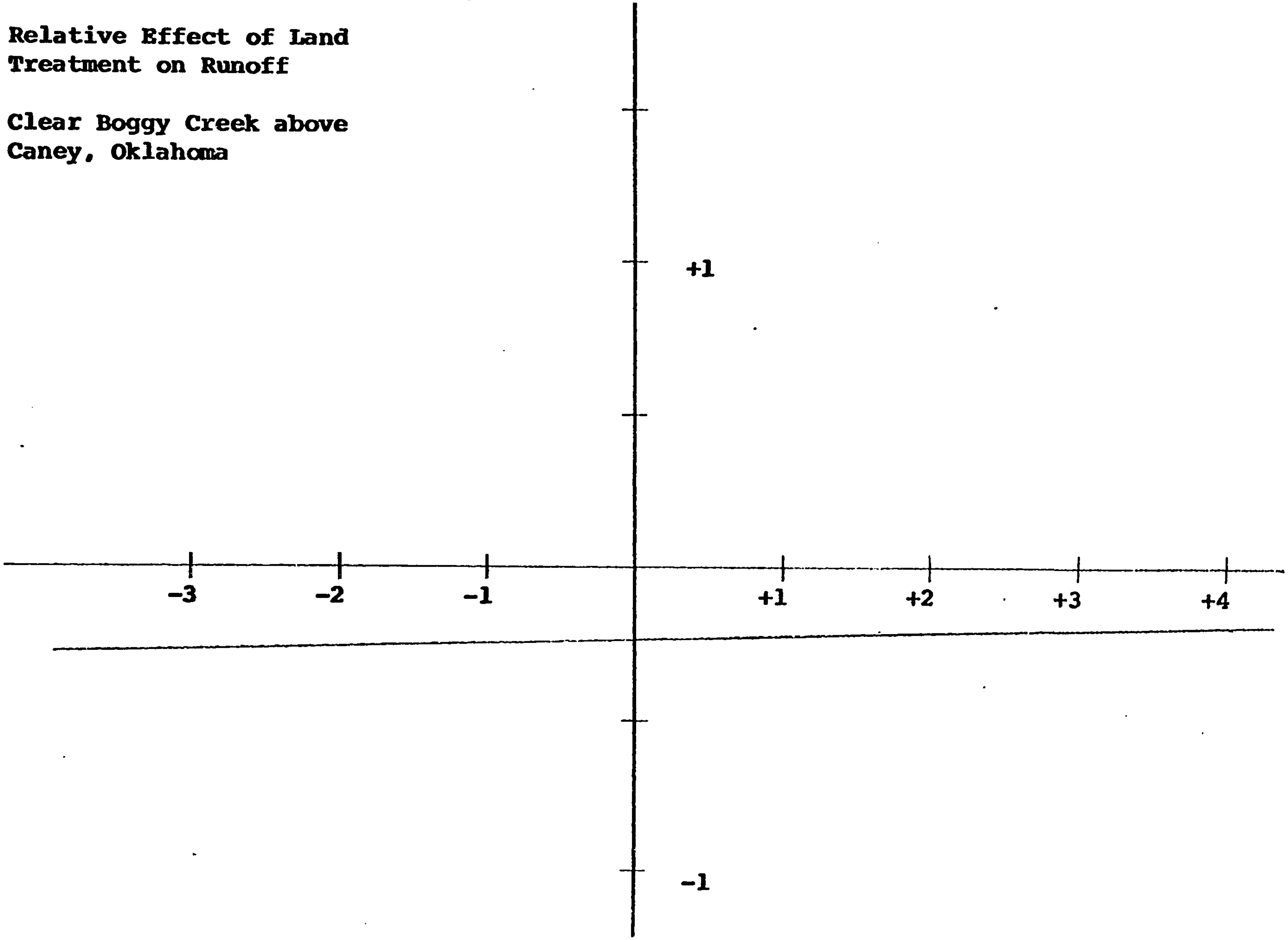

Relative Effect of Land Treatment on Runoff
Clear Boggy Creek above Caney, Oklahoma
+1
-3
-2
-1
+1
+2
+3
+4
-1

Figure 4c

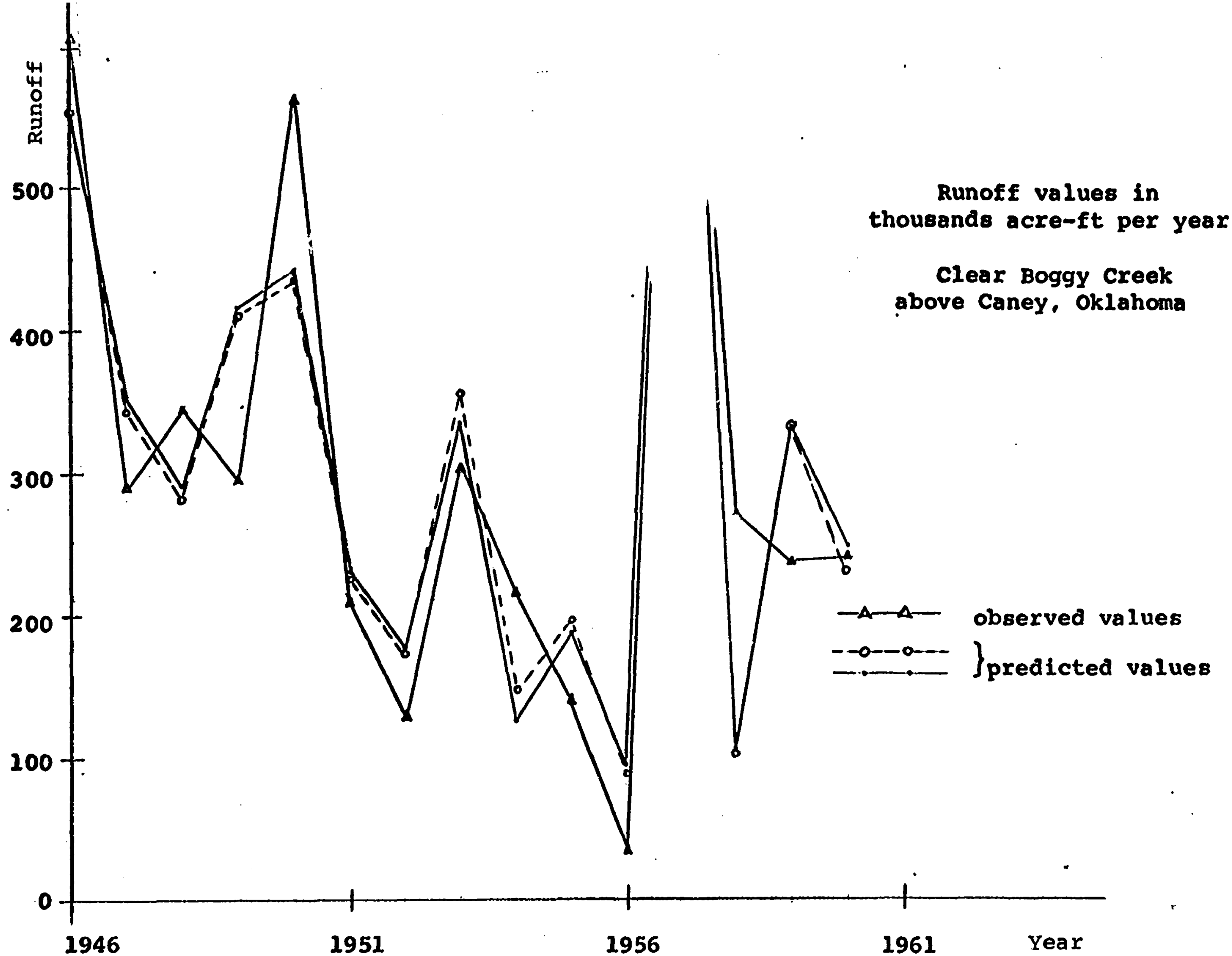

Runoff
500
400
300
200
100
0
1946
1951
1956
1961
Year
Runoff values in thousands acre-ft per year
Clear Boggy Creek above Caney, Oklahoma
observed values
predicted values

Figure 4d

Equation		R^2
5.	$Y = -\ 0.1617 + 0.8413\ X_1 - 0.2473\ X_2 + 0.1110\ X_3 + 0.0264\ X_3^{\ 3}$	0.8444
6.	$Y = -\ 0.2421 + 0.7400\ X_1 - 0.3528\ X_2 + 0.0107\ X_3 + 0.2388\ X_1^{\ 2} + 0.0005\ X_3^{\ 2}$	0.8778

To estimate the effect of farm ponds on runoff, equation 1 gives the best result. This equation also best describes the relationship of precipitation and runoff. In the case of agricultural practices, equation 3 gives the best description. The closeness of the predicted values and the observed values is shown in Figure 4d.

Using these equations as our model a sensitivity analysis is performed, and the results shown in Figures 4a, 4b, and 4c.

Identification of the First Principal Component

A correlation matrix of the agricultural practices and their principal component is computed for each case. It was hoped that from these matrices the first principal component in each case could be identified in terms of the agricultural practices.

Results:

<u>Case I</u>

$$\begin{array}{l} \text{Contour farming} \\ \text{Cover cropping} \\ \text{Seeding} \\ Z_1 \end{array} \begin{bmatrix} 1.0000 & 0.3646 & 0.7479 & 0.9697 \\ 0.3646 & 1.0000 & -0.1561 & 0.2854 \\ 0.7479 & -0.1561 & 1.0000 & 0.8703 \\ 0.9697 & 0.2854 & 0.8703 & 1.0000 \end{bmatrix}$$

Correlation between contour farming and Z_1 is very high, 0.9697. Hence, Z_1 could be best identified as contour farming.

<u>Case II</u>

$$\begin{array}{l} \text{Contour farming} \\ \text{Cover cropping} \\ \text{Seeding} \\ \text{Crop residue management} \\ Z_1 \end{array} \begin{bmatrix} 1.0000 & 0.6184 & 0.8852 & 0.7931 & 0.9686 \\ 0.6184 & 1.0000 & 0.2479 & 0.1312 & 0.4929 \\ 0.8852 & 0.2479 & 1.0000 & 0.9761 & 0.9631 \\ 0.7931 & 0.1312 & 0.9761 & 1.0000 & 0.9129 \\ 0.9686 & 0.4929 & 0.9631 & 0.9129 & 1.0000 \end{bmatrix}$$

As in Case I the highest correlation was found to be that of Z_1 and contour farming.

<u>Case III</u>

$$\begin{array}{l} \text{Contour farming} \\ \text{Cover cropping} \\ \text{Crop residue management} \\ Z_1 \end{array} \begin{bmatrix} 1.0000 & 0.4845 & 0.7986 & 0.8375 \\ 0.4845 & 1.0000 & 0.8760 & 0.8771 \\ 0.7986 & 0.8760 & 1.0000 & 0.9902 \\ 0.8375 & 0.8771 & 0.9902 & 1.0000 \end{bmatrix}$$

Different from the earlier two cases, in this case the correlation between Z_1 and crop residue management is the highest.

Future Projections

The number of farm ponds in 1980 and 2000 are estimated to be 63 and 85 percent higher than the present number respectively. The following table shows the estimate value in our water resources region (1).

TABLE 5a

Estimated Number of Ponds and Other Reservoirs in Upper Arkansas, White and Red River Water Resources Region

Year	Item	Number	Surface in acres
1959		125900	62950
1980	Ponds	205000	102500
2000		232950	116475
1959	Other	500	2000
1980	Reservoirs	2150	8600
2000		2500	10000

Table 5a shows that the number and surface area of farm ponds are much larger than those of other reservoirs, and are therefore much more significant. Based on this information the estimated number of ponds in our three watersheds are:

TABLE 5b

Estimated Number of Ponds

	Case I	Case II	Case III
1980	5050	1020	2635
2000	5630	1200	2975

The acreage of agricultural practices in the year 1980 and 2000 is estimated to be 10 and 15 percent higher than the present value respectively.

TABLE 5c

Estimated Acreage of Agricultural
Practices in Percents

	Case I	Case II	Case III
1980	35	38	15
2000	37	40	17

Results of future projections are shown in Figures 5a, 5b, and 5c.

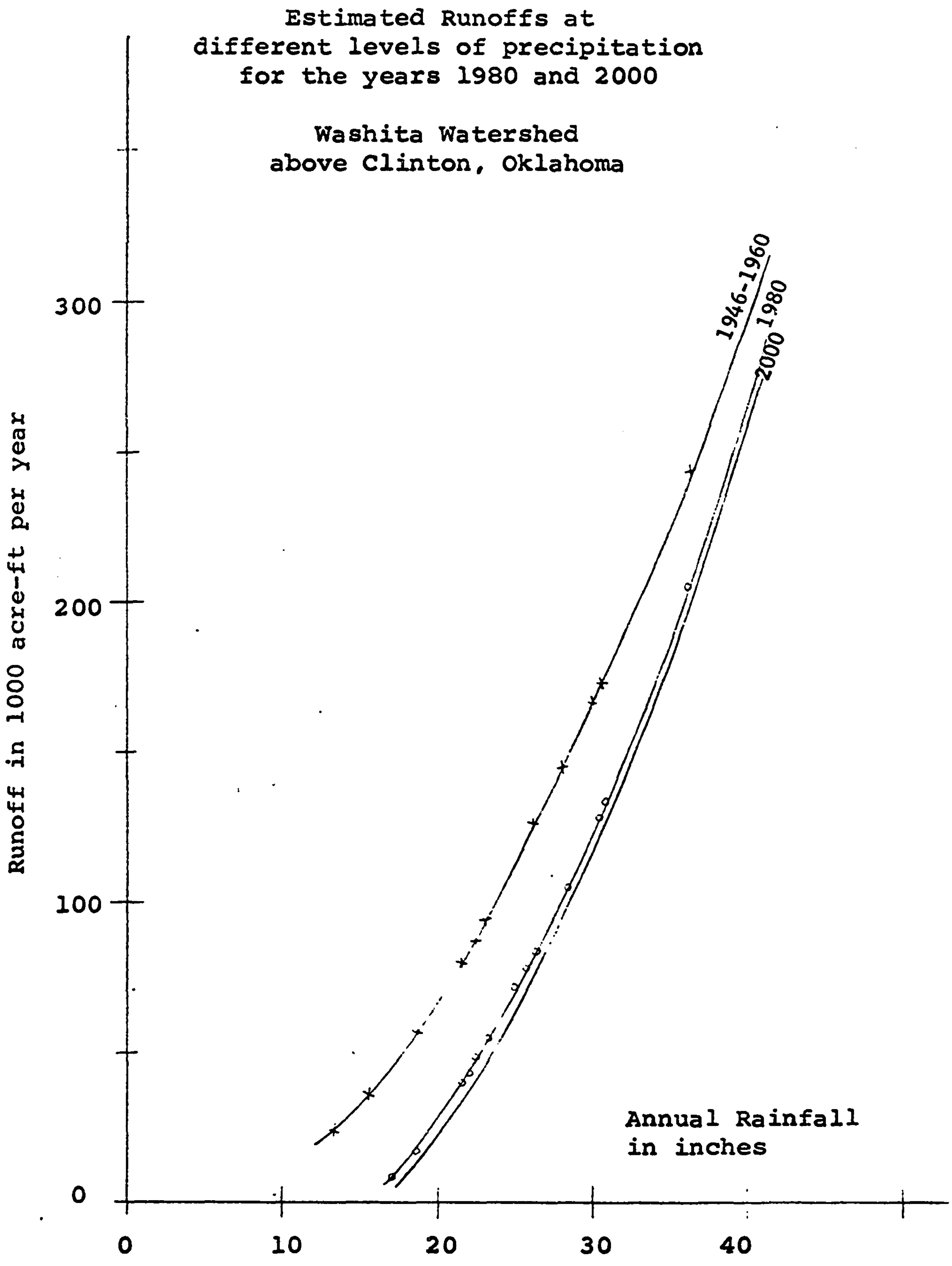

Estimated Runoffs at
different levels of precipitation
for the years 1980 and 2000
Washita Watershed
above Clinton, Oklahoma
Runoff in 1000 acre-ft per year
300
200
100
0
0
10
20
30
40
1946-1960
1980
2000
Annual Rainfall
in inches

Figure 5a

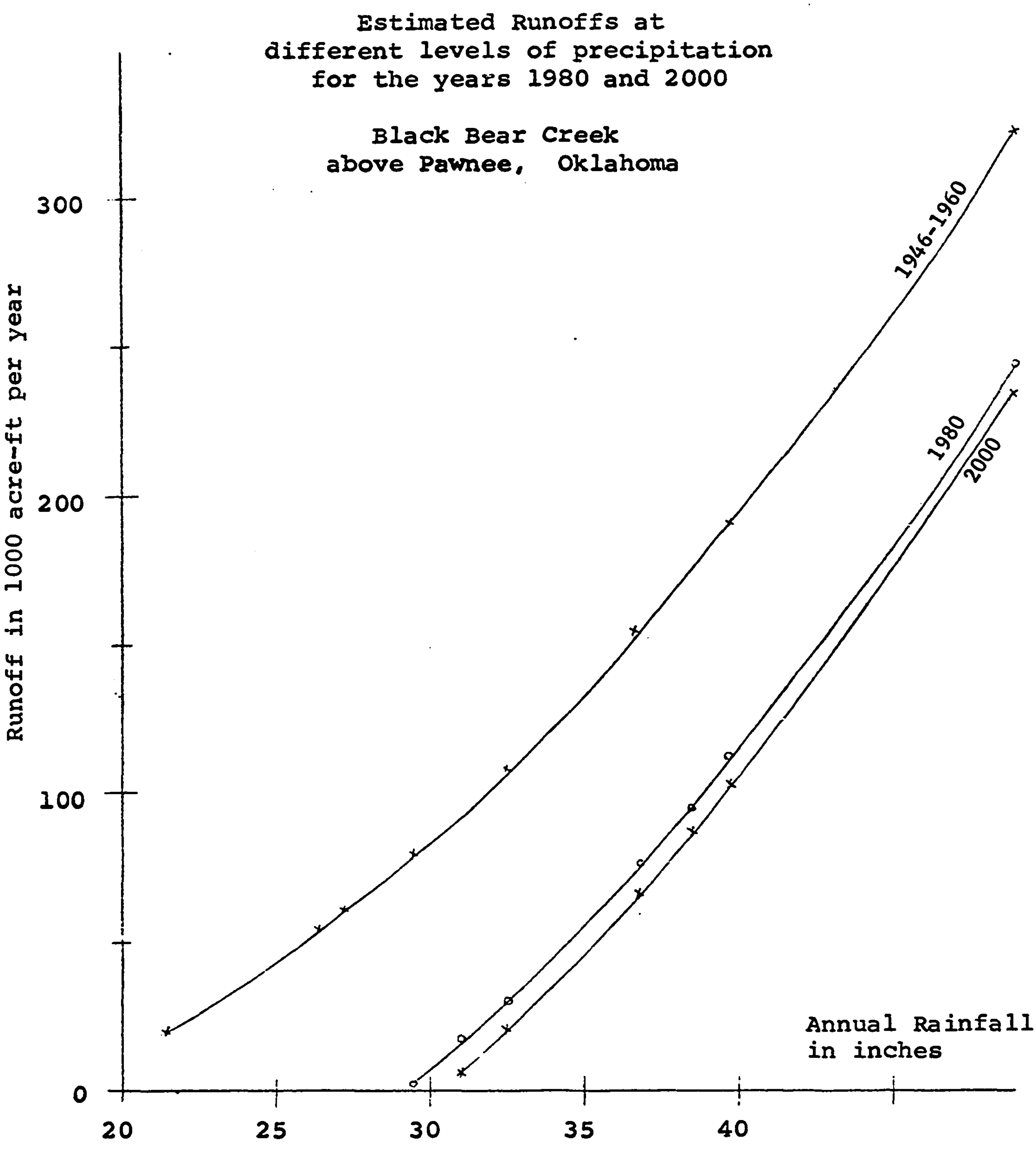

Estimated Runoffs at
different levels of precipitation
for the years 1980 and 2000
Black Bear Creek
above Pawnee, Oklahoma
Runoff in 1000 acre-ft per year
300
200
100
0
1946-1960
1980
2000
Annual Rainfall
in inches
20
25
30
35
40

Figure 5b

Estimated Runoffs at
different levels of precipitation
for the years 1980 and 2000

Clear Boggy Creek
above Caney, Oklahoma

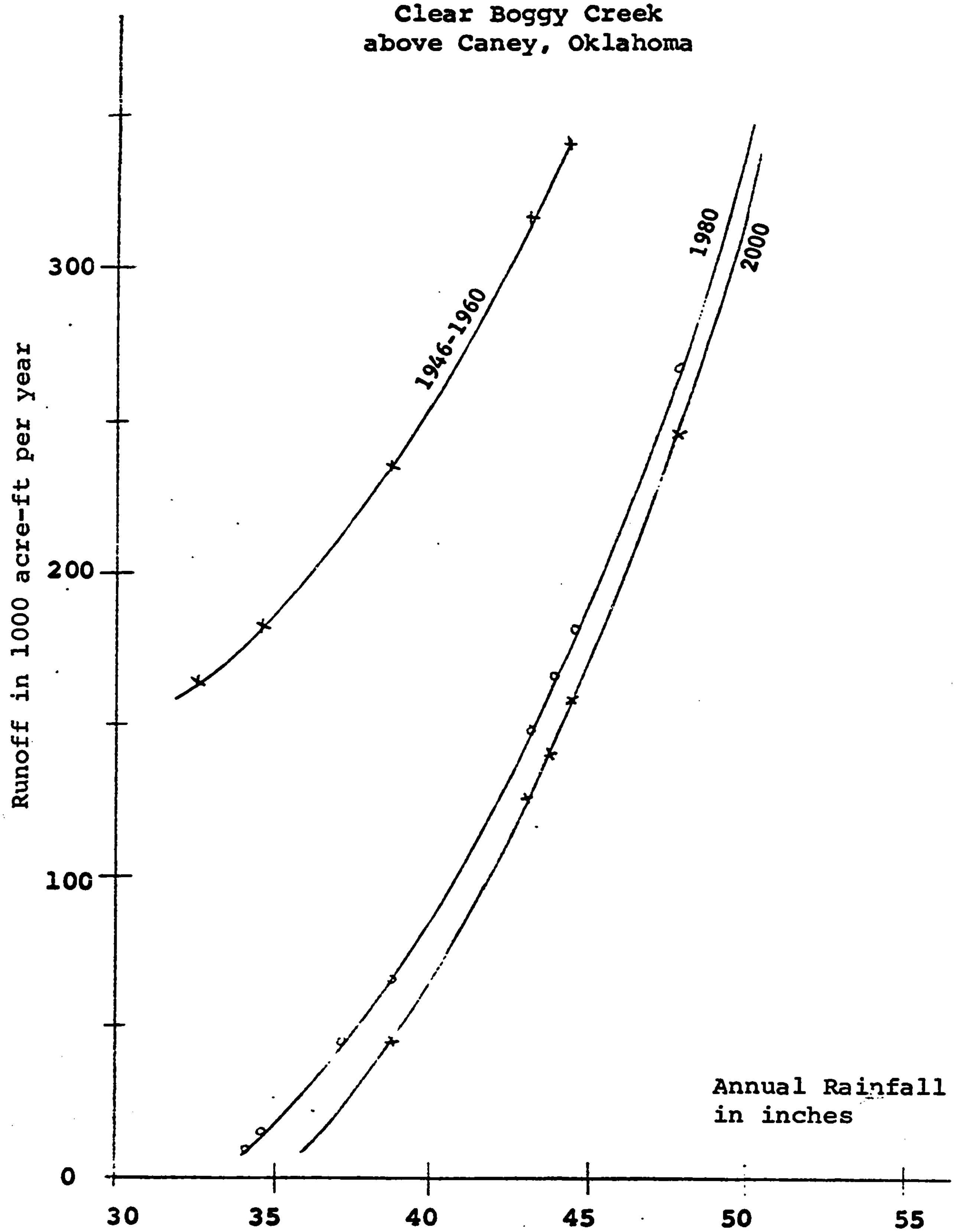


Figure 5c

CHAPTER V

SUMMARY AND CONCLUSION

Results in this study show that the effect of farm ponds on water yield is significant. Water yield could be increased by limiting the number of farm ponds in a watershed, and by improvement of agricultural practices such as contour farming and cover cropping.

Multiple linear regression, with 3 "independent" variables, accounts for about 85% of the variances of water yield. The relationships of runoff and rainfall are the same in all three cases. The relative effect of farm ponds on runoff in Case I is a little different from those in Case II and III when the numbers of ponds are small, but they are all the same when the numbers of ponds are large. The relative effect of agricultural practices on runoff are the same in Case I and II.

Due to the sample size being limited to 15 years, the developed regressions do not provide a rigorous formula for estimation. They are presented rather to show that this approach provides a "simple" method by which effect of land use and treatment can be estimated. The results suggest that a large-scale study could provide a better accuracy.

It would be difficult to compare results of this method with those of the Rational Procedure Method developed co-operatively by the Bureau of Reclamation, Soil Conservation Service, and Agricultural Research Service. In this study the regression equations found for predicting runoff from precipitation, number of ponds, etc., are mainly intended to predict future values rather than to determine the effect of each element on runoff as done by the Rational Procedure Method. Only time will determine how accurate these future projections are.

Land practices are greatly influenced by economic factors. Thus estimates of projected land practices could be determined by finding the trend of the economic indexes.

BIBLIOGRAPHY

1. U.S. Department of Agriculture, "A Report of Estimated Water Requirements for Agricultural Purposes and Their Effect on Water Supplies", Washington, D.C., December 1959.

2. U.S. Department of Agriculture, "Yearbook of Agriculture", Washington, D.C., 1955.

3. Sparkman, W. B., "The Yield of Water from Surface Run-off in Oklahoma", Unpublished Master's Thesis, University of Oklahoma.

4. U.S. Department of Commerce, Weather Bureau, "Climatological Data of Oklahoma".

5. Assenzo, J. R., "Use of Multiple Regression Techniques for Estimating Municipal Sewage Treatment Costs" Unpublished Ph.D. Dissertation, University of Oklahoma, 1963.

6. Graybill, F. A., "An Introduction to Linear Statistical Models", Vol. I, McGraw-Hill Book Company, Inc., New York, 1961.

7. Mood, A. M., and Graybill, F. A., "Introduction to the Theory of Statistics", McGraw-Hill Book Company, Inc., New York, 1963.

8. Keeping, E. S., "Introduction to Statistical Inference", D. Van Nostrand Company, Inc., Princeton, 1962.

9. Kendall, M. G., "A Course in Multivariate Analysis", Charles Griffin and Company Limited, London, 1957.

10. Kendall, M. G., "A Course in the Geometry of n-Dimensions", Charles Griffith and Company Limited, London, 1961.

11. Tintner, G., "Econometrics", John Wiley and Sons, Inc., New York, 1952.

12. U.S. Department of Agriculture, "Development of a Procedure for Estimating the Effects of Land and Watershed Treatment on Streamflow", Washington, D. C., March 1966.

APPENDIX

Data on number of ponds and agricultural practices were obtained from the Department of Agriculture Offices. A sample of the data sheet is shown in this appendix. Data on streamflow and rainfalls were obtained from the monthly publications of the Weather Bureau in Oklahoma.

Method of Data Collection of Farm Ponds

A census of farm ponds in Cleveland County was made. Based on the result of this trial study further steps were then planned for the collection of more data from other parts of the state to determine the average surface area, storage capacity and depth of farm ponds.

Analysis of the Data

We assume that the sampled populations of the counties have equal variances within a fixed range. This range is determined from the desirable precision of the measurements. It seems reasonable that we can measure the depth with one foot of precision.

From the collected data of Cleveland County, an attempt is made to determine the size of samples which should be collected from other counties in order to get a reasonably high confidence.

Sample Size Determination

If $\overline{X}$ is the mean of a random sample of size n taken from a normal population having mean μ and variance σ^2, then

$$t = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

is the value of a random variable having the Student-t distribution with n-1 degree of freedom. Suppose the confidence width is 2d,

$$d = t\,\sigma_{\overline{X}}$$

We can use this to solve for n since $\sigma_{\overline{X}}$ is a function of n. However, when using this equation to obtain a sample size one needs an estimate of the variance.

Suppose we were sampling from an infinite population,

$$\sigma_{\overline{X}}^2 = \frac{\sigma^2}{n}$$

Thus from the previous equation,

$$n = \frac{t^2\,\sigma^2}{d^2}$$

where σ^2 = variance of the random variable,

t = random normal variate, or if only an estimate of σ^2, say s^2, is available, t is the Student-t value,

d = desired precision.

Based on the above method, data from four other counties, Woodward, Pushmataha, Kiowa, and Rogers County, were collected. We decided to divide the state of Oklahoma into 5 regions: Northwest, Northeast, Southwest, Southeast, and Central. The means for the four variables of interest for each region are:

	Drainage Area (acres)	Surface Area (acres)	Storage (acre-ft)	Depth (ft)
Southeast	5.21	0.55	1.75	9.31
Northwest	117.53	2.21	11.61	16.27
Northeast	29.19	0.81	2.70	8.74
Southwest	43.78	2.23	5.53	12.82
Central	53.65	2.05	10.20	12.30

OK-504 (REV. 1-63)
STILLWATER, OKLA.

COMPARISON OF ACCOMPLISHMENTS

U.S. DEPT. OF AGRICULTURE
SOIL CONSERVATION SERVICE

AREA V
WORK UNIT Norman DATE 2/8/66

	UNIT	Goals FY 1966	Accom- plished	Percent Accom- plished			
100 DISTRICT COOPERATORS	No	75	36	48			
108,112,116 BASIC PLANS	No	75	37	49			
110,114,119 REVISIONS	No	100	35	35			
102 FARMS SERVICED	No	720	369	51			
106 FARMS APPLYING PRACTICE	No	540	304	56			
104 SERVICES TO OPERATORS	No						
170 ACP REFERRALS, RECEIVED	No						
172 ACP REFERRALS, SERVICED	No						
184 STD.SOIL SURVEYS,MAPPING	Ac						
186 STD.SOIL SURVEYS, CONV.	Ac						
196 SOIL CONSERV. SURVEYS	Ac						

CONSERVATION PRACTICES APPLIED

	UNIT	Goals FY 1966	Accom- plished	Percent Accom- plished			
314 BRUSH CONTROL (R)	Ac	2500	1070	43			
328 CONS. CROP. SYSTEM (A)	Ac	10000	6408	64			
330 CONTOUR FARMING (R)	Ac	2000	1134	57			
340 COVER & GREEN MANURE (R)	Ac	5000	2899	58			
342 CRITICAL AREA PLANTING (E)	Ac	125	34	27			
344 CROP RESIDUE USE (R)	Ac	10000	4530	45			
352 RANGE DEF. GRAZING (R)	Ac	8000	5740	72			
362 DIVERSIONS (E)	Ft	15840	9285	58			
378 FARM PONDS (E)	No	60	47	78			
388 IRRIG. FIELD DITCHES (E)	Ft	---	---	--			
398 FISHPOND STOCKING (E)	No	50	40	80			
410 GRADE STAB. STRUCTURE (E)	No	10	2	20			
412 GRASSED WATERWAYS (E)	Ac	20	4	20			
420 HAYLAND PLANTING (E)	Ac	---	---	--			
443 IRRIG.SYS.,SPRINKLER (E)	No	---	---	--			
449 IRRIG. WATER MGT. ( )	Ac	---	---	--			
460 LAND CLEARING (E)	Ac	400	215	54			
480 DRAINAGE MAINS & LAT. (E)	Ft	---	---	--			
511 PASTURE & HAYLAND REN. (E)	Ac	500	273	55			
512 PASTURE PLANTING (E)	Ac	1500	671	45			
526 PASTURE PROPER USE (A)	Ac	---	---	--			
528 RANGE PROPER USE (A)	Ac	20000	10120	81			
549,550 RANGE SEEDING (E)	Ac	100	4	4			
588 STUBBLE MULCHING (R)	Ac	---	---	--			
590 DRAINAGE FIELD DITCHES (E)	Ft	---	---	--			
600,602,604 TERRACES (E)	Ft	5280	3415	65			
612 TREE PLANTING (E)	Ac	---	---	--			
645 WILDLIFE HABITAT DEV. (E)	Ac	100	0	0			
656 WOODLAND INTER.CUTTING ( )	Ac	---	---	--			
666 WOODLAND SEEDING (R)	Ac	---	---	--			