8-1966

An analysis of factors affecting water yield of small agricultural watersheds

Guillermo Díaz

To the Graduate Council:

I am submitting herewith a thesis written by Guillermo Díaz entitled "An analysis of factors affecting water yield of small agricultural watersheds." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Biosystems Engineering.

J.I. Sewell, Major Professor

We have read this thesis and recommend its acceptance:

J.J. McDow, B.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.)

August 8, 1966

To the Graduate Council:

I am submitting herewith a thesis written by Guillermo Diaz entitled "An Analysis of Factors Affecting Water Yield of Small Agricultural Watersheds." I recommend that it be accepted for nine quarter hours of credit in partial fulfillment of the requirements for the degree of Master of Science, with a major in Agricultural Engineering and a minor in Water Resources.

John I. Sewell
Major Professor

We have read this thesis and recommend its acceptance:

John J. McDow

[signature]

Accepted for the Council:

Hilton A. Smith
Dean of the Graduate School

AN ANALYSIS OF FACTORS AFFECTING WATER YIELD OF SMALL AGRICULTURAL WATERSHEDS

A Thesis

Presented to

the Graduate Council of

The University of Tennessee

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Guillermo Diaz

August 1966

ACKNOWLEDGMENTS

The author wishes to express his sincere appreciation to:

Dr. J. I. Sewell, major professor, for his invaluable guidance and unreserved help throughout this investigation.

Professor C. H. Shelton, for his guidance in procuring, selecting, and interpreting the data used in this investigation.

Dr. J. J. McDow, Head, Agricultural Engineering Department for serving on the graduate committee.

Dr. B. A. Tschantz, minor professor, for his help and suggestions while serving on the graduate committee.

Dr. C. C. Thigpen, Head, Department of Statistics, for his guidance and unreserved help in selection and utilizing the statistical techniques employed in this investigation.

The Agricultural Experiment Station, for providing financial assistance for this program of study.

The University of Tennessee Water Resources Research Center and the United States Department of the Interior for providing financial assistance for this program of study.

The personnel of the University Computing Center for helping to make available the services of the IBM 7040 Digital Computer.

The writer is deeply grateful to his wife for her loyalty and assistance and to his parents-in-law and other members of his family for their continual patience and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

THE PROBLEM

I. INTRODUCTION

Water is a natural resource of immeasurable value to any community, state, or nation. In many areas it is a limiting factor to agricultural, industrial, recreational, and population expansion or development.

The United States has experienced such a large increase in commerce and population that the local water supplies are often failing to meet the demand (48, 49).* This demand has quintupled from 1900 to 1950, and this demand is expected to double by 1970 (43). This has brought about a steadily growing concern and need for controlled use and conservation of the soil, water, and other natural resources.

Concern over water supplies is evidenced by studies of water use and conservation by Presidential Commissions, the Congress, Federal agencies, states, local groups, and individuals. Further, this concern is becoming more acute because of the recent acceleration of planning and construction of soil and water conservation facilities by Federal, state, and local agencies (9, 22, 23, 43, 50).

The question of the effects of such practices on stream flow and water yield has arisen. This question is becoming a major factor in

---

*Numbers in parentheses refer to similarly numbered items in the List of References.

planning future water supplies. The extent to which these soil and water conservation facilities will affect water yield can only be determined by accurate predictions of the water yields of given watersheds during specific periods of time. These yield predictions are not only needed to estimate the effects due to conservation programs upon water yield, but also they are essential in the effective planning and design of a multitude of hydraulic engineering undertakings, large and small.

Yield predictions are needed to estimate minimum amounts of water available for urban, industrial, and agricultural use. They are also needed for estimating future dependable supplies for hydroelectric power, navigation, and flood control projects under varying patterns of rainfall. Where these predictions are accurate, a satisfactory feasibility study of a proposed project is usually possible.

At present there are several methods of estimating water yield, each one of which seems to be an improvement on the previous methods; however, none of these can be said to be truly accurate. Some of these methods estimate yields by utilizing only climatic factors, while others will use only geographic locations. Others use a combination of geographic and climatic factors, and still others combine some watershed characteristics with climatic and geographic factors. Since water yield depends not only on the climatic factors but also the physiographic factors, a dearth of information exists concerning water yield under the combined effects of the various factors affecting it.

Water yield studies are an attempt to provide another measuring technique useful in planning for the development of water resources.

Improved methods for predicting yield are needed for optimum and unified development.

Hydrologists, hydraulic engineers, agricultural engineers, and others engaged in the development of water supplies have for many years been concerned with the prediction of runoff and water yield. The basic tools which have been utilized are past records of runoff and rainfall in or near the watershed and some of the watershed characteristics.

The rainfall-runoff records are generally of short duration at best. Some of the watershed characteristics used are of a qualitative nature described with numerical scales having hydrologic significance based on the limited knowledge of the complex plant-soil-water relationships. Therefore, the classic problem has been that of obtaining the best possible statistical estimates of the hydrologic characteristics of a watershed based on inadequate data.

It is known that the climate over a watershed, its geology, topography, antecedent moisture, vegetative cover, land-use practices, and the status of aquifers all influence the watershed performance and have their effect on modulating the relationship between precipitation, runoff, and water yield.

Many of the factors that are of concern in watersheds cannot be measured satisfactorily. It is difficult to measure and to assign specific numbers to the effects of different vegetative covers and land-use practices. It is known that the rate of infiltration is a fundamental factor in watershed performance. Even though infiltration can be measured at a given point in time and space, this information is inadequate to make projections of the net result of the variations in infiltration that prevail over the watershed and to predict changes that take place with time.

Planning for adjustments in water resources through land-use changes requires quantitative data. Before reliable quantitative data can be obtained, the relationship between land-use and hydrologic response must be known; and in addition, inventories of land use must be available. At the present time, adequate relations between land-use and hydrologic response are not known. Adequate land-use inventory data are just becoming available. This urgent need for planning and design data has inspired the majority of the quantitative methods currently used in watershed hydrology.

Most of these quantitative methods are only approximations. Some of them are so oversimplified in their prediction technique that their use will give unsound and misleading results (6, 40). The complexity and amount of calculations of some others make their use prohibitive.

Improved and simplified methods of predicting water yield and runoff are needed. Since water yield and runoff are interrelated with climatic physiographic factors, improved methods of assigning numerical scales based on hydrologic principles to quantitative and qualitative watershed characteristics are also needed.

II. OBJECTIVES OF INVESTIGATION

This thesis presents the first phase of an overall project attempting to develop relations between water yield and land-use.

This work was done under a project entitled "Factors Affecting Water Yields from Small Watersheds in Tennessee" which is a contributing project to Regional Project S-53.

Approximately ten states in the South are participating in similar projects under Regional Project S-53. The combined results of all the projects will be used in future resource development plans.

The objectives of the first phase follow:

1. To determine factors affecting the water yield of selected watersheds through the utilization of available programmed statistical techniques.
2. To determine the magnitude of the effects caused by these identified factors upon the hydrologic behavior of the selected watersheds.

CHAPTER II

HYDROLOGY OF NATURAL WATERSHEDS

I. RELATING THE HYDROLOGIC CYCLE TO WATERSHEDS

The following is a brief discussion of the hydrologic cycle emphasizing the different conditions existing within a watershed in which all precipitation is assumed to occur as rainfall.

Consider a small homogeneous area within a watershed. Precipitation over this area may be intercepted by vegetative material thus preventing it from reaching the ground. Precipitation which reaches the ground may penetrate the soil as infiltration, may be detained in puddles, ditches, and other depressions in the soil surface as depression storage, or may become surface runoff. The infiltration capacity depends mainly on the soil type and soil moisture levels.

As precipitation continues, the infiltration capacity decreases allowing the excess rainfall to overflow the depression storages thus permitting increasing amounts of interflow and surface runoff to move toward the stream channel. Some of the infiltration that is not retained as soil moisture may either move to the stream as interflow or may percolate to ground water and eventually contribute to base flow in the stream.

As soon as the overland flow reaches the stream as direct runoff, the storage in the stream channel is increased causing an increase in flow rate which moves from tributaries into the main stream and is registered at gaging stations. In many watersheds the effect of a change in the rate of inflow into the stream channels is first evident in the tributaries where lateral inflow is a relatively large part of the total streamflow. In the main channel, lateral inflow may be negligible compared to flow rates from the upper watershed. Therefore, as direct runoff decreases, a decrease in streamflow will again move through the channel system to the gage (30). Recorded streamflow at the outlet of the watershed often declines from one maximum while a high-intensity rainfall which will later cause a second maximum flow is occurring. The time delay between the maximum rate of direct runoff and the maximum streamflow recorded at the outlet of the watershed is a measure of the time required for a flood wave to move through the watershed. This time delay varies with the velocities of overland flow and streamflows, and with the quantities of channel and surface storage (30).

When the rainfall ends, intercepted water evaporates, and depression storage is either evaporated, used by vegetation, or infiltrates into the soil; none appears as surface runoff. In some areas a part of the infiltrated water that was stored in the soil zone may move to lower zones and enter ground water to reappear later as stream flow.

The moisture content of the soil profile is reduced by evapo-transpiration, first near the surface and then at greater depths. Point to point variations in available or accessible moisture for evapo-transpiration must be taken into account to obtain accurate estimates of evapo-transpiration for an entire watershed.

The sum of incremented volumes of interception, infiltration, depression storage, and direct runoff from all parts of the basin constitute the total volumes within the entire watershed during a time interval. To make accurate estimates of direct runoff volume, point-to-point rainfall intensity, initial soil moisture profiles, soil type, surface slope, and cover must be taken into account.

II. WATER YIELD

The terms direct surface runoff, runoff, and water yield are often interchanged since they have generally been used synonymously or with close meanings by various writers. However, yield is usually considered in terms of total volumes per year or as average flow for long periods of time whereas the other two terms ordinarily are applied to instantaneous rates or to average rates for short periods.

Water yield is not the same as direct surface runoff or as runoff. Direct surface runoff persists for only a short time after the rain stops and it includes only that water which reaches the stream channel. It does not include that which percolates into the water table. Runoff includes all the water flowing in a stream channel (direct runoff plus ground water flow) past any given point on the stream.

Water yield includes all of the water that can be held in the watershed and which can be made available for future use in the watershed. As can be seen, surface runoff is only a part of runoff and runoff is only a part of water yield.

According to this concept and the hydrologic cycle, water yield can be expressed by either of the following two equations:

$$Y = P - E - T - Kd$$

where Y is water yield, P is precipitation, E is evaporation, T is transpiration, and Kd is deep percolation; or

$$Y = RO + SM + GW$$

where Y is water yield, RO is runoff, SM is soil moisture storage, and GW is ground water storage.

Factors affecting any one of the parameters on the right side of the equations will affect water yield. A list of the principal factors affecting water yield is given in Appendix A.

CHAPTER III

REVIEW OF LITERATURE

In order to make the proper selection of the statistical techniques used in this study and to be able to evaluate objectively the results obtained, it was necessary to make a literature review. This review was concerned with not only the factors that influence the hydrologic behavior of watersheds and have effect on the relationship between precipitation, runoff, and water yield; but also with several of the methods, including statistical techniques, used in developing these relationships.

The literature of the past several years includes many papers dealing with the subject of watershed hydrology and factors affecting it, but only a few of these studies and their findings will be summarized here.

In 1944 Copley et al. (11) reported that in the study of eight years of record of rainfall and runoff on bare plots at Statesville, North Carolina, rainfall characteristics had an effect on runoff. In this study it was found that about 29 percent of the total precipitation occurred in storms of 0 to 1 inch, whereas approximately 23 per cent of the total runoff resulted from such storms. At the other extreme, 10 per cent of the precipitation occurred in storms of over 3 inches, but 13 per cent of the runoff resulted from such storms.

In 1945 Smith et al. (51) studied the relationship between the maximum rate of runoff from an 8.03-acre cultivated watershed and the rainfall intensity. The 5-, 15-, and 30-minute rainfall intensities from seventy-nine storms covering an 8-year period were analyzed. He found that the 30-minute rainfall intensity had the greatest effect on the maximum rate of runoff.

Studies conducted by Barnett (4) on the effects of rainfall intensity on runoff and soil erosion indicate correlation between the expected runoff amount and rainfall intensity on duration. The existence of this correlation agrees with the studies made by Sharp et al. (37) of several watersheds with different areas, soils, and cover conditions. Of the several variables used by Sharp in his statistical analysis, he found that only those variables related to rainfall characteristics were significantly related to runoff.

A runoff study from 15 years of records of several plots located near Watkinsville, Georgia, was performed by Hendrikson et al. (24) in 1963. From this study it was concluded that, on the average, a total annual rainfall of 48.85 inches will produce 10.8 inches of runoff for that locality.

Harrold (21), in the Ohio River Basin watershed studies, concluded that watershed size may determine the season at which high runoff may be expected to occur. He observed that on watersheds in the Ohio River Basin, 99 per cent of the floods from drainage areas of one square mile occurred in May through September, and that 95 per cent of the floods on drainage areas of 100,000 square miles occurred in October through April. In one of his latest reports (20) on the analysis of 46 years of data on agricultural watersheds of 29 to 17,540 acres located near Coshocton, Ohio, it was concluded that both runoff volumes and rates increase as watershed size increases. However, both rate and volume per unit of watershed area decrease as the runoff area increases.

The dynamic watershed area concept was presented in 1964 (56). From the analysis of the hydrologic behavior of two watersheds, it was concluded that the area contributing to runoff is dynamic. The dimensions of this area vary during the storm. It is small when the rain begins, then expands as the rain continues, and finally shrinks back to a localized area after the rain has ended. Betson (7), based on some of the concepts presented in (56), found that storm runoff from a small test watershed in pasture frequently occurred from a small but consistent part of the watershed area. He concluded that this phenomenon also seems to be true for larger watersheds with complex vegetation.

Spreen (53) correlated the mean seasonal precipitation with watershed elevation, slope, orientation, and exposure for western Colorado. He found that elevation alone accounts for 30 per cent of the variation in precipitation and that the five parameters together accounted for 85 per cent of this variation.

Several investigations on the effects of watershed soil, soil moisture, vegetation, and land-use on runoff and water yield have been conducted. Duley (13) pointed out that a rapid reduction in the rate of water intake through the soil surface is due to a partial surface sealing caused by the beating action of raindrops and the water flowing over the surface. Several studies (47) have found that vegetation greatly reduces surface-sealing.

In a study of the factors affecting the infiltration capacity of different soil types, Lewis and Powers (38) found that vegetative cover and surface conditions often have more influence on infiltration rates than do the soil type and texture. Duley (14) in a similar study made the same findings.

In the study of the influence of land-use on the hydrology of small watersheds at Coshocton, Ohio, Harrold et al. (23) reported that mixed-cover watershed had an increase in infiltration potential. The effect of this increase in infiltration potential was reflected by reduced peak rates of runoff for most storms and reduced peak volumes of runoff at high rates of flow.

Ursic and Thames (61) studied the effect of cover types and soils on runoff in three watersheds in Mississippi. They found that surface runoff and peak flows were greatest from abandoned fields, intermediate from depleted upland hardwood forest and least from a 20-year-old pine plantation that had been established on eroding land. They also found that the presence of a shallow hardpan more than doubled the amount of surface runoff and increased peak flows. On a morphologically comparable soil in Missouri, Fletcher and McDermott (15) found that the hardpan restricted water transmission and also limited the depth of root penetration.

In 1928 Bates and Henry (5) showed that the removal of the forest cover of an area in the Rio Grande National Forest near Wagon Wheel Gap, Colorado, increased annual runoff by about 0.96 inches per year, or about 15 per cent. Peak discharges also were increased by as much as 35 per cent.

Hoyt and Troxell (31) reported that complete destruction by fire of the forest cover in the Fish Creek Basin in California resulted in an average annual increase of 1.55 inches, or 29 per cent, and an increase of 0.19 inches, or 475 per cent for the summer months.

The denudation of an experimental area in Coweeta Experimental Forest, North Carolina, was reported by Hoover (27) to have increased annual runoff by 17 inches in the first year following cutting of all brush and trees, and by 13 inches in the second year after some regrowth had occurred. Hoover also reported that no significant changes occurred either before or after treatment, and all runoff was controlled by ground water conditions.

A study conducted cooperatively by the Tennessee Valley Authority and North Carolina State University at Raleigh (59) shows that cover changes exert a large influence upon how water moves off the land, but these do not seem to affect the total outflow materially. The authors of this study concluded that the density of plant cover and the physical condition of the soil itself are the two principal factors in controlling the rate of runoff as well as the amount of erosion produced by a given storm pattern.

Harrold (23) reported that the integrated effects of land-use practices at the Coshocton, Ohio, watersheds are reflected in stream flow. In this study it was found that in every watershed where land-use practices were improved as compared with previous practices for that watershed, some reduction or an indication of some reduction in stream flow occurred. Several reports with similar findings have been published (8, 19, 41).

In reference to the studies made by a group of investigators under the "Cooperative Water Yield Procedures Study Project," Sharp (45) reported that the group made many extensive and intensive studies of precipitation-streamflow relations in creek and river basins in the Great Plains and the Southwest. Every statistical method thought applicable was employed in the analysis of all available research data. In these studies, statistically significant results were obtained in only two cases.

Thus, the information reported by Sharp suggests that hydrologic data from natural watersheds undergoing conservation programs lack sufficient accuracy to detect the individual effects of conservation practices on water yield. It does, however, give some idea of the magnitude of any change in the volume of water yield.

Since water yield and runoff determinations are probably the most challenging problems in hydrology and engineering, a large number of empirical formulas for predicting runoff have been developed.

Munson (42) listed twenty-six formulas and two sets of curves that were in standard use in the determination of runoff. He classified them under three general headings, as follows:

1. Runoff formulas for metropolitan districts and streams (7 formulas)

2. Runoff from small areas, swamps, and wet lands (6 formulas)

3. Flood flows from watersheds over 200 square miles in area (13 formulas and 2 sets of curves)

The various methods of estimating floods were discussed by Jarvis et al. (32) under the following headings: extreme-flood formulas, flood-frequency formulas, statistical (or probability) methods, methods dependent on relation of rainfall to runoff, and methods estimating peak flow from 24-hour average flow.

An examination of these methods reveals that many of them were developed based on short periods of record and for specific localities. Some are further limited with respect to the size and conditions of the drainage basins to which they apply. Still others are applicable to runoff and floods of certain frequencies. It is interesting to note that most of the mathematical expressions listed by Munson under "flood-frequency formulas and statistical methods" contain only one variable, the size of the drainage area.

In developing the various formulas referred to above, the paramount objective apparently was to find some simple device which would give answers to the troublesome problems and would involve the determination of only the simplest factors, such as the area or the slope of the drainage basin. Although some of these mathematical relationships are being improved as more rainfall-runoff data from watersheds of various characteristics is becoming available, none of these adequately recognizes all the complications of the runoff process. Linsley et al. (40)

stated that "formulas of this type have no place in modern engineering design."

The so-called "rational formula" deserves particular mention not only because this formula (6, 16, 28) is probably responsible for much of the advance made toward the discovery of some of the important factors affecting runoff (51), but also because it still is used in the design of some hydraulic structures on comparatively small drainage areas (16).

The rational method assumes that rainfall occurs at a uniform intensity over the entire area of the watershed for a duration at least equal to the time of concentration of the watershed. Under these conditions, all parts of the watershed would be contributing simultaneously to the discharge at the outlet. This assumption is literally true for small areas; however, it is not true for areas of a few acres or larger. Selecting a runoff coefficient from a table of verbal descriptions can also introduce error. In choosing the rainfall intensity for a known return period, it is presumed that frequency is accounted for. If the runoff coefficient and the rainfall intensity are statistically independent and if the return period of the computed maximum runoff is to be the same as that for the rainfall intensity, then the runoff coefficient should have a return period of unity (1 year). In practice, the runoff coefficient is related only to type of terrain without regard to frequency (6). These discussions show rather clearly how complex and indeterminate the rational method is. In the past, the rational method has served us well when better prediction techniques were not available.

In the hands of unqualified users, the rational method can become a rather dangerous and misleading device (40).

In the past several years, statistical methods have been increasingly used in hydrology. This was probably brought about by a general realization of research workers and engineers that statistics, including confidence levels, must be applied in all prediction techniques related to hydrology.

Statistical techniques have been used by Alter (1) who in 1940 demonstrated the feasibility of using precipitation data from relatively low-level stations in forecasting runoff from mountain basins. In 1943 Clyde and Work (10) demonstrated that high correlations can sometimes be found with precipitation stations remote from the basin of interest. Light and Kohler (39) in 1943 described a statistical solution developed in pilot studies conducted by the United States Weather Bureau.

A more reliable method was developed by Kohler and Linsley (37) in 1949. Their procedure is based on a statistical analysis of precipitation and streamflow data, but it adheres carefully to pertinent hydrologic aspects.

Since multiple regression is one of the few numerical techniques which permits a simultaneous evaluation of the effect of several causative factors, its increased use as a tool in hydrology is readily understandable. Hydrologists have used this technique because they are working essentially with uncontrolled experiments where they cannot hold the causative factors constant. For this reason, they are forced into an attempt to evaluate their influence on experimental results.

In 1956 Koelser and Ford (36) reported on an exhaustive regression study to determine reliable equations for use in forecasting runoff. In trying to discover the independent effects of the causative factors, they stated that high correlation between independent variables might result in an apparent lack of consistency.

The limitations of the multiple regression approach to water yield studies were discussed by Sharp et al. (46) in 1960. Their presentation included results of several analyses of annual and monthly streamflow of the Delaware River Basin in Kansas. These results were used as a background for an examination of the method. They also pointed out how hydrologic data, in general, and factors affecting water yield, in particular, may not fit the premises upon which the multiple regression method of analysis is based: (1) There are no errors in the independent variables; errors occur only in the dependent variable, (2) the variance of the dependent variable (runoff) does not change with changing levels of the independent variables (precipitation, land-use, and so forth), (3) the observed values of the dependent variable are uncorrelated random events. They show how hydrologic data may not fit the assumption implicit in tests of significance of multiple correlation and regression coefficients. They concluded that although the multiple regression approach will result in a line of best fit and best estimating equation for hydrologic data, it is not safe to place too much reliance on values estimated by such equations, particularly at levels far removed from the mean, despite very high correlation coefficients. They also suggested an investigation of the more modern statistical procedures that may be better tools than the multiple regression approach for evaluating effects of watershed parameters on water yield.

In 1961 Harris *et al.* (18) discussed the difficulties ordinarily encountered in the application of multiple regression analysis to hydrologic data, then they presented a mathematical development of a statistical model that avoids some of these difficulties. This model, however, does not eliminate assumptions intrinsic in the multiple regression approach. The chief virtue of their model lies in its ability to evaluate the importance of many individual variables successively after the effects of previously selected variables have been removed.

Some possibilities for multivariate analysis in hydrologic studies were discussed by Snyder (52) in 1962. He reported comparative results of multiple regression and multivariate analysis for three applications. From the first two applications he concluded that multivariate analysis offers the more satisfactory solution to the problem of estimating independent effects when the independent variables are correlated. In the third application he shows that the use of multivariate analysis improves the convergence to a solution by the iterative technique of nonlinear least squares.

During the past several years the Tennessee Valley Authority has published a series of papers concerned with hydrology (54, 55, 56, 57, 58). One paper (55), "A Water Yield Model for Analysis of Monthly Runoff Data," is an interim report on hydrologic model building and evaluation. Analyses were performed by fitting the model to ten test sets of data by composite methodology of nonlinear least squares and by using the multivariate technique of component analysis. It was demonstrated that the use of this method of numerical analysis allows the identity of the structural coefficients of the model to be retained. Tennessee Valley Authority Technical Paper Number 5 (54) presented the results obtained from a factor analysis of hydrologic condition survey data from seven watersheds located throughout the Tennessee Valley. The paper also includes a brief discussion of factor analysis, its uses, and a comparison with regression analysis. The writers of the paper concluded that the method of factor analysis is not limited to the type of data presented in the paper. Any set of variables which are thought to be interrelated may be analyzed by this method.

CHAPTER IV

FACTOR ANALYSIS

Factor analysis is a branch of multivariate analysis which deals with the internal structure of matrices of covariances and correlations. Factor analysis was developed mainly by psychologists, and was primarily concerned with hypotheses relating to the organization of mental ability suggested by an examination of matrices of correlation between cognitive test variates. Gradually the statistical theory of the subject was developed.

Since the mathematical techniques inherent in factor analysis are not limited to psychological applications, the use of factor analysis has spread to disciplines other than psychology. Because of the computations involved in factor analysis, the advent of high-speed electronic computers has facilitated the use of this multivariate statistical technique. Thus factor analysis has become the most widely used of the multivariate techniques (29).

I. FACTOR MODELS

The factor method is based upon the correlation coefficient which is a mathematical statement of the degree of agreement between two series of measurements. It is assumed that correlation is produced because similar influences are at work. These influences are called factors.

The principal concern of factor analysis is the resolution of a set of variables linearly in terms of a small number of categories or factors. This resolution can be accomplished by an analysis of the correlations among the variables.

Since there are an infinite number of factorizations of a correlation matrix which may account for the observed data equally well, the preferred types of factor solutions are determined on the basis of two general principles: (1) statistical simplicity, and (2) psychological meaningfulness in the case of psychology. In turn, each of these requires interpretation and each has been applied variously to yield several distinct schools of factor analysts.

A preferred type of factor solution based entirely upon statistical considerations would be the principal component solution. This solution is not only a statistically optimal solution, but it is also unique in the mathematical sense.

The procedure usually recommended by psychologists is to initiate the analysis of a correlation matrix by means of some arbitrary solution and then rotate it to some more meaningful solution.

Since the factors can be rotated to different positions, clearly the subjective element in it is large; and as a consequence, different investigators might interpret the same data differently. This possibility is generally thought to be undesirable; and to avoid it, various empirical techniques for rotating factors have been proposed. Thurstone's concept of "simple structure" (17) is the best known technique.

In general, the rotation of axes (factors) into a simple structure is an attempt to reduce the complexity of the variables. The ultimate objective would be a uni-factor solution where each variable is of complexity one.

Many specific proposals for analytical procedures for the attainment of simple structure have been made. Kaiser's (34) "varimax" method is one of the best known. According to Harmon (17), "This procedure not only does a better job of approximating the classical simple-structure principles, but it also tends to lead to factorial invariant solutions."

II. PRINCIPAL COMPONENT SOLUTION

The principal component (factor) solution is considered to be a preferred type of solution, or an excellent reduction of the correlation matrix which provides a basis for rotation to some other form of solution usually of the multiple-factor type (35).

The technique of component analysis deals with the determination of all truly independent components of variation in an array of variables. It is a relatively straight-forward method of arranging a correlation matrix into a set of orthogonal components or axes equal in number to the number of variates concerned. These correspond to the roots and accompanying vectors of the characteristic equation of the matrix. The resulting roots and vectors, which are sometimes called eigenvalues and eigenvectors, define new and mutually independent variates.

In this method, the roots are extracted in descending order of magnitude; each successive eigenvalue accounts for decreasingly less variance among the observations which is important if only a few of the components are to be used for summarizing the data. Although a few components may extract a large percentage of the total variance of the variates, all components are required to reproduce the correlations between the variates exactly.

The principal component pattern, a set of equations expressing each variable as a linear combination of one or more components,which results from the principal component solution is such that all the components (factors) are usually general. General factors have nonzero coefficients, called loadings, for all the variables. The first component is a general component with positive loadings where the solution is based upon a positive correlation matrix; the remaining components are bi-polar. A bi-polar component has positive and negative loadings. Variables with high loadings, in a bi-polar component, can be considered as measures on a single scale in opposite directions.

III. USES OF FACTOR ANALYSIS

Factor analysis can be used for one or more of the following purposes: (1) to condense a set of variables by expressing them in terms of a relatively small number of linearly independent hypothetical variables (factors), (2) to discover the underlying influences that operate to produce the measurement of the variables, (3) to test a hypothesis concerning the underlying influences, or (4) to supplement, and perhaps simplify, conventional statistical techniques and computations.

The uses of factor analysis in hydrologic investigations are numerous, but it lends itself best to those hydrologic investigations where the variables under analysis are correlated; and specially, if the numerical structure of the solution is of primary importance. Multiple regression may produce unsatisfactory results in applications of this kind (46).

A multivariate approach to the above problem is based upon recognition of the correlations among the independent variables. A component analysis can be made of the correlation matrix of the independent variables to identify all the truly independent, or orthogonal, variations that are present. This is done by determining all the significant roots and associated vectors of the characteristic equation of the matrix. These vectors can then be used to compute the component of each of the original variables that is present in the orthogonal. Kendall (35) has shown how this technique may be used to evaluate equations of relationship of the original variates. The coefficients of the variables in these equations are then actual estimates of the independent contributions of these variables.

CHAPTER V

PROCEDURE

The review of literature led to the selection of two statistical techniques for conducting the analysis of the data for the selected twenty-two watersheds. The principal component method was used to identify underlying factors which operated to produce the RO/P measurements of the variables. Hereafter runoff is designated by RO, and precipitation is designated by P. Multiple regression analysis was used for the evaluation of the combined effects of the identified factors, independent variables, on the RO/P measurements which were used as the dependent variable.

The monthly and annual precipitation and runoff data were punched on IBM data processing cards. Computer programs were developed to convert these data into RO/P ratios and to transform the data into the format required by the principal component analysis program used. The same general data conversion procedure was followed for the stepwise regression program used to compute the multiple regression analyses. All the statistical analyses were made on an IBM 7040 Digital Computer at The University of Tennessee Computing Center. See Appendix D for program developed to convert the data.

I. DATA

Selection

The criteria for the selection of the watersheds were based on standardized data, physiographic location, length of record, and watershed size.

Uniformity of data, watersheds in which the same procedure was followed for collecting and reporting data, was imposed in the selection of the watersheds. This was done not only to avoid the difficulty of interpreting different sets of data, but also to reduce the error inherent in the analysis of data with various degrees of accuracy.

As many physiographic locations as possible, but with similar annual rainfall to Tennessee, were required. This criterion was included because some of the effects between the watershed characteristics and the components could be unique to a given area. A period which included the same years, and for at least fifteen consecutive years of record, was considered to be of sufficient duration to investigate effectively any relationships which might exist among the watersheds.

Since the primary interest of this study concerned small agricultural watersheds, the selection of the watersheds was restricted to those having areas from 2 to 350 acres. Since the reliability of any statistical measure varies directly with the number of individuals on which the measure is based, the number of selected watersheds was intended to be as large as possible.

Watersheds

Twenty-two watersheds from three agricultural experiment stations were selected for this study. Some of these watersheds had over 30 years of continuous record, but the period from 1946 to 1961 was selected to obtain the largest number of watersheds meeting the criteria discussed earlier. The monthly and annual precipitation and runoff data as well as the physical features of the selected watersheds were secured from existing records (25, 26, 60).

The selected watersheds were:

Number	Location	Name
14	Coshocton, Ohio	5, 10, 129, 131, 135, 169, 172, 177, 183, 185, 187, 188, 192, 196
7	Riesel (Waco), Texas	W-1, W-2, W-6, W-10, Y, Y-2, Y-4
1	Watkinsville, Georgia	W-1

The Coshocton watersheds are found on hilly to steep topography in the Allegheny-Cumberland Plateau. The predominant soil, Muskingum silt loams, is a well-drained residual soil. The watersheds selected vary from 2.05 to 349 acres. Average annual rainfall totals 37 inches. The watersheds were in either improved pasture, cultivation, or mixed cover.

The Riesel (Waco) watersheds are located in the Blacklands of the Texas Coastal Plains on deep, fine-textured and slowly permeable soils. The soils are predominantly Houston black clay which is in cultivation or mixed land uses. The watersheds selected were from 19.7 to 309 acres. The average annual rainfall on these watersheds was 32 inches.

The Watkinsville watershed is in the igneous and schist area of the Southern Piedmont Plateau. The terrain is moderately sloping and well drained and it was largely under cultivation. The watershed area is 19.2 acres. The average annual precipitation was 46 inches. Additional information on the watersheds may be found in Agricultural Research Service Hydrologic Data (25, 26, 60).

II. PROGRAMS

Selection of Parameters

The selection of the parameters used in this study was based on statistical and hydrologic considerations as related to the purpose of the investigation. Since precipitation is the most important single factor affecting water yield, it was decided to attempt to relate precipitation and runoff in some way that could serve as an indicator of the effects of the remaining factors affecting the hydrologic behavior of the watersheds.

In several investigations using multiple regression techniques to predict runoff from precipitation and watershed physical characteristics, the variance of the dependent variable (runoff) changed with changing levels of the independent variables, especially with precipitation variations. The ratio of runoff to rainfall was selected as the main parameter in both programs. This ratio would tend to remove the main climatic effect from the other watershed characteristics affecting water yield leaving mainly the physical characteristics to be analyzed. Also, by using the ratio as the dependent variable in the multiple regression analysis, the changes in variance of the dependent variable would be reduced.

Two watershed factors were identified from the component analysis. The main concern was not with the mathematical model but with the evaluation of the effects of the identified factors upon the RO/P measurements. For this reason, twenty-one area and slope variables and interactions were arbitrarily selected.

Principal Component Analysis Program

The principal component analyses were computed on the IBM 7040 Digital Computer using the BMD01M Principal Component Analysis (12) program. BMD01M computes the principal components of standardized data and rank orders each standardized case by the size of each principal component separately. Output from this program includes: correlation coefficients, eigenvalues, cumulative proportion of the total variance, and eignevectors or principal components of standardized data.

The parameters used with this program were the monthly and annual RO/P measurements for twenty-two watersheds as the twenty-two variables, and the sixteen years of record represent the number of cases.

Stepwise Regression Program

The multiple regression analyses were computed on the IBM 7040 Digital Computer using the BMD01R Stepwise Regression (12) program.

This program computes a sequence of multiple linear regression equations in a setpwise manner. At each step one variable is added to the regression equation. The variable added is the one which makes the greatest reduction in the error sum of squares. It is the variable which has the highest partial correlation coefficient with the dependent variable partialed on the variables which have already been added; thus it is the variable which, if it were added, would have the highest F value. The multiple correlation coefficient R and analysis of variance table are outputs from this program at each step,

The parameters used with this program were the RO/P ratio for the sixteen years of record of the Ohio and Texas watersheds and twenty area and slope combinations for each watershed. The area and slope combinations were used as the independent variables. A list of the variables is presented in Table IV in Appendix B.

CHAPTER VI

RESULTS AND DISCUSSION

Statistical techniques were primarily utilized in the correlation of the RO/P parameters with the watershed factors designated by the component analysis program. It was accomplished in three steps. The first step was an analysis of the precipitation-runoff data by the principal component technique to establish a set of orthogonal eigenvectors which represented the original data. In the second step an attempt was made to relate the factors designated by the component analysis program to some physical watershed characteristics. In the last step the correlations for RO/P and the related watershed physical factors were computed by the multiple regression technique.

I. FINDINGS

The first principal component solution was obtained using the annual precipitation and runoff data to damp out seasonal effects on the hydrologic behavior of the watersheds. The principal-component pattern consisted of twenty-two eigenvectors since the data was standardized to zero mean and unit variance and unities were planned in the principal diagonals of the correlation matrix. Of these twenty-two eigenvectors, the first eleven accounted for the total variance of the variables; therefore, the remaining eleven eigenvectors did not contribute to the explanation of the variations among the watersheds. Thus, all the information in the set of variables was contained in the first eleven orthogonal components.

Following Kaiser's (33) recommendation for finding the number of common factors that are necessary for the explanation of the correlations among the variables, the first four components (factors) were selected to be analyzed. Kaiser's recommendation was that "the number of common factors should be equal to the number of eigenvalues greater than one for a correlation matrix having unity in the principal diagonal."

The correlations among the watersheds are presented in Table V in Appendix C. Only half of this correlation matrix is given since it is symmetric. The first five principal components for the twenty-two watersheds, including their eigenvalues and cumulative proportion of total variance, are given in Table I.

The first four eigenvalues explained at least 90 per cent of the total variance of the original variables. Fifty-one per cent of the total variance was explained by the first or principal component.

When the principal component analysis is based on a matrix having negative correlations, the first component is, generally, a bi-polar component. However, it may be observed (Table I) that all the watersheds in the first component had uniformly low negative loadings. In general, as component loadings become lower, more variables are required to describe the component. In Table I, the first component accounted for slightly more than half of the total variance of the watersheds. All of these considerations suggest that the first factor is an important one

TABLE I

FIRST FIVE COMPONENTS FOR TWENTY-TWO WATERSHEDS

Watershed	Name	P1[a]	P2	P3	P4	P5
1. Ohio	129	-.2157	-.0227	.3295	.2102	-.6187
2. Ohio	135	-.2389	-.0616	.2305	.4066	-.1053
3. Ohio	131	-.2318	-.0583	-.0496	.2259	-.1561
4. Ohio	188	-.2253	.1542	.1064	.1132	.2919
5. Ohio	185	-.2270	.0213	.2689	.4151	.2364
6. Ohio	187	-.2198	.1301	-.2449	-.1592	.0029
7. Ohio	192	-.2094	.2108	.0993	.1182	.4360
8. Ohio	172	-.2501	.1741	-.0478	-.2016	-.0712
9. Ohio	169	-.2471	.2105	-.0476	.0051	-.0834
10. Ohio	177	-.2474	.1448	-.3314	.0040	-.0314
11. Ohio	183	-.2388	.2147	-.1034	-.1473	.0121
12. Ohio	196	-.2598	.1455	-.1225	-.0709	-.9320
13. Ohio	10	-.2563	.1663	-.0148	-.1638	-.4900
14. Ohio	5	-.2716	.0963	.0231	-.2057	.0882
15. Texas	W-1	-.1566	-.3332	-.0215	-.0444	.1889
16. Texas	W-2	-.1581	-.3159	-.0217	-.1352	.2379
17. Texas	W-6	-.1554	-.3278	-.0785	.0705	.0337
18. Texas	W-10	-.1375	-.3166	-.1842	-.1946	-.2924
19. Texas	Y	-.1888	-.3048	-.0163	-.0666	.0655
20. Texas	Y-2	-.1858	-.3092	-.0033	-.0673	.0476
21. Texas	Y-4	-.1792	-.3159	-.0387	.0121	.0667
22. Georgia	W-1	-.0681	-.0030	.7063	-.5534	.0029
EIGENVALUES		11.30387	6.19969	1.14377	1.13422	0.66751
Cumulative Proportion of Total Variance		0.51	0.80	0.85	0.90	0.93

[a]P designates principal components.

compounded of unique characteristics of the variables, which when intercorrelated, helped to explain half of the total variance.

The first eigenvector contributed 51 per cent to the total variance of the variables. The second, third, and fourth eigenvectors were bi-polar and contributed 29 per cent, and 5 per cent, respectively, to the total variance of the variables. Since 80 per cent of the information in the set of variables was contained in the first two components and the second component had the largest group of watersheds with relatively high loadings, the second component was selected for an analysis of its loadings. Hereafter descriptive terms for the magnitude of the loadings as "high" or "low" refer only to the numerical value.

In general a bi-polar component having a group of variables with high loadings can be named after the characteristics relative to that group; and if there is a second group with a relatively high loading of the opposite sign, this is a good indication that the component is measuring the same characteristics associated with the data, but at different levels.

The loadings for the Texas watersheds were characterized by high negative values. With the exception of three which were low negative, the Ohio watershed loadings were low positive. Because the Georgia watershed exhibited a value unlike the others and very close to zero, this watershed loading apparently had no relationship with the other watershed characteristics associated with this component.

The watersheds were grouped according to the sign and magnitude of their loadings. This grouping was done in an attempt to identify the watershed characteristics which made the watersheds behave in a similar manner as related to the component.

The watersheds were classified in the following three groups: those with high positive loading, those with high negative loadings, and watersheds with low positive and negative loadings. All the available data for each group was tabulated, and group comparisons of this data were made. From all the group comparisons only one gave positive results; the others were considered unsuccessful.

For the case with positive results, the watershed data associated with the highest positive and negative loadings are compared. This comparison suggested that the area and slope were represented by the behavior of the components.

Texas watersheds W-1, W-6, and W-10 had the highest negative loading. Ohio watersheds 183, 192, and 169 had the highest positive loadings (Table I, page 35). These Texas watersheds have slopes ranging from 0 to 6 percent with the largest percent of the total area in slopes from 0 to 3 percent. The Ohio watershed with the three highest loadings have slopes ranging from 6 to 35 percent; with the largest percentage of total area in slopes from 6 to 18 percent. It was found that the numerical value of the loadings decreased as the area for those watersheds decreased and as the percentage of the total in flatter slopes increased. The watershed data is summarized in Table II.

The results obtained from the principal component analysis program using monthly data yielded little insight into the problem. This was probably caused by the effects of seasonal variations. From the

TABLE II

COMPONENT LOADINGS AND SLOPE CLASSES FOR THREE TEXAS AND OHIO WATERSHEDS

Texas		Ohio		Percent of Area in Slope Class			
				Texas Slope Class[a]		Ohio Slope Class	
Name	Loading	Name	Loading	0-3	3-6	6-18	18-35
W-1	-0.3332	183	0.2147	86	14	78	22
W-6	-0.3278	192	0.2108	99	1	82	18
W-10	-0.3167	169	0.2105	99	1	82	18

[a]Slope class in percent.

twenty-two eigenvectors found, seventeen were needed to account for the total variance of the variables, an increase of six over the first solution. The first two components could account for 68 percent of the total variance, while in the first solution, they accounted for 80 percent. This was probably caused by the increase in error associated with monthly precipitation and runoff measurements and by the increased number of watershed characteristics reflecting their influences in monthly runoff.

The evaluation of the combined effects of area and slope on the RO/P parameter was done for the Ohio and Texas watersheds separately and combined as a group. The Georgia watershed was excluded from this analysis because it behaved differently from all the other watersheds used in the principal component analysis.

The variables selected for use in the stepwise regression analysis are presented in Table IV in Appendix B. The results of the stepwise regression program were summarized in Table III where the order of the five most important independent variables when related to the RO/P ratio, the multiple correlation coefficient R, and corresponding F-values and degrees of freedom for determining the significance of the last variable ranked, are given.

For the Ohio watersheds, step one shows that SQRTA was the most highly correlated with the RO/P measurements having a multiple correlation coefficient R of 0.809. Step two shows A · (12-16S) as the next most important variable after taking into account SQRTA. The addition of A · (12-16S) improved the over-all correlation with RO/P measurements

TABLE III

ORDER OF INDEPENDENT VARIABLES WHEN RELATED TO RUNOFF PRECIPITATION RATIO WITH CORRESPONDING STATISTICAL INFORMATION

Rank	Variable[a]	R[b]	df[c] Lesser MS[d]	df Greater MS	F[e]
			Ohio		
1	SQRTA	.809	222	1	419**
2	A · (12-16S)	.885	221	2	399
3	A · (25-35S)	.892	220	3	284
4	25-35S	.894	219	4	219
5	12-16S	.895	218	5	175**
			Texas		
1	0-3S	.036	1	110	9+
2	SQRTA · (0.3S)	.042	2	109	9
3	A · (3-6S)	.061	3	108	9
4	SQRTA · (3-6S)	.081	4	107	5
5	A · (0-3S)	.128	5	106	3
			Ohio and Texas		
1	A · (16-25S)	.684	334	1	293**
2	16-25S	.713	333	2	173
3	SQRTA · (16-25S)	.766	332	3	157
4	0-3S	.780	331	4	128
5	SQRTA · (6-12S)	.795	330	5	114**

[a]For description of the variable see Table IV, Appendix B.

[b]R signifies the multiple correlation coefficient.

[c]df signifies degrees of freedom.

[d]MS signifies mean square.

[e]F is the value used in the F test for statistical significance.

**Significant at the 1 percent level.

+Not significant at the 25 percent level.

as indicated by the increase in R- values from 0.809 to 0.885. The next three variables entered by the program were A · (25-35S), 25-35S, and 12-16S. Using these five variables, the R- value was 0.895 with an F- value of 175 which is significant at the 1 percent level.

For the Texas watersheds, none of the variables was significantly correlated with the RO/P measurements as is indicated by the F- values. The first variable, the most highly correlated with the RO/P measurement, had an R- value of 0.036 and an F- value of 9 which is not significant at the 25 percent level.

For the Texas and Ohio watersheds as a group, A · (16-25S) was the most highly correlated variable with RO/P measurements having an R- value of 0.684 and an F- value of 293 which is significant at the 1 percent level. The next four variables entered by the program were: 16-25S, SQRTA · (16-25S), 0-3S, and SQRTA · (6-12S). The R- value for these five variables was 0.795 with an F- value of 114 which is significant at the 1 percent level.

The smaller multiple correlation coefficient obtained for the Ohio and Texas watersheds as a group can probably be explained by the fact that the area and slope of the Texas watersheds used in this analysis were not significantly correlated with the RO/P measurements.

Considering the results of this analysis, it appears that area is the most significant factor of those examined, affecting RO/P for the Ohio watersheds used. The SQRTA and A · (12-16S) accounted for 78.3 per cent of the variation in RO/P measurements. This suggests that for the

Ohio data the RO/P measurements might be related to the square root of the watershed area.

In the Texas watersheds, the area and slope did not appear to be significantly correlated with the RO/P measurements. If time had permitted, a further analysis would have been done using soil type and vegetal cover as variables.

II. RECOMMENDATIONS FOR FUTURE STUDIES

Further investigations utilizing the techniques of principal component analysis, factor analysis, and multiple regression could give more insight into precipitation-runoff relationships.

From the principal component solution of the data in this investigation, only the area and slope could be found to be represented in the second component. When the evaluation of their combined effect on RO/P was performed, it was found that the area and slope were significantly related to the behavior of the Ohio watersheds but not to the Texas watersheds. This implies that some other important watershed factors that were present in the component were not identified.

It is felt that if more factors could have been identified and used with the multiple regression analysis, the relation of area and slope to the watersheds would probably have been different from the one obtained.

Due to time limitations, this study was terminated before fully exploring all the techniques necessary to accomplish the above. The available data was inadequate in quantitative descriptions of several watershed characteristics such as soils, cover, and land-use.

Further studies attempting to relate the watershed factors to principal components would be in order. The following is a suggested procedure for developing a prediction equation to be used on ungaged watersheds:

1. Select watersheds with more years of record than those used in this study. Those watersheds should have more exact quantitative data describing the watershed characteristics with primary attention to soil, geology, and land-use.
2. Analyze the precipitation-runoff data by a principal component analysis method.
3. Determine the number of significant components that could represent the data.
4. Analyze the precipitation-runoff data by a factor analysis method. The number of factors rotated should be equal to the number of significant components found in step three.
5. Analyze the component and factor loadings to identify the watershed characteristics represented in them.
6. Analyze the effectiveness as predictors of those identified characteristics by a multiple regression technique.
7. Select the most important variables from those identified and found to measure essentially the same information. Those variables should be representative of the components

or factors, and they should have the greatest effectiveness in prediction.

8. Utilize the selected variables from step seven in a prediction equation determined by a multiple regression technique.

9. Check the prediction equation for various degrees of accuracy by adding variables that were identified but were not used, or by removing some of the ones already used in the model.

This set of selected variables will constitute a reduced group from the original variables for the selected watershed.

By developing several prediction equations for watersheds with different characteristics, a set of curves relating the prediction models and the various characteristics can also be developed. From this set of curves, a general prediction model could be used for ungaged watersheds by means of the curves and the measurable watershed characteristics.

CHAPTER VII

SUMMARY AND CONCLUSIONS

This study was conducted to determine factors affecting the water yield of selected watersheds and to determine the magnitude of the effects caused by the identified factors upon the hydrologic behavior of the selected watersheds. Twenty-two watersheds were selected from three agricultural experiment stations, and two programmed statistical techniques were used.

The study was conducted in three steps. The first step was an analysis of the precipitation and runoff data by a principal component technique to establish a set of orthogonal variables representing the original data. In the second step, an attempt was made to relate the factors designated by the component analysis program to some physical watershed characteristics. In the last step, the correlations of the runoff to precipitation ratio and the related watershed physical factors were computed by a multiple regression technique.

A summary of the findings of this study follows:

1. The area and slope were found to be represented in the second component obtained from the principal component analysis.
2. An analysis of Ohio watersheds suggested that for the Ohio data the RO/P measurements might be related to the square root of the watershed area.

3. In the Texas watersheds, the area and slope did not appear to be significantly correlated with RO/P measurements.

The techniques of this study could be applied to determine those watershed variables which contribute significantly to understanding the general precipitation-runoff relationship for a selected watershed. Determining these variables might enable a reduction in the number of watershed factors necessary in water-yield predictions. Eliminating the unimportant interrelated watershed variables measuring the same basic element of information would reduce superfluous information in subsequent statistical analyses, thus increasing the assurance of stability in the final analysis.

Also, a reduction of watershed variables might enable a reduction in field data collection and data and computer processing. This could result in substantial economy.

The major contribution of this study lies in the approach and statistical techniques applied in the analyses.

In similar studies which might be conducted in the future, the utilization of multivariate factor analysis should be given consideration because this technique might give additional insight into the effects of various factors on water yield. See "Recommendations for Future Studies" in Chapter VI.

LIST OF REFERENCES

LIST OF REFERENCES

1. Alter, C. J. "The Mountain Snow Survey: Its Genesis, Exodus, and Revelation," Transaction of the American Geophysical Union, 21:892-900, 1940.

2. Anderson, T. W. Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons, 1958.

3. Arneston, E. P. "Rational Runoff Theory Applied to Design of Road Ditches and Channel Changes," The Texas Engineer, July, 1932.

4. Barnett, A. P. "How Intense Rainfall Affects Runoff and Soil Erosion," Agricultural Engineering, 39:703-707, 711, 1958.

5. Bates, C. G., and A. J. Henry. Forest and Streamflow Experiment at Wagon Wheel Gap, Colorado. Monthly Weather Review, United States Weather Bureau Supp. 30, 1928.

6. Bernard, M., R. L. Gergory, and C. E. Arnol. "Discussion of Runoff-Rational Runoff Formulas," Transactions of the American Society of Civil Engineers, 96:1038-45, 1932.

7. Betson, R. P. "What is Watershed Runoff?," Journal of Geophysical Research, 69:1541-1552, 1964.

8. Brakensiek, D. L., and C. R. Amerman. "Evaluating the Effects of Land Use on Stream Flow," Agricultural Engineering, 41:158-161, 167, 1960.

9. Carreker, J. R. "Research Needs in Soil and Water Conservation." Paper presented at meeting of the Soil Conservation Section. Association of Southern Agricultural Workers, Birmingham, Alabama, February 1-3, 1960. (Mimeographed.)

10. Clyde, G. D., and R. A. Work. "Precipitation-Runoff Relationship as a Basis for Water-Supply Forecasting, with discussion by Bernard," Transactions of the American Geophysical Union, 243:43-55, 1943.

11. Copley, T. L., et al. Effects of Land Use and Season on Runoff and Soil Loss. North Carolina Agricultural Experiment Station, Bulletin 347, May, 1944.

12. Dixon, W. J. (ed.). Biomedical Computer Programs, Health Science Computing Facility, University of California, Los Angeles. September 1, 1965.

13. Duley, F. L. "Surface Factors Affecting the Rate of Intake of Water by Soils," Soil Science Society of American Proceedings, 4:60-64, 1939.

14. ______, and L. L. Kelly. Effect of Soil Type, Slope, and Surface Conditions on Intake of Water. Nebraska Agricultural Experiment Station Research Bulletin 112, May, 1939.

15. Fletcher, P. W., and R. E. McDermott. "Moisture Depletion by Forest Cover on a Seasonally Saturated Ozark Ridge Soil," Soil Science Society of American Proceedings, 21:547-550, 1957.

16. Frevert, R. K. Soil and Water Conservation Engineering. New York: John Wiley and Sons, Inc., 1955.

17. Harman, H. H. Modern Factor Analysis. Chicago: The University of Chicago Press, 1960.

18. Harris, B., et al. "An Improved Statistical Model for Evaluating Parameters Affecting Water Yield of River Basin," Journal of Geophysical Research, 66:3319-3329, 1961.

19. Harrold, L. L. "Effect of Vegetation on Soil Moisture," Transactions of the American Society of Agricultural Engineers, 1:6-8, 11, 1958.

20. ______. "Minimum Water Yield from Small Agricultural Watersheds," Transactions of the American Geophysical Union, 38:201-208, 1957.

21. ______. "The Role of Agriculture in the Hydrologic Cycle," Water Conservation Symposium, Abstracts, Hoosier Chapter 29, SCSA, Purdue University, 1950.

22. ______, et al. Evaluation of Agricultural Hydrology by Monolith Lysimeters, 1944-55. United States Department of Agriculture Technical Bulletin Number 1179, August, 1958.

23. ______, 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 Number 1256, January, 1962.

24. Hendrikson, B. H., et al. Runoff and Erosion Control Studies on Cecil Soil in the Southern Piedmont. Agricultural Research Service, United States Department of Agriculture Technical Bulletin 1281, 1963.

25. Hobbs, H. W. (ed.). Hydrologic Data for Experimental Agricultural Watersheds in the United States, 1956-59. Soil and Water Conservation Research Division, Agricultural Research Service, United States Department of Agriculture Miscellaneous Publication No. 945, November, 1963.

26. ______. Hydrologic Data for Experimental Agricultural Watersheds in the United States, 1960-61. Soil and Water Conservation Research Division, Agricultural Research Service, United States Department of Agriculture Miscellaneous Publication No. 994, May, 1965.

27. Hoover, M. D. "Effect of Removal of Forest Vegetation on Water Yields," Transactions of the American Geophysical Union, 25:969-977, 1944.

28. Horner, W. W., and F. L. Flynt. "Relation Between Rainfall and Runoff for Small Urban Areas," Transactions of the American Society of Civil Engineers, 101:140-146, 1936.

29. Hotelling, H. "The Relation of the Newer Multivariate Statistical Methods to Factor Analysis," British Journal of Statistical Psychology, 10:69-79, 1957.

30. Hoyt, W. G. An Outline of the Runoff Cycle. Pennsylvania State College, School of Engineering Technical Bulletin 27:57-67, 1942.

31. ______, and H. C. Troxell. "Forest and Streamflow," Transactions of the American Society of Civil Engineers, 99:1-30, 1934.

32. Jarvis, C. S., et al. Floods in the United States, Magnitudes and Frequency. United States Geological Survey Water-Supply Paper Number 771, 1936.

33. Kaiser, H. F. "The Application of Electronic Computers to Factor Analysis," Educational and Psychological Measurements, 20:141-51, 1960.

34. ______. "The Varimax Criterion for Analytic Rotation in Factor Analysis," Psychometrica, 23:187-200, 1958.

35. Kendall, M. G. A Course in Multivariate Analysis. New York: Hafner Publishing Co., 1957.

36. Koelzer, V. A., and P. M. Ford. "Effect of Various Hydroclimatic Factors on Snowmelt Runoff," Transactions of the American Geophysical Union, 37;578-587, 1956.

37. Kohler, M. A., and R. K. Linsley. "Recent Developments in Water Supply Forecasting from Precipitation," Transactions of the American Geophysical Union, 30:427-437, 1949.

38. Lewis, M. R., and W. L. Powers. "A Study of the Factors Affecting Infiltration," Soil Science Society of American Proceedings, 3:334-339, 1938.

39. Light, L., and M. A. Kohler. "Forecasting Seasonal Runoff by Statistical Methods," Transactions of the American Geophysical Union, 24: 719-736, 1943.

40. Linsley, R. K., M. A. Kohler, and J. L. H. Paulhus. Hydrology for Engineers. New York: McGraw-Hill Book Company, Inc., 1958.

41. McGuinness, J. L., et al. "Some Effects of Land Use and Treatment on Small Single Crop Watersheds," Reprinted from Journal of Soil and Water Conservation Vol. 15, No. 2, March, 1960. (Mimeographed.)

42. Munson, T. A. "Formulas and Methods in General Use by Engineers to Determine Runoff from Watershed Areas," The Texas Engineer, April, 1934.

43. Picton, W. L. "Water Use in the United States, 1900-1980." A Report by the United States Department of Commerce.(Washington: Government Printing Office, March, 1960), p. 8.

44. Ramser, C. E. Flow of Water in Drainage Channels. United States Department of Agriculture Technical Bulletin 149, 1929.

45. Sharp, A. L. "Evaluating Effects of Upstream Programs on Downstream Water Yields." Paper presented at the annual meeting of the American Society of Civil Engineers, Omaha, Nebraska, May, 1962. (Mimeographed.)

46. ______, et al. "Application of the Multiple Regression Approach in Evaluating Parameters Affecting Water Yield of River Basins," Journal of Geophysical Research, 65:1273-1286, 1960.

47. ______, H. N. Holtan, and G. W. Musgrave. Infiltration in Relation to Runoff in Small Watersheds. Soil Conservation Service, United States Department of Agriculture Technical Bulletin 81, 1949.

48. Select Committee on National Water Resources, United States Senate. Water Resources Activities in the United States: Water Supply and Demand. Committee Print Number 32, 86th Congress, Second Session, 1960.

49. ______. Water Resources Activities in the United States: Estimated Water Requirements per Agricultural Purposes and Their Effects on Water Supplies. Committee Print Number 13, 86th Congress, Second Session, 1960.

50. Smerdon, E. T. "Research Needs in Long-Range Water Resource Planning." Paper presented at the joint meeting of the Southeast and Southwest Section of the American Society of Agricultural Engineers, Dallas, Texas, February 1, 1965. (Mimeographed.)

51. Smith, D. D., et al. Investigations in Erosion Control and Reclamation of Eroded Shelby and Related Soils at the Conservation Station, Bethany, Missouri, 1930-1942. The United States Department of Agriculture Technical Bulletin 883, 1945.

52. Snyder, W. M. "Some Possibilities for Multivariate Analysis in Hydrologic Studies," Journal of Geophysical Research, 67:721-729, 1962.

53. Spreen, W. C. "Determination of the Effect of Topography Upon Precipitation," Transactions of the American Geophysical Union, 28: 285-290, 1947.

54. Tennessee Valley Authority, Division of Water Control Planning, Hydraulic Data Branch. Design of a Hydrologic Condition Survey Using Factor Analysis. Research Paper No. 5, Knoxville, Tennessee, November, 1965.

55. ______, Office of Tributary Area Development. A Water Yield Model for Analysis of Monthly Runoff Data. Research Paper No. 2, Knoxville, Tennessee, February, 1963.

56. ______. Bradshaw Creek-Elk River. A Pilot Study in Area-Stream Factor Correlation. Research Paper No. 4, Knoxville, Tennessee, April, 1964.

57. ______. TVA Computer Programs for Hydrologic Analyses. Research Paper No. 3, Knoxville, Tennessee, June, 1963.

58. ______. Matrix Operations in Hydrograph Computations. Research Paper No. 1, Knoxville, Tennessee, December, 1961.

59. ______ in Cooperation with North Carolina State College of Agriculture and Engineering. Hydrology of Small Watersheds in Relation to Various Crop Covers and Soil Characteristics. Cooperative Research Project in Western North Carolina, 1960.

60. United States Department of Agriculture, Agricultural Research Service, Soil and Water Conservation Research Branch. Monthly Precipitation and Runoff for Small Watersheds in the United States. Washington: Government Printing Office, June, 1957, 20402.

61. Ursic, S. J., and J. L. Thames. "Effects of Cover Types and Soils on Runoff in Northern Mississippi," Journal of Geophysical Research, 65:663-668, 1960.

APPENDIXES

APPENDIX A

FACTORS AFFECTING WATER YIELD

The factors affecting water yield from a watershed may be divided into those associated with the climatic characteristics of the watershed. Probably the most important are those related to the precipitation and those associated with the physical characteristics of the watershed. A listing of these factors follows:

A. Principal Climatic Factors
   1. Precipitation
      a. Amount
      b. Intensity
      c. Duration
      d. Distribution (areal)
      e. Form
   2. Air temperature
   3. Wind speed
   4. Evaporation
   5. Relative humidity
   6. Solar radiation
B. Principal Physiographic Factors
   1. Soil
      a. Type
      b. Degree of erosion
      c. Percolation

d. Infiltration

e. Water-holding capacity

f. Temperature

2. Soil moisture

3. Topography

a. Watershed size

b. Watershed shape

c. Watershed slope

d. Drainage-way density

e. Surface condition

f. Orientation

4. Watershed elevation (mean sea level)

5. Land use

a. Vegetative type

b. Vegetative stage of growth

c. Vegetative density

d. Vegetative coverage

e. Conservation practice

f. Water-control structures

6. Geology

a. Mechanical analysis of soil material to parent material

b. Degree of disintegration from parent material

c. Substrata characteristics

7. Ground water elevations

8. Surface runoff

    a. Rates

    b. Amounts

APPENDIX B

TABLE IV

VARIABLES USED IN THE STEPWISE REGRESSION PROGRAM
FOR THE OHIO AND TEXAS WATERSHEDS

Symbol	Description
RO/P	Runoff precipitation ratio
A	Area of watershed in acres
SQRTA	Square root of the watershed area
0-3S	Slope class ranging from 0 to 3 percent slope
3-6S	Slope class ranging from 3 to 6 percent slope
6-12S	Slope class ranging from 6 to 12 percent slope
12-16S	Slope class ranging from 12 to 16 percent slope
16-25S	Slope class ranging from 16 to 25 percent slope
25-35S	Slope class ranging from 25 to 35 percent slope
A · (0-3S)	Interaction of area and 0 to 3 percent slope class
A · (3-6S)	Interaction of area and 3 to 6 percent slope class
A · (6-12S)	Interaction of area and 6 to 12 percent slope class
A · (12-16S)	Interaction of area and 12 to 16 percent slope class
A · (16-25S)	Interaction of area and 16 to 25 percent slope class
A · (25-35S)	Interaction of area and 25 to 35 percent slope class
SQRTA · (0-3S)	Interaction of SQRTA and 0 to 3 percent slope class
SQRTA · (3-6S)	Interaction of SQRTA and 3 to 6 percent slope class
SQRTA · (6-12S)	Interaction of SQRTA and 6 to 12 percent slope class
SQRTA · (12-16S)	Interaction of SQRTA and 12 to 16 percent slope class
SQRTA · (16-25S)	Interaction of SQRTA and 16 to 25 percent slope class
SQRTA · (25-35S)	Interaction of SQRTA and 25 to 35 percent slope class

NOTE; The slope classes were measured in percentage of the total area laying in the specific slope range. When the data for the Ohio and Texas watersheds were run as separate groups, only the slope classes present in the group were used as variables.

APPENDIX C

TABLE V

CORRELATION COEFFICIENT MATRIX OF TWENTY-TWO WATERSHEDS FOR 16 YEARS OF ANNUAL RUNOFF/PRECIPITATION MEASUREMENTS

Watershed	Name	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. Ohio	129	1.000																					
2. Ohio	135	.811	1.000																				
3. Ohio	131	.595	.720	1.000																			
4. Ohio	188	.508	.546	.557	1.000																		
5. Ohio	185	.660	.837	.560	.685	1.000																	
6. Ohio	187	.371	.490	.538	.598	.423	1.000																
7. Ohio	192	.371	.562	.450	.791	.683	.644	1.000															
8. Ohio	172	.553	.489	.542	.776	.530	.732	.769	1.000														
9. Ohio	169	.581	.594	.595	.794	.611	.785	.858	.951	1.000													
10. Ohio	177	.499	.505	.548	.763	.587	.806	.697	.879	.889	1.000												
11. Ohio	183	.449	.493	.531	.761	.551	.858	.807	.928	.948	.885	1.000											
12. Ohio	196	.638	.577	.569	.702	.636	.736	.723	.930	.920	.923	.919	1.000										
13. Ohio	10	.560	.520	.610	.746	.611	.740	.778	.950	.932	.865	.938	.955	1.000									
14. Ohio	5	.593	.599	.579	.772	.631	.743	.779	.914	.864	.822	.883	.906	.934	1.000								
15. Texas	W-1	.336	.507	.493	.112	.360	.126	-.016	.078	-.009	.142	-.010	.146	.118	.314	1.000							
16. Texas	W-2	.319	.477	.380	.109	.362	.171	.020	.126	.008	.178	.028	.186	.133	.355	.968	1.000						
17. Texas	W-6	.379	.549	.626	.120	.336	.159	-.067	.065	.019	.161	-.012	.127	.101	.243	.957	.876	1.000					
18. Texas	W-10	.409	.382	.405	-.041	.121	.174	-.207	.109	-.010	.188	-.004	.205	.099	.266	.879	.879	.862	1.000				
19. Texas	Y	.457	.592	.559	.178	.425	.231	.063	.230	.134	.269	.114	.286	.239	.404	.969	.964	.938	.899	1.000			
20. Texas	Y-2	.455	.573	.579	.167	.415	.191	.045	.223	.120	.242	.095	.275	.243	.403	.974	.955	.945	.895	.994	1.000		
21. Texas	Y-4	.429	.581	.617	.164	.418	.167	.022	.175	.089	.220	.067	.240	.209	.366	.979	.935	.969	.877	.979	.991	1.000	
22. Georgia	W-1	.278	.121	.059	.192	.122	.125	.151	.260	.146	-.088	.203	.116	.280	.326	.133	.174	.046	.071	.176	.183	.105	1.000

APPENDIX D

DOCUMENTATION OF DEVELOPED PROGRAM

Purpose

The program converts the precipitation and runoff data into a ratio of runoff to precipitation. The output of the data is in the format required by the BMD01M Principal Component Analysis (12) program.

Usage

For FORTRAN PROGRAM purposes, the names of the variables follow:

P is the input array for precipitation values.

RO is the input array for runoff values.

R is a variable containing the dimensions of the input arrays.

NVAR is the number of variables or watersheds.

NCLASS is the total number of precipitation or runoff observations per watershed.

Coding Information

The program is written in FORTRAN IV language with the exception of the READ and PRINT statements which are in FORTRAN II language, although they are processed by the FORTRAN IV computer for compatability with FORTRAN II.

The P and RO are dimensioned to the total number of observations per watershed. R is dimensioned to the total number of observations and watersheds.

The input data will be punched on cards, sixteen values per card, starting in card column number 1 and with format of (16F5.0). For each watershed the precipitation values will be punched first, and the runoff values will be punched second. The first input card will have the number of watersheds and the number of observations with a format of (2I5).

The output data will be the RO/P ratio with a format of (10F8.4). These ratios will be punched in the format required by the BMD01M program. Specifically, the variables of watersheds are in row order, and the cases or observations are in column order.

For each case, the ratio for the variables will be punched in row order.

Example

An illustration of the program follows:

```
      DIMENSION P (208), RO (208), R (208,22)
      READ 1, NVAR, NCLASS
    1 FORMAT (2I5)
      DO 3 J = 1, NVAR
      READ 2, (P(I), I = 1, NCLASS)
    2 FORMAT (16F5.0)
      DO 3 I = 1, NCLASS
    3 R (I,J) = RO (I) P (I)
      DO 11 I = 1, NCLASS
   11 PUNCH 10, (R(I,J),J = 1, NVAR)
```

```
10  FORMAT (10F8.4)
    END
ENTRY
    22 208
   (rest of data)
```

The number of watersheds is 22, and the number of observations is 208.