WATER RESOURCES RESEARCH

VOLUME 1 FOURTH QUARTER 1965 NUMBER 4

Multivariate Statistical Methods in Hydrology—A Comparison Using Data of Known Functional Relationship[1]

JAMES R. WALLIS

Institute of Engineering Research
University of California, Berkeley

Abstract. Conventionally hydrologists have used regression analysis for solving their multivariate problems. Recently other multivariate statistical methods have been advocated. This paper discusses and compares the effectiveness of six methods of analysis: regression, principal component, varimax, oblimax, key cluster, and object. Strengths and weaknesses of each method are discussed, and the combination of principal component regression with varimax rotation of the factor weight matrix is recommended for an initial analysis of multifactor hydrologic problems.

INTRODUCTION

Hydrology includes many problems of prediction based upon complex interactions of many variables and processes. Traditionally such problems have been approached by multiple regression analysis [*Anderson,* 1957; *Ford,* 1959]. But now that large computers have become generally available many other statistical techniques are being tried. This paper compares the relative effectiveness of six of these methods by solving by each method a problem using identical data of known functional relationship.

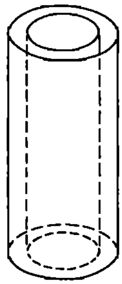
Those readers who require a more technical presentation than is given here should consult the cited references (in particular, the books by *Johnston* [1960] on regression and by *Horst* [1964] on factor analysis); those who want more specific information on how to make a particular type of analysis should consult the footnoted computer program write-ups.

[1] Paper presented at the Western National Meeting, American Geophysical Union, December 30, 1964.

THE HOLLOW CYLINDER EXAMPLE

If a person did not know the formula for calculating the weight of hollow cylinders he might collect many cylinders, measure some of their characteristics, choose a model, and subject the resulting data to a statistical analysis. If he does know the formula for calculating the weight of hollow cylinders, he still can use this procedure. By either adding random errors or specifying an incorrect model he can compare the relative effectiveness of prediction equations made by different statistical techniques. The approach is used in this paper.

A population of 75 synthetic cylinders was generated. To each cylinder 4 scaled random numbers were assigned: height (H); density of cell wall material (D); outside radius (RO); and inside radius (RI), with (RO) greater than (RI). From these 4 initial variables 11 other parameters describing each cylinder were generated (Figure 1). Prediction equations were made by regression and other multivariate methods for the fifteenth variable (weight), using the other 14 variables as predictors.



Variable	Symbol	Function
–	K	Constant (?)
1	H	Height
2	HH	$(\text{Height})^2$
3	2KROH	Outside curved surface
4	2KRIH	Inside curved surface
5	D	Density
6	DD	$(\text{Density})^2$
7	DDIAGO	Density times diagonal of outside cylinder
8	DDIAGI	Density times diagonal of inside cylinder
9	RO	Radius of outside cylinder
10	KRORO	End area of outside cylinder
11	DIAGO	Diagonal of outside cylinder
12	RI	Radius of inside cylinder
13	KRIRI	End area of inside cylinder
14	DIAGI	Diagonal of inside cylinder
15	W	Weight

Fig. 1. Variables and their symbols for the hollow cylinder test problem.

Because a linear model is the best initial assumption in many research studies, it was the model selected for the cylinder data. Only the first 60 of these cylinders were used to develop the prediction equations; the remaining 15 observations were set aside to compare independently the merits of the different prediction equations. Appendix 1 lists the raw data cards for these analyses.

TRANSFORMATIONS

Bartlett [1947] has published a useful summary of data transformations and discussion of why and when they are necessary. Often the most important criteria to be satisfied by transformation are: (1) the variance of the transformed variates should not be affected by changes in the mean level of the variables, and (2) the transformed scale should be one for which real effects are linear and additive. The lack of a necessary transformation or inclusion of an unnecessary one will alter both the prediction equation coefficients and the validity of significance tests. Supporting literature [*Benson,* 1960; *Leopold,* 1962] shows that a logarithmic transformation is most likely to be appropriate for hydrologic data, but each data set should be carefully analyzed to see that the chosen transformation is appropriate.

The 15-variable hollow cylinder test data used here were not transformed before the statistical analyses. This limitation meant that there was an error in the specification of the model (linear rather than multiplicative) and that the R squared values (coefficients of multiple determination) for the different prediction methods could not be expected to reach unity. It also meant that the customary t test of significance was no longer completely valid, although its values have been included for the regression analyses.

MULTIPLE REGRESSION

The technique of multiple regression analysis (Program G2 BC MPRV, University of California Computer Center Library, Berkeley, using zero F level for retaining variables) and its underlying assumptions have been adequately discussed by *Johnston* [1960]. In many hydrologic studies, two of the underlying assumptions of regression analysis are often violated.

First, errors exist in the dependent as well as the independent variable, but regression analysis assigns all of the errors (e) to the dependent variable. This procedure introduces bias into the least squares estimates of the a and b terms of equation 1.

$$Y = a + b_1X_1 + \cdots b_nX_n + e \qquad (1)$$

Customarily the bias resulting from errors in the independent variables is assumed to be small and is ignored.

Second, the residual errors of the transformed variables are probably not independent and normally distributed (autocorrelation), and conventional F and t significance tests are then in jeopardy. The Durbin and Watson statistic can be used to test for autocorrelation [*Durbin and Watson,* 1951], and evasive measures can be taken when it has been detected [*Johnston,* 1960]. Note that many research data are nonprogressive, and that the observations may therefore have to be reordered before tests of autocorrelation are made.

In many studies we wish to understand the underlying functional relationship, but when high intercorrelations exist between the predictor variables (multicollinearity), the regression β coefficients become unstable [*Johnston,* 1960, p. 201–207]. *Anderson* [1954] has demonstrated that this pitfall can be minimized by choosing variables and data carefully. The developed prediction equations should be used carefully, however, and should not be extrapolated far beyond the range of the data used in their formation.

Equation 2 resulted from a multiple regression of the hollow cylinder test problem:

$$
\begin{aligned}
W = {} & +956.0 + 89.7(H) + 7.82(HH) + 4.55(2KROH) - 7.22(2KRIH) \\
& \qquad t = 1.0 \quad t = 1.7 \quad t = 2.8 \quad t = 5.9 \\
& - 1550.(D) + 589.0(DD) + 90.3(DDIAGO) + 3.50(DDIAGI) \\
& \qquad t = 2.8 \quad t = 2.1 \quad t = 3.4 \quad t = 0.2 \\
& - 334.0(RO) + 2.94(KRORO) + 57.6(DIAGO) + 595.0(RI) \\
& \qquad t = -1.2 \quad t = 1.3 \quad t = 0.4 \quad t = 4.8 \\
& + 1.78(KRIRI) - 312(DIAGI) \\
& \qquad t = 0.7 \quad t = -3.7
\end{aligned} \tag{2}
$$

For this equation the coefficient of multiple determination (R^2) was 0.92 (throughout the text the R^2 values given have not been corrected for degrees of freedom). But multicollinearity has led to coefficients that are unstable and hard to interpret in an underlying functional relationship. In equation 2 the variables (D) and (DD) were significantly correlated with weight (t's greater than 2.0), but their coefficients have received opposite signs.

By stepping outside the analysis and using our additional knowledge of the system, we can realize that both (D) and (DD) should have positive coefficients. An additional problem with equation 2 is that we have oversubscribed the number of parameters needed to describe the system. Using too many variables has the same effect as omitting an important parameter from the system: the coefficients become unstable. For instance, equation 3 resulted from removing the variable (RI) and re-analyzing the data. All coefficients for the variables of equation 3 are very different from those of equation 2; several have even changed signs. (For example, the coefficient for the variable ($DIAGI$) has changed from -312 to $+78$.)

$$
\begin{aligned}
W = {} & 763.8 + 17.3(H) + 2.96(HH) + 1.64(2KROH) - 1.76(2KRIH) \\
& \qquad t = 0.2 \quad t = 0.5 \quad t = 0.90 \quad t = 3.0 \\
& - 1486.(D) + 277.1(DD) + 114.2(DDIAGO) - 7.12(DDIAGO) \\
& \qquad t = 2.2 \quad t = 0.9 \quad t = 3.6 \quad t = 0.3 \\
& + 142.0(RO) + 5.28(KRORO) - 218.1(DIAGO) \\
& \qquad t = 0.5 \quad t = 1.9 \quad t = 1.2 \\
& - 5.52(KRIRI) + 77.7(DIAGI) \\
& \qquad t = 2.6 \quad t = 2.7
\end{aligned} \tag{3}
$$

The underlying functional relationship for the weight of hollow cylinders is reflected far better by equation 4 (whose β coefficients were all similar: +0.53, +0.41, +0.64, and −0.50) than it is by either equation 2 or 3. The principal disadvantage with equation 4 appears to be the loss in the accuracy of fit ($R^2 = 0.72$).

$$W = -783. + 74.6(H) + 633.(D) + 3.18(KRORO) - 3.38(KRIRI)$$

$$t = 7.2 \quad t = 5.7 \quad t = 8.1 \quad t = -6.3 \qquad (4)$$

The variables used in equation 4 could completely explain all variations in cylinder weight if they had been assembled into the correct functional relationship. This situation is analogous to many research results that are based upon an imperfect knowledge of the underlying system.

STEPWISE MULTIPLE REGRESSION

A variation of multiple regression known as stepwise multiple regression (Program G2 BC MPRV, University of California Computer Center Library, Berkeley) has also been used for hydrologic data analysis [*Ralson and Wilf*, 1960; *Fritts*, 1962]. Given a wide choice of variables, a stepwise multiple regression tends to pick variables that confound several independent effects and to build models that are hard to interpret in terms of the real world. Its chief advantage seems to be that it produces an equation that uses a small number of predictor variables and has a comparatively high R^2 value. *Furnival* [1964] has described limitations of model building using stepwise regression.

Equation 5 resulted from using a stepwise multiple regression (95% significance level to bring a variable into the equation and 90% to eliminate it from the equation):

$$W = -310 + 2.78(2KROH) - 2.70(2KRIH) + 37.9(DDIAGO)$$

$$t = 12.0 \quad t = 9.2 \quad t = 7.6 \qquad (5)$$

R^2 for equation 5 was 0.82, but the variables selected by the stepwise procedure and their coefficients leave much to be desired in ease of understanding of the underlying system.

PRINCIPAL COMPONENT ANALYSIS

Principal component analysis, originated by Karl Pearson in the early 1900's (Program G2 BC PCPE, Computer Center Library, University of California, Berkeley) now forms one of the foundation stones of many forms of factor analysis [*Harman*, 1960]. This method has been discussed in numerous books [for example, *Anderson*, 1958, pp. 272–287; *Thurstone*, 1947, pp. 473–510], monographs [*Kendall*, 1961, pp. 70–75; *Burket*, 1964], and papers [*Snyder*, 1962; *Fiering*, 1964].

For some interpretation purposes it is useful to convert the eigenvalues and eigenvectors to a p times r matrix of factor weights (appendix 2, equation 11) by multiplying each of the r columns of the eigenvector matrix by the square root of its corresponding eigenvalue [*Harman*, 1960, p. 182]. The chief disadvantage of the principal component factor weight matrix is the difficulty in assigning names to the concepts represented by the factor loadings of each column of the factor weight matrix [*Thurstone*, 1947, p. 508]. Figure 2 illustrates a visual interpretation of the factor loadings that result from a 2-cluster system of variables projected onto the first and second principal component axes. It can be seen from the figure that the first component has high positive loadings on all variables; the second has high positive and negative loadings with comparatively few intermediate values. Variable loadings similar to those in Figure 2 are the rule rather than the exception when making principal component analyses of correlation matrices.

When the 14 predictor variables of the hollow cylinder problem were subjected to a principal component analysis, 8 dimensions accounted for 0.995 of the intercorrelations of the covariance matrix. The resulting factor weight

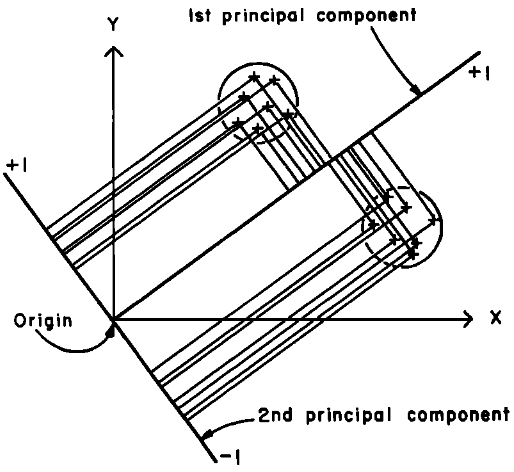


Fig. 2. High positive loadings for all variables on the first principal component and bipolar variable loadings upon the second component, for a two-cluster system of variables.

matrix (Table 1) is hard to interpret, but by considering just those variables of each dimension that have high loadings it appears that the first dimension is a general 'bigness' dimension, the second is 'squatness-slenderness,' the third is 'density,' and the fourth appears to be indeterminate and for want of a better name will be called a 'wall thickness' dimension. The other 4 dimensions have such low factor loadings that it is impossible to characterize the concept they represent. These dimensions are an unsatisfactory way of defining cylinders, but by rotating the factor weight matrix (Appendix 2, equation 12) the dimensions can be made to coincide more closely with groups of variables. The results achieved by two commonly used rotation methods, varimax and oblimax, are illustrated below.

Kendall [1961, pp. 72–73] has given an example of regression of a criterion (dependent variable) based upon a principal component solution of predictors (independent variables). Prediction equations based upon this method have three advantages over those from a multiple regression. First, the β coefficients tend to be stable even if intercorrelations are high. Second, the method does not capitalize on errors in the criterion observations; therefore it tends to give more reliable results than regression when the prediction equations are used with different populations [*Burket,* 1964]. Third, the rank of the predictor correlation matrix is determined by the number of positive eigenvalues (Appendix 2, m). Knowing this rank and observing the factor contributions and factor weight matrix, it is sometimes possible to estimate which variables are important and which are merely repetitious or irrelevant. The prediction equations obtained by this method are identical to those of a multiple regression if there is no collinearity and if enough dimensions are retained to stabilize the coefficients.

The chief disadvantage of principal component regression is that the prediction equation coefficients vary as the number of dimensions used is reduced. This variation is disconcerting if the goal is to estimate functional relationship, although it may be of slight importance

TABLE 1. Matrix of Factor Weights Resulting from a Principal Component Analysis of the 14 Hollow Cylinder Variables

Variable Symbol	Factor Loadings on Each of the Eight Dimensions 1	2	3	4	5	6	7	8
H	+513	+758	−163	+349	+047	−041	+042	+000
HH	+534	+723	−136	+354	+150	+120	−077	−003
2KROH	+741	+331	−233	+482	−128	−151	+105	−007
2KRIH	+817	+368	−302	−134	−258	+097	−107	+000
D	+447	+004	+887	−037	−032	−009	+035	+055
DD	+471	+029	+858	+011	−029	+112	+096	−132
DDIAGO	+665	−193	+698	+123	+012	−002	−028	+089
DDIAGI	+820	+084	+471	−218	+061	−137	−141	−010
RO	+539	−752	−193	+316	−040	+019	+021	−006
KRORO	+520	−755	−199	+330	+023	−031	−032	−017
DIAGO	+668	−560	−238	+411	+073	+080	−027	−001
RI	+746	−156	−290	−561	−027	+070	+079	+082
KRIRI	+738	−168	−304	−558	+038	−070	−028	−102
DIAGI	+850	+156	−316	−347	+145	+007	+104	+023

with some sets of data and for some prediction purposes. *Fiering* [1964] has shown that for generating synthetic hydrologic data for use with a computer simulator, all dimensions with positive eigenvalues are significant and should be included in the principal component regression. Psychologists concerned with functional relationship have usually suggested retaining a smaller number of dimensions and have developed numerous arbitrary rules of thumb for deciding how many eigenvalues to retain (for example, all eigenvalues greater than unity). The different rules of thumb do not always give the same cutoff point. If used they should be applied with care (an alternative procedure that allows for greater flexibility and that appears to work is suggested in the section on varimax rotation).

Prediction equation 6 resulted from a principal component regression of the first 8 dimensions (Table 1) of the hollow cylinder test data:

$$W = -781.2 + 50.9(H) - 3.06(HH) + 1.90(2KROH) - 2.09(2KRIH) + 588.0(D) - 373.0(DD) + 31.6(DDIAGO) - 9.35(DDIAGI) + 23.2(RO) + 0.364(KRORO) - 2.44(DIAGO) + 20.6(RI) - 2.96(KRIRI) + 27.5(DIAGI) \quad (6)$$

The R^2 for equation 6 was 0.81. Because of the orthogonal structure of the factor weight matrix, it is possible to allocate this explained variance among the 8 orthogonal dimensions. Factor contributions for the 8 dimensions of Table 1 were: 0.17, 0.00, 0.10, 0.43, 0.00, 0.04, 0.05, and 0.02. That is, only the 'bigness,' 'density,' and 'wall thickness' dimensions were effective contributors when explaining variance in weight.

That principal component regression coefficients are stable is demonstrated by comparing equations 6 and 7. Equation 7 was formed by removing variable (*RI*) and re-analyzing the remaining hollow cylinder data:

$$W = -701.2 + 56.4(H) - 3.06(HH) + 1.84(2KROH) - 2.02(2KRIH) + 581.4(D) - 366.9(DD) + 35.9(DDIAGO) - 15.48(DDIAGI) + 10.62(RO) + 0.735(KRORO) - 4.29(DIAGO) - 1.914(KRIRI) + 23.8(DIAGI) \quad (7)$$

The desirable structure which principal component regression is reputed to produce [*Kendall*, 1961] is not evident in either equation 6 or 7. However, when only the first 4 dimensions of Table 1 were used, 4 (*RI, KRIRI, DIAGI, DDIAGI*) of the 5 variables that might be expected to be measuring the hollowness of the cylinders received negative coefficients, whereas the remaining 12 variables had positive coefficients. In addition, the absolute values of the standardized regression coefficients for 12 of the variables were between 0.10 and 0.23, the remaining two (*DDIAGI* and *DIAGI*) being much smaller (0.028 and 0.023). Such a result is remarkably good when it is considered that the initially chosen model was incorrect (additive rather than multiplicative).

VARIMAX ROTATION OF A FACTOR WEIGHT MATRIX

Varimax rotation of a factor weight matrix (BC TRY system of factor and cluster analysis, Computer Center Library, University of California, Berkeley) has been used in hydrology [*Wong*, 1963; *Wallis*, 1965] and discussed in detail elsewhere [*Kaiser*, 1956]. A summation of the principles of varimax rotation (Appendix 2, *T*) has been provided by *Harman* [1960, pp. 301–308]. The method simplifies the columns of the factor weight matrix while maintaining an orthogonal structure. The effect of such a rotation can be visualized for 2 clusters of variables and 2 dimensions by referring to Figure 2 and imagining the factor loadings that would result from rotating the planes of the first and second principal components to the *X* and *Y* positions. Such a rotation tends to produce correspondence between the factor dimensions and the variables, resulting in fewer problems in naming the dimensions. Another advantage of varimax rotation is the great stability of the resulting dimensions when predictor variables are omitted from the analysis [*Kaiser*, 1956].

The factor weight matrix of the hollow cylinder that resulted from a varimax rotation of the principal component solution is given in Table 2. There are 4 orthogonal factors. By looking at the high loadings of each column, we can name the factors as 'height,' 'density,' 'outside radius,' and 'inside radius.' The factor loadings of Table 2 resulted from rotating all 8 dimensions of Table 1, but when the number of dimensions was curtailed before rotation the factor loadings remained essentially constant until the framework had been specified as 3-dimensional. Under these circumstances the dimensions became height, density, and outside-inside radius. Table 3 gives the factor loadings that result from performing a varimax rotation on only the first 2 dimensions of Table 1. It is apparent that varimax rotation of too few dimensions will give a nonorthogonal factor weight matrix even if true independence does exist among the variables; if too many dimensions are used the distortion is much less noticeable.

William Meredith of the University of California at Berkeley has recently formulated a principal component regression program that gives regression on criteria variables based upon the correlations with the rotated predictor factor weight matrix. The chief advantage of his formulation over that of principal component regression programs is that the factor contributions are assigned on the basis of the rotated factor weight matrix (Appendix 2, equation 13). (Access to this program is through subroutine SMIS (Symbolic Matrix Interpretive System) used as a component of BCTRY-AUX, Computer Center Library, University of California, Berkeley.) For the hollow cylinder data the rotated factor contributions were 0.221 for height, 0.206 for density, 0.180 for outside radius, and 0.098 for the inside radius dimension. In other words height, density, and outside radius were of about equal importance for predicting cylinder weight, whereas inside radius was of lesser importance.

TABLE 2. Matrix of Varimax Rotated Factor Weights for the 14 Variable Hollow Cylinder Problem

Variable Symbol	Factor Loadings on Dimensions			
	1	2	3	4
H	+980	+062	−120	+108
HH	+964	+096	−080	+107
2*KROH*	+872	+111	+360	+182
2*KRIH*	+636	+094	+086	+679
D	+007	+994	−030	+014
DD	+067	+980	−004	+007
DDIAGO	+097	+924	+329	+098
DDIAGI	+245	+786	+080	+500
RO	−063	+081	+972	+184
KRORO	−064	+065	+976	+162
DIAGO	+191	+092	+951	+207
RI	+035	+092	+222	+961
KRIRI	+030	+072	+234	+954
DIAGI	+418	+103	+155	+880

TABLE 3. Matrix of Factor Weights That Result from Varimax Rotation of the First Two Principal Component Dimensions of the Hollow Cylinder Test Problem

Variable Symbol	Factor Loadings on Dimension	
	1	2
H	881	−248
HH	875	−207
2*KROH*	780	226
2*KRIH*	861	246
D	344	285
DD	378	282
DDIAGO	384	577
DDIAGI	681	465
RO	−074	922
KRORO	−090	912
DIAGO	149	859
RI	469	600
KRIRI	455	605
DIAGI	750	430

Rules for model building using principal component analysis and varimax rotation of the factor weight matrix are still in an embryonic state of development. The tentative procedure suggested below will doubtless be modified.

(1) Know as much as possible about the system being investigated.

(2) Use only parameters that can reasonably be expected to relate to a single underlying process.

(3) Transform parameters so that they approach multivariate normal distribution.

(4) Make a principal component analysis of

the predictor variables using a high percentage of explained variance (0.995?).

(5) Make a varimax rotation of the principal component factor weight matrix.

(6) Retain no more than 2 defining variables per dimension (preferably only 1). Defining variables are those that have the highest factor loadings (0.900?). If 3 or more definers are in a dimension, pick the variable with the highest loading and pair the variable with the definer with which it has the lowest simple correlation.

(7) Make a principal component analysis with varimax rotation on the remaining variables. For this analysis dimensionality is set equal to the number of single defining variables plus ½ the number of paired defining variables.

(8) If variables have factor loadings of greater than 0.40 on 2 or more dimensions, attempt to redefine them to eliminate this confounding or look for additional observations where such confounding does not exist.

(9) Investigate the defining variable or variables of each dimension to see whether or not further transformation would increase its value as a predictor.

(10) If the model is still too complex, eliminate variables whose dimensions have low factor contributions for a given criterion (less than 0.05?).

(11) Check for autocorrelation.

The 11-step procedure given above will work as long as some measure of independence exists among the predictor variables. If none exists, then varimax rotation gives a factor weight matrix that is hard to interpret and is unstable [*Tryon*, 1961]. Data with oblique structures can be analyzed by key cluster analysis or by oblique rotation of a factor weight matrix.

Steps 4 through 8 of the above suggested procedure when used with the hollow cylinder data resulted in equation 8 and Table 4.

TABLE 4. Varimax Rotated Factor Weight Matrix for Eight of the Hollow Cylinder Variables

Variable Symbol	Factor Loadings 1	2	3	4
H	+987	−041	−032	+059
HH	+987	+070	+011	+056
D	+031	+995	−057	+030
DDIAGO	+092	+935	+311	+119
KRORO	−148	+092	+961	+198
DIAGO	+120	+109	+956	+227
RI	+064	+083	+194	+965
KRIRI	+057	+057	+210	+963
Factor Contributions for Criterion Variable				
W	0.164	0.224	0.203	0.091

different heights (7.17 and 1.37) and weights ($W = 401$ against $W = 77$). By observing that these cylinders had about equal height/weight ratios it could be truly concluded that the effect of height upon weight is multiplicative rather than additive. Transformation of the data could be made to allow for this observation, and the data could then be re-analyzed. Continuations of such procedures probably would lead to an estimating equation with correct functional relationship.

OBLIQUE SOLUTIONS (OBLIQUE ROTATION AND KEY CLUSTER ANALYSIS)

Oblique rotation (Program G2 BC FA80, University of California Computer Center Library, Berkeley) of a principal component factor weight matrix has been used in a study of watershed characteristics [*Aschenbrenner et al.*, 1963; *Maxwell*, 1964]. To visualize the effect of an oblique rotation of a principal component factor weight matrix, imagine the factor loadings on each component that would result from rotating the first and second component axes of

$$W = 695.6 + 27.1(H) + 2.96(HH) + 281.8(D) + 22.5(DDIAGO) + 1.111(KRORO) + 30.4(DIAGO) - 44.3(RI) - 1.64(KRIRI) \qquad (8)$$

Further deductive model building is possible with this equation (step 9 above). For example, two cylinders (18 and 29) had similar densities, inside radii, and outside radii, but widely

Figure 2 by different amounts so that each axis bisected a cluster of variables.

This procedure has two disadvantages. First, because the number of dimensions has already

TABLE 5. Matrix of Oblimax Rotated Factor Weights for the 14 Variable Hollow Cylinder Problem

Variable Symbol	Factor Loadings on Dimensions 1	2	3	4
H	+925	+158	−068	+351
HH	+914	+193	−034	+357
2KROH	+905	+246	+426	+491
2KRIH	+879	+245	+229	+849
D	+129	+978	+062	+127
DD	+178	+957	+083	+136
DDIAGO	+263	+959	+422	+290
DDIAGI	+523	+865	+241	+644
RO	+102	+180	+993	+347
KRORO	+090	+164	+988	+326
DIAGO	+338	+217	+978	+433
RI	+449	+230	+392	+961
KRIRI	+439	+215	+399	+954
DIAGI	+751	+262	+320	+972

TABLE 6. Matrix of Key Cluster Factor Weights for the 14 Variable Hollow Cylinder Problem

Variable Symbol	Factor Loadings on Dimensions 1	2	3	4
H	+998	+176	−071	+248
HH	+940	+211	−032	+256
2KROH	+895	+271	+416	+406
2KRIH	+740	+284	+235	+779
D	+103	+974	+054	+100
DD	+159	+943	+078	+106
DDIAGO	+217	+962	+417	+266
DDIAGI	+394	+880	+242	+612
RO	+027	+199	+990	+377
KRORO	+022	+183	+985	+358
DIAGO	+279	+240	+980	+440
RI	+191	+272	+406	+974
KRIRI	+185	+258	+414	+966
DIAGI	+557	+306	+333	+959
W	+502	+487	+395	−077

been set in the previously run orthogonal solution, rotation may result in more oblique factors than are needed to describe the system. Second, the method will give an oblique solution even if there is a true orthogonal functional relationship.

Table 5, which resulted from an oblique rotation of the factor weight matrix of Table 1, illustrates this second point. The four factors height, density, outside radius, and inside radius are comparable to those found by varimax rotation (Table 2), but the structure is not as clean as that of Table 2. For instance, the variables (*RI*) and (*KRIRI*) have loadings on the height factor of 449 and 439, but our external knowledge of the system tells us that this is an unreasonable property for a height factor.

An alternative method of obtaining an oblique solution called key cluster analysis has been suggested [*Tryon,* 1958, 1959]. The most independent subsets of variables are first selected and are then used to define the clusters (factors). The method gives fewer and stronger groupings of variables (dimensions) and answers that are usually easier to interpret than those derived by oblique rotation of a principal component solution. Table 6 gives the matrix of factor weights (correlations of variables with oblique cluster domains) obtained when the 60 test hollow cylinders were subjected to a key cluster analysis. The results are similar to those given by the oblique rotation (Table 5), although the factors tend to have fewer undesirable intermediate factor loadings.

Simple sum factor scores[2] for defining variables are easy to accumulate and can be used in subsequent regression analysis of criteria variables. Equation 9 gives the result of regressing observed cylinder weight (W) against simple sum factor scores with means of 50 and standard deviations of 10. (For this analysis defining variables were considered to be those whose key cluster factor coefficients were greater than 0.87. See BC TRY User's Manual for other alternatives.) The relative magnitude of the regression coefficients can be interpreted as approximate factor contributions to the explained variance in cylinder weight. Total R^2 for this equation was 0.69.

$$\begin{aligned} W = -1487.0 &+ 21.6(\text{height}) + 17.5(\text{density}) \\ & \quad t = +6.5 \qquad t = 5.3 \\ &+ 19.2(\text{outside radius}) - 22.5(\text{inside radius}) \\ & \quad t = +5.6 \qquad t = -6.3 \end{aligned} \tag{9}$$

[2] A simple sum factor score is made by summing the appropriate standardized definer variables. After standard scores are made for all the subjects on a given dimension, the scores are restandardized.

OBJECT ANALYSIS AND PREDICTION

Psychologists have long used a classification of objects based upon the similarities in test profiles for individuals. This technique has been referred to as Q-Mode analysis [*Imbrie,* 1963] and as O-analysis [*Tryon,* 1955, 1958]. Tryon's terminology will be used in this paper.

The groupings resulting from O-analysis can also be used to predict criteria variables not used in the preceding analysis of variables. For an excellent description of the mechanics and theory of the methodology, the reader should consult the sections called EUCO, OMARK, and PREDICT in the BC TRY User's Manual; for a practical example, refer to *Tryon* [1955].

For predictions based upon O-analysis, statistics of goodness of fit (ETA) and significance (F test) are available [*Quinn,* 1962]. By examining differences between the ability of O-analysis groups to predict individual criteria, it may be possible to evaluate the threshold effectiveness of certain predictor variables. With further refinements this procedure may lead to an understanding of the different relationship that exists for large and small events.

Prediction procedures based upon O-analysis were not effective with the hollow cylinder data. As might be expected, no natural clusters existed among the cylinders. As a result each of the 10 arbitrarily selected group definers was associated with numerous individuals that were comparatively dissimilar to the definer (i.e., that had oblique factor coefficients of less than 0.900). Under these circumstances, prediction of values for criteria variables is inaccurate.

DISCUSSION

The principal justification for generating a prediction equation from a set of observations is that it is expected that another set of observations of the predictor variables will become available and that the criterion variable can then be predicted from the equation rather than having to be measured. But if we also measure the criterion variable of the new population we can compare it with its predicted value and from this result assess the validity of the prediction equation.

Fifteen hollow cylinder test data not used in the original analyses were used for an independent test of the prediction equations developed by the different multivariate methods discussed above. The correlations between predicted and observed weights for some of the different equations were

0.77 for the all-variable regression (equation 2),
0.68 for the 4-variable regression (equation 4),
0.76 for the 3-variable stepwise regression (equation 5),
0.75 for the all-variable principal component regression (equation 6),
0.70 for the 8-variable principal component regression (equation 8),
0.69 for the key cluster regression (equation 9),
0.50 for the object analysis.

Except for the last result, these values are all quite similar to each other. As might have been expected, the much better fits—higher R^2 values originally displayed by equations 2, 5, and 6—have now largely evaporated.

The 8-variable principal component (equation 8), and key cluster regression solutions (equation 9) work reasonably well with the 15 new cylinders, but they also have two other advantages. First, it is possible to understand the line of thought that underlies the equations. It has been pointed out that scientific predictions must have this property [*Theil,* 1961]. Second, both equations have identified the true components of the underlying relationship, and it is therefore possible to start some further deductive model building. These two advantages also apply to the 4-variable regression (equation 4), but this equation depended upon an initial intuitive selection of the correct 4 variables.

CONCLUSION

It is evident that the different methods of generating prediction equations do not give identical answers, even when the same data are used. Furthermore, no panacea exists for all problems and data; therefore the method that will be most suitable for each specific problem should be selected. The combination of principal component regression with varimax rotation of the factor weight matrix is recommended for an initial analysis of multifactor hydrologic problems. If many observations are available, an object analysis based upon key cluster groupings may be regarded as a logical second step for the analysis.

APPENDIX 1. LISTING OF HOLLOW CYLINDER DATA
5 Variables to a Line, 3 Lines to an Observation

(H, HH, 2KROH, 2KRIH, D/DD, DDIAGO, DDIAGI, RO, KRORO/DIAGO, RI, KRIRI, DIAGI, W)

1	1	0.48080	0.23117	29.93040	24.30870	0.55283	1	1
1	2	0.30562	10.95760	8.90080	9.90760	308.38040	1	2
1	3	19.82100	8.04670	203.41610	16.10050	27.89950	1	3
2	1	7.19240	51.73062	398.09810	20.86470	0.15444	2	1
2	2	0.02385	2.93890	1.11990	8.80920	243.79380	2	2
2	3	19.02990	0.46170	0.66960	7.25140	270.06090	2	3
3	1	5.64100	31.82088	293.17750	262.51540	0.09551	3	1
3	2	0.00912	1.66930	1.51390	8.27170	214.95090	3	2
3	3	17.47870	7.40660	172.34060	15.85090	22.95720	3	3
4	1	3.88510	15.09400	91.20110	38.51780	0.47991	4	1
4	2	0.23031	4.04170	2.40200	3.73610	43.85170	4	2
4	3	8.42180	1.57790	7.82180	5.00530	67.17760	4	3
5	1	9.14280	83.59079	540.15230	492.58140	0.94893	5	1
5	2	0.90047	19.84240	18.44180	9.40280	277.75650	5	2
5	3	20.91030	8.57470	230.98710	19.43430	405.76560	5	3
6	1	4.00830	16.06647	151.97570	17.33470	0.28404	6	1
6	2	0.08068	3.61210	1.20370	6.03440	114.39780	6	2
6	3	12.71700	0.68830	1.48830	4.23810	128.54950	6	3
7	1	9.34190	87.27110	523.48230	98.49920	0.99542	7	1
7	2	0.99086	20.04280	9.88100	8.91840	249.87550	7	2
7	3	20.13510	1.67810	8.84670	9.92640	2241.35385	7	3
8	1	0.37040	0.13720	20.68120	12.60550	0.09454	8	1
8	2	0.00894	1.68060	1.02470	8.88640	248.08560	8	2
8	3	17.77660	5.41640	92.16610	10.83910	5.45990	8	3
9	1	7.81810	61.12269	458.61810	333.94010	0.30702	9	1
9	2	0.09426	6.21500	4.81520	9.33620	273.83570	9	2
9	3	20.24300	6.79810	145.18600	15.68370	308.79950	9	3
10	1	4.59720	21.13425	229.84990	23.11670	0.41796	10	1
10	2	0.17469	6.92370	2.03450	7.95740	198.92630	10	2
10	3	16.56540	0.80030	2.01210	4.86780	378.35990	10	3
11	1	6.08980	37.08566	380.01230	63.89590	0.65448	11	1
11	2	0.42834	13.59710	4.54560	9.93150	309.87000	11	2
11	3	20.77550	1.66990	8.76050	6.94540	1200.11760	11	3
12	1	4.03980	16.31998	182.43130	172.99140	0.74904	12	1
12	2	0.56106	11.18410	10.64880	7.18720	162.28160	12	2
12	3	14.93120	6.81530	145.92160	14.21660	49.50470	12	3
13	1	6.82450	46.57380	291.09670	115.98070	0.76011	13	1
13	2	0.57777	11.55060	6.61940	6.78870	144.78480	13	2
13	3	15.19600	2.70480	22.98370	8.70840	631.82750	13	3
14	1	2.70190	7.30026	150.74140	141.02910	0.62264	14	1
14	2	0.38768	11.18450	10.48080	8.87940	247.69490	14	2
14	3	17.96310	8.30730	216.80510	16.83280	51.96610	14	3
15	1	9.98680	99.73617	484.01370	451.05800	0.65869	15	1
15	2	0.43387	12.10500	11.53030	7.71350	186.91870	15	2
15	3	18.37730	7.18830	162.33120	17.50490	161.74120	15	3
16	1	8.79040	77.27113	349.50070	285.36560	0.01222	16	1
16	2	0.00015	0.18820	0.16570	6.32790	125.79660	16	2
16	3	15.40910	5.16670	83.86410	13.56650	4.50430	16	3
17	1	6.88510	47.40460	405.40140	361.89020	0.66444	17	1
17	2	0.44148	13.26680	12.02110	9.37120	275.89270	17	2
17	3	19.96700	8.36540	219.84830	18.09210	256.38810	17	3
18	1	7.17420	51.46915	351.72540	202.74650	0.43769	18	1
18	2	0.19157	7.51760	5.03600	7.80280	191.27170	18	2
18	3	17.17560	4.49780	63.55500	11.50600	401.04000	18	3
19	1	0.27420	0.07519	14.69980	2.65780	0.53842	19	1
19	2	0.28990	9.18910	1.66770	8.53230	228.70830	19	2
19	3	17.06680	1.54270	7.47670	3.09750	32.66140	19	3
20	1	5.61810	31.56305	194.16170	106.24810	0.97190	20	1
20	2	0.94459	12.00520	8.00270	5.50040	95.04700	20	2

APPENDIX 1. (Continued)

20	3	12.35230	3.00990	28.46120	8.23410	353.57350	20	3
21	1	5.53550	30.64176	307.26890	242.07970	0.66599	21	1
21	2	0.44354	12.33130	9.97690	8.83450	245.19620	21	2
21	3	18.51580	6.96020	152.19250	14.98060	342.86630	21	3
22	1	7.89610	62.34839	338.97350	142.30390	0.17823	22	1
22	2	0.03177	2.81280	1.73950	6.83240	146.65480	22	2
22	3	15.78210	2.86830	25.84630	9.75990	170.01640	22	3
23	1	9.26540	85.84764	575.46149	171.73200	0.67211	23	1
23	2	0.45173	14.67430	7.38260	9.88490	306.96890	23	2
23	3	21.83320	2.94990	27.33780	10.98430	1741.36559	23	3
24	1	1.37570	1.89255	69.46570	37.78190	0.55988	24	1
24	2	0.31347	9.03180	4.95470	8.03650	202.90080	24	2
24	3	16.13170	4.37100	60.02210	8.84950	110.04890	24	3
25	1	0.57760	0.33362	34.90930	0.62310	0.09876	25	1
25	2	0.00975	1.90080	0.06630	9.61910	290.68240	25	2
25	3	19.24680	0.17170	0.09260	0.67190	16.57630	25	3
26	1	7.10390	50.46539	416.47690	190.09200	0.53434	26	1
26	2	0.28552	10.66950	5.92640	9.33070	273.51320	26	2
26	3	19.96780	4.25880	56.98020	11.09120	821.93690	26	3
27	1	6.16720	38.03436	266.02410	112.94750	0.84450	27	1
27	2	0.71318	12.71120	7.16670	6.86520	148.06630	27	2
27	3	15.05180	2.91480	26.69110	8.48630	632.14610	27	3
28	1	4.29880	18.47968	252.33960	91.24570	0.74651	28	1
28	2	0.55728	14.31270	5.97800	9.34240	274.19950	28	2
28	3	19.17290	3.37820	35.85250	8.00800	764.87859	28	3
29	1	1.36520	1.86377	66.60920	47.54670	0.60427	29	1
29	2	0.36514	9.42080	6.74950	7.76530	189.43760	29	2
29	3	15.59040	5.54300	96.52490	11.16970	76.64820	29	3
30	1	5.31860	28.28751	200.54650	175.81410	0.46438	30	1
30	2	0.21565	6.09630	5.47500	6.00120	113.14250	30	2
30	3	13.12800	5.26110	86.95660	11.79000	64.67520	30	3
31	1	4.78900	22.93452	276.39930	144.14390	0.58708	31	1
31	2	0.34466	11.14590	6.28820	9.18570	265.07840	31	2
31	3	18.98530	4.79040	72.09300	10.71100	542.58340	31	3
32	1	6.56540	43.10448	283.91430	212.73870	0.27896	32	1
32	2	0.07782	4.25420	3.41070	6.88250	148.81340	32	2
32	3	15.25050	5.15710	83.55270	12.22640	119.52390	32	3
33	1	0.03240	0.00105	1.79060	0.48300	0.97213	33	1
33	2	0.94504	17.10150	4.61320	8.79590	243.05820	33	2
33	3	17.59180	2.37270	17.68620	4.74550	7.09850	33	3
34	1	4.82440	23.27483	257.14780	153.23620	0.26486	34	1
34	2	0.07015	4.67180	2.96700	8.48320	226.08370	34	2
34	3	17.63890	5.05520	80.28350	11.20240	186.30200	34	3
35	1	6.40810	41.06374	340.71590	200.39430	0.88831	35	1
35	2	0.78909	16.07560	10.51620	8.46220	224.96570	35	2
35	3	18.09690	4.97710	77.82200	11.83840	837.59789	35	3
36	1	6.06170	36.74421	221.33370	90.36850	0.39900	36	1
36	2	0.15920	5.23020	3.07160	5.81130	106.09530	36	2
36	3	13.10830	2.37270	17.68620	7.69820	213.82790	36	3
37	1	7.28600	53.08579	181.44160	17.64790	0.88421	37	1
37	2	0.78183	9.51990	6.47830	3.96340	49.34980	37	2
37	3	10.76660	0.38550	0.46680	7.32660	314.92130	37	3
38	1	0.46310	0.21446	25.21320	9.36520	0.56596	38	1
38	2	0.32031	9.81170	3.65260	8.66510	235.88320	38	2
38	3	17.33630	3.21860	32.54490	6.45380	53.29410	38	3
39	1	0.68840	0.47389	28.12510	14.44050	0.15018	39	1
39	2	0.02255	1.95570	1.00800	6.50240	132.83030	39	2
39	3	13.02300	3.33860	35.01690	6.71250	10.11230	39	3
40	1	2.66110	7.08145	126.74400	21.58740	0.42983	40	1
40	2	0.18475	6.61610	1.59380	7.58030	180.51880	40	2
40	3	15.39230	1.29110	5.23680	3.70790	200.49120	40	3
41	1	9.95310	99.06420	111.48500	77.10200	0.19605	41	1

APPENDIX 1. (Continued)

41	2	0.03844	2.07270	2.01020	1.78270	9.98400	41	2
41	3	10.57240	1.23290	4.77530	10.25390	10.16370	41	3
42	1	6.48820	42.09674	343.86180	241.10970	0.09310	42	1
42	2	0.00867	1.68270	1.25600	8.43490	223.51650	42	2
42	3	18.07440	5.91440	109.89330	13.49130	68.63420	42	3
43	1	8.95520	80.19561	433.89310	260.62950	0.82626	43	1
43	2	0.68271	14.73550	10.64610	7.71130	186.81210	43	2
43	3	17.83400	4.63200	67.40420	12.88470	883.53780	43	3
44	1	1.61440	2.60629	25.31530	19.19570	0.46064	44	1
44	2	0.21219	2.41650	1.89540	2.49570	19.56740	44	2
44	3	5.24590	1.89240	11.25060	4.11470	6.18480	44	3
45	1	1.05890	1.12127	43.33270	23.75680	0.82935	45	1
45	2	0.68782	10.83870	5.98740	6.51300	133.26370	45	2
45	3	13.06890	3.57070	40.05490	7.21940	81.85580	45	3
46	1	1.71410	2.93814	88.15030	34.50060	0.74525	46	1
46	2	0.55540	12.26610	4.94250	8.18480	210.45820	46	2
46	3	16.45900	3.20340	32.23830	6.63210	227.66400	46	3
47	1	1.04740	1.09705	55.04940	18.90390	0.27481	47	1
47	2	0.07552	4.60650	1.60480	8.36490	219.82210	47	2
47	3	16.76250	2.87250	25.92200	5.83960	55.81140	47	3
48	1	6.55160	42.92346	70.74600	42.78260	0.43308	48	1
48	2	0.18756	3.20410	2.97670	1.71860	9.27890	48	2
48	3	7.39850	1.03930	3.39330	6.87340	16.69950	48	3
49	1	9.00600	81.10803	469.33290	247.88220	0.31997	49	1
49	2	0.10238	6.03950	4.02020	8.29410	216.11670	49	2
49	3	18.87520	4.38060	60.28600	12.56440	449.04930	49	3
50	1	9.35720	87.55719	330.39910	159.62890	0.88285	50	1
50	2	0.77942	12.91130	9.55120	5.61970	99.21470	50	2
50	3	14.62460	2.71510	23.15900	10.81860	628.29580	50	3
51	1	6.39680	40.91905	367.23280	0.77970	0.20203	51	1
51	2	0.04082	3.91150	1.29230	9.13690	262.26930	51	2
51	3	19.36100	0.01940	0.00110	6.39690	338.94110	51	3
52	1	5.41800	29.35472	304.08620	200.61800	0.42627	52	1
52	2	0.18171	7.95790	5.52950	8.93260	250.67180	52	2
52	3	18.66860	5.89320	109.10690	12.97200	326.94870	52	3
53	1	7.69120	59.15456	353.12690	252.11750	0.03426	53	1
53	2	0.00117	0.56570	0.44400	7.30730	167.75040	53	2
53	3	16.51480	5.21710	85.50820	12.96250	21.67080	53	3
54	1	1.52480	2.32501	77.36140	40.87560	0.64603	54	1
54	2	0.41735	10.47950	5.59980	8.07480	204.83930	54	2
54	3	16.22140	4.26650	57.18640	8.66810	145.44790	54	3
55	1	8.88940	79.02143	335.03870	334.04450	0.87977	55	1
55	2	0.77400	13.13620	13.11110	5.99850	113.04070	55	2
55	3	14.93140	5.98070	112.37090	14.90290	5.23880	55	3
56	1	7.41350	54.95998	237.29450	188.28720	0.59516	56	1
56	2	0.35422	7.49910	6.52820	5.09430	81.53020	56	2
56	3	12.60020	4.04220	51.33160	10.96890	133.24280	56	3
57	1	4.10490	16.85020	57.31980	6.77550	0.25267	57	1
57	2	0.06384	1.52870	1.04560	2.22240	15.51650	57	2
57	3	6.05030	0.26270	0.21680	4.13830	15.86860	57	3
58	1	5.72230	32.74472	151.35310	87.99080	0.52239	58	1
58	2	0.27289	5.31780	3.93360	4.20960	55.67130	58	2
58	3	10.17970	2.44730	18.81580	7.53000	110.17090	58	3
59	1	4.88030	23.81733	118.30710	67.06480	0.59520	59	1
59	2	0.35426	5.43420	3.90070	3.85820	46.76480	59	2
59	3	9.13010	2.18710	15.02750	6.55360	92.18910	59	3
60	1	9.60520	92.25986	278.18300	264.61000	0.53984	60	1
60	2	0.29143	7.18700	7.02110	4.60940	66.74800	60	2
60	3	13.31330	4.38450	60.39340	13.00590	32.95020	60	3
61	1	6.15480	37.88156	352.14880	122.48880	0.86071	61	1
61	2	0.74082	16.54630	7.60210	9.10610	260.50410	61	2
61	3	19.22400	3.16740	31.51770	8.83230	1213.05489	61	3

APPENDIX 1. (Continued)

62	1	4.10720	16.86909	146.45840	1.02450	0.22093	62	1
62	2	0.04881	2.66680	0.90750	5.67530	101.18760	62	2
62	3	12.07080	0.03970	0.00490	4.10790	91.81350	62	3
63	1	9.72260	94.52895	273.11620	87.52810	0.61712	63	1
63	2	0.38084	8.15160	6.25510	4.47080	62.79430	63	2
63	3	13.20910	1.43280	6.44940	10.13610	338.06990	63	3
64	1	3.77190	14.22723	209.82150	206.45140	0.47309	64	1
64	2	0.22381	8.56480	8.43330	8.85340	246.24640	64	2
64	3	18.10400	8.71120	238.39970	17.82600	14.00200	64	3
65	1	8.59450	73.86543	525.68188	471.51370	0.18903	65	1
65	2	0.03573	4.02290	3.67910	9.73470	297.71100	65	2
65	3	21.28190	8.73160	239.51760	19.46350	94.54210	65	3
66	1	8.30830	69.02785	380.95360	7.13080	0.77025	66	1
66	2	0.59329	12.93570	6.40290	7.29760	167.30540	66	2
66	3	16.79420	0.13660	0.05860	8.31270	1070.29037	66	3
67	1	0.11840	0.01402	2.76040	1.52670	0.02697	67	1
67	2	0.00073	0.20010	0.11070	3.71060	43.25510	67	2
67	3	7.42210	2.05230	13.23210	4.10630	0.09580	67	3
68	1	7.79830	60.81348	345.94170	31.73120	0.00216	68	1
68	2	0.00000	0.03480	0.01700	7.06030	156.60150	68	2
68	3	16.13080	0.64760	1.31750	7.90510	2.61560	68	3
69	1	1.66400	2.76890	85.43690	50.61370	0.78495	69	1
69	2	0.61615	12.89500	7.71130	8.17170	209.78510	69	2
69	3	16.42780	4.84100	73.62410	9.82390	177.84760	69	3
70	1	9.78590	95.76384	429.10280	317.16640	0.27698	70	1
70	2	0.07672	4.72150	3.93850	6.97880	153.00700	70	2
70	3	17.04630	5.15830	83.59160	14.21950	188.15010	70	3
71	1	1.06900	1.14276	33.66420	22.75890	0.42176	71	1
71	2	0.17788	4.25160	2.89350	5.01200	78.91720	71	2
71	3	10.08080	3.38840	36.06940	6.86050	19.31840	71	3
72	1	3.46290	11.99168	195.30230	50.15880	0.26907	72	1
72	2	0.07240	4.91940	1.55150	8.97610	253.11920	72	2
72	3	18.28310	2.30530	16.69570	5.76620	220.29060	72	3
73	1	1.71200	2.93094	71.70490	9.61980	0.19859	73	1
73	2	0.03944	2.66930	0.49160	6.66600	139.59840	73	2
73	3	13.44140	0.89430	2.51250	2.47580	46.60720	73	3
74	1	2.61910	6.85968	160.03240	117.96690	0.38744	74	1
74	2	0.15011	7.60340	5.64660	9.72470	297.09970	74	2
74	3	19.62490	7.16850	161.43820	14.57420	137.66170	74	3
75	1	4.77040	22.75672	198.72300	52.15950	0.97253	75	1
75	2	0.94581	13.70480	5.74280	6.63000	138 09460	75	2
75	3	14.09190	1.74020	9.51360	5.90500	596.53310	75	3

APPENDIX 2

Given:

Rxx Correlation matrix of predicter variables (n, n).

Rxy Correlation matrix of the predictor variables with the criteria variables.

Δ Diagonal matrix of eigenvalues of the *Rxx* matrix whose diagonal elements, $\lambda_1 > \lambda_2 > \lambda_3 \cdots > \lambda_m > 0$, are what mathematicians refer to as 'characteristic roots' of the *Rxx matrix* $(m \leq n)$.

Q Is the eigenvector matrix associated with Δ in equation 10.

A Principal component factor weight matrix.

B Varimax rotated factor weight matrix.

T Orthonormal transformation matrix for detailed computational equations of the Varimax transformation matrix (see equations 18.4.1 to 18.4.10 in *Horst*, 1964).

β Matrix of square roots of rotated factor contributions for each of *m* dimensions on the *y* criteria variables.

Then:

$$Rxx = Q \Delta Q' \quad (10)$$

$$A = \sqrt{\Delta}\, Q \quad (11)$$

$$B = AT \quad (12)$$

$$\beta = (BB')^{-1} B' Rxy \quad (13)$$

Note that in equation 13 prime refers to the transpose of the matrix and −1 to the inverse of the matrix.

Acknowledgment. The research reported in this paper was supported by the Pacific Southwest Forest and Range Experiment Station, Forest Service, U. S. Department of Agriculture, Berkeley, California.

REFERENCES

Anderson, H. W., Suspended discharge as related to streamflow, topography, soil, and land use, *Trans. Am. Geophys. Union, 35,* 268–281, 1954.

Anderson, H. W., Relating sediment yield to watershed variables, *Trans. Am. Geophys. Union, 38,* 921–924, 1957.

Anderson, T. W., *Introduction to Multivariate Statistical Analysis,* 374 pp., John Wiley and Sons, New York, 1958.

Aschenbrenner, B. C., W. Burgess, F. Doyle, J. Imbrie, and J. Maxwell, *Applicability of Certain Multifactor Computer Programs to the Analysis, Classification, and Prediction of Landforms,* 114 pp., Defense Documentation Center, AD. 428030, 1963.

Bartlett, M. S., The use of transformations, *Biometrics, 3,* 39-52, 1947.

Benson, M. A., Areal flood-frequency analysis in a humid region, *Intern. Assoc. Sci. Hydrol., 19,* 1960.

Burket, G. R., A study of reduced rank models for multiple prediction, *Psychometric Monograph, 12,* 66 pp., 1964.

Durbin, J., and G. S. Watson, Testing for serial correlation in least squares regression, 2, *Biometrika, 38,* 159–178, 1951.

Fiering, Myron B, Multivariate technique for synthetic hydrology, *Proc. Am. Soc. Civ. Eng.,* pp. 43–60, 1964.

Ford, P. M., Multiple correlation in forecasting seasonal runoff, *U. S. Bur. of Reclamation, Engineering Monographs, 2,* 41 pp., 1959.

Fritts, H. C., An approach to dendroclimatology. Screening by means of multiple regression techniques, *J. Geophys. Res., 67*(4), 1413–1420, 1962.

Furnival, George M., More on the elusive formula of best fit, Paper presented at the Society of American Foresters annual meeting, Denver, Colorado, 1964.

Harman, H. H., *Modern Factor Analysis,* 471 pp., University of Chicago Press, Chicago, 1960.

Horst, Paul, The factor analysis of data matrices, 177 p. (Mimeographed), Defense Documentation Center for Scientific and Technical Information, Cameron Station, Alexandria, Va., 1964.

Imbrie, J., Factor and vector programs for analyzing geologic data, *Tech. Rept. No. 6, Task No. 389–135, ONR Contract No. 1228(26),* 1963.

Johnston, J., *Econometric Methods,* 300 pp., McGraw-Hill Book Company, New York, 1960.

Kaiser, Henry F., The varimax method of factor analysis, Ph.D. thesis, University of California, Berkeley, June, 1956.

Kendall, M. G., *A Course in Multivariate Analysis,* 2nd ed., 185 pp., Hafner Publishing Company, New York, 1961.

Leopold, L. B., Rivers, *Am. Scientist, 50*(4), 511–537, 1962.

Maxwell, J. C., Comparison of watershed and other terrain data by multifactor analysis, Paper presented at 45th annual meeting of American Geophysical Union, Washington, D. C., April 23, 1964.

Quinn, McNemar, *Psychological Statistics,* 451 pp., John Wiley and Sons, New York, 1962.

Ralson, A., and H. Wilf, *Mathematical Methods for Digital Computers,* 293 pp., John Wiley and Sons, New York, 1960.

Snyder, W. M., Some possibilities for multivariate analysis in hydrologic studies, *J. Geophys. Res., 67*(2), 721–729, 1962.

Theil, H., *Economic Forecasts and Policy,* 567 pp., North Holland Publishing Company, Amsterdam, 1961.

Thurstone, L. L., *Multiple-Factor Analysis,* 535 pp., University of Chicago Press, Chicago, 1947.

Tryon, R. C., Identification of social areas from cluster analysis, University of California publication in *Psychology, 8*(1), 1–100, 1955.

Tryon, R. C., General dimensions of individual differences, cluster analysis vs. multiple factor analysis, *Educational Psychology Measurement, 18,* 477–495, 1958.

Tryon, R. C., Domain sampling formulation of cluster and factor analysis, *Psychometrika, 24,* 113–135, 1959.

Tryon, R. C., Cluster structure vs. simple structure, Paper presented at the annual meeting of Psychonomics Society, Columbia University, New York, September 1961.

Wallis, James R., A factor analysis of soil erosion and stream sedimentation in northern California, Ph.D. thesis, University of California, Berkeley, January 1965.

Wong, S. T., A multivariate statistical model for predicting mean annual flood in New England, *Ann. Assoc. Am. Geog., 53*(3), 298–311, 1963.

(Manuscript received May 14, 1965.)