3-1972

Applying factor analysis and multiple linear regression analysis in developing water yield models for small Tennessee watersheds

Rex Duane Haren

To the Graduate Council:

I am submitting herewith a dissertation written by Rex Duane Haren entitled "Applying factor analysis and multiple linear regression analysis in developing water yield models for small Tennessee watersheds." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Biosystems Engineering.

John I. Sewell, Major Professor

We have read this dissertation and recommend its acceptance:

John S. Bradley, John J. McDow, Curtis H. Shelton, Robert R. Shrode, Bruce A. Tschantz

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

142

November 27, 1972

To the Graduate Council:

I am submitting herewith a dissertation written by Rex Duane Haren, entitled "Applying Factor Analysis and Multiple Linear Regression Analysis in Developing Water Yield Models for Small Tennessee Watersheds." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Agricultureal Engineering.

John I. Sewell
Major Professor

We have read this dissertation
and recommend its acceptance:

Robert R. Shrode

John S. Bradley

Curtis H. Shelton

John J. McDow

Accepted for the Council:

Hilton A. Smith
Vice Chancellor for
Graduate Studies and Research

APPLYING FACTOR ANALYSIS AND MULTIPLE LINEAR REGRESSION ANALYSIS IN DEVELOPING WATER YIELD MODELS FOR SMALL TENNESSEE WATERSHEDS

A Dissertation

Presented to

the Graduate Council of

The University of Tennessee

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Rex Duane Haren

March 1972

ACKNOWLEDGEMENTS

The author wishes to express his appreciation to the following for their assistance in making this study possible:

Dr. John I. Sewell, Associate Professor, Agricultural Engineering, for his constructive criticism and guidance while serving as major professor on the author's committee.

Dr. John S. Bradley, Professor, Mathematics Department; Dr. John J. McDow, Professor and Head of the Agricultural Engineering Department; Mr. Curtis H. Shelton, Associate Professor, Agricultural Engineering Department; Dr. Robert R. Shrode, Professor, Animal Science Department; and Dr. Bruce A. Tschantz, Associate Professor, Civil Engineering Department for their guidance to the author while serving on his committee.

The University of Tennessee Agricultural Experiment Station Project H-204, "Factors Affecting Water Yields from small Watersheds in Tennessee," and Cooperative Regional Project S-53, "Factors Affecting Water Yields from Small Watersheds and Shallow Ground Aquifers" for field data and much of the data processing and reduction necessary for the completion of this study.

ABSTRACT

The uncertainty that now exists in predicting water yield requires that large factors of safety be incorporated into the design of hydraulic structures. If a mathematical model could be developed from data that is readily available or easily measured, that would predict the water yield with greater accuracy, this might allow a reduction of the safety factors thereby lowering the costs of these projects.

This study was designed to examine the feasibility of using factor analysis and multiple linear regression techniques in the development of mathematical models that would predict water yield from small watersheds in Tennessee on a seasonal and an annual basis.

Twelve parameters were initially selected for study by use of factor analysis. Of these 12 parameters one was deleted by factor analysis. Multiple linear regression analyses were then performed using various combinations of data from watershed parameters and various time periods. From these analyses the following conclusions were drawn:

1. Factor analysis can be used to screen superfluous parameters and thereby reduce the number of parameters needed to characterize the hydrologic properties of watersheds.
2. Watersheds must be grouped using similar hydrologic characteristics and especially similar geologic characteristics.
3. Many of the prediction equations of this study indicate that as area increases, runoff decreases which is contrary to that which is generally reported.

4. Prediction equations can be derived from different parameters for the same watersheds, and these equations often produce satisfactory predictions as long as the data used for the prediction are near the mean values of the parameters used in deriving the equations. The best results are obtained using data collected over a long time period.

TABLE OF CONTENTS

LIST OF TABLES

TABLE PAGE

LIST OF FIGURES

CHAPTER I

INTRODUCTION

I. IMPORTANCE OF THE STUDY

In designing hydraulic structures and in planning water supply systems in agricultural areas, the ability to predict the volume and rate of water yield from watersheds is essential. The capability of describing the hydrologic cycle from the time rainfall strikes the earth's surface until it is utilized or until it runs off the surface of the watershed in question would be most desirable. By using known or accurately estimated conditions within a watershed, the development of such a method for predicting with a high degree of reliability the total water yield from a watershed should be possible.

To realize the importance of such a water-yield model in planning for water supplies in rural areas, it is only necessary to look at a few examples of water consumption. To produce the feed for a steer for beef, 3,750 gallons of water are required for each pound of meat. In homes, the water being used is increasing due to the introduction of many conveniences such as the automatic washing machine, etc. Kahler (24) and Buie (6) suggest that the water used in homes, on farms, and in factories will continue to increase during the next several years.

The development of a reliable prediction equation that is applicable to watersheds of various sizes could reduce the cost of construction of hydraulic projects on a watershed. The uncertainty that now exists in predicting runoff requires that large factors of safety be incorporated into the engineering design of hydraulic structures. If a model could be developed that would predict the water yield with greater accuracy, this might allow a reduction of the safety factors thereby lowering the costs of these projects.

II. OBJECTIVES

The objectives of this study were (1) to examine the feasibility of using factor analysis and multiple linear regression techniques in the development of mathematical models that would predict water yield from selected small agricultural watersheds in Tennessee on a seasonal and an annual basis, and (2) to investigate the ability of the models developed to predict the water yields from the various watersheds for periods of varying lengths.

CHAPTER II

REVIEW OF THE LITERATURE

I. FACTORS AFFECTING WATER YIELD

Much literature dealing with watershed hydrology and factors affecting it has been written during the past several years. Only those articles and subjects which are most closely related to the area of this study are included in this review.

Harrold (18), on the analysis of 46 years of data collected on watersheds varying in size from 29 to 17,450 acres located near Coshocton, Ohio, concluded that both runoff volume and rate increase as the watershed size increases. He found also that under Coshocton conditions runoff rate and volume per unit area decrease as the area increases.

Upon correlating mean seasonal precipitation with watershed elevation, slope, orientation and exposure in Western Colorado, Spreen (34) found that elevation accounted for 30 percent of the variation in precipitation. He found also that five variables accounted for 85 percent of the variation.

Several studies have been made on the effect of vegetation and land use on runoff from watersheds. Ursic and Thames (36) found in their study of Northern Mississippi watersheds that runoff and peak flows were greatest from abandoned fields, less from upland hardwood forest and least from pine plantations. Harrold, et al. (19) found that mixed cover on a watershed increased the infiltration potential, thereby reducing the peak rates of runoff.

In a joint study conducted by North Carolina State University and the Tennessee Valley Authority (21), vegetative cover changes were found to affect the method of flow of water off the land, but not to affect materially the total outflow. The density of the cover and the physical condition of the soil were concluded to be the two most important factors in controlling the runoff rate.

Strahler (35) has listed several methods of qualification of landform description. He designated stream length according to: order 1 being the smallest finger-tip tributaries, order 2 being the confluence of two first order channels, order 3 as the junction of two second order channels, etc. The main stream into which all channels flow is the stream segment of highest order. Other factors listed are stream length, drainage basin area, drainage density and texture ratio, valley side slopes, relief ratio, and hypsometric analysis. Several values may be obtained from the hypsometric curve, and these may be used for comparisons with other drainage basins. These include the relative area lying below the curve, the slope of the curve at the inflection point and the degree of sinuosity.

II. RUNOFF FORMULAS

Chow (7) made a study of several methods of computing runoff volumes and peak discharge rates. Each method investigated employed either empirical formulas or a combination of empirical and theoretical formulas. The rational formula may still be the most widely used formula for predicting peak rates of runoff for relatively small areas. The main reason for its widespread use is the relative simplicity of the method. Basically, this formula states that the peak runoff is a function of the drainage area, the rainfall intensity and a runoff coefficient based on some of the physical characteristics of the watershed.

Another method that is used when peak flow is needed is Cook's method (9). This formula predicts runoff using the watershed characteristics of relief, infiltration, vegetal cover, surface storage, and watershed area and precipitation as the independent variables. This formula, as well as the rational formula, is relatively easy to apply. It utilizes families of curves from which appropriate numbers are obtained to compute runoff rates.

A much more comprehensive procedure for predicting runoff is a method developed by the U. S. Soil Conservation Service (22). This method produces a runoff hydrograph that depends upon the duration and intensity of rainfall. Four major hydrologic soil groups are utilized. The basis for the classification of these soils is the infiltration rate that occurs at the end of storms of long duration. These four soil groups are rated from A through D with A having least runoff and D having the greatest.

As suggested by Amorocho (1), two general methods are often used to predict the peak rate and runoff volume from a watershed. One methods, sometimes called the probabilistic method, considers runoff as a chance occurrence, and then uses the maximum values of runoff for a period of historical record. An attempt is made to calculate the probability of the occurrence of a given event.

The other general method, the deterministic, is by application of correlation analyses. In using this method, an attempt is made to describe the functional relationship existing between the factors that affect the runoff given that a certain amount of rainfall has occurred. These factors are formed into a mathematical model. Multiple regression techniques are employed to establish the coefficients of the factors which will yield the highest correlation between the values of record and the predicted values obtained from the model.

III. HYDROLOGIC MODELS

Hydrologic models developed to predict runoff usually take into account many hydrologic phenomena such as overland flow, interflow, base flow, ground water, evaporation, soil properties, cover, infiltration and precipitation characteristics. These models differ from runoff formulae in that the runoff formulae usually consider fewer parameters, tend to be more empirical, and are often based on approximation and averages.

Recently considerable work has been directed toward developing hydrologic models. Many models have been developed to synthesize and predict runoff. Among the developers of these models are Gray (17) and Reich (29) who used mathematical models to represent a unit hydrograph. Gray's was a dimensionless hydrograph, and Reich used a three-parameter function to obtain a unit hydrograph.

Amorocho and Orlob (2) described a physical model of the hydrologic cycle. In their model rainfall represented inflow, and the outflow was composed of overland flow, interflow, base flow, groundwater flow, and evaporation. This representation permitted a general continuity equation to be written that included all the usually recognized elements of the hydrologic cycle.

Crawford and Linsley (10) developed a method called the "Stanford Model" of synthesizing streamflow records from hydrologic data by modeling the hydrologic cycle on a digital computer. They attempted to describe the complete hydrologic process that occurs within a watershed by using a large number of parameters related to the various components of the watershed output. This method requires a five-year or longer period of detailed records from a watershed to be synthesized.

Huggins and Monke (20) developed a model to simulate the surface runoff of small watersheds based on the subdivision of the watershed to be modeled into small independent elements. A runoff hydrograph for the entire watershed was obtained by applying the equation of continuity to integrate the responses from each element. The model was supposed to do the following: (1) readily simulate complex watershed conditions both from temperature and space distribution, (2) eliminate the "lumped parameter" coefficients that represent "effective averages" for varied parameters, (3) provide independence between the model developed and the relationships used to estimate the different parts of the hydrologic process, and (4) allow the hydrologic process to be broken into independent parts that can be applied to each of the small elements.

Miller and Viessman (27), using data collected from four small urban watersheds, developed a runoff model for small urban watersheds where rainfall and runoff data were plotted for each watershed, and a least-squares regression line was found. They observed that the regression coefficient of the rainfall-runoff relationship increased as the percent impervious area increased. This relationship could be described by the equation M = b(I-a) where M is the predicted regression coefficient, I is the percent impervious area in the watershed, b is the slope of the least squares regression line as previously determined, and a is the abscissa intercept for M. The linear prediction equation thus obtained was of the form R = M(P-Ia) where R is the runoff in inches, P is the rainfall in inches, and Ia is the initial abstraction in inches.

Beasley (4), using constants based upon certain parameters, developed for Missouri conditions a model that predicts the peak rates of runoff from watersheds of less than 200 acres. The parameters used in this model were: (1) peak rate of runoff from a watershed with a specific set of watershed conditions, (2) watershed location, (3) soil infiltration, (4) topography, (5) shape, (6) vegetative cover, (7) surface storage, and (8) runoff frequency.

The use of linear regression was adopted early by hydrologists since many problems faced by hydrologists consist of the determination of the relationship of one or more independent variables to a dependent variable. Linear regression is the mean curve defined by a scatter diagram. In its simplest form it defines the linear relationship between two variables. Its equation is of the form y = a + hx where y is the dependent variable and x is the independent variable, "a" is the regression constant and "h" is the regression coefficient. While in multiple linear regression, the linear regression of the dependent on more than one independent variable is defined. The form of this relationship is

$$y = a + h_1x_1 + h_2x_2 + \ldots, \quad \text{where}$$

"a" and "h_i" again are the y intercept and the regression coefficients, respectively. The evaluations of the constants of the equations are done by using the equations as outlined by Diskin (15). To obtain "h" use the equation

$$h_i = \frac{\Sigma(x_i y_i) - N(\bar{x}_i)(\bar{y})}{\Sigma(x_i{}^2) - N(\bar{x}_i)2} \quad ;$$

to obtain "a", use $a = \bar{y} - \Sigma h_i \bar{x}_i$ where $\bar{x}_i$ and $\bar{y}$ are the means of the independent variable and the mean of the dependent variable, respectively, and N is the number of observations. A more detailed discussion of multiple linear regression may be found in Linsley (26) and Johnstone and Cross (23), as well as in many other books on statistical methods. Also Beard (3) gives a discussion and sample calculations made to illustrate the procedure for determining a multiple linear regression equation as used specifically in hydrology.

Sharp, et al. (31) discuss the limitations of regression analysis and point out that correlation and regression methods of analysis are based upon normally distributed data for each season; however, hydrologic data seem rarely to be normally distributed. Also three assumptions are inherent in the application of multiple regression methods of analyses to the hydrologic problem: (1) that no errors exist in the independent variables, that is, errors can occur only in the dependent variable; (2) that the variance of the runoff is independent of the other variables; and (3) that the measured values of runoff are uncorrelated random variables. They conclude that the multiple regression method of analysis will give a 'best fit' equation and that limited reliance should be placed on values obtained from such equations, especially at values far removed from the mean. It should be noted that hydrologic data infrequently fit the assumptions enumerated above.

Diskin (15), in his study of the regression equation, found that the regression equation is not applicable in cases where carryover or lag between rainfall and runoff is appreciable. He contended that it is applicable to watersheds where more than one distinct season is present each year. This change in season allows for a minimum of carryover from one year to the next.

Other authors have used linear regression, with varying degrees of success, to obtain a functional relationship between runoff and various hydrologic parameters. Among them are Amorocho and Orlob (2), Linsley, et al. (26), Johnstone and Cross (23), and the Tennessee Valley Authority (TVA) (5).

In an attempt to reduce the number of parameters used in constructing a mathematical model of water yield, many researchers have turned to multivariate statistical methods. One of these statistical methods is factor analysis. It is a technique used primarily to reduce the number of variables in a problem. It is a measure of the interrelationships of two or more variables. Using this technique one can reduce the number of overlapping variables in a problem to a smaller number of linearly independent variables, thus reducing the number of variables that need to be measured.

The computations begin with a n x n correlation matrix, as is discussed by Shelton and Sewell (33) or any text on factor analysis, which is developed from observed data. These correlation coefficients are expressions of correlations of each variable with each of the other variables. Using this symmetric correlation coefficient matrix, a solution is found by solving the equation $|R - \lambda I| = 0$ for the eigenvalues. R is the matrix of the correlation coefficients, λ is the eigenvalue (characteristic root) and I is the unit or identity matrix. Performing the indicated operations on the n x n matric equation gives an equation of the form $c_1\lambda^n + c_2\lambda^{n-1} + \ldots + c_n = 0$ where the c_i values are constants. The solution of this equation gives the characteristic roots.

The eigenvectors can be found from the equation $[R - \lambda_i I][V_i] = 0$ where V_i is the eigenvalue ($i = 1, 2, 3, \ldots, n$) and n is equal to the number of eigenvalues. The eigenvectors are linear combinations of the observed variables and of such magnitude that the product of the eigenvector and its transpose must equal one.

Wallis (38) recommended this method for use in the initial analysis of hydrologic problems. Other researchers who have used this method follow: The Tennessee Valley Authority (13) applied factor analysis to 44 variables measured in a hydrology study conducted in 1965 when 22 of the 44 variables studied were eliminated by use of factor analysis; Webb and Briggs (39) used principal component analysis to screen 125 chemical analyses of biotites; Dawdy and Feth (11) used factor analysis in their study of ground water quality; and Diaz, Sewell, and Shelton (14), in a study of various watershed factors recorded for 14 watersheds near Coshocton, Ohio and 7 watersheds from Riesel, Texas, found by using factor analysis that 96.8 percent of the covariance could be explained by seven factors from the Coshocton studies, and 97.5 percent of the covariance could be explained by three factors in the Riesel studies.

Shelton and Sewell (33), in a study of factors affecting water yield, used four gauged watersheds located in the Valley and Ridge Province of East Tennessee. They restricted their study to 20 variables that were measured on each of the watersheds. These variables were: area, form, compactness, mean elevation, hypsometry, total relief, median elevation, mean slope distribution of 0-5 percent, 5-10 percent, 10-20 percent, and 20-40 percent, drainage density, stream order 1, stream order 2, stream order 3, stream slope, stream length, runoff volume, and groundwater level. The watersheds selected were apparently geomorphologically similar with variations in water-yield values. Seventeen of the 20 variables used were measures of selected physical conditions on the watersheds. This study illustrates how, by factor analysis, the 17 variables selected were reduced to 6 variables which would supposedly reveal the underlying relationships of the various parameters that affect water yield. The authors concluded that factor analysis was useful in screening overlapping variables and perhaps watersheds, thereby making possible the use of a smaller number of variables to describe the underlying relations and influences. They concluded also that in the selection of any variable, a knowledge of the watershed, relevance of the variable, and the quality of measurement should be considered.

Dawdy, Lichty, and Bergman (12), in their rainfall-runoff simulation model, considered three components of the hydrologic cycle, antecedent moisture conditions, infiltration, and surface runoff. Basically, their method consists of maintaining a water budget for a watershed.

CHAPTER III

METHODS AND PROCEDURES

I. WATERSHED DESCRIPTIONS

Ten watersheds located in Tennessee and ranging in size from 9.3 to 198.7 acres were selected for this study. The location of each watershed is given in Figure 1. Typical weir and H-flume construction, as well as watershed conditions, are shown in Figure 2 through Figure 4. Also, a topographic map of each watershed is shown in Figure 9 through Figure 18 of Appendix A. Table XI in Appendix B gives 19 basin characteristics that have been determined for each of the 10 watersheds. Other parameters used which are not included in Table XI are evapotranspiration (ET), temperature (T), soil cover index (SCI), depth of topsoil (DI), precipitation (P), and water yield (Q). These additional parameters are not included because they vary with climatic conditions and the seasons; however, total annual values and means are given in Tables XI through XIV (Appendix B). The author realizes that this study was done with a minimum amount of data for evaluating parameters, but no more data were available. It is also recognized that attempting to characterize almost as many parameters as watersheds available for study is undesirable from a statistical standpoint.

Watersheds I through IV are located near Oak Ridge in the Valley and Ridge Province. The soils of the four watersheds are predominantly Fullerton cherty silt loam. The depth to the ground water table at the base of Watersheds I and II is relatively shallow varying throughout

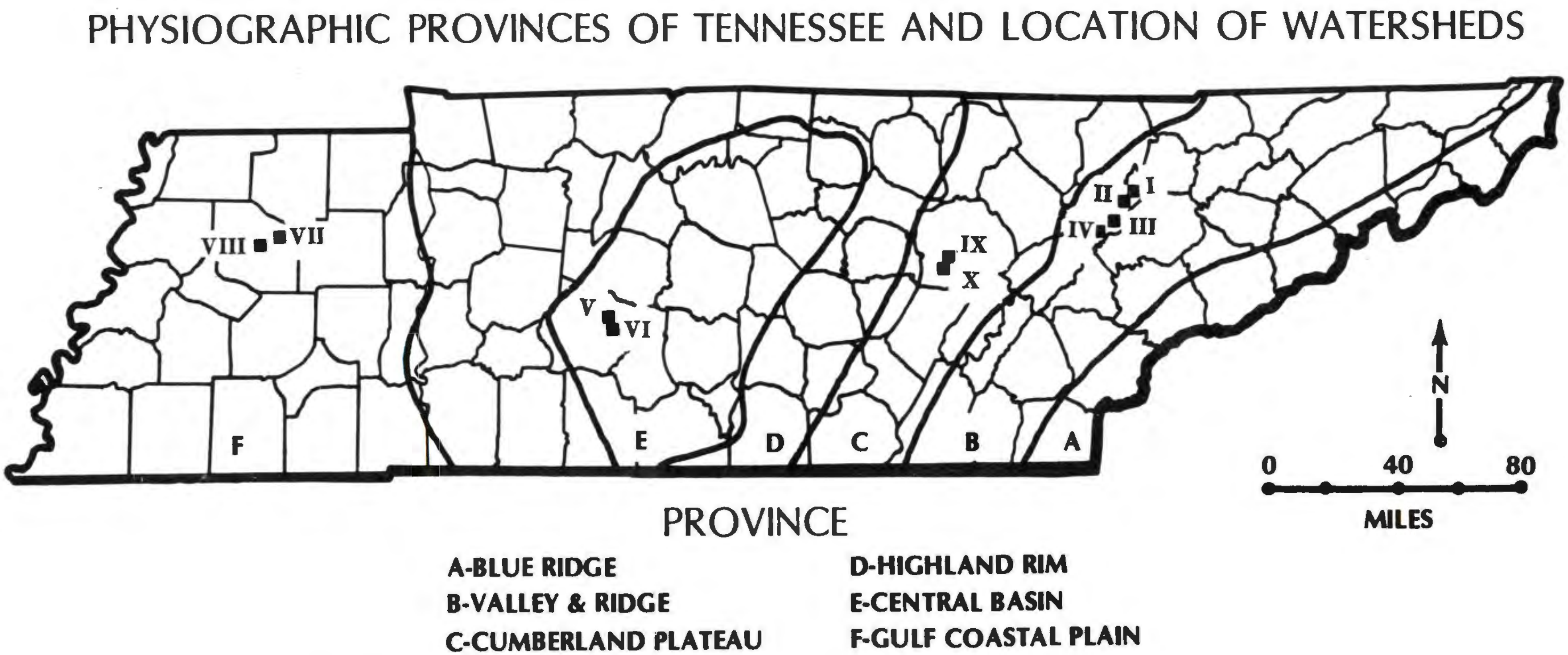

Figure 1. Watershed Locations.

Figure 2. H-flume on Watershed I.

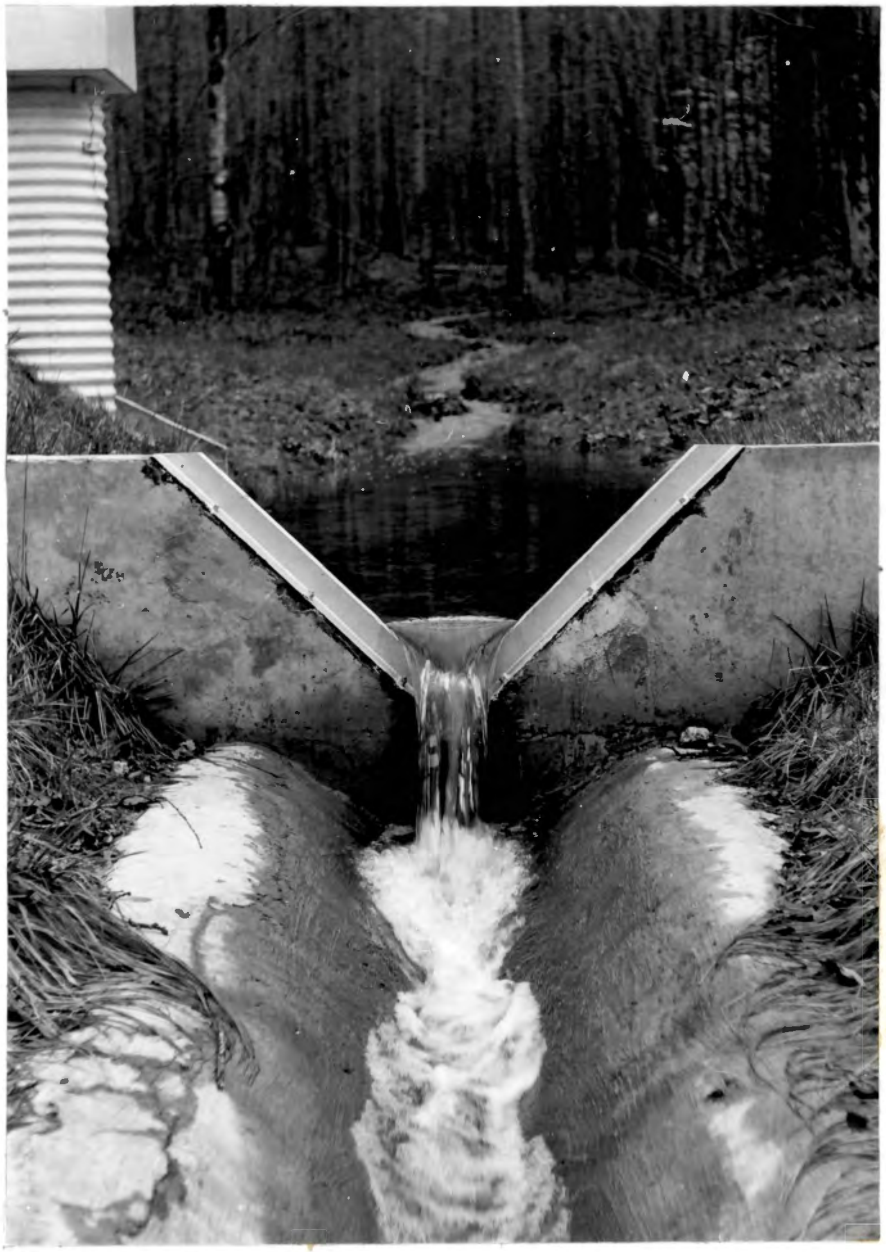

Figure 3. Sharp crest weir on Watershed II.

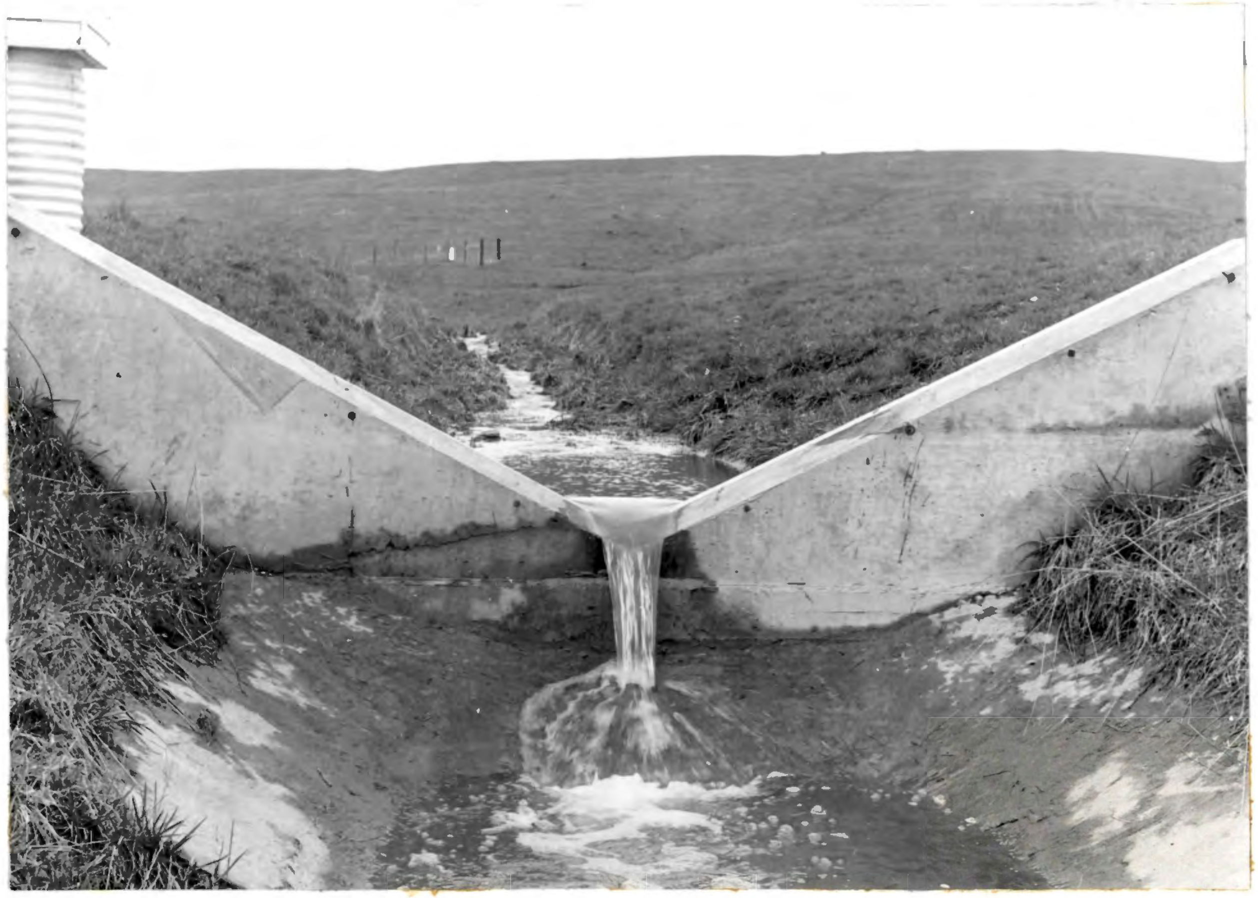

Figure 4. Broad crest weir on Watershed III.

the year from 0 to nearly 4 feet. The depth to the groundwater table of Watersheds II and IV is relatively deep varying throughout the year up to 20 feet. The flow from Watersheds I and II is continuous, and the flow from Watersheds III and IV is ephemeral. The drainage is good on all watersheds except for less than one acre of sinks in each of Watersheds I, III, and IV.

A summary of soil characteristics as determined from soil borings on Watersheds I through IV is shown in Table I. This table indicates that the borings extended downward to a depth of 20 feet on each of the watersheds where bedrock was not encountered. These borings indicate that the geology of the watersheds is somewhat similar; pervious cherty strata were encountered in all watersheds; but Watersheds III and IV have more cherty strata than do I and II. Also, Watershed IV appeared to have less clay at the 2- to 3-foot depth than did the other watersheds. Two factors, a less pronounced clay layer and more chert being present in the soils of Watershed IV, might explain the apparent inconsistency of the hydrologic behavior of Watersheds III and IV.

Watersheds V and VI are located in the Central Basin Province four miles south of Spring Hill. The predominant soil of both watersheds is Maury. It is composed largely of silt with low plasticity from 0 to 3 feet in depth. The depth to bedrock varied from 8 to 25 feet. The depth to the water table at the weir location varied from 0 to 9 feet. The flow from both watersheds is ephemeral. The drainage is good except for a few sinks along the lower reaches of the drainage channel of Watershed VI.

TABLE I

SUMMARY OF SOIL BORINGS OF OAK RIDGE WATERSHEDS

	Watersheds			
	I	II	III	IV
Number of Borings	18	8	14	6
Number of Mechanical Analyses	48	18	48	13
Average Depth to Impenetrable Layer (ft.) when Encountered	9.1	8.4	8.4	8.4
Average Depth when Bedrock not Encountered (ft.)	18.0	18.3	16.5	19.0
Bedrock Encountered (Percent of borings)	83.3	62.5	50.0	83.3
Percent of Borings Containing Clay Layer at 2-4 ft.	50.0	87.5	78.5	32.0

Watersheds VII and VIII are located in the Gulf Coastal Plain Province two miles northeast of Milan. The surface drainage is good on both watersheds except for approximately three acres of depressions and small ponds. The soil borings did not reach the water table at the 20-foot level where the borings stopped. The predominant soil of the 0- to 4-foot stratum for Watershed VII is Grenada. It is silt with low plasticity. Bedrock was not encountered for any drillings on either Watershed VII or VIII. The predominant soil on Watershed VIII is Calloway. It is silt with low plasticity for the 0- to 7-foot profile.

Watersheds IX and X are located in the Cumberland Plateau Province approximately eight miles south of Crossville. The depth to the water table fluctuates rapidly from 0 to 4 feet for both watersheds due to the presence of fractured sandstone. The surface drainage is good on both watersheds except for about a one-acre sink and a two-acre pond on Watershed X. The dominant soil on both watersheds is Hartsells. The soil is silt in the 0- to 2-foot profile. From the 2-foot depth to bedrock the soil is poorly graded sand. The descriptions of the watersheds were taken from Lillard, et al. (ed.) (25).

II. PARAMETER SELECTIONS AND DEFINITIONS

The parameters were selected for analysis according to the following criteria: (1) that they be generally accepted as variables which affect runoff, (2) that they be hydrologically relevant, and (3) that the parameters be easily determinable from data generally available or from field investigations of ungauged watersheds.

Summaries of parameters used are presented in Tables XI through XIV (Appendix B). Using the above criteria, the following parameters were selected for the initial analyses:

1. Area (A) in acres.
2. Form Factor (FF), the ratio of average width to the axial length of the basin. FF is a dimensionless parameter.
3. Mean elevation (ME) in feet. Using contour area method, ME is determined according to

   $$ME = \Sigma(a \frac{h_1 + h_2}{2})/A \qquad \text{where}$$

   a = area between any two contours at elevations h_1 and h_2, respectively.
4. Precipitation (P) in inches, the accumulated precipiration over a specified time period.
5. Water yield (Q) in inches for a time period.
6. Total stream channel length (L) in feet which was determined from contour maps of the watersheds.
7. Soil cover index number (SCI) is a dimensionless parameter.

The computation of Parameter 7 was done by the method as outlined in The U.S. Soil Conservation Service Handbook, Hydrology Supplement A (22), as used by The Tennessee Valley Authority (5) and as modified by Associate Professor Curtis H. Shelton of the University of Tennessee Department of Agricultural Engineering. An average of the runoff curve numbers for each land use, by hydrologic soil groups, was computed using values listed in the Soil Conservation Service Handbook (22). Meadow of soil group A had a low SCI value of 30 for all the curves. The number 29 was used as a reference level, and it was subtracted from values listed in the Soil Conservation Service Handbook. The number 29 was selected as a reference to insure that all values used in computations would equal or exceed one. To determine the weighted soil cover index number (SCI) for a particular watershed, the percent of area comprising the different SCI classifications was computed. The SCI value used was weighted according to respective land areas.

8. Hypsometric curve factor (HYP) was determined by finding the average slope of the hypsometric curve for each watershed. The area bounded by the horizontal and vertical axis and the curve itself was determined by a planimeter. The hypsometric curves were made available by Tennessee Agricultural Experiment Station Project H-204. Using this area, the height of triangle of equal area was found. Then knowing the horizontal distance of the hypsometric curve, the slope for each curve was computed. For example, the area A between the curve and the horizontal and vertical axes for Watershed I was 12.27 square units. The height h equals 2A/b where b is the length of the base. Thus, $h = 2(12.27)/5 = 4.91$ units. Then the slope becomes h divided by b, or slope $= 4.91/5 = 0.98$.

9. Average depth of soil (DT) in feet is the average depth of top soil.

In later analyses, three additional parameters were considered:

10. Mean sea level elevation (MSL) in feet.

11. Evapotranspiration (ET) in inches.

12. Average air temperature (T) in degrees F.

The basic data for parameters 1 through 5 and 10 were furnished by the University of Tennessee, Knoxville Agricultural Engineering Department, and parameters 6 and 8 were determined from contour map studies. Parameter 7 was determined as previously described, and parameter 9 was developed from geologic test borings made by U.S. Soil Conservation Service personnel. Parameter 11 was determined by the Penman method and was obtained from the Tennessee Valley Authority (28 and 37). The data for parameter 12 were taken from U.S. Department of Commerce Climatological Data (8).

III. PROCEDURE

After selecting the variables according to the criteria previously listed, they were tested to ascertain whether or not they were quantitative measures of the same hydrologic factors. This was done by using rotated factor analysis. Factor analysis, as pointed out in the Review of Literature, is used in an effort to show the interrelationships between the different parameters and to facilitate the selection of parameters that act independently.

The precipitation (P) data were obtained from charts of automatic recording gages located on the watersheds. The amount of precipitation was read from the scaled charts by the following procedure. The height of the curve was measured at each major change in slope, and the time and date of each point read were recorded. These data were then punched on standard 80-column IBM cards. These cards were processed by the University of Tennessee, Knoxville IBM 360-65 Computer programmed such that the printout would give the month, day, and year of each rainfall event, its duration (hours) and intensity (inches per hour), the interval accumulation (inches), the storm accumulation (inches), the daily accumulation (inches), and the monthly accumulation (inches).

The runoff (Q) data were collected from all the watersheds for various time periods. They were measured on Watershed I and Watershed IV by a Type H-flume, on Watershed II and Watershed V by a sharp crest weir, and on Watershed III and Watershed VI through Watershed X by a 2 to 1 broad crest weir. The Q data were recorded by a Stevens A-35 Water Level Recorder on Watersheds I through III and Watersheds V through X. The Q data for Watershed IV were recorded by a Belfort FW-1 Liquid Level Recorder.

The charts were read and recorded in the same manner as was described for P data. The data were then punched on standard 80-column IBM cards and processed by the computer such that the printout would give the average runoff rate (inches per hour), interval runoff (inches), accumulated daily runoff (inches), and accumulated monthly runoff (inches).

A series of factor analyses were performed using the BMD03M Rotated Factor Analysis Program (16) available at the University of Tennessee, Knoxville Computer Center. The first series of factor analyses were done using the parameters previously listed and the data from the four seasons for each of the years 1967 and 1968 for all 10 watersheds. A study of these analyses suggested the deletion of the parameter stream length (L) from further analysis. A discussion of the factor analysis and rationale employed is found in Chapter IV.

Using the remaining parameters A, FF, ME, SCI, HYP, DT, P, and Q, a stepwise multiple linear regression analysis was performed employing BMDO2R Multiple Linear Regression Program (16) in an attempt to develop a prediction equation for each of the three-month periods. The multiple linear regression analyses were done in a pattern, the purpose of which was to test the practicality of determining an equation for different periods of the calendar year.

The order in which the multiple linear regression analyses were performed is listed below. A comprehensive discussion of the equations developed for analysis is given in Chapter IV. All regressions obtained in this study are given in Table II.

Regression 1 was developed from the monthly averages of P and Q data for the year 1967. Regression 3 was for the monthly averages of P and Q for the year 1968. Then Regression was developed using the monthly average values of P and Q for the combined years 1967 and 1968. For Regressions 4, 5, 6, and 7, the quarterly averages of monthly P and Q data were used for the periods of January through March, April through June, July through September, and October through December of 1967. The quarterly procedure was repeated for the year 1968 giving Regressions 8, 9, 10, and 11. Then the two years of data were combined, and Regressions 12, 13, 14 and 15 were obtained using the two-year quarterly monthly averages of the P and Q data. Regressions

TABLE II

LIST OF REGRESSION EQUATIONS

Regression Number	Regression Equation	Multiple R_2	Years of Record	Watersheds
1	Q = 28.449 - 5.147(FF) - 0.025(A) - 0.059(ME) - 2.160(SCI) + 1.229(HYP) - 0.253(DT) + 0.017(P)	0.895	1967	I-X
2	Q = - 33.489 + 0.036(HYP) + 9.123(P) - 0.082(SCI) - 0.103(A) + 0.025(ME) + 0.172(DT)	0.952	1968	I-X
3	Q = 27.196 - 5.053(FF) - 0.019(A) - 0.051(HYP) + 0.864(DT) - 3.226(P)	0.700	1967-1968	I-X
4	Q = - 49.439 + 5.068(P) - 0.057(ME) - 2.332(DT) + 0.276(HYP) + 17.185(FF) + 0.033(A) + 0.051(SCI)	0.964	Jan.-Mar. 1967	I-X
5	Q = 14.237 - 0.024(ME) - 0.007(A) + 0.109(DT) - 0.057(HYP) - 3.036(FF) + 0.153(SCI) - 2.031(P)	0.668	Apr.-June 1967	I-X
6	Q = 23.045 - 0.022(A) + 2.189(DT) - 0.294(SCI) - 0.093(HYP) - 5.578(FF) - 0.373(P) + 0.022(ME)	0.962	July-Sept. 1967	I-X
7	Q = 22.673 - 9.399(FF) - 0.010(A) - 0.058(ME) -0.079(HYP) + 0.881(DT) - 0.179(SCI)	0.801	Oct.-Dec. 1967	I-X
8	Q = - 27.737 + 6.676(P) - 0.004(A) + 1.207(DT) + 0.023(ME) + 0.017(SCI) + 0.010(HYP)	0.972	Jan.-Mar. 1968	I-X
9	8.818 - 0.041(HYP) - 0.013(A) - 5.620(FF) + 0.382(DT) - 0.022(ME)	0.578	Apr.-June 1968	I-X

TABLE II (Continued)

Regression Number	Regression Equation	Multiple R^2	Years of Record	Watersheds
10	Q = - 10.392 + 1.538(P) + 0.002(A) + 0.038(ME) + 0.350(DT) + 4.582(FF) + 0.040(HYP) - 0.076(SCI)	0.990	July-Sept. 1968	I-X
11	Q = 3.113 + 0.279(DT) + 0.019(HYP) + 0.009(ME) - 4.118(P) + 0.185(SCI) + 0.009(A) + 1.845(FF)	0.989	Oct.-Dec. 1968	I-X
12	Q = - 64.724 + 13.025(P) - 0.035(ME) + 0.568(SCI + 0.013(A) - 0.582(DT) - 3.538(FF) - 0.015(HYP)	0.997	Jan.-Mar. 1967-1968	I-X
13	Q = 7.693 - 0.019(ME) - 0.013(A) - 5.364(FF) + 0.254(DT) - 0.048(HYP) + 0.053(SCI)	0.628	Apr.-June 1967-1968	I-X
14	Q = 15.752 + 0.036(ME) - 0.014(A) - 1.629(P) - 0.228(SCI) + 1.051(DT) - 0.019(HYP) - 0.325(FF)	0.993	July-Sept. 1967-1968	I-X
15	Q = 5.334 - 0.051(HYP) - 7.971(FF) - 0.036(ME) - 0.014(A) + 0.279(DT) + 1.251(P) + 0.044(SCI)	0.704	Oct.-Dec. 1967-1968	I-X
16	Q = 189.11 - 132.35(FF) - 0.435(ME) + 0.361(P) - 1.095(HYP) - 0.185(A) + 8.701(DT)	0.757	1967	I-X
17	Q = - 397.19 + 0.427(P) + 8.986(HYP) - 0.964(SCI) - 0.122(A) + 0.303(ME) + 2.072(DT)	0.953	1968	I-X
18	324.99 - 60.399(FF) - 0.228(A) + 0.605(HYP) + 10.404(DT) - 1.796(SCI) - 3.113(P)	0.698	1967-1968	I-X

TABLE II (Continued)

Regression Number	Regression Equation	Multiple R^2	Years of Record	Watersheds
19	Q = − 40.112 + 0.059(MSL) + 1.020(SCI) + 0.231(HYP) − 1.356(P)	0.958	1966-1968	I-VIII
20	Q = −3.4434 + 0.2852(SCI)	0.361	All years	I-VIII
21	Q = 0.1325 + 0.2805(P)	0.122	1964-1968	I and II
22	Q = − 150.40 − 0.009(MSL) − 0.005(A) − 0.048(SCI) + 29.239(ET) + 2.39(T) + 0.272(P)	0.999	Jan.-Mar. All years	I-VIII
23	Q = − 246.79 − 0.020(HYP) + 23.231(ET) + 0.05(MSL) − 0.133(SCI) − 0.008(A) + 2.195(T)	0.967	Apr.-June All years	I-VIII
24	Q = − 36.294 + 0.017(SCI) + 0.561(MSL) − 0.010(HYP) − 0.429(P) − 0.734(ET)	0.962	July-Sept. All years	I-VIII
25	Q = − 2.706 − 0.004(MSL) − 0.507(P) − 0.010(HYP) − 0.034(SCI) − 0.001(A) + 0.219(T)	0.925	Oct.-Dec. All years	I-VIII
26	Q = − 951.530 − 0.107(A) + 0.087(MSL) − 1.047(SCI) + 11.936(ET) + 6.923(T) + 2.453(P)	0.998	All years	I-VIII
27	Q = − 220.32 + 0.197(HYP) − 0.066(A) + 1.758(ET) + 3.231(P) + 0.002(MSL)	0.845	All years	I-VIII

16 and 17 were done using the accumulated yearly values of P and Q for the years 1967 and 1968, respectively. The average yearly values of P and Q were found for the combined years of 1967 and 1968, and Regression 18 was developed from the combined averages of the two years of data.

Eighteen regression analyses were performed using the eight parameters and the data collected from 10 watersheds for the periods specified above. Equation 8 was considered to have satisfactory statistical indices. It was the equation obtained by using the average monthly values of P and Q for the period of January through March of the year 1968. A summary of the P and Q values used in this study is given in Table XII of Appendix B.

At this point in the analyses the number of parameters was reduced, and mean sea level elevation (MSL) was substituted for mean elevation (ME), leaving six parameters, A, MSL, SCI, HYP, P, and Q. Also, the data from Watersheds IX and X were eliminated. Justification for this change in the parameters was based on the results of a study of a rotated factor analysis which is described in more detail in Chapter IV. The parameter MSL was substituted for ME because it was felt that the former was the better measure of the physiographic characteristics of the watersheds. The parameter, depth of soil (DT), was dropped because of the difficulty in obtaining this parameter in the field.

Watersheds IX and X were deleted from further analysis because the geologic conditions of the two watersheds are very different from those of the other watersheds under study. The internal drainage is poor with an impermeable sandstone stratum lying from 2 to 7 feet below the soil surface. Also, the altitude of the watersheds was at least 800 feet higher than that of any of the other watersheds. This difference in altitude is accompanied by a lower temperature for the winter months. Therefore, snow is retained for longer periods than on the other watersheds. Often, for the January through March period, Q was greater than P due to snowmelt.

Regression 19 was performed using the reduced number of parameters (A, MSL, SCI, HYP, P, and Q) for the average P and Q values for the years 1966, 1967 and 1968 for watersheds I through VIII. Yearly P and Q values are shown in Table XIII, Appendix B. The equation obtained was considered acceptable from a statistical viewpoint.

Regression 20 was done using the same six parameters including the P and Q data from all the years of record (Appendix B, Table XIII) from Watersheds I through VIII. The equation obtained from this analysis included only the parameters SCI, and the statistical indices were not satisfactory.

Regression 21 was performed using the six parameters including the average P and Q data from Watersheds I and II for the years 1964 through 1968.

At this point in the study the parameters average evapotranspiration (ET) and temperature (T) were added to the six used in Regressions 19 through 21. With these additions, the number of parameters used in the regression analyses was increased to eight. The independent parameters were A, MSL, SCI, HYP, ET, T and P, and the dependent parameter was Q. The next series of regression analyses was executed using all the data available at the time. This included data for Watersheds I through IV for the period 1964 through 1969 with 1967 excluded because the P and Q values greatly exceeded the normal levels for the data available for study. For Watersheds II and IV, which are in the same area as I and III, the annual P and Q values for the year 1967 were 74.82 and 32.03 inches, respectively for Watershed II. For Watershed IV in 1967 the annual P was 72.73 inches, and the annual Q value was 3.43 inches; whereas, the averages of P and Q for the other years of record were 47.14 inches and 8.15 inches, respectively, for Watershed II. For Watershed IV the average P value for the other years was 45.19 inches, and the average Q value for the same period was 0.28 inches. For Watersheds V through VIII, the P and Q values for the period 1966 through 1969 were used. Table XIII, Appendix B contains yearly observed P and Q values.

Four analyses were done using the average monthly values of P and Q for the four quarters of the year for the period of record 1966 through 1969. Regression 22 was obtained using the average monthly P and Q values for the quarter January through March for the period of record. For Regression 23, April through June average P and Q values were used; and for Regression 24, July through September values were used. For Regression 25, values for the period October through December were used. From this series one regression, Regression 22, was obtained which was considered satisfactory from a statistical standpoint.

All of the annual data were then combined, and Regression 26 was obtained from the average yearly P and Q values for the period of record for the eight watersheds. The equation obtained using this data had good statistical indices.

The water yield (Q) was found to be very sensitive to the parameter T. In an attempt to eliminate the sensitivity of the water yield equation to T, Regression 27 was done using all the parameters discussed above with T being eliminated, thus leaving parameters A, MSL, ET, SCI, HYP, P and Q for analysis. The results were not satisfactory from a statistical viewpoint.

CHAPTER IV

RESULTS AND DISCUSSION

I. RESULTS OF FACTOR ANALYSES

Factor analysis, as stated in Chapter III, was used to test the selected parameters for redundance and irrelevance. Eight factor analyses were made using the BMD03M program (16) with the parameters area (A), form factor (FF), mean elevation (ME), stream length (L), soil cover index (SCI), hypsometric slope factor (HYP), depth of top soil (DT), precipitation (P), and water yield (Q). These parameters were selected using the criteria previously described in Chapter III. Also, it was felt at this point in the study that they would characterize all the parameters available. The data for the four seasons January through March, April through June, July through September, and October through December from 10 watersheds for the years 1967 and 1968 were used to develop Factor Analyses 1 through 8. Table III gives the rotated factor matrix obtained for each factor, and the contributions of the loading corresponding to each parameter are given in Table IV. The data presented in Tables III and IV are typical of the eight rotated factor analyses obtained from the data described above.

Table III taken from a 9 by 9 matrix gives the rotated factor matrix obtained from a rotated factor analysis of data collected from Watersheds I through VIII for the period of January through March for the calendar year 1967.

TABLE III

ROTATED FACTOR MATRIX OF LOADINGS DEVELOPED FROM DATA FOR JANUARY THROUGH MARCH, 1967

Parameter	Symbol[a]	Factor Number				
		1	2	3	4	5
1	A	0.043	-0.989	0.087	0.084	0.061
2	FF	0.530	0.354	0.631	0.320	-0.269
3	ME	-0.964	0.062	-0.059	0.053	0.225
4	L	0.001	-0.994	0.041	-0.061	0.061
5	SCI	0.780	-0.014	0.323	0.324	-0.357
6	HYP	0.038	0.207	-0.955	-0.010	-0.203
7	DT	0.381	0.138	-0.198	0.269	-0.851
8	P	-0.829	0.018	0.227	-0.485	0.036
9	Q	-0.182	0.028	-0.102	-0.956	0.205
Factor[b] Contribution (%)		29.86	23.98	16.99	16.02	11.85

[a]Refer to Chapter III for definiti-n of symbols.

[b]Percent of total variance of rotated factor matrix explained by factor; the five most significant of the nine factors are given.

TABLE IV

CONTRIBUTIONS OF LOADING TO THE VARIANCE OF THE FACTORS FOR ROTATED FACTOR ANALYSES FOR DATA OF JAUNARY THROUGH MARCH 1967 (PERCENT)

Factor	Parameters[a]								
	A	FF	ME	L	SCI	HYP	DT	P	Q
1	0.07	10.46	34.61	0.0	22.64	0.06	5.39	25.55	1.24
2	45.06	5.80	0.18	45.83	0.0	1.98	0.88	0.01	0.04
3	0.46	26.06	0.23	0.11	6.82	59.69	2.55	3.39	0.68
4	0.49	7.10	0.19	0.26	7.27	0	5.02	16.33	63.34
5	0.35	6.78	4.72	0.35	11.92	3.88	67.91	0.12	3.96

[a]See Chapter III for definition of symbols.

The entries in Table III are known as factor loadings. They are measures of the degree to which a loading is related to a parameter. The loadings are expressed as decimals varying between -1.00 and +1.00. The closer the values are to +1.00 or to -1.00, the closer the loading is related to the parameter. A factor containing loadings that have high absolute values for more than one parameter indicates that the parameters might give similar information. Table III indicates the contribution of each factor toward the total variance of the rotated factor matrix. To find the contribution of each factor in percent, the sum of the squares of each parameter's contribution to each factor is computed along with the total sums of squares of the parameters' contributions to all the factors. The percent contribution of each factor is then found. For example, in factor 1 the sum of the squares of the contributions of each variable is 2.688 and the total sum of squares is 9.001. The percent contribution of factor 1, therefore, is

$$\frac{2.688}{9.001}\ (100) = 29.86\,.$$

In this case 98.70 percent of the variance of the rotated factor matrix is accounted for by the first five of the nine factors.

Factors 1, 3 and 4 of Table III are significantly associated with only one parameter. Here significant association is defined as a factor loadings exceeding an absolute value of 0.900. The factor loading for the other parameters are here considered relatively low (less than an absolute value of 0.900). This might indicate, under these conditions, that the parameters represented in these factors are relatively independent and should probably be retained for further analysis. Factor 2 contains two parameters, area and stream length, with significant loadings. This indicates, since the parameters A and L are logically related, that one of the parameters could be omitted from further analyses.

Table IV shows the percent contribution of each loading to each factor of Table III. To find the contribution of each loading to each factor, the sum of the squares of the contribution of each parameter to each factor is found. Then the square of each parameter's contribution is divided by the sum of squares of the contribution of each parameter to each factor. For example, the contribution of the loading for the parameter area (A) to Factor 1 Table III, is

$$\frac{(0.043)^2}{2.688} (100) = 0.070.$$

In Factor 2 of Table III, it is noted that the loadings corresponding to A and L make high and practically equal contributions to the total variance of Factor 2. This, also, is probable supporting evidence that one of the parameters is superfluous and could be deleted from further analyses. Using the information presented in Tables III and IV and the reasoning based on the interpretation of a rotated factor matrix, the parameters A, FF, ME, SCI, HYP, DT, P and Q were retained for further analyses. The parameter A was selected over stream length SL because it is more readily available, or it is relatively easy to measure.

At this point in the study, the parameters ET and T were added and regression equations were obtained containing parameters A, MSL, SCI, HYP, ET, T, P and Q. Then factor analyses were utilized to check for independence of the parameters. Table V is taken from the 8 by 8 rotated factor matrix of Factor Analysis 9. This matrix was obtained using the above parameters and the data collected from Watershed I through Watershed VIII for the periods listed below. For Watershed I through Watershed IV, data were collected for the years 1964 through 1969 with the exception of 1967. The year 1967 was excluded because the annual precipitation and annual runoff greatly exceeded the norm for the data available for this study. For Watershed V through Watershed VIII, data were included for the period 1966 through 1969.

For Factor 1 of the matrix (Table V), the loadings for T and SCI were high and nearly equal. This, however, does not indicate that one of the parameters could be eliminated from further analysis because the two parameters do not appear sufficiently closely related to indicate that one is a measure of the other. In Factor 2, P and HYP have high absolute values, but they are not logically related.

Table VI gives the contribution in percent of each loading to the variance of each factor of Table V. Here T and SCI make high and nearly equal contributions to the same factor. But since the two variables do not appear related from a hydrologic sense, neither of the two variables was eliminated from consideration.

To further check the independence of the parameters A, MSL, SCI, HYP, ET, T, P and Q used for the final series of regression analyses as discussed later in this chapter, a series of factor

TABLE V

ROTATED FACTOR MATRIX OF DATA FROM WATERSHEDS I THROUGH VIII FOR ALL YEARS OF RECORD

Parameter	Symbol[a]	Factor Number 1	2	3	4	5
1	A	0.159	-0.117	-0.189	0.962	0.004
2	MSL	-0.842	-0.490	-0.205	-0.073	0.049
3	SCI	0.979	0.088	0.152	-0.049	0.044
4	HYP	0.083	0.962	0.210	-0.020	0.152
5	ET	0.829	0.466	0.293	0.089	-0.031
6	T	0.963	0.030	0.083	0.235	-0.050
7	P	-0.328	-0.896	0.028	0.207	0.214
8	Q	0.291	0.146	0.919	-0.223	0.005
Factor contribution (%)		43.86	27.87	13.53	13.61	0.96

[a]See Chapter III for definitions of symbols.

TABLE VI

CONTRIBUTIONS OF LOADINGS TOWARD THE VARIANCE OF THE FACTORS FOR ROTATED ANALYSIS OF ALL YEARS OF RECORD (PERCENT)

	Parameters[a]							
Factor	A	MSL	SCI	HYP	ET	T	P	Q
1	0.72	20.22	27.33	0.19	19.58	26.45	3.07	2.41
2	0.61	10.75	0.34	41.51	9.76	0.04	36.01	0.96
3	3.30	3.87	2.14	4.07	7.91	0.63	0.07	78.00
4	84.84	0.49	0.38	0.04	0.72	5.04	3.93	4.54
5	0.0	3.14	2.49	30.14	1.18	3.28	59.76	0.0

[a]See Chapter III for definition of symbols.

analyses was performed using all data of the period of record for various combinations of parameters of the eight watersheds. Factor Analysis 10 was performed using all of the parameters except T. The factor loadings of the rotated factor matrix obtained by using all parameters except T indicated that none of the parameters were sufficiently closely related to permit eliminations of some of them. Factor Analysis 11 was performed using all parameters except SCI. In Factor Analysis 12, all parameters except T and SCI were considered. Factor Analysis 13 was performed with all parameters except ET. Factor Analysis 14 took into account all parameters except MSL. Factor Analysis 15 included all parameters except ET and MSL. In Factor Analyses 10 through 15 the results indicated, by the factor loadings of the rotated factor matrix, that none of the parameters were so related that some could have been eliminated.

II. MULTIPLE LINEAR REGRESSION EQUATIONS

The results obtained from the factor analyses as shown in Tables III and IV (pages 35 and 36), were further analyzed by multiple linear stepwise regression methods. Twenty-seven multiple linear regression analyses were performed using various parameters and data collected for various periods of time from the watersheds under study. From these 27 analyses, four regression equations were obtained which had acceptable statistical indices. The discussion of these four equations follows. A complete listing of all regression equations obtained is given in Table II (page 27). Part of the results of this study have been reported by Shelton, Haren and Sewell (32).

The results of Regression 8 are shown in Table VII. Eight parameters, area (A), form factor (FF), mean elevation (ME), soil cover index (SCI), hypsometric slope (HYP), depth of topsoil (DT), average monthly precipitation (P), and average monthly water yield (Q) for the period of January through March for the year 1968 for Watersheds I through X were used. Table VII shows that only the parameter form factor (FF) failed to appear in the final equation. Table VII shows also the relative predictive value (RPV), the multiple R^2, the standard error of estimate (S) in inches of water yield, and the significance level in percent with the convention of labeling the 5 percent probability level as significant. The RPV is a measure of the amount of improvement in the predictive value as determined by the multiple R. The RPV of an equation is determined according to

$$RPV = \frac{\text{Multiple R of equation n.}}{\text{Multiple R of equation 1}}$$

In Equation 1 of Table VII with only one parameter P entered, the relative predictive value was 1.00 (as a reference), the multiple R^2 was 0.833, S was 0.745 inches, and the significance level was 5 percent. The multiple R^2 values indicate that with the addition of a single parameter, no great increase resulted in the total variation accounted for; but with the inclusion of all six independent parameters, 98.6 percent of the variation was accounted for. The relative predictive value increased from 1.00 in Equation 1 to 1.09 in Equation 8. The standard error of estimate decreased from 0.745 inches in Equation 1 to 0.404 inches in Equation 8 of Regression 8. The significance level of

TABLE VII

STEPWISE REGRESSION 8 FOR PREDICTING WATER YIELD DEVELOPED FROM JANUARY THROUGH MARCH 1968 DATA FOR WATERSHEDS I THROUGH X

Equation	Independent Parameters[a]								Statistics[b]			
Number	C	A	ME	SCI	HYP	P	FF	DT	RPV	R^2	S	Sig
1	-19.878					5.479			1.000	0.833	0.745	5
2	-20.015	-0.007				5.605			1.039	0.899	0.627	5
3	-17.912	-0.009				5.324	-1.660		1.067	0.947	0.495	5
4	-18.029	-0.009				5.306	-2.073	0.160	1.075	0.962	0.471	5
5	-24.195	-0.006	0.017			6.361	-0.717	0.289	1.087	0.984	0.349	5
6	-25.010	-0.006	0.019	0.018		6.428	-1.065	0.269	1.088	0.985	0.415	5
7	-27.392	-0.005	0.023	0.018	0.009	6.641	-0.159	0.211	1.088	0.986	0.571	>5
8[c]	-27.737	-0.004	0.023	0.017	0.010	6.676		1.207	1.088	0.986	0.404	5

[a]See Chapter III for definition of symbols. C is the regression coefficient.

[b]RPV is the relative predictive value; R^2 is multiple R squared; S is the standard error of estimate in inches; and Sig is the probability level of the equation in percent.

[c]Equation 8 written in conventional form is: $Q = -27.737 - 0.004(A) + 0.023(ME) + 0.017(SCI) + 0.010(HYP) + 6.676P + 1.207DT$.

each of the equations is 5 percent with the exception of Equation 7 which was not significant. The probable reason for Equation 7 dropping below the specified level of significance was that the degrees of freedom of the lesser mean square dropped to 1 with the degree of freedom of the greater mean square being 7. In Equation 8 after the parameter FF was removed, the degrees of freedom of the greater and lesser mean squares were 6 and 2, respectively. With this increase in the degrees of freedom of the lesser mean square, the significance level of the equation raised above the specified level.

The measured value of mean water yield versus the computed value of water yield for the January through March period as calculated by Regression 8 is given in Figure 5, where the "equal" regression line is shown as an even-dash line, and the solid line is the best-fit line which was obtained using all data on which the regression was based. The broken line shows the best-fit line determined from data collected from all watersheds except Watersheds III and IV which have a cover of good permanent pasture. The Q data for the two watersheds were deleted because the predicted values far exceeded the measured values. Runoff from watersheds with good permanent pasture cover is usually expected to be greater than that for good hardwood forest cover on Group B soils according to the U.S. Soil Conservation Service (22), and Schwab, et al. (30). Therefore, it is believed by the author that this difference in measured and computed values of Q is due to geologic conditions which are different from that of the other watersheds used in the development of the prediction equations. The equation of the

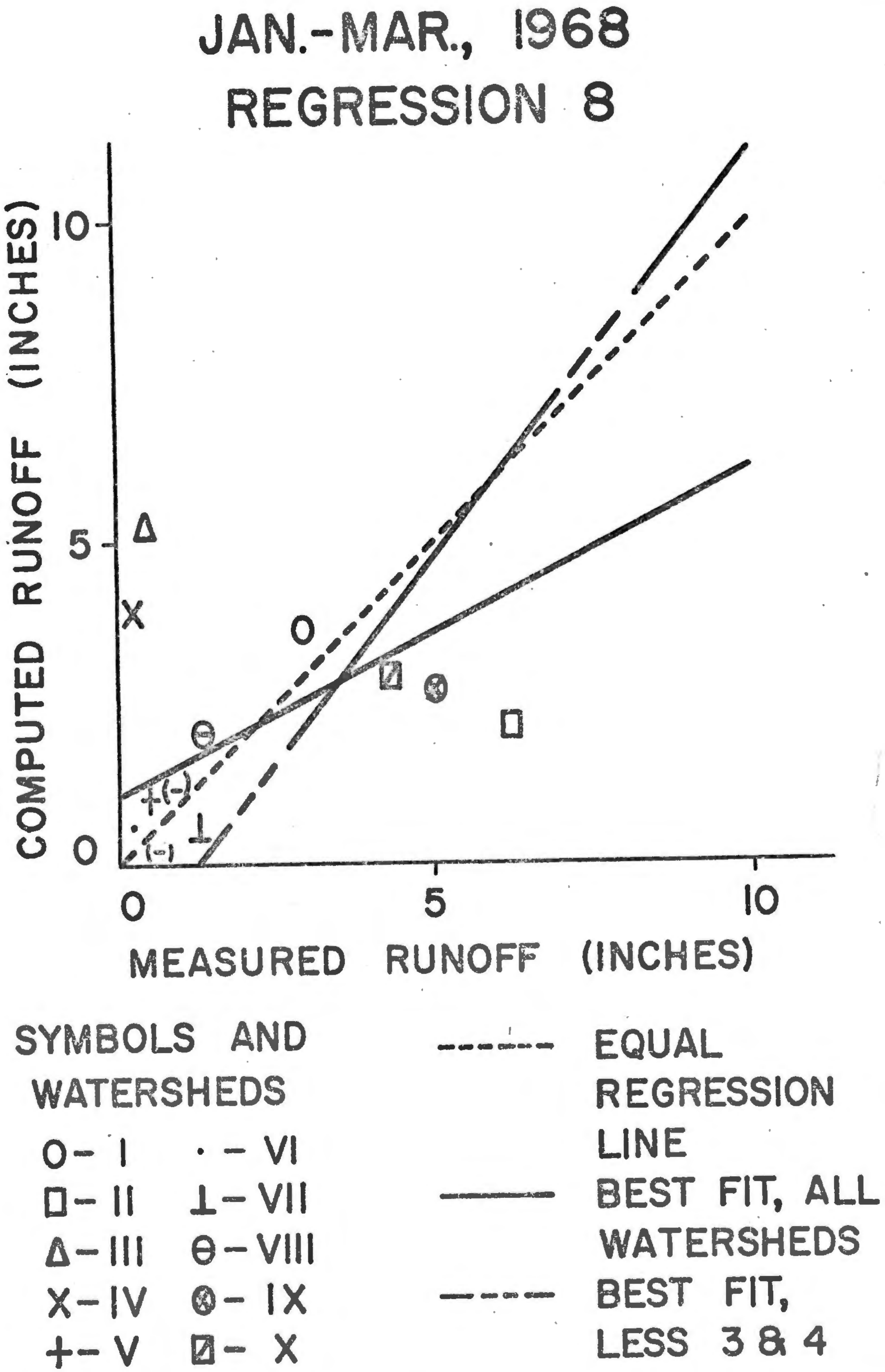

Figure 5. Computed and Measured Water Yield for January Through March, 1968.

best-fit line (solid line) for all watersheds shown in Figure 5 is

$$CQ = 1.098 + 0.500\ MQ.$$

where CQ is the computed water yield and MQ is the measured water yield.

Figure 5 suggests that the equation obtained from the data of all the watersheds overpredicts Q for Watersheds III and IV and underpredicts it for Watershed II. Figure 5 suggests also that the computed values of Q for Watersheds V and VI are small negative values. These negative values were not considered to invalidate the regression because the measured values of average water yield are small positive values. The calculated numerical differences for MQ and CQ for Watersheds V and VI differed by less than two inches.

Watersheds III and IV were eliminated, and a best-fit regression line was computed the equation of which is

$$CQ = 1.46 + 1.29\ MQ.$$

The graph of this equation is shown as the broken line in Figure 5. Watersheds III and IV were deleted because the observed Q values were much lower than the computed values. An explanation for this is given with the discussion of Table I, in the Methods and Procedures Chapter. Figure 5 indicates also that, when the best-fit regression line is computed with the deletion of Watersheds III and IV, the best-fit line more nearly approaches the "equal" regression line. The computed water yields were calculated from the same data on which the regressions were based. The author realizes this procedure is undesirable from a statistical point of view; however, no other data were available at the time.

Table VIII, presenting the data of Regression 22, lists the equations obtained by performing a stepwise regression analysis on the parameters A, MSL, SCI, ET, mean monthly T, mean monthly P, and mean monthly Q for the January-through-March period for all years of record for Watersheds I through VIII.

In Equation 1 of Table VIII where ET is the first variable entered, the RPV is 1.00 and the standard error of the estimate is 0.80 inches. The significance level of this equation is 25 percent, and only 24 percent of the total variation was accounted for by ET. With the addition of the variables MSL and T, the RPV doubled, the multiple R^2 increased to 0.954, and the standard error of estimate decreased to 0.24 inches. The significance level of the equation after the addition of the variables MSL and T was 5 percent. With the addition of variables A, SCI and P, the RPV increased slightly to 2.047, the multiple R^2 also had a slight increase to 0.996, and the standard error of estimate decreased to 0.06 inches. The significance level of Equations 4, 5, and 6 was 5 percent.

The following observations are made from Table VIII using multiple R^2 as a measure of variance explained:

1. Little was contributed by the parameters SCI and P toward explaining the variance.
2. Approximately 98 percent of the variance was accounted for by the parameters ET, MSL, T, and A.

TABLE VIII

STEPWISE REGRESSION 22 PREDICTING WATER YIELD DEVELOPED FROM JANUARY THROUGH MARCH DATA FOR ALL YEARS OF RECORD FOR WATERSHEDS I THROUGH VIII

Equation	Independent Parameters[a]								Statistics[b]			
Number	C	A	MSL	SCI	HYP	ET	T	P	RPV	R^2	S	Sig
1	- 2.800					2.250			1.000	0.239	0.804	25
2	- 13.051		0.003			6.665			1.437	0.493	0.719	25
3	-126.02		0.011			26.899	1.805		2.000	0.954	0.243	5
4	-129.79	-0.003	0.011			27.068	1.899		2.026	0.978	0.191	5
5	-150.86	-0.005	0.009	-0.046		29.741	2.398		2.043	0.995	0.111	5
6[c]	-150.40	-0.005	0.009	-0.043		29.239	2.387	0.272	2.047	0.996	0.059	5

[a]See Chapter III for definition of symbols. C is the regression coefficient.

[b]See Table VI for definition of symbols.

[c]Equation 6 written in conventional form is: Q = - 150.40 - 0.005(A) + 0.009(MSL) - 0.043(SCI + 29.239(ET) + 2.387(T) + 0.272(P).

3. The parameters MSL, T, and P tended to increase the computed annual water yield.

The algebraic sign for the parameter A of Equation 6 of Regression 22 appears to be contradictory to that which is generally considered logical from a hydrologic sense. Many investigators feel that as area increases so should the mean annual water yield.

The apparently inverse effect of area (A) on the water yield equations of this study could be due to the fact that some of the smaller watersheds are cultivated while most of the larger watersheds have vegetative cover. Thus the smaller watersheds tended to produce a greater water yield per unit area than did the larger watersheds.

Equation 6 of Regression 22 was tested using all the data available from all watersheds. The results are shown in Figure 6. Here, also, the "equal" regression is shown as an even-dash line; the best-fit line developed from data for all watersheds is the solid line; and the best-fit line determined from data excluding Watersheds III and IV is shown as the broken line. The equation of the best-fit line for all watersheds is

$$CQ = 3.468 + 0.195MQ$$

where CQ is the computed mean monthly water yield and MQ is the mean monthly measured water yield for the period January through March for all years of record. The equation of the best-fit line without data from Watersheds III and IV is

$$CQ = 2.73 + 0.35MQ$$

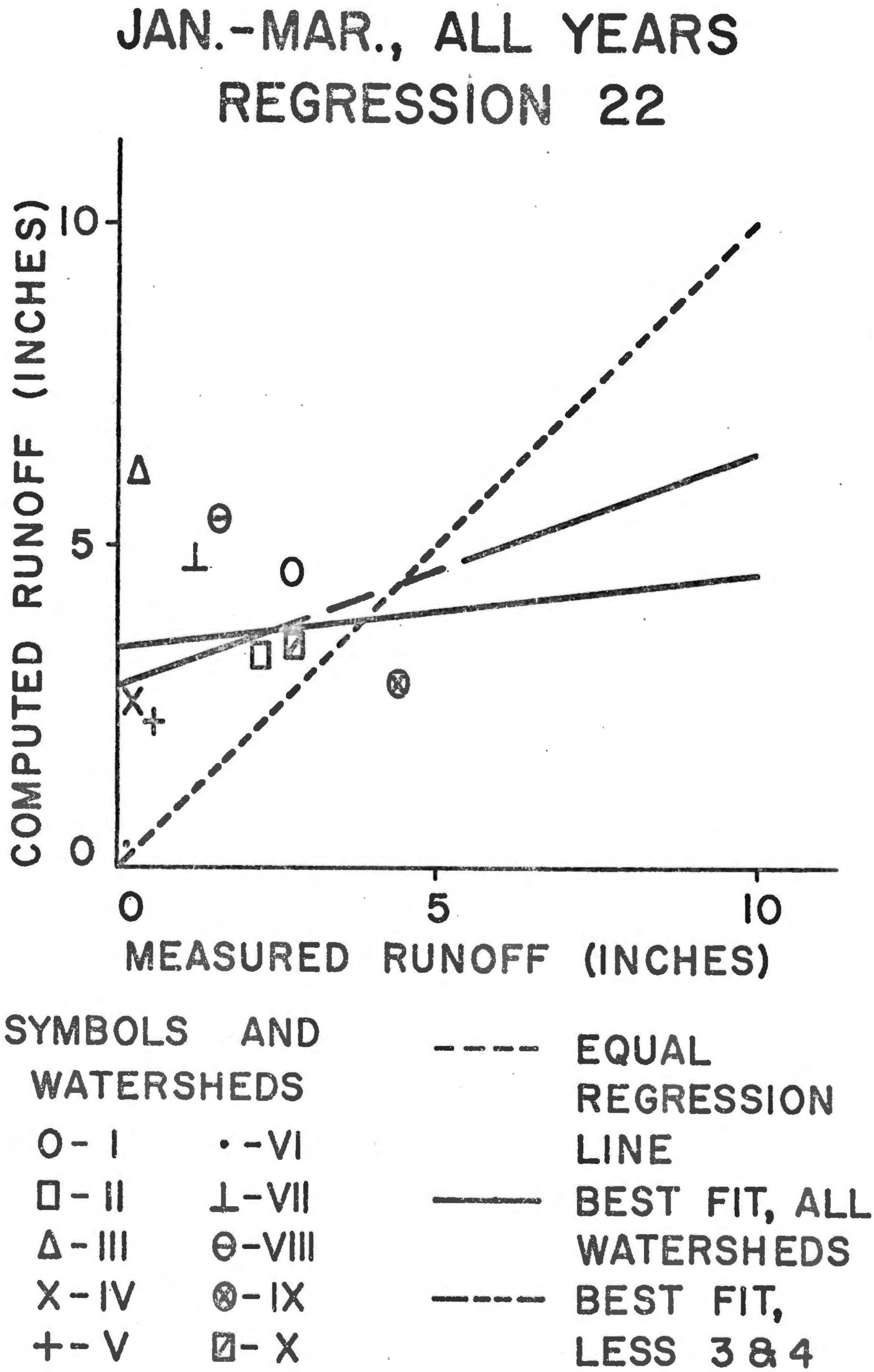

Figure 6. Computed and Measured Water Yield for January Through March For All Years.

The deletion of Watersheds III and IV has been previously discussed. As shown in Figure 6 and based on Regression 22, the regression line of best-fit was improved slightly with the deletion of Watersheds III and IV.

Table IX gives the equations obtained in the development of Regression 19. A stepwise regression analysis was performed on data for the parameters A, MSL, SCI, HYP, P, and Q for the years 1966, 1967, and 1968 for Watersheds I through VIII. The equations of Table IX are presented to show some of the typical results obtained from the regression analyses that were performed and to illustrate why many of the equations obtained were rejected.

From a statistical viewpoint, Equations 4 through 6 of Table IX (Regression 19), appear to be satisfactory. The relative predictive value varies from 1.00 for Equation 1 to 1.30 for Equations 5 and 6. Equation 1 indicates that 57 percent of the total variation was accounted for by the precipitation parameter. With the addition of SCI, HYP, and MSL, 96 percent of the variation was explained. All of the equations were significant at the 10 percent level except Equation 5 which was significant only at a very high level. The standard error of estimate decreased from 5.07 inches for Equation 1 to 2.51 inches for Equation 6.

Equation 6 suggests that increases in the values of the parameters MSL, SCI, and HYP tend to increase annual water yield predictions. According to Equation 6, annual water yield is not dependent upon the area of the watershed, and annual precipitation affects water yield in an apparently inverse manner.

TABLE IX

STEPWISE REGRESSION 19 PREDICTING AVERAGE ANNUAL WATER YIELD DEVELOPED FROM ANNUAL DATA OF WATERSHEDS I THROUGH VIII FOR 1966 THROUGH 1968

Equation	Independent Parameters[a]						Statistics[b]			
Number	C	A	MSL	SCI	HYP	P	RPV	R^2	S	Sig
1	58.861					-1.062	1.000	0.570	5.069	5
2	63.601	-0.044				-1.117	1.146	0.749	4.331	10
3	79.241	-0.049		-0.259		-1.229	1.238	0.874	3.545	10
4	76.357	-0.043		-0.293	-0.774	-1.308	1.269	0.911	3.517	20
5	-67.283	-0.012	0.073	1.323	0.268	-1.366	1.300	0.960	3.475	>>5
6[c]	-40.112		0.059	1.020	0.231	-1.356	1.300	0.958	2.509	10

[a]See Chapter III for definition of symbols. C is the regression coefficient.

[b]See Table VI for definition of symbols.

[c]Equation 6 written in conventional form is: $Q = -40.112 + 0.059(MSL) + 1.020(SCI) + 0.231(HYP) - 1.356(P)$.

Water yield predictions for each watershed were made using Equation 6 and the data from which Equation 6 was developed. The yield computed by Regression 19 along with the observed yields are presented in Figure 7 for the years 1966, 1967, and 1968. The "equal" regression line is shown as an even-dash line; the best-fit line as determined from data for Watersheds I through VIII is the solid line; and the best-fit line is shown as the broken line in Figure 7. The equation of the best-fit line shown in Figure 7 excluding Watersheds III and IV is

$$CQ = 1.319 + 0.565\ MQ$$

where CQ indicates the computed annual water yield and MQ indicates the measured annual water yield. The equation of the best-fit line for all watersheds is

$$CQ = 9.677 - 0.096\ MQ.$$

A much closer relationship between the equal regression line and the line of best-fit as shown in Figure 7 was obtained by deleting predictions for Watersheds III and IV. Watersheds III and IV were deleted because the observed Q values were much lower for Watersheds III and IV than for the others as is explained in the discussion of Table I.

Notwithstanding the elimination of the parameter A from the final regression Equation 6 and the apparently inverse relationship between precipitation and water yield, the best-fit line excluding

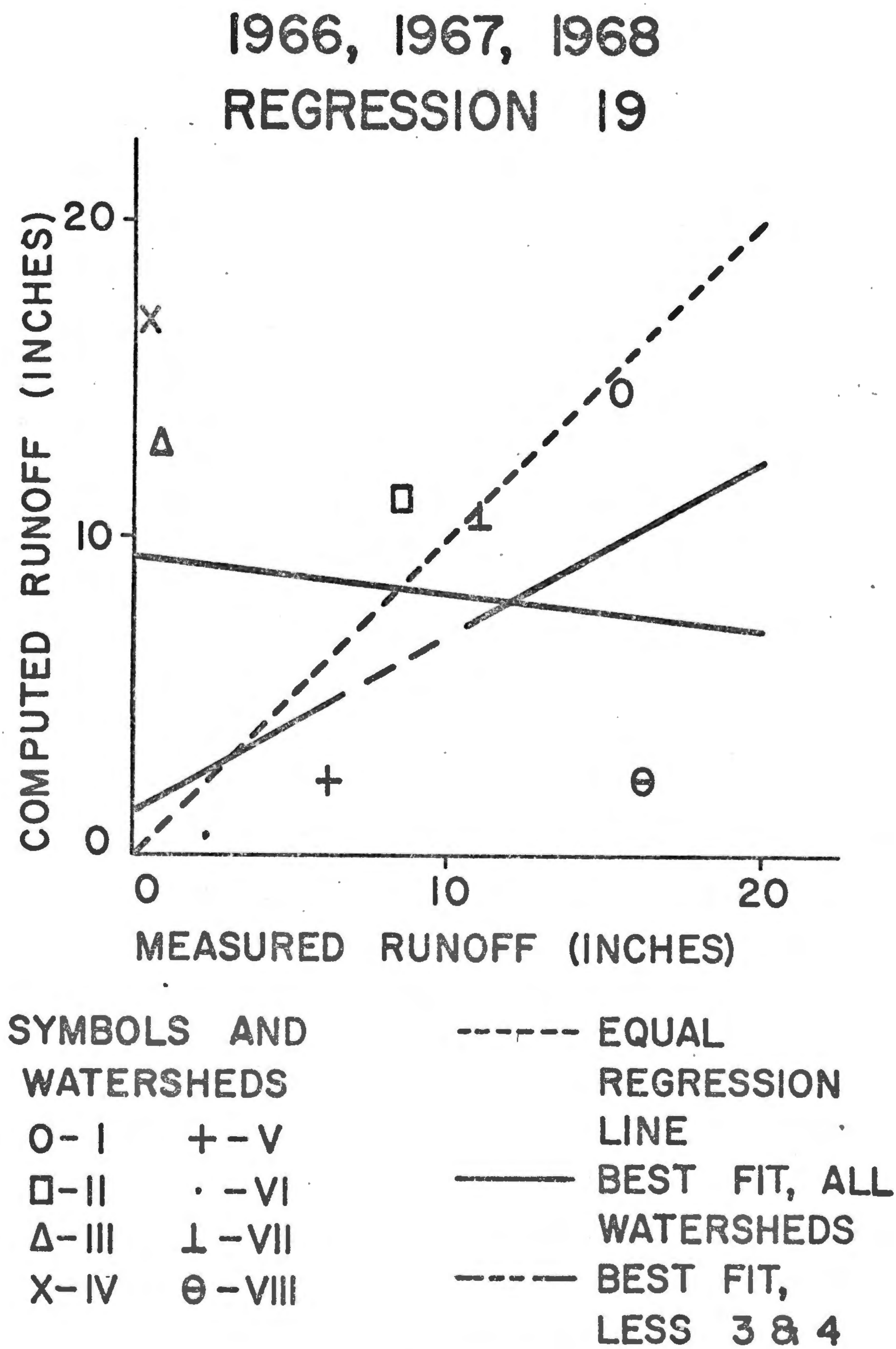

Figure 7. Computed and Measured Annual Water Yield for Years 1966, 1967, 1968.

Watersheds III and IV still agrees reasonably well with the equal regression line.

Regression 26 which was developed by using annual averages of all years of record for Watersheds I through VIII is given in Table X. Of the seven parameters (A, MSL, SCI, HYP, ET, T, and P) which were entered as independent parameters in the regression analysis, only the hypsometric slope parameter failed to make a significant contribution and was not entered into the multiple linear regression equations. Also shown in Table X is the relative predictive value (RPV), the multiple R^2, the standard error of estimate (S) in inches and the significance level in percent.

In Equation 1 of Table X (Regression 26) where the first parameter ET was entered into the regression, the relative predictive value is 1.00, the R^2 value is 0.312, and the standard error of estimate is 5.675 inches at a significance level of 25 percent. With the addition of three other parameters, A, MSL, and SCI, the relative predictive value increased to 1.661, the R^2 became 0.860, and the standard error of estimate dropped to 3.615 inches. The significance level of the regression equation at this point was 10 percent. Equation 6 shows that with the addition of P and T the relative predictive value reached 1.790, the multiple R^2 value became 0.998, the standard error of estimate decreased to 0.29 inches and the significance level was 5 percent suggesting that this regression explains almost all of the variance.

Using the multiple R^2 as a measure of the variance explained, the following observations may be made from Table X. Little was

TABLE X

STEPWISE REGRESSION 26 PREDICTING AVERAGE ANNUAL WATER YIELD FOR WATERSHEDS I THROUGH VIII FOR ALL YEARS OF RECORD

Equation	Independent Parameters[a]							Statistics[b]			
Number	C	A	MSL	SCI	ET	T	P	RPV	R^2	S	Sig
1	- 33.332				1.168			1.000	0.312	5.675	25
2	-539.580		0.132		12.691			1.479	0.682	4.224	10
3	-534.780	-0.041	0.131		12.649			1.651	0.851	3.234	5
4	-539.260	-0.042	0.129	-0.103	12.934			1.661	0.860	3.615	10
5	-425.540	-0.090	0.015	-0.806	7.690		4.156	1.776	0.985	1.472	5
6[c]	-951.530	-0.107	0.087	-1.047	11.936	6.923	2.453	1.790	0.998	0.290	5

[a]See Chapter III for definition of symbols. C is the regression coefficient.

[b]See Table VI for definition of symbols.

[c]Equation 6 written in conventional form is: $Q = -951.530 - 0.107(A) + 0.087(MSL) - 1.047(SCI) + 11.936(ET) + 6.923(T) + 2.453P$.

contributed toward explaining the variance by the parameters SCI and T, and the parameters ET, MSL, and A accounted for about 85 percent of the variation in mean annual water yield. The parameter P accounted for about 12 percent; therefore, the parameters ET, MSL, A, and P accounted for about 98 percent of the variation in the mean annual water yield.

Regression 26 was tested against data collected from the same watershedsas were used to develop the equation. The results are shown in Figure 8 where measured annual water yield and computed annual water yield are plotted. The even-dash line indicates the "equal" regression line and the solid line is the best-fit curve obtained by least squares technique. The equation of the best-fit curve is

$$CQ = 0.923 + 1.075\ MQ.$$

Figure 8 suggests that Equation 6 of Regression 26 predicts the annual water yield for the period of record rather well. The equation slightly overpredicted water yield for Watersheds I, IV, V, and VIII.

Since the equations were developed, P and Qm (measured values of precipitation) data from Watershed V, 46.67 inches and 5.06 inches, respectively, and P and Qm for Watershed VI, 46.67 and 2.72 inches, respectively, have become available for 1970. Using Regression 19, the computed annual water yields (Qc) were 14.09 and 3.29 inches for Watersheds V and VI, respectively. Using Regression 26 with the long-term average for mean annual temperature, Qc was 5.60 and -2.28 for

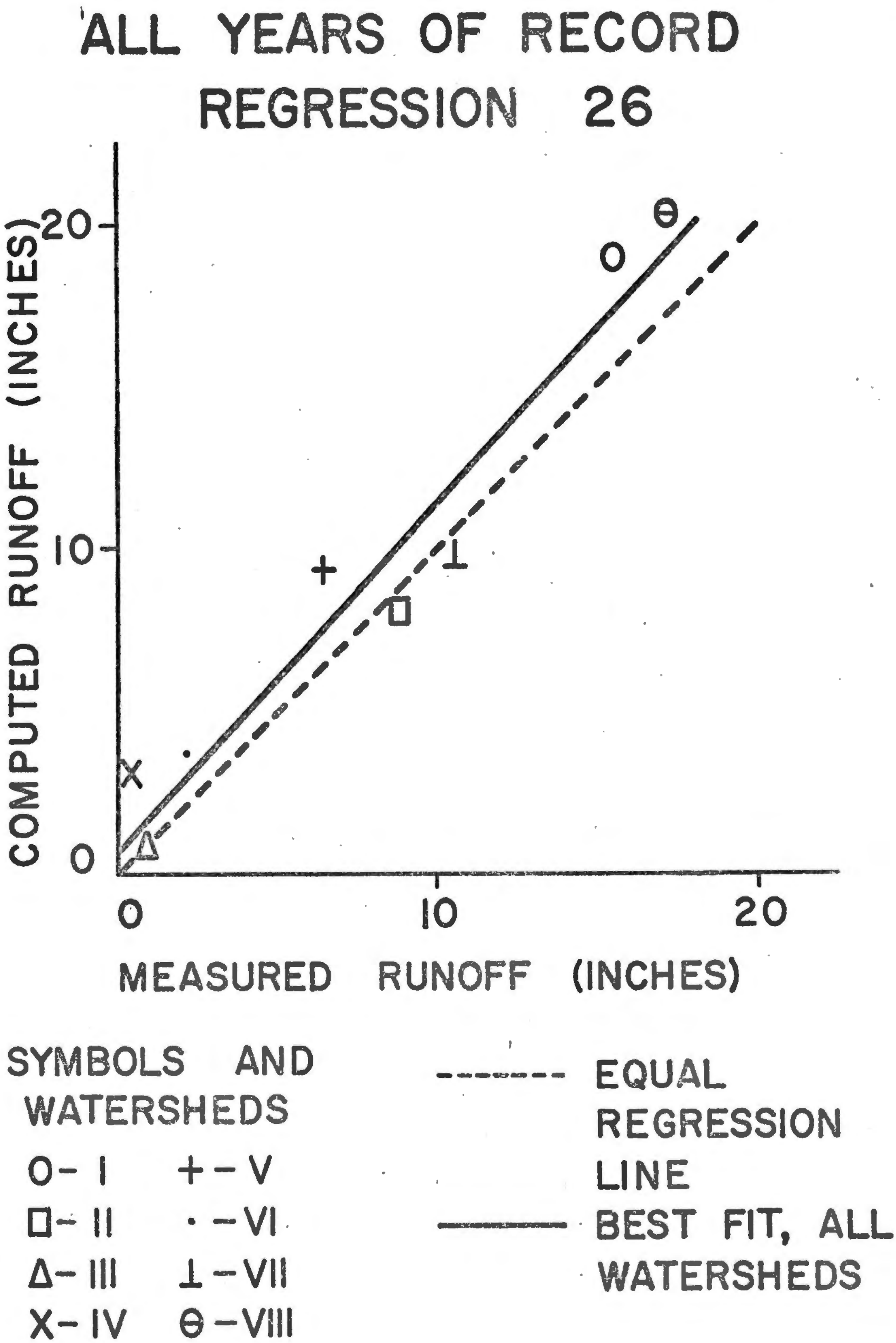

Figure 8. Computed and Measured Annual Water Yield For All Years Of Record.

Watersheds V and VI, respectively.

The Qc values do not closely agree with the Qm values. This might suggest that a regression developed from long-term averages of various parameters would be expected to give better predictions based on long-term averages of data rather than on one year's data. The data used in making predictions must lie near the means of those data on which the prediction technique used was developed.

CHAPTER V

SUMMARY AND CONCLUSIONS

I. SUMMARY

The objectives of this study were to: (1) examine the feasibility of using factor analysis and multiple linear regression techniques in the development of equations to predict water yield from selected small watersheds in Tennessee on a seasonal and annual basis; and (2) to investigate the ability of the models developed to predict the water yields from the various watersheds for periods of varying lengths.

To achieve these objectives, data which had been collected for various periods of time from 10 Tennessee Watersheds were first analyzed by factor analysis. From the 15 parameters available from 10 watersheds at the beginning of this study, the following nine: area (A), form factor (FF), mean elevation (ME), precipitation (P), water yield (Q), stream length (L), soil cover index (SCI), hypsometric slope (HYP), and depth of top soil (DT) were selected for initial investigation by factor analysis. This analysis suggested that A and L were measures of the same characteristic of the watersheds and that one could be deleted from further analysis. Area was selected as the parameter to retain because of its ease of determination.

Eighteen regression analyses were performed using the eight parameters A,FF, ME, SCI, HYP, and DT with P and Q data for time periods of varied lengths. Regression 8,

$$Q = -27.737 + 6.676(P) - 0.004(A) + 1.207(DT) + 0.023(ME) + 0.017(SCI) + 0.010(HYP),$$

which was obtained from this series of analyses was considered to have satisfactory statistical indices. It was derived using the mean monthly P and Q data for the period of January through March of the year 1968. Its standard error of estimate was 0.40 inches, and its significance level was 5 percent. The multiple R^2 value indicated that 98.6 percent of the variation of the Q data was explained by the regression.

This equation, under conditions existing at the time the data for this study were collected, overpredicts Q for Watersheds III and IV. By deleting the Q data from Watersheds III and IV, the best-fit line obtained by least squares techniques more nearly approached the "equal" regression line. The Q data for the two watersheds were deleted because the predicted values far exceeded the measured values. This is believed to be caused by geologic conditions which are different from the other watersheds used in the development of the prediction equations since runoff from good permanent pasture is usually expected to be greater than that for good hardwood forest on Group B soils.

In later analyses the number of parameters was reduced to six (A, mean sea level (MSL), SCI, HYP, P, and Q), and MSL was substituted for ME. Also the data from Watersheds IX and X were dropped because the geologic conditions of the two watersheds are very different from those of the other watersheds under study. Three regressions were performed using these six parameters, and Regression 19,

$$Q = -40.112 - 1.356(P) + 1.020(SCI) + 0.231(HYP) + 0.059(MSL),$$

was obtained from mean annual P and Q values for the years 1966, 1967 and 1968. The equation has satisfactory statistical indices in that the multiple R^2 was 0.96, the standard error of estimate was 2.51 inches, and the regression was significant at the 10 percent level.

This regression suggests that for the relatively small watersheds of this study, mean annual water yield is independent of the area of the watershed and that mean annual precipitation affects water yield in an apparently inverse way. In this case, parameters other than area probably exerted the greatest effect on water yield. Even with the absence of the area parameter and the apparently inverse relationship between precipitation and water yield, the best-fit line developed from data excluding that of Watersheds III and IV agrees reasonably well with the equal regression line.

Evapotranspiration (ET) and mean annual temperature (T) were added to the parameters, thus increasing the number of parameters to eight which were A, MSL, SCI, HYP, ET, T, P, and Q. Six regression analyses were performed using these eight parameters for time periods of various lengths, and all data available at this time were used except that for the year 1967 for Watersheds I through IV which were eliminated because the P and Q data far exceeded the normal for the data available for this study.

Two regressions with satisfactory statistical indices were obtained using the mean monthly values of P and Q for the January through March period for all the years of record except 1967. Regression 22,

$$Q = -150.40 + 29.239(ET) + 0.009(MSL) + 2.387(T) - 0.005(A) - 0.046(SCI) + 0.272(P).$$

was considered to have satisfactory statistical indices in that the standard error of estimate was 0.06 inches, the significance level of the regression was 5 percent, and the multiple R^2 was 0.996. This regression indicates, under the conditions that existed when the data were collected, that very little was contributed by the parameters SCI and P toward explaining the variance of water yield. The regression suggests that water yield variance was affected mostly by the parameters MSL, T, and ET. This regression indicates also that the parameter A affects Q in a manner contrary to that which is generally accepted by hydrologists. This could have been because most of the smaller watersheds with high water yields were cultivated and therefore more water yield per unit area would occur on them.

Computations of Q based on this equation show that water yields for Watersheds III and IV were overpredicted and those for Watersheds II were underpredicted. Deleting Q data from Watersheds III and IV made the regression line of best-fit more nearly approach the "equal" regression line.

The second regression, Regression 26,

$$Q = -951.530 - 0.107(A) + 0.087(MSL) - 1.047(SCI) + 11.936(ET) + 6.923(T) + 2.452(P),$$

was developed from the average annual data for Watersheds I through VIII for all years of record.

The statistical indices of the equation appear to be satisfactory. The standard error of estimate was 0.290 inches, and the multiple R^2 was 0.998. The relative predictive value increased from a value of 1.00 in the first stepwise regression equation to 1.79 in the final stepwise regression equation. The significance level of this equation was 5 percent. The multiple R^2 values indicate that ET, MSL, A, and P accounted for about 98 percent of the variation. The parameter A again appears to affect water yield in an inverse manner. A possible explanation has been previously given in the Summary.

The regression equation was tested against all data available from all watersheds including the year 1967. The equal regression line of measured Q versus predicted Q and the regression line of best-fit obtained by the least-squares technique have almost equal slopes, and intercepts differ by less than one inch. Thus the regression developed from the average annual data from the eight watersheds predicted rather accurately the observed water yields.

II. CONCLUSIONS

The results of this study show that equations predicting water yield can be developed from a variety of parameters associated with the watersheds under study. These equations have been shown, in many cases, to give satisfactory results when the data on which predictions are based fall within the means of the data used for development of the respective prediction equations. Attempting to make predictions based on data falling outside these means or attempting to use a prediction technique developed for dissimilar watersheds has been demonstrated to give misleading and incorrect estimates.

A. Mathematical Models

The following specific conclusions concerning the mathematical models were drawn from the results of this study:

1. Factor analysis can be used to assist in screening superfluous parameters thereby reducing the number of parameters required to characterize the hydrologic properties of watersheds.

2. Prediction equations were derived using different parameters for the same watersheds, and these equations often produced satisfactory predictions as long as the data used in making the predictions were near the mean values of the parameters used in developing the prediction equations. The best results were obtained using mean data collected over a long period of time.

3. Watersheds must be grouped only with those having similar hydrologic characteristics and, especially, similar geologic characteristics.

B. Predictive Ability

Based on the ability of the models to predict water yields and parameter selection, the following conclusions were drawn:

1. The best applications of linear regression modeling in this study, from a seasonal standpoint, appeared to be for those

periods of the year where the hydrologic conditions of the watersheds were relatively uniform. This was especially true for the January-through-March period.

2. Due to the sensitivity of the parameters evapotranspiration, temperature, and especially precipitation, if used in a prediction equation, these parameters must be based on precise field measurements over small incremental areas.

3. The parameter precipitation affected the predicted water yield in a positive manner for the January through March period in the majority of regressions; however, no trends were evident for the other seasons of the year.

4. The hypsometric factor as evaluated in this study either did not enter the regression equation or it made only a very small contribution toward explaining water yield based on the coefficient of determination, R^2.

5. The parameter, form factor, did not enter the equation, or it accounted for a very small amount of variation.

In studying and applying the regressions developed in this study, the absolute values and algebraic signs of individual regression coefficients probably should not be examined without taking into account the implications of the entire regression.

III. RECOMMENDATIONS FOR FURTHER STUDY

Future studies on the use of multiple linear regression techniques in deriving prediction equations should be concerned with the following:

1. Since the equations developed were quite sensitive to the parameters evapotranspiration and temperature, the data for these parameters should be determined for each watershed location.

2. An increase in the number of watersheds having similar hydrologic and geologic characteristics would increase the amount of data available for study; therefore, better statistical results would be expected due to increased degrees of freedom available.

3. Other methods of characterizing the factors affecting water yield should be attempted. That is, water yield is not necessarily directly related to the measure of the parameters, but it may be related by some other function.

4. In order to obtain a water-yield equation for the growing season, the author feels that the above-normal rainfall intensity and duration parameters should be characterized and taken into account in the analysis.

5. A study of the effect of each parameter upon water yield should be made in order to develop a model that will adequately describe the contributions of each parameter to water yield.

REFERENCES

REFERENCES

1. Amorocho, J. and W. E. Hart, A critique of current methods in hydrologic systems investigations. Transactions of American Geophysical Union, Vol. 45, No. 2, pp. 307-321, 1964.

2. Amorocho, J., and G. T. Orlob, An evaluation of the inflow-outflow relations in hydrologic studies. Water Resources Center Contribution No. 14, University of California, Berkeley, Calif. 1961.

3. Beard, L. R., Statistical methods in hydrology. U.S. Army Engineer District, Corps of Engineers, Sacramento, Calif. 1962.

4. Beasley, R. P., Estimating peak rates of runoff from small watersheds, Science and Technology Guide, University of Missouri-Columbia Extension Division. Columbia, Missouri. 1969.

5. Bradshaw Creek-Elk River, A pilot study in area-stream factor correlation. Tennessee Valley Authority Office of Tributary Area Development, Knoxville, Tennessee. Research Paper No. 4. 1964.

6. Buie, E. C., L. M. Glymph, and W. H. Heneberry, Development and the role of water, A good life for more people. The Yearbook of Agriculture. United States Department of Agriculture, United States Printing Office, Washington, D. C. pp. 345-350. 1971.

7. Chow, V. T., Hydrologic determination of waterway areas for the design of drainage structures in small drainage basins. Illinois Engineering Experiment Station, Urbana, Illinois. Bulletin No. 462. 1962.

8. Climatological Data National Summary for Years 1964 through 1969. Weather Bureau, United States Department of Commerce, Environmental Science Service Administration, Environmental Data Service, United States Printing Office, Washington, D. C. Vol. 15 through vol. 20.

9. Cook, H. L., The infiltration approach to the calculation of surface runoff. Transactions of the American Geophysical Union, Vol. XXVII. pp. 726-743. 1946.

10. Crawford, N. H. and R. K. Linsley, Digital Simulation in Hydrology: Stanford Watershed Model IV, Civil Engineering Department, Stanford University, Stanford, California. Technical Report No.39. 1066.

11. Dawdy, D. R. and J. H. Feth, Application of factor analysis in the study of chemistry of ground water quality. Mojave River Valley, California. Water Resources Research, 3, No. 2, pp. 505-510. 1967.

12. Dawdy, D. R., R. W. Lichty and J. M. Bergmann, A rainfall-runoff simulation model for estimation of flood peaks for small drainage basins--synthesis in hydrology. United States Geological Survey Professional Paper 427-A. 1963.

13. Design of a hydrologic condition survey using factor analysis. Division of Water Control Planning, Hydraulic Data Branch, Tennessee Valley Authority, Knoxville, Tennessee. Research Paper No. 5. 1965.

14. Diaz, G., J. I. Sewell, and C. H. Shelton, An application of principal component analysis and factor analysis in the study of water yield. Water Resources Research, Vol. 4, No. 2, pp. 299-306. 1968.

15. Diskin, M. H., Definition and uses of the linear regression model, Water Resources Research, Vol. 6, No. 6, pp. 1668-1673. 1970.

16. Dixon, W. J. (ed.), Biometrical computer programs. Health Services Computing Facility, University of California, Los Angeles. 1965.

17. Gray, D. M., Derivation of hydrographs for small watersheds from measurable physical characteristics. Iowa State University, Department of Agricultural Engineering, Ames, Iowa. Research Bulletin 506. 1962.

18. Harrold, L. L., Minimum water yield from small agricultural watersheds. Transactions of the American Geophysical Union, Vol. 38, No. 2, pp. 201-208. 1957.

19. Harrold, L. L., et al., Influence of land use and treatment on the hydrology of small watersheds at Coshocton, Ohio, 1938-1957. United States Department of Agriculture. Technical Bulletin No. 1256. 1962.

20. Huggins, L. F., and E. J. Monke, The mathematical simulation of the hydrology of small watersheds. Purdue University Water Resources Research Center, Lafayette, Indiana. Technical Report No.1. 1966.

21. Hydrology of small watersheds in relation to various covers and soil characteristics. Cooperative Research Project in Western North Carolina, Tennessee Valley Authority and North Carolina State College of Agriculture and Engineering. Tennessee Valley Authority, Knoxville, Tennessee. 1960.

22. Hydrology Supplement A, Engineering Handbook, Soil Conservation Service, United States Department of Agriculture. U.S. Printing Office, Washington, D. C. 1957.

23. Johnstone, D., and W. P. Cross, Elements of applied hydrology: The Roland Press Co., New York, New York, pp. 236-269. 1949.

24. Kahler, K. O., Trends in the utilization of water. Water the Yearbook of Agriculture, United States Department of Agriculture, U. S. Printing Office, Washington, D. C., pp. 35-38. 1955.

25. Lillard, J. H., C. H. Shelton, and T. V. Wilson (eds.), Hydrologic data summaries from small agricultural watersheds in the southern region, 1962-1967, Southern Cooperative Series, South Carolina Agricultural Experiment Station, Clemson, South Carolina. Bulletin No. 147. 1969.

26. Linsley, R. K., Jr., M. A. Kohler, and J. L. Paullus, Hydrology for Engineers: McGraw-Hill Book Company, Inc., pp. 311-321. 1958.

27. Miller, C.R., and W. Viessman, Jr., Runoff volumes from small urban watersheds, Water Resources Research, 3, pp. 429-434. 1972.

28. Occurrence of drought in the Tennessee Valley, Division of Water Control Planning, Hydraulic Data Branch, Tennessee Valley Authority, Knoxville, Tennessee. T 58-2AE. 1958.

29. Reich, B. M., Design hydrograph for small watersheds from rainfall. Unpublished Doctoral Dissertation, Civil Engineering Section, Colorado State University, Fort Collins, Colorado. 1962.

30. Schwab, G. O., R. K. Frevert, T. W. Edminster, and K. B. Barnes, Soil and water conservation engineering. New York: John Wiley and Sons, Inc., pp. 98 and 627. 1966.

31. Sharp, A. L., A. E. Gibbs, W. S. Owens, and B. Harris, Application of the multiple regression approach in evaluating parameters affecting water yield of river basins. Journal of Geophysical Research, Vol. 65, pp. 1273-1286. 1960.

32. Shelton, C. H., R. D. Haren, and J. I. Sewell, Water yield models for small Tennessee watersheds. Paper presented at the 1972 annual meeting of the American Society of Agricultural Engineers Hot Springs, Arkansas.

33. Shelton, C. H., and J. I. Sewell, Parameter Screening for watershed Analysis. Transactions of American Society of Agricultural Engineers, Vol. 12, No. 4, pp. 533-536. 1969.

34. Spreen, W. C., Determination of the effect of topography upon precipitation. Transactions of the American Geophysical Union, Vol. 28, No. 2, pp. 285-290. 1947.

35. Strahler, A. N., Quantitative analysis of watershed geomorphology, Transactions of the American Geophysical Union, Vol. 38, No. 6, pp. 913-920. 1957.

36. Ursic, S. J., and J. L. Thames, Effects of cover types and soils on runoff in Northern Mississippi. Journal of Geophysical Research, Vol. 65, No. 2, pp. 663-667. 1960.

37. Van Bavel, C. H. M., et al., Agricultural drought in South Carolina. South Carolina Agricultural College, Clemson, S. C., Bulletin 447. 1957.

38. Wallis, J. R., Multivariate statistical methods in hydrology--a comparison using data of known functional relationships. Water Resources Research. Vol. 1, No. 4, pp. 447-461. 1965.

39. Webb, W. M. and L. I. Briggs, The use of principal component analysis to screen mineralogical data. Journal of Geology, Vol. 74, No. 5, pp. 716-720. 1966.

APPENDICES

APPENDIX A

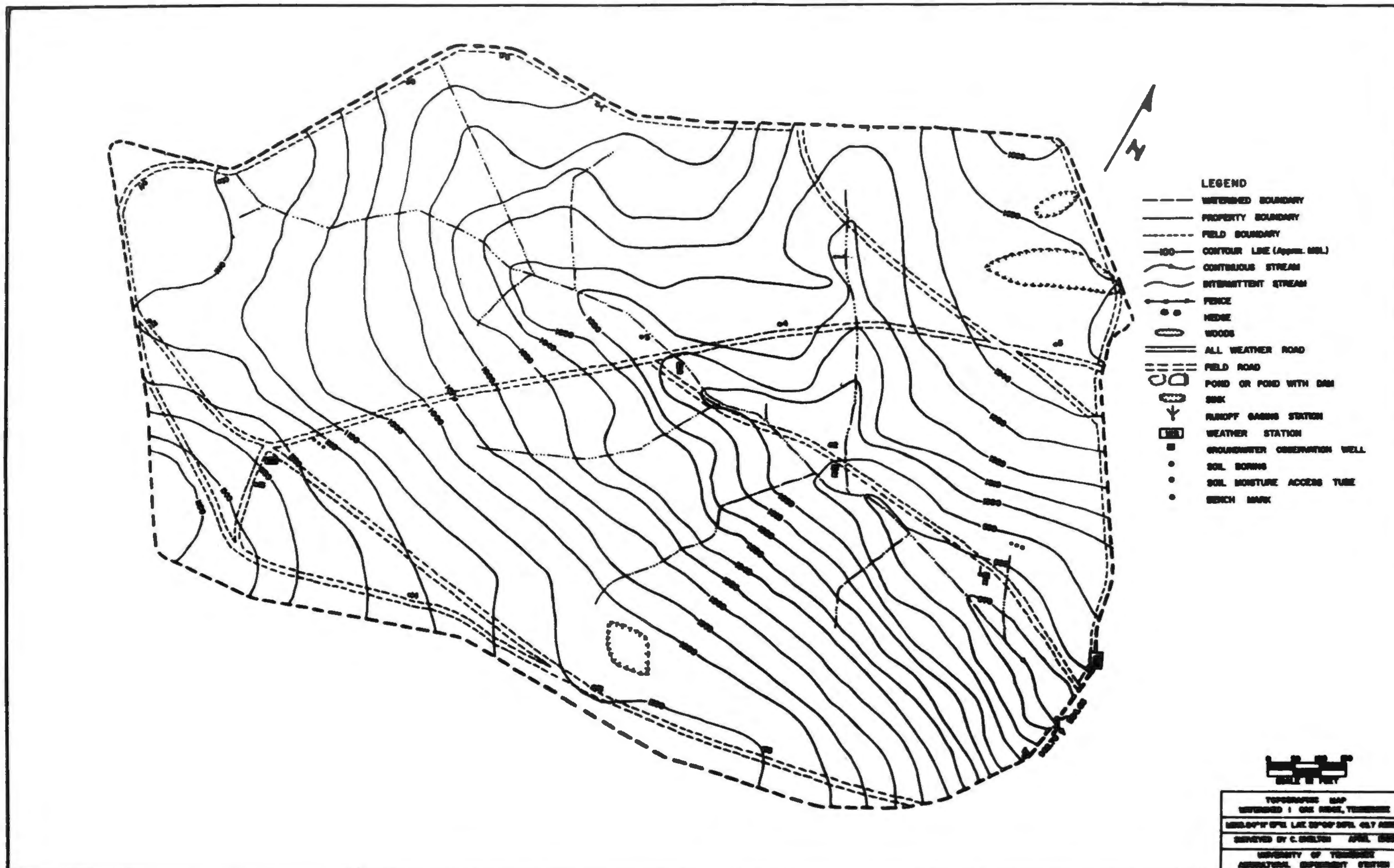

Figure 9. Topographic map of Watershed I.

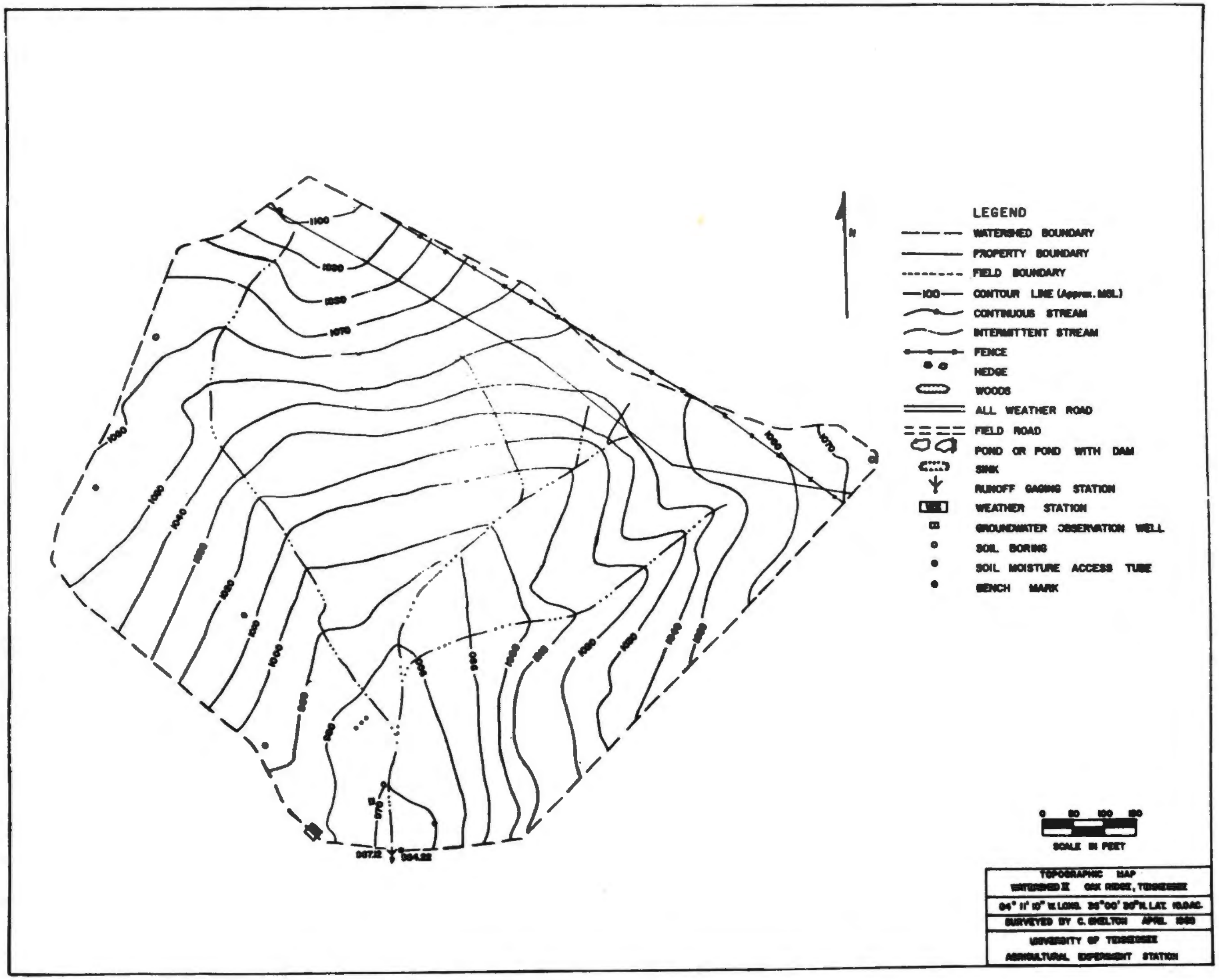

Figure 10. Topographic map of Watershed II.

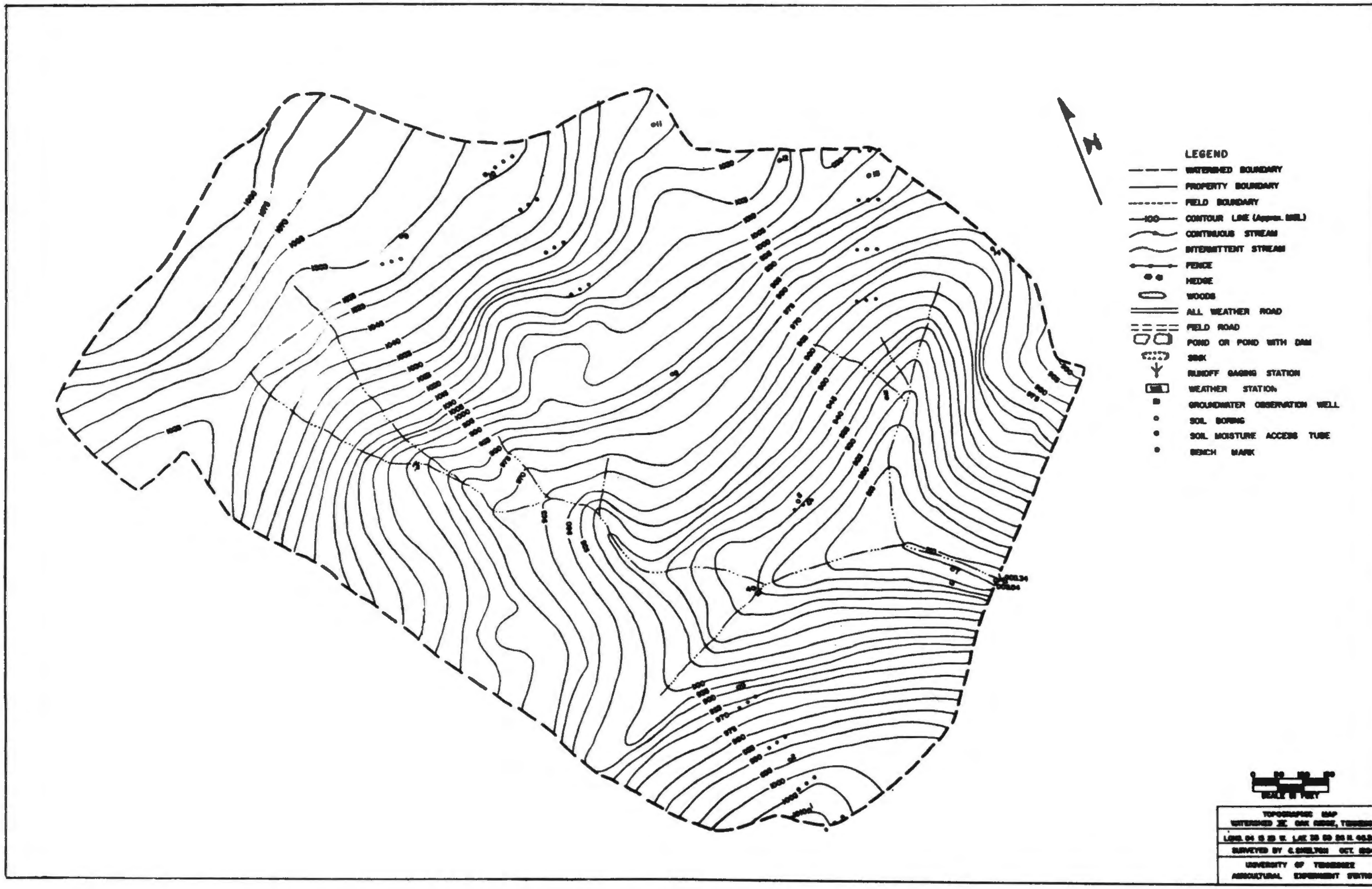

Figure 11. Topographic map of Watershed III.

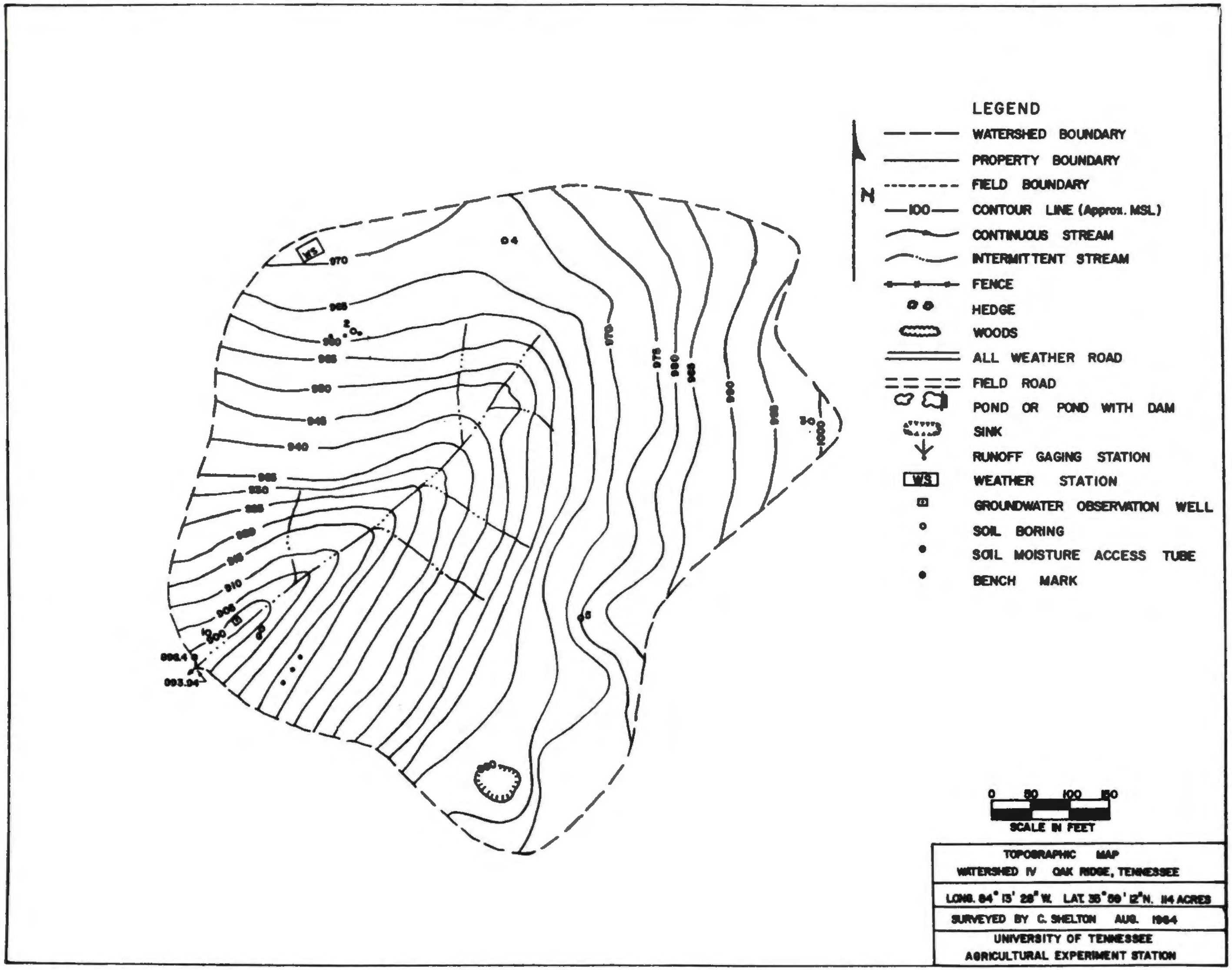

Figure 12. Topographic map of Watershed IV.

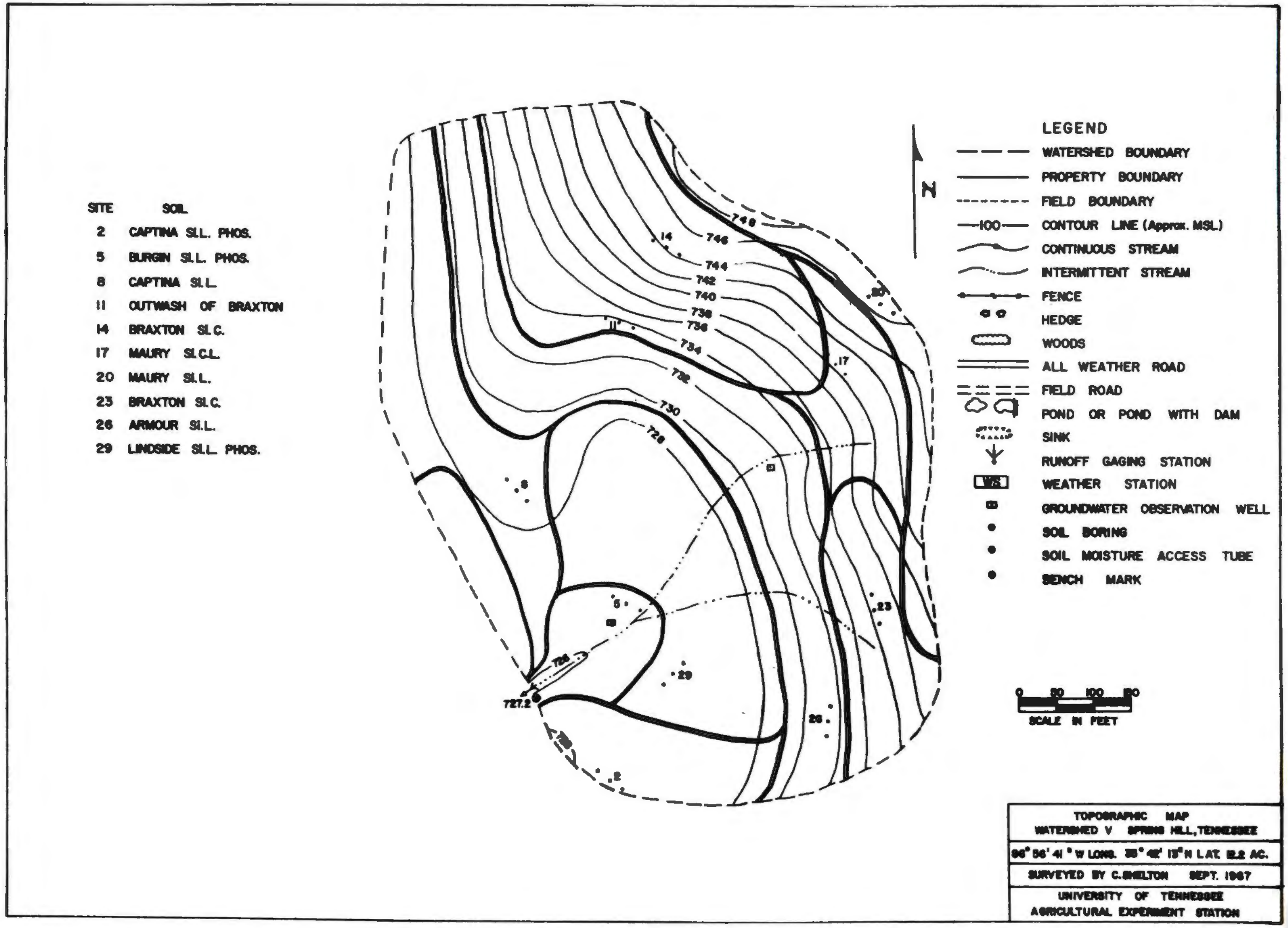

Figure 13. Topographic map of Watershed V.

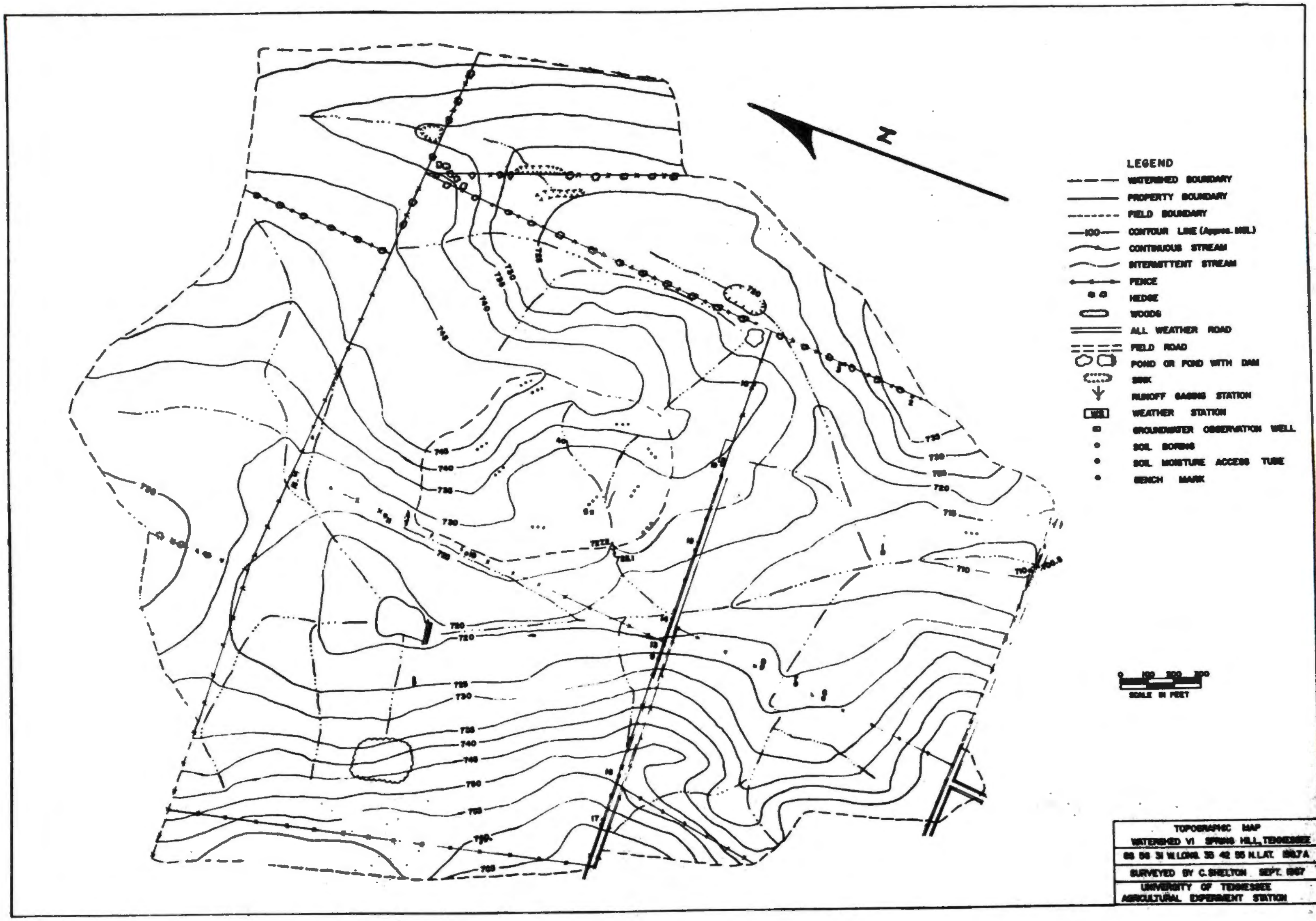

Figure 14. Topographic map of Watershed VI.

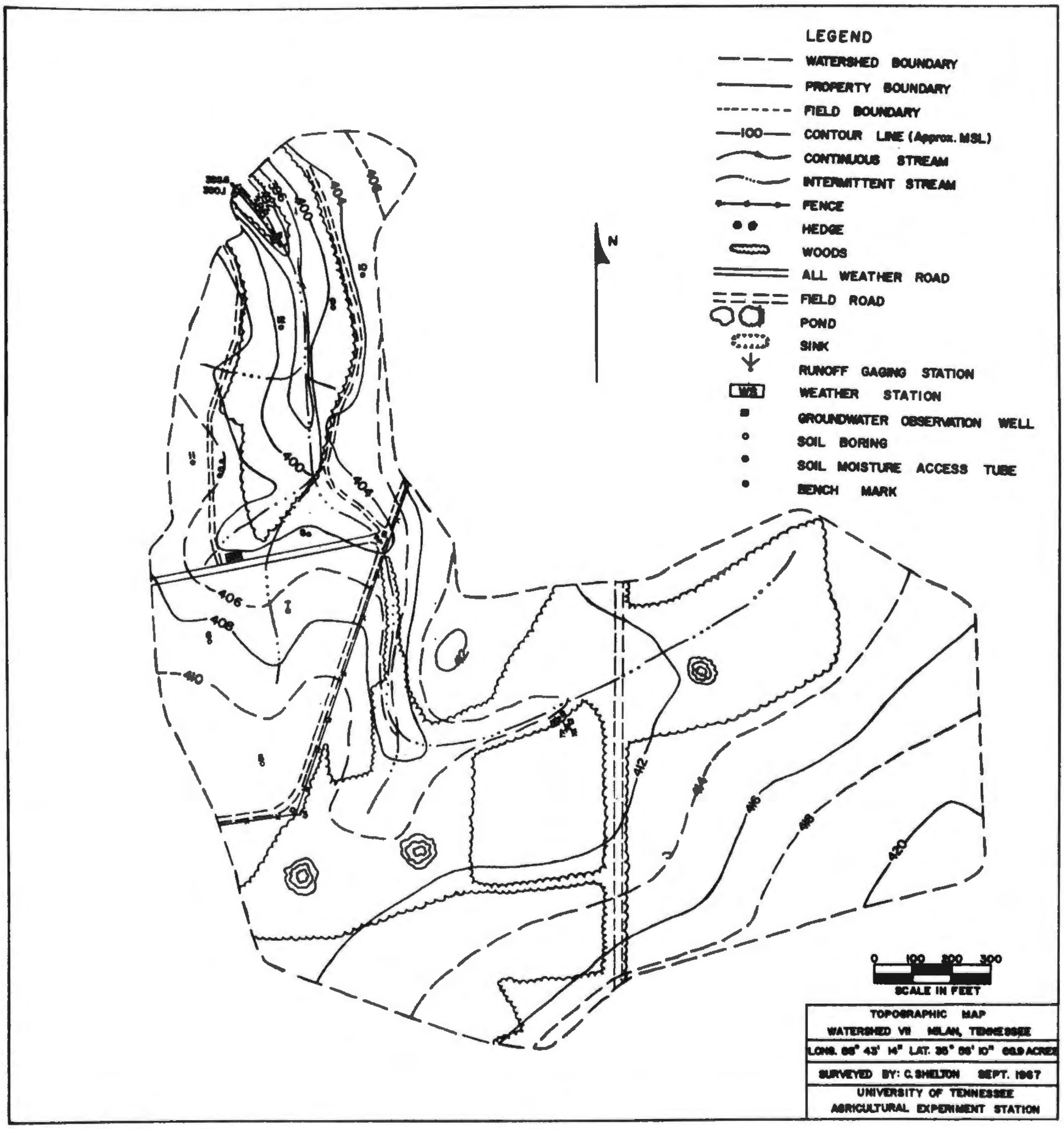

Figure 15. Topographic map of Watershed VII.

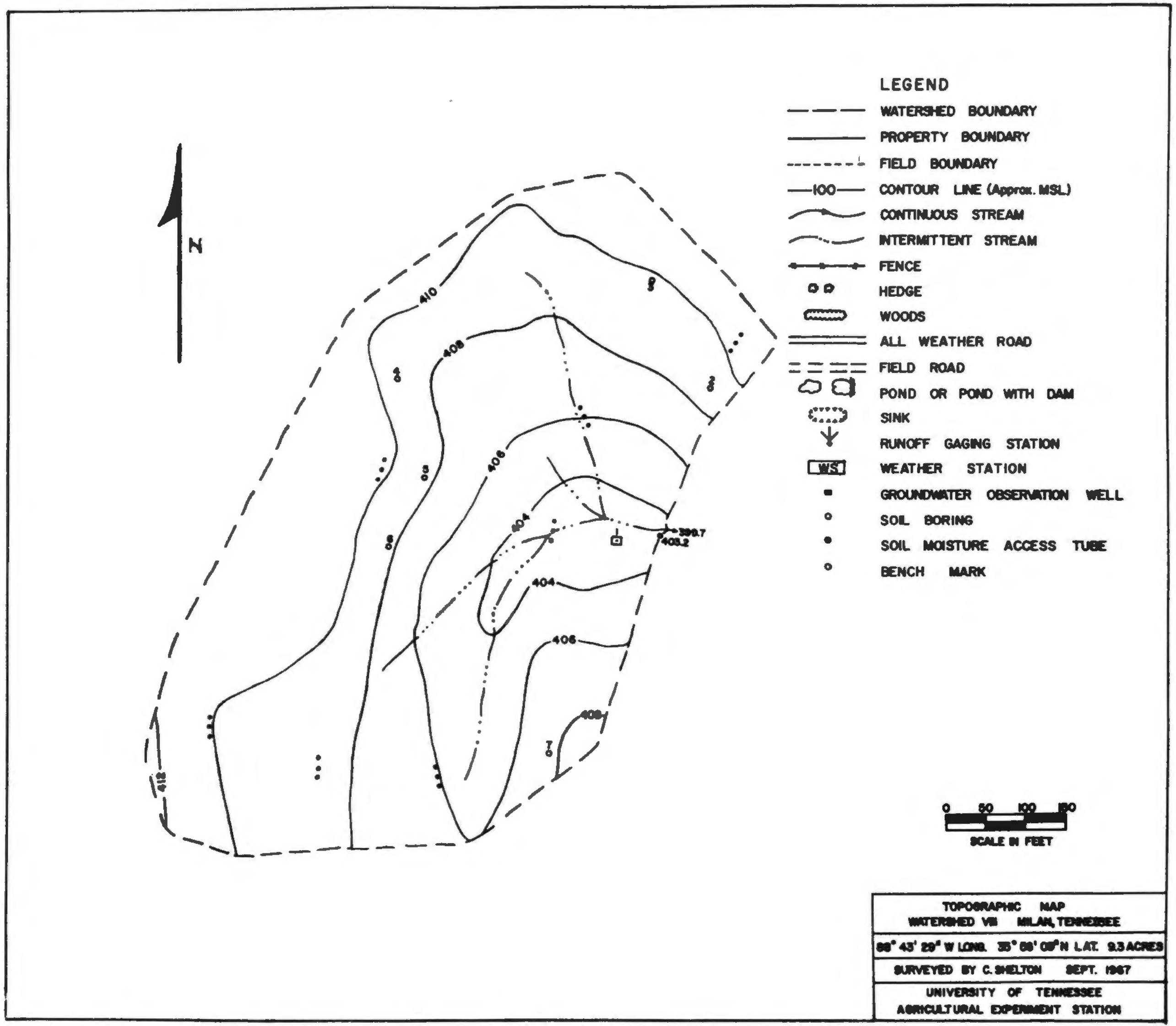

Figure 16. Topographic map of Watershed VIII.

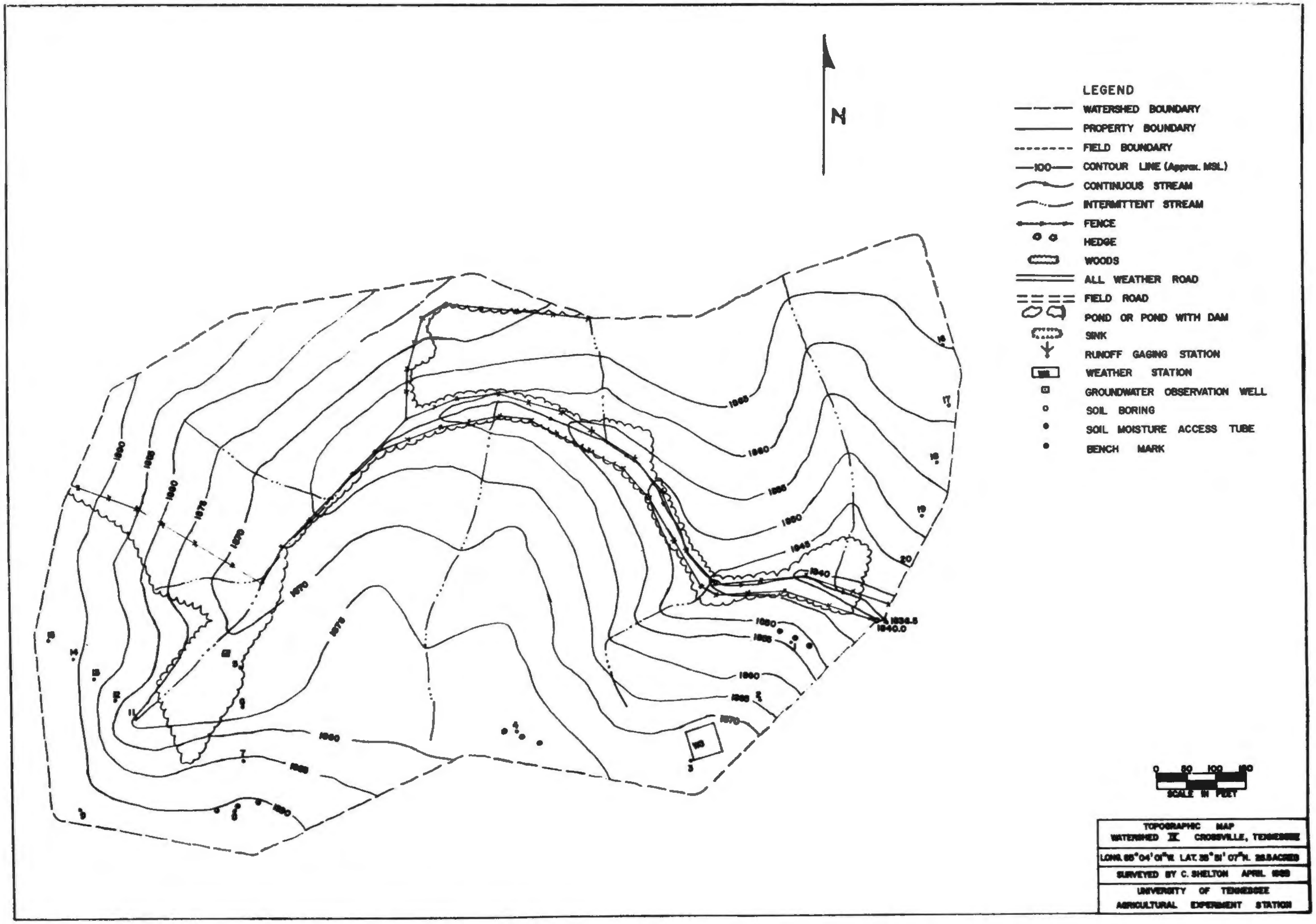

Figure 17. Topographic map of Watershed IX.

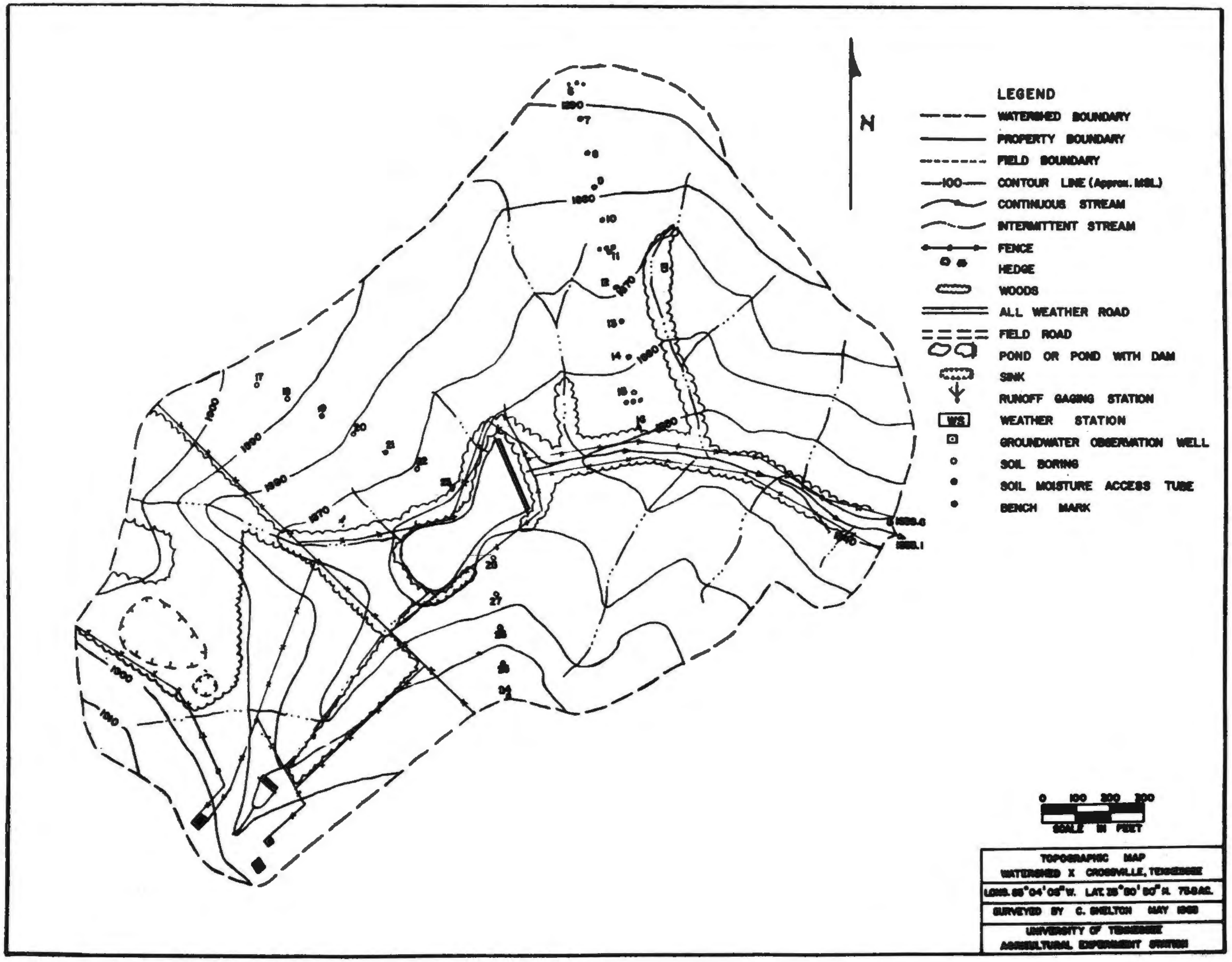

Figure 18. Topographic map of Watershed X.

APPENDIX B

TABLE XI

WATERSHED CHARACTERISTICS

No.	Symbol	Name	Unit	Watershed Number I	II	III	IV	V	VI	VII	VIII	IX	X
1	A	Area	AC	46.7	19.0	46.3	11.4	11.4	198.9	66.9	9.3	26.5	75.8
2	FF	Form	Ft^2/L^2	.38	.57	.41	.51	1.07	.46	.35	.74	.32	.44
3	L	Axial Length	Mi	.43	.21	.41	.19	.13	.82	.55	.14	.35	.52
4	MSL	Elevation, (msl)	Ft	963	962	905	894	725	707	387	400	1837	1835
5	H	Total Relief	Ft	193.6	148.2	183.4	108.9	24.1	61.0	34.4	12.7	57.0	79.0
6	R	Relief Ratio	Ft/mi	450	706	447	573	185	74	63	91	163	152
7	HYP	Hypsometry (a/A at mean h)	Ft/Ft	98.2	91.6	99.3	116.8	68.9	81.5	143.0	133.3	114.1	102.6
8	C	Compactness Co-efficient	-	1.14	1.12	1.14	1.10	1.11	1.14	1.34	1.11	1.18	1.16
9	B	Perimeter	Ft	5775	3610	5750	2725	2840	11940	8110	2525	4375	7450
10	S	Mean Slope	%	15.6	16.3	13.6	14.9	3.8	4.0	1.8	2.4	7.5	6.1
11	L	Channel Length	Ft	4986	3415	3718	1339	915	17100	4390	1170	3785	8780
12	D	Drainage Density	Ft/ac	107	180	80	118	75	85	66	126	143	115
13	G	Channel Gradient	%	10.2	10.0	9.2	12.2	1.5	1.7	1.8	1.6	3.9	3.7
14	O	First Order Streams	No	11	7	8	6	2	17	6	4	7	13
15	I	Depth to Impervious Boundary	Ft	12.5	12.1	13.2	10.2	10.0	16.5	5.6	5.0	2.4	2.4
16	E	Dominant Erosion	Class	1	1	2	2	2	2	2	2	1	1
17	N	Dominant Soil	Name	Fuller-ton	Fuller-ton	Fuller-ton	Fuller-ton	Maury	Maury	Grenada	Callo-way	Hart-sells	Hart-sells
18	K	Permeability	Class	Mod. Rapid	Mod. Rapid	Rapid	Rapid	Mod. Slow	Mod.	Slow	Slow	Mod. Slow	Mod. Slow
19	U	Land Use	Name	Woods	Woods	Pas-ture	Pas-ture	Cult.	Cult. Past.	Cult. Woods	Cult.	Past. Woods	Past. Woods

TABLE XII

QUARTERLY PRECIPITATION AND RUNOFF
JANUARY - MARCH

Year	Watershed Number																			
	I		II		III		IV		V		VI		VII		VIII		IX		X	
	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q
1964	4.99	2.17	4.99	1.65	2.02	0.26	2.02	0.003												
1965	6.04	3.09	6.04	2.56	5.72	0.63	5.97	0.26												
1966	3.80	1.13	3.80	0.85	3.72	0.16	3.72	0.01	3.22	0.16	3.22	0.16	3.84	1.00	3.84	1.69				
1967	4.75	2.89	4.75	2.23	4.46	0.33	4.46	0.17	3.01	0.49	3.01	0.49	3.01	0.18	2.46	0.55	3.95	4.52	3.95	3.42
1968	3.91	2.55	3.91	1.93	3.62	0.01	3.62	0.00	3.92	0.78	3.92	0.35	4.00	1.91	4.00	2.62	4.41	4.56	4.41	4.01
1969	4.00	4.75	4.00	1.08	3.76	0.00	3.76	0.0	3.88	0.73	3.88	0.30	2.86	0.93	2.86	1.42				
Mean	4.58	2.76	4.58	1.71	3.88	0.23	3.92	0.07	3.51	0.54	3.51	0.25	3.29	1.08	3.29	1.57	4.18	4.54	4.18	3.71

TABLE XIII

YEARLY PRECIPITATION AND RUNOFF

	Watershed Number																			
	I		II		III		IV		V		VI		VII		VIII		IX		X	
Year	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q	P	Q
1964	46.96	11.96	46.09	8.18	44.32	0.92	44.68	0.06												
1965	47.52	13.96	47.52	9.64	43.34	2.01	44.34	0.83												
1966	50.96	10.89	52.77	7.67	50.41	1.54	49.54	0.27	42.94	1.31	42.94	0.73	43.11	8.72	43.11	12.37				
1967	74.84	34.00	74.84	32.03	72.73	11.26	72.73	3.43	49.49	11.31	49.49	2.71	47.68	15.41	47.68	17.11	57.11	38.21	57.11	30.69
1968	41.55	12.60	41.55	8.67	39.96	0.01	39.96	0.01	45.88	5.62	45.88	1.56	44.63	10.03	44.63	20.32	43.85	20.56	43.85	21.23
1969	47.78	29.31	47.78	6.58	47.43	0.04	47.43	0.24	51.70	6.92	51.70	4.19	42.03	8.20		18.14				
Mean	51.60	18.65	51.75	12.12	49.69	2.63	49.78	0.81	47.50	6.29	47.50	2.29	44.36	10.59	45.15	16.98	50.48	29.38	50.48	25.94

TABLE XIV

MEANS OF CLIMATIC AND COVER PARAMETERS

Parameter	Symbol	Watershed Number									
		I	II	III	IV	V	VI	VII	VIII	IX	X
Soil Cover Index	SCI	31	31	29	29	58	44	58	54	34	32
Depth of Topsoil (ft.)	DT	1.7	2.8	2.0	1.7	2.6	2.6	6.1	6.1	1.6	1.5
Average Annual Temp., ° F	T	58.2	58.2	58.2	58.2	60.4	60.4	60.3	60.3	53.1	53.1
Evapotranspiration (in./yr.)	ET	32.4	32.4	32.4	32.4	35.4	35.4	39.3	39.3	32.4	32.4

VITA

Rex Duane Haren was born in Green County, Tennessee, on October 10, 1926. He attended elementary schools in Green and Washington Counties. He was graduated from Sulphur Springs High School in 1944. He spent one and one-half years in the U.S. Army. He entered the University of Tennessee in 1947, and in December of 1949 he received a Bachelor of Science degree in Agricultural Engineering. He worked for the United States Department of Agriculture for a short time. In the fall of 1952 he was employed as an instructor by Emory and Henry College. He received the Master's degree in Agricultural Engineering from the University of Tennessee in 1961.

He entered the Graduate School at The University of Tennessee in June 1968, and began studying for a Doctor of Philosophy degree with a major in Agricultural Engineering.