SELECTED HYDROLOGICAL STUDIES USING MULTIVARIATE STATISTICAL TECHNIQUES

Edited by
A. S. Hope

Hydrological Research Unit
University of Zululand

Research Funded by The Water Research Commission
Republic of South Africa

September 1984

WRC Report 66/3/84 ISBN 908356 20 X

--oOo--

PREFACE

In recent years increasing attention has been given to the application of multivariate statistical techniques to hydrological problems. One of the research objectives of the Zululand Hydrology Project is to develop physiographically based models to calculate runoff from small catchments. The potentially large number of physiographical variables which need to be investigated in this study resulted in multivariate statistical techniques being selected as the suitable statistical approach to identify variables which are related to catchment runoff. Dr. E. Seyhan of the Institute of Earth Sciences, Free University, Amsterdam, was invited to spend seven weeks in South Africa as guest of the Zululand Project, this scientist having extensive experience in applying multivariate statistical techniques to hydrological problems. During his stay in South Africa, Dr Seyhan presented a number of lectures on this topic and assisted in laying the foundation for future studies of the Zululand Project based on multivariate techniques.

Although the development of physiographically based equations to calculate runoff was the major research objective during Dr Seyhan's visit, this report also deals with two additional hydrological studies which included multivariate procedures. The three studies may be regarded as preliminary investigations which were intended to illustrate the use of these statistical procedures in hydrological analyses. This report contains three papers which emanated from the research conducted while Dr Seyhan was in South Africa. These papers have been submitted to scientific journals for publication.

Most of the research presented in this report was conducted at the University of Natal in the Department of Agricultural Engineering and

thanks are due to Professor P. Meiring for placing the facilities of the Department at the authors' disposal. The funding of Dr Seyhan's visit to South Africa by the Water Research Commission is also gratefully acknowledged.

CONTENTS

I

CALCULATING RUNOFF FROM CATCHMENT PHYSIOGRAPHY IN SOUTH AFRICA

CALCULATING RUNOFF FROM CATCHMENT PHYSIOGRAPHY IN SOUTH AFRICA

E. Seyhan and A.S. Hope

ABSTRACT

Multivariate statistical techniques and regression analysis are used to develop runoff equations for small catchments ($< 100\ km^2$) in two environmentally different regions of South Africa, the winter rainfall region and the summer rainfall region east of the 800 mm isohyet. Fourteen physiographical and three rainfall variables are selected as independent variables while the dependent runoff variables are the mean annual runoff, mean annual flood and 10-year return period flood. The statistical procedures are described and results are presented and discussed. The catchments of the two regions are shown to be physiographically and hydromorphometrically distinct. Regression equations for these two regions are also notably different and despite inadequacies in the runoff data are considered to be highly satisfactory.

INTRODUCTION

Conventional methods of collecting data for hydrological studies are generally time consuming and costly. In recent decades hydrologists have turned increasingly to mathematical and statistical techniques in order to reduce to a minimum the potentially large number of variables which may be required for a particular investigation (Seyhan and Keet, 1981). Greater use is now being made of multivariate statistical techniques to study multidimensional hydrological problems and to develop uncorrelated components which are able to define the given multivariate population in a more simplified structure. Numerous multivariate statistical techniques exist along with a wide variety of solution procedures and similarity measures which may be used in these procedures.

Some of the most widely used multivariate techniques are factor analysis, principal components analysis, cluster analysis, canonical correlation analysis and discriminant analysis while examples of solution procedures are principal components, principal factors, centroid and maximum-likelihood. Similarity measures which are adopted in most analyses are square Euclidean distance, product-moment correlation coefficient, variance, matching coefficient and average distance. A description of these multivariate statistical techniques is given by Seyhan (1981a).

Runoff from a catchment is largely a function of catchment variables which operate on the primary input, rainfall. In attempting to develop empirical equations to calculate catchment runoff the first step is to identify and quantify those climate and catchment variables which influence runoff. Since these variables are generally highly interrelated it is necessary to establish the underlying dimensions of interrelatedness.

The first thorough quantitative study of river systems in a catchment was undertaken by Horton (1945). Following Horton, pioneering work in the field of quantitative geomorphology directed at the formulation of catchment variables was conducted by, <u>inter alia</u>, Miller (1953), Schumm (1956), Strahler (1957), Morisowa (1958) and Maxwell (1960). These early works showed that the catchment variables which are most significantly related to runoff are : climate, catchment area, mean catchment altitude above the outlet, mean distance to the outlet, channel slope, catchment lag, surface storage, vegetation and soil. However, none of these authors made any attempt to study the interdependency of the variables in their multiple regression analyses.

Research conducted by Hope (1979) on small catchments (<100 km^2) in Natal, South Africa indicated that readily measurable physiographical and climatic variables could be used to calculate selected runoff characteristics, such as the mean annual runoff, of these catchments. The purpose of this investigation is to extend the research initiated by Hope (1979) to include a larger sample of catchments, catchments from different environmental regions and to incorporate multivariate statistical techniques in the development of predictive equations.

AIMS AND PROCEDURES

The mean annual runoff, mean annual flood and 10-year return period flood are runoff characteristics which are widely used in the design of water storage and control structures in small catchments. This study had, as its central aim, the objective of developing physiographically based regression equations to calculate these three runoff variables for catchments in two regions of South Africa. The two selected regions were the winter rainfall region of the South-West Cape and summer rainfall region east of the 800 mm isohyet (Figure 1). Background information to these two regions is given in Table 1.

In order to achieve the objective of this study three specific aims were defined, viz., to

a) determine whether the catchments of the winter and summer rainfall regions differed in terms of the measured runoff, rainfall and physiographical variables (Q-mode analysis),

b) establish the underlying dimensions and interrelationships of the catchment variables in the two regions (R-mode analysis) and

c) screen and select independent physiographical and rainfall variables to be included in regression equations for calculating the selected runoff characteristics.

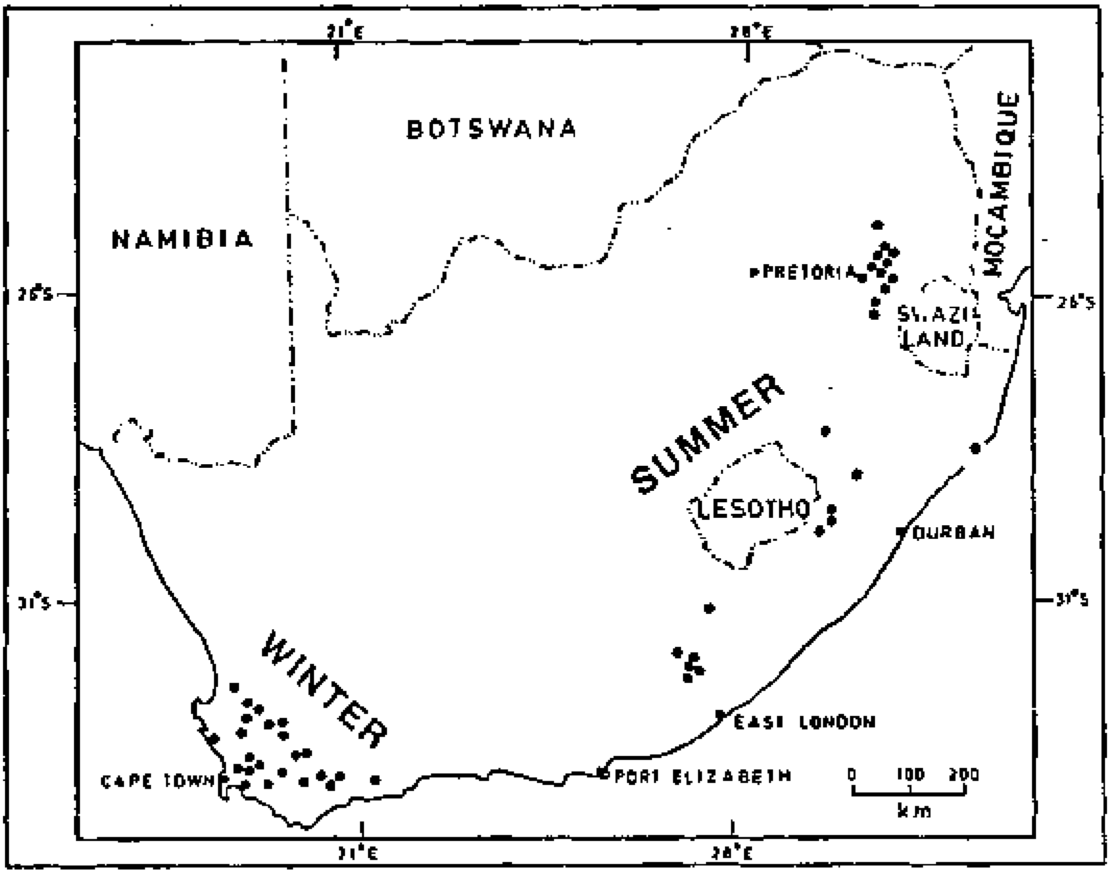


Figure 1 : Study catchments in the summer and winter rainfall regions of South Africa

Table 1 : Background information to the winter rainfall region and summer rainfall region

Variable	Rainfall Region	
	Winter	Summer
Mean Annual Rainfall (mm)	300 - 1000	600 - 1000
Mean Daily Temperature (°C)	15,0 - 17,5	12,5 - 22,5
Free Water Evaporative Losses (mm/yr)	< 1250 - 1600	< 1250 - 1750
Major Natural Vegetation	Mediterranean	Temperate Forest, Sub-Tropical Coastal Forest, Temperate Grasses
Geology	Sandstone, quartzite, unconsolidated superficial deposits	Shale, mudstone, sandstone limestone, tillite, granite

STATISTICAL PROCEDURES

The selection of suitable multivariate statistical techniques for analysing a given body of data is not determined by set rules. Generally, such a selection is based on the past experience of the investigator who should take cognisance of the sampling procedures applied, measurement errors and the statistical assumptions involved as well as the statistical and mathematical structure of the selected multivariate techniques. .

Factor analysis is the most widely used multivariate statistical method in hydrology. The primary objective of this technique is to represent a given group of variables in terms of several factors (Hotelling, 1933; Kaiser, 1958; Harman, 1968 and Seyhan, 1981a). In factor analysis an attempt is made to reduce the original group of variables to a smaller number of factors which will account for the observed variance in the given data. Various types of factor analysis exist and are described by Seyhan (1981a). These methods differ in terms of the format of the input data matrix, the procedure for extracting the initial factors, the type of rotation applied to established factors and the method of computing factor scores (Seyhan, 1981a).

In order to determine whether the selected catchments for this study could be classified into two significantly distinct groups (Q-mode analysis), linear discriminant analysis was selected as the required statistical procedure. Discriminant analysis provides a numerical criterion for classifying the cases (catchments) into two or more statistically distinguishable groups (categories or classes). Discriminant analysis may be regarded as a special type of factor analysis that extracts orthogonal factors to show the differences among several groups. Mathematically, discriminant analysis can be viewed as the development of discriminant functions (of the discriminant axes) that best separate the multivariate samples (Nie, Hull, Jenkins, Steinbrenner and Bent, 1975).

Identification of the independent underlying dimensions in the physiographical and hydromorphometrical data of the selected catchments (R-mode analysis) was made using principal components factor analysis. The selection of this statistical technique was based on the findings of Seyhan and Keet (1981) who found it to be an appropriate technique for R-mode analysis of data from catchments in Italy. Principal components factor analysis was applied to the correlation matrix of the input data (using Pearson's product-moment correlation coefficients) with variance orthogonal rotation criterion (Seyhan, 1981a).

The procedural steps by which the R-mode and Q-mode analysis were conducted arepresented in Figure 2. As illustrated in Figure 2 allowance had to be made for positive or negative interpretation of the results. In the case of a negative or meaningless interpretation the statistical selection of significant catchment variables could have been repeated with different selection criteria. Alternatively additional catchment variables may have had to be selected and the analysis repeated.

STUDY CATCHMENTS AND DATA

Forty-eight catchments were selected for this study, 25 from the summer rainfall region and 23 from the winter rainfall region. Catchment selection was based on a number of criteria, viz. :

a) Only catchments with areas in the range four to 100 km^2 were considered. These limits ensured that all the required physiographical information could be derived from 1:50 000 scale maps and hence standardise the source of this information.

b) A minimum of 10 years of corresponding rainfall and runoff data were required for each catchment.

d) No substantial hydraulic structures or diversions were within the catchments.

d) The catchments were not underlain by complex geological features such as large scale faulting or karstification.

Most of the potentially suitable catchments for this study were rejected on the basis of the quality or duration of the runoff record.

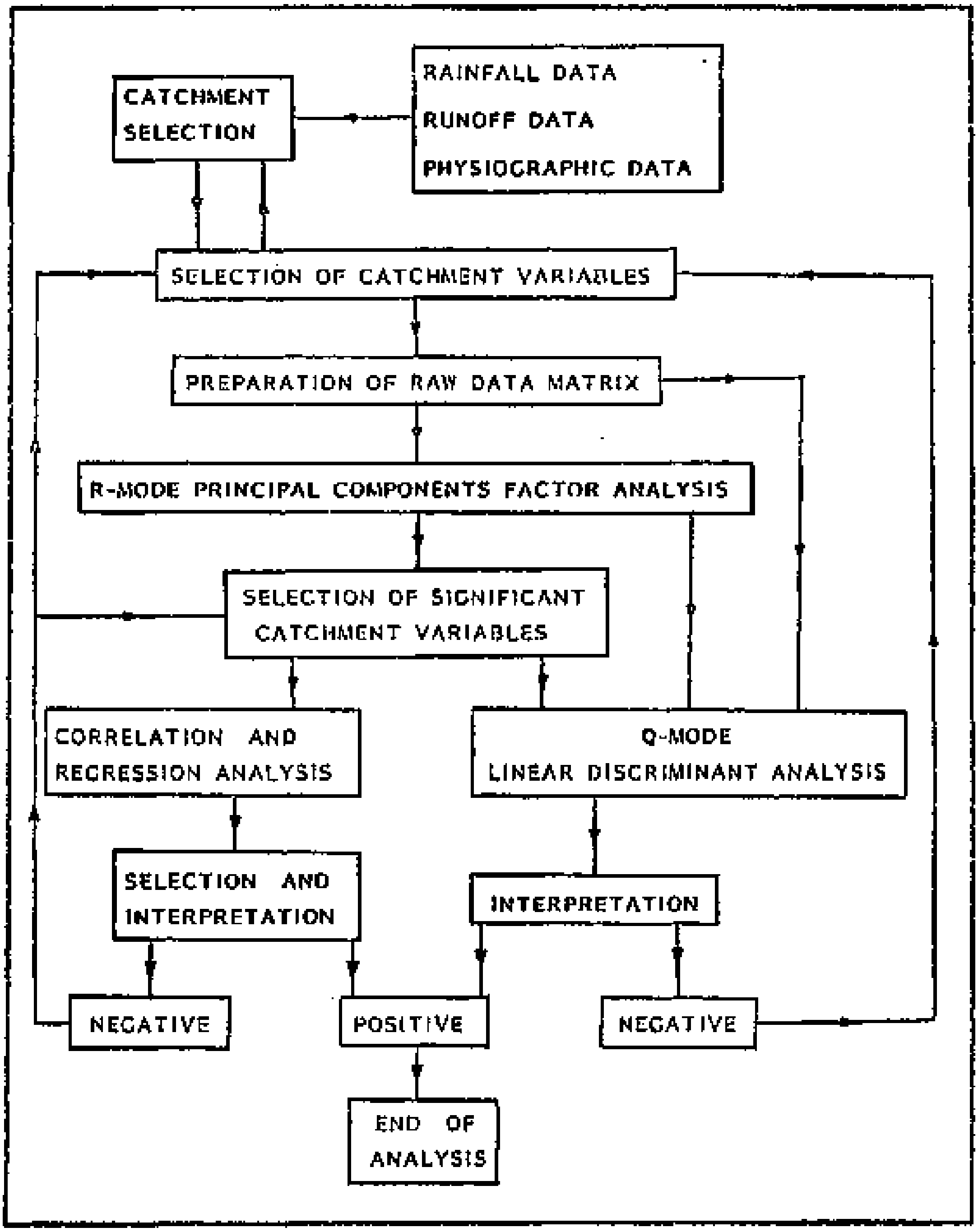


Figure 2 : Procedural steps for R-mode and Q-mode analyses

Numerous variables in a catchment may influence, directly or indirectly, the transfer of rainfall to runoff. Some of these variables, such as catchment area, determine the potential volume of runoff while others, such as land-use or soils, modify the output. Attempts to relate long term runoff characteristics to catchment and climatic variables by authors such as Seyhan (1976), Hope (1979), Seyhan (1981b) and Seyhan and Keet (1981) have revealed that in most areas physiographical variables have the best association with the runoff variables. The selection of independent variables for this study was based on these earlier findings and were of two types, namely, rainfall variables and physiographical variables. The dependent variables were the mean annual runoff, mean annual flood and the 10-year return period flood. The independent and dependent variables tested in this study are given in Table 2.

Procedures for calculating the physiographical variables are described by inter alia Seyhan (1976), Seyhan (1977) and Seyhan and Keet (1981). Mean annual precipitation values were obtained from the standard 1: 250 000 rainfall maps of South Africa published by the Government Printer while the 2,33-years, 24 hour duration rainfall and 10-years, 24 hour duration rainfall were obtained from maps published by Schulze (1982). Runoff variables were calculated from data obtained from the Department of the Environment, Directorate of Water Affairs, Pretoria. The mean and standard deviation values for each of the rainfall, physiographical and runoff variables for the two regions selected catchments are given in Table 3. Analyses were conducted using linear data and then repeated using logarithmically transformed data.

Table 2 : Selected dependent and independent variables

Runoff variables:

MAR	=	mean annual discharge ($x10^6$ m^3 s^{-1})
MAF	=	mean annual flood (m^3 s^{-1})
MA10	=	10-year return period flood (m^3 s^{-1})

Physiographical variables:

A	=	area of catchment (km^2)
S85	=	mean slope of main channel by 85-10 slope factor (m km^{-1})
D	=	drainage density (km^{-1})
P	=	drainage perimeter (km)
W	=	width of watershed (km)
LB	=	main channel length (km)
HM	=	maximum basin relief (m)
T	=	topographic factor (km^3 m^{-1})
RH	=	Schumm's relief ratio (m km^{-1})
RHP	=	Melton's relief ratio (m km^{-1})
RF	=	form factor (-)
RC	=	circularity ratio (-)
RE	=	elongation ratio (-)
C	=	compactness ratio (-)

Rainfall variables:

MAP	=	mean annual precipitation (m)
I2,33	=	intensity of 2,33-years 24 hour duration rainfall (mm day^{-1})
I10	=	intensity of 10-years 24 hour duration rainfall (mm day^{-1})

Table 3 : Mean ($\bar{x}$) and standard deviation (s) values of runoff, physiographical and rainfall variables for catchments in the winter rainfall region (N = 23) and summer rainfall region (N = 25)

Variable	Winter Rainfall Region		Summer Rainfall Region	
	$\bar{x}$	s	$\bar{x}$	s
MAR	23,306	36,854	11,625	12,306
MAF	30,493	59,229	14,464	22,246
MA10	58,029	92,514	51,982	96,978
A	29,114	17,583	45,026	31,628
S85	86,795	60,485	43,412	40,187
P	2,248	0,805	2,073	0,663
D	26,032	10,766	29,778	13,779
W	2,856	1,088	3,426	1,251
LB	9,951	4,004	12,035	6,331
HM	1041,755	417,935	664,912	390,615
T	1,458	0,971	2,528	1,782
RH	102,377	55,371	52,966	36,191
RHP	43,199	21,374	25,966	18,705
RF	0,319	0,140	0,348	0,188
RC	0,565	0,174	0,602	0,126
RE	0,622	0,143	0,645	0,161
C	1,403	0,336	1,309	0,138
MAP	0,976	0,563	0,915	0,195
I2,33	60,217	22,711	73,720	13,183
I10	94,000	36,692	115,040	27,388

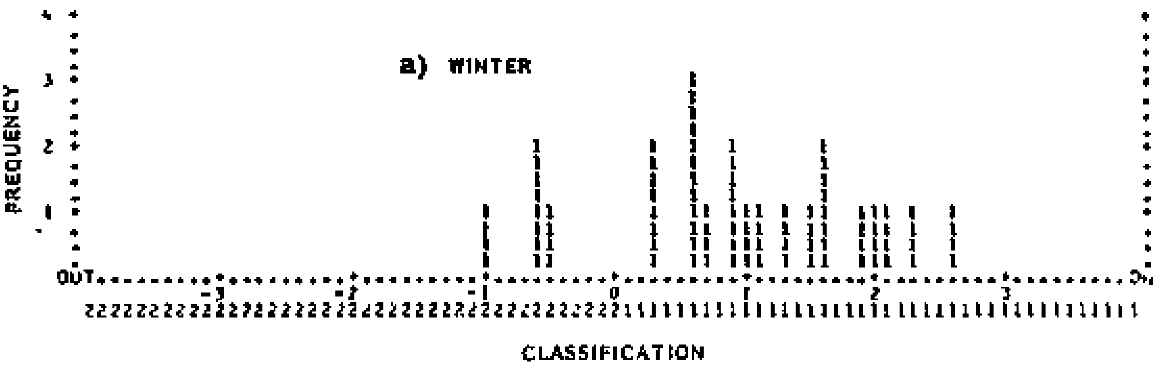
RESULTS AND DISCUSSION

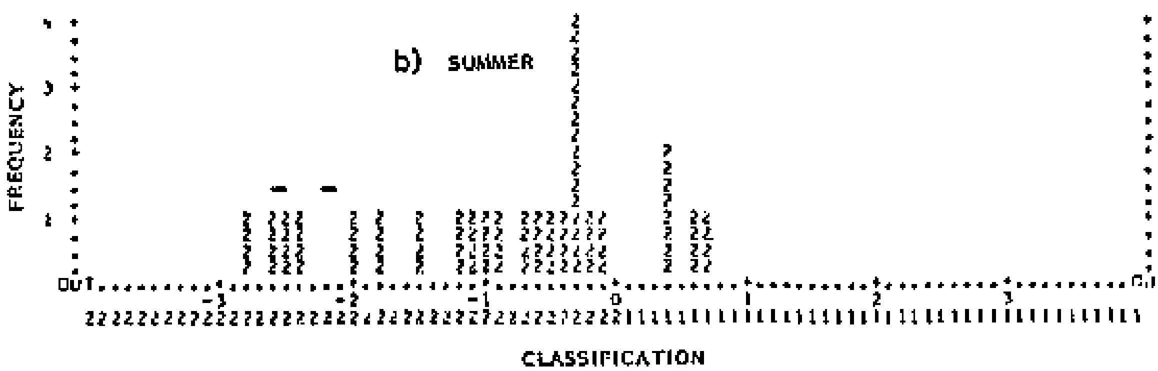
In all analyses where linear data were used the results were markedly poorer than the results obtained using logarithmic data. This observation was in keeping with the findings reported by Seyhan and Keet (1981) and Seyhan and Keet (1982). Only the results of analysis conducted using logarithmic data are presented.

Q-mode linear discriminant analysis

Linear discriminant analysis was used to determine whether the catchments of the winter rainfall region and summer rainfall region could be considered as statistically different in terms of (i) physiography and (ii) hydromorphometry. Fourteen physiographical variables (Table 2) were used for the first analysis and the results of the discriminant analysis are given in Figure 3.

The histograms presented in Figure 3 illustrate the number of catchments classified into either group 1 (winter rainfall region) or group 2 (summer rainfall region) according to the discriminant function developed for catchments from these two regions using 14 physiographical variables (Table 2). The histogram presented in Figure 3a indicates that the discriminant function for the winter rainfall region predicted correctly the group membership of 82,6% of the catchments in this region. The proportion of catchments from the summer rainfall region which were classified correctly was 89,0% (Figure 3b). The minimal overlap in the histograms of Figure 3 reveals that the catchments in the two regions are, on the basis of physiography, two distinct groups.





RAINFALL REGION	NO. OF CATCHMENTS	PREDICTED GROUP MEMBERSHIP	
		WINTER	SUMMER
WINTER	23	19(82,6%)	4(17,4%)
SUMMER	25	4(16,0%)	21(89,0%)

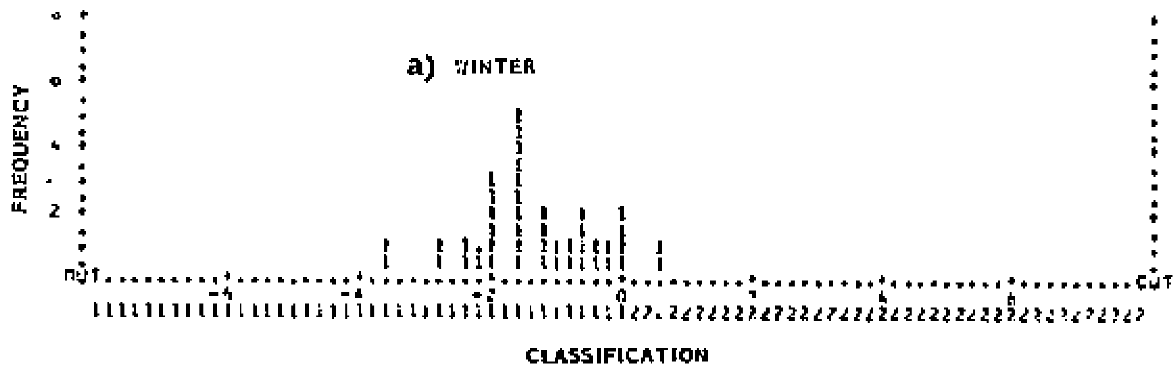
PERCENT CLASSIFIED CORRECTLY = 85,8

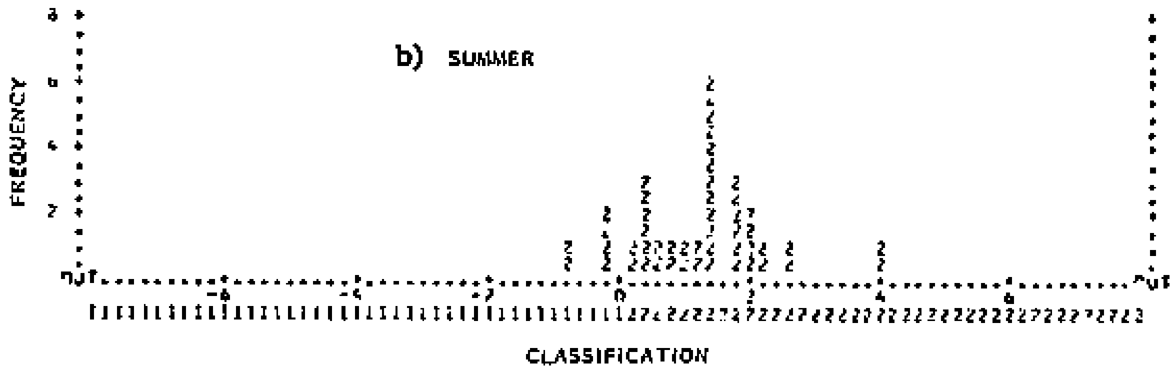
Figure 3 : Classification histogram based on discriminant functions for the a) winter and b) summer rainfall regions using physiographical data

The inclusion of rainfall and runoff variables (Table 2) in the discriminant analysis also resulted in a distinct hydromorphometrical classification of the catchments into the winter and summer rainfall regions (Figure 4). Eighty seven percent of the catchments of the winter rainfall region and 88% of the catchments in the summer rainfall region were classified correctly by the discriminant functions for these two region. The results from both these analyses thus substantiated the hypothesis that the underlying physiographical and hydromorphometrical dimensions of catchments in the two regions were different and that they should be considered independently in developing runoff equations.

R-mode principal components factor analysis

The R-mode principal components factor analysis (Seyhan, 1981a) was adopted to analyse the interrelationships between the selected independent and dependent variables. Analyses were conducted using linear data and then repeated using data with a logarithmic transformation. For each of these raw data matrices a symmetrical correlation matrix of Pearson's product-moment correlation coefficients was developed. Factor identification was based on this correlation matrix, these factors explaining the variance of the input data. The degree of association between an individual variable and other variables contained in a factor is given by the factor loadings with the sign indicating a direct (+) or inverse (-) relationship. These factor loadings may be taken as having a meaning similar to that of a correlation coefficient in a correlation matrix. According to Seyhan (1981a) it is difficult to interpret a factor matrix which contains all the factor loadings. Thus, an arbitrary factor loading is selected and only variables having loadings greater than or less than this value are considered. The selection of this critical factor loading is, to a large extent, based on the





RAINFALL REGION	NO. OF CATCHMENTS	PREDICTED GROUP MEMBERSHIP	
		WINTER	SUMMER
WINTER	23	20(87,0%)	3(13,0%)
SUMMER	25	3(12,0%)	22(88,0%)

PERCENT CLASSIFIED CORRECTLY = 87,5

Figure 4 : Classification histogram based on discriminant functions for the a) winter and b) summer rainfall regions using hydromorphometrical data

experience of the researcher. However, in most analyses, there is a sharp transition between variables having high factor loadings and those having low loadings which assists in the selection of the threshold value.

The results of each analysis includes the eigenvalue (explained variance) for each factor as well as the cumulative percentage of total explained variance. The communality, h_i^2, gives the percentage of variance of the ith variable in the matrix of m factors.

The simplified rotated matrix for the physiographical data of catchments in the winter rainfall region is given in Table 4. This analysis reveals that there are three major physiographical dimensions which may be identified for catchments of this region. These dimensions are catchment shape (RD, RC, RE and C), length (A, P, W and LB) and relief (S85, HM, RH and RHP). Only factors with eigenvalues greater than one were included in Table 4, an arbitrary decision recommended by Seyhan (1981a). This practice was adopted to simplify the results since, theoretically, the total number of factors is equal to the number of variables.

Although drainage density (D) was not reflected as a factor in Table 4, there was evidence to suggest that D also constituted an independent factor. The three factors given in Table 4 explain a mere 23,2% of the variance of D (communality) with the rest of the variance being explained by the remaining factors.

By including the rainfall and runoff variables in the factor analysis for the winter rainfall region, the same independent physiographical dimensions are identified as in the preceding analysis viz. shape, length, relief and drainage density (Table 5). Two additional factors are included in Table 5, these being the dimensions of rainfall (MAP, I2,33 and I10) and runoff (MAR, MAF and MA10). Together, these five factors explain

Table 4 : Simplified rotated factor matrix of physiographical data from catchments in the winter rainfall region ($N = 23$, factor loadings $\geq 0,70$ and $\leq -0,70$ retained)

Variables	Factors I	Factors II	Factors III	h_i^2(%)
A		0,982		96,83
S85			0,748	80,43
D				23,27
P		0,836		91,79
W		0,817		96,55
LB		0,775		95,34
HM			0,832	90,80
T				59,90
RH			0,753	72,87
RHP			0,682	89,45
RF	-0,971			95,22
RC	-0,875			82,66
RE	-0,971			95,22
C	0,715			54,58
Eigenvalue	5,30	3,57	2,28	
Explained Variance (Cumulative %)	37,85	63,35	79,65	

Table 5 : Simplified rotated factor matrix of hydromorphometrical data from catchments in the winter rainfall region (N = 23, factor loadings ≥ 0,65 and ≤ -0,65 retained)

Variables	Factors I	II	III	IV	V	h_i^2(%)
MAR					0,656	72,83
MAF					0,854	95,97
MA10					0,838	96,00
A				0,949		99,35
S85			0,920			89,11
D·'					0,819	72,86
P				0,821		98,56
W				0,798		99,11
LB				0,733		97,67
HM			-0,771			95,00
T						93,80
RH			-0,771			80,58
RHP			-0,949			91,17
RF	-0,976					96,60
RC	-0,890					89,99
RE	-0,976					96,62
C	0,888					89,90
MAP		0,779				69,53
I2,33		0,914				95,95
I10		0,907				89,37
Eigenvalue	7,00	4,59	3,74	1,71	1,06	
Explained Variance (Cumulative %)	35,01	57,98	76,66	85,19	90,50	

most of the variance contained in the correlation matrix (90,5%). Since drainage density is included in factor V (Table 5), it may be concluded that this is the most important independent variable associated with runoff in these catchments.

The factor analysis of physiographical data from catchments in the summer rainfall region isolates the same three principal dimensions as reported for the winter rainfall region (Table 6). However, in comparing the results of Table 4 with those of Table 6, two major differences are notable. First, drainage density (D) is correlated directly with the relief factor (II) in the summer rainfall region and secondly the order of the three factors for this region differs to the order in the winter rainfall region. This latter finding indicates that the relative importance of these factors in explaining the variance in the correlation matrix differs in these two regions. This is also evident in comparing the cumulative percentages of explained variance for each factor in the two regions. Furthermore, the three factors identified explain more variance in the physiographical data of the summer rainfall region (90,72%) than they do for the corresponding data of the winter rainfall region (79,64%).

Factor analysis of the hydromorphometrical variables (physiographical, rainfall and runoff) from catchments in the summer rainfall region (Table 7) resulted in a different classification to that for catchments in the winter rainfall region (Table 5). Factor I in Table 7 reflects the intercorrelation of runoff (MAR, MAF and MA10), catchment size (A, LB and P), shape (REF, RC, RE and C) and the topographical factor (T). Factor II defines the association of catchment relief (S85, HM, RH and RHP), length (W) and drainage density (D). The independent dimensions of rainfall intensity (I2,33 and I10) and mean annual precipitation (MAP) are revealed in factors III and IV respectively. The four factors identified for the hydromorpho-

Table 6 : Simplified rotated factor matrix of physiographical data from catchments in the summer rainfall region (N = 25, factor loadings ≥ 0,70 and ≤ -0,70 retained)

Variables	Factors			h_i^2(%)
	I	II	III	
A	0,899			98,10
S85		0,890		84,69
D		0,763		70,71
P	0,870			99,85
W	0,951			99,60
LB	0,711			96,77
HM		0,787		93,35
T				95,22
RH		0,868		88,52
RHP		0,876		93,37
RF			0,972	96,20
RC			0,752	78,82
RE			0,971	96,20
C			-0,750	78,75
Eigenvalue	6,90	4,48	1,32	
Explained Variance (Cumulative %)	49,27	81,26	90,72	

Table 7 : Simplified factor matrix of hydromorphometrical data from catchments in the summer rainfall region (N = 25, factor loadings $\geq$ 0,63 and $\leq$ −0,63 retained)

Variables	Factors				h_i^2(%)
	I	II	III	IV	
MAR	0,706				89,01
MAF	0,636				79,76
MA10	0,638				80,26
A	0,832				91,35
S85		0,846			76,92
D		0,766			71,86
P	0,833				93,65
W		−0,689			70,01
LB	0,954				96,50
HM		0,741			93,56
T	0,782				94,17
RH		0,841			85,28
RHP		0,958			92,87
RF	−0,886				86,94
RC	−0,910				82,91
RE	−0,882				87,15
C	0,909				82,81
MAP				0,770	68,61
I2,33			0,943		92,55
I10			0,888		86,38
Eigenvalue	8,85	5,15	1,66	1,36	
Explained Variance (Cumulative %)	44,25	69,95	78,26	85,08	

metrical data of catchments in the summer rainfall region are, however, not as clearly definable as those described for the winter rainfall region.

Having identified the major physiographical and hydromorphometrical dimensions of the catchments in the two selected regions, attention is now turned to developing runoff equations for these two regions.

Regression analysis

The results of the Q-mode analysis indicated that the catchments of the winter rainfall region could be considered to be physiographically and hydromorphometrically distinct from the catchments of the summer rainfall region. Furthermore, the major hydromorphometrical factors established using factor analysis were also found to be different for these two regions. In view of these findings separate regression analyses were conducted for the winter and summer rainfall regions.

A fundamental requirement of regression analysis is that the independent variables should be completely independent. This constraint was satisfied by referring to the factors identified in the R-mode analysis and selecting a single independent variable from each factor. Using correlation analysis, multiple linear regression and step-wise multiple linear regression, a suite of equations was developed to calculate the mean annual runoff (MAR), mean annual flood (MAF) and 10 year return period flood (MA10) for catchments in the two regions. The standard error of estimate (S_y) and multiple correlation coefficient (r) for each equation was determined. The significance of the regression coefficients ($\beta_{1-\gamma}$) was determined using Student's t-test while the significance of the multiple correlation coefficient was calculated using an F-test (Snedecor and Cochran, 1972).

The significant runoff equations for the winter rainfall region are given in Table 8. The most significant equations for this region are for the mean annual flood (MAF) and 10-year return period flood (MA10). These peak discharges are principally a function of drainage density, catchment shape and rainfall intensity (Table 8). The multiple correlation coefficients of these peak discharge equations are all in excess of 99% significance and the significance of the regression coefficients is greater than 95%. However, the standard error of estimate values range from 34,1% to 54,5%.

The selected regression equations for the summer rainfall region are notably different to those of the winter rainfall region (Table 9). While the equations for calculating mean annual runoff are less significant than equations for calculating mean annual runoff are less significant than equations 6 and 7 in Table 9 for calculating peak discharges, the peak discharge equations have markedly higher standard errors of estimate than mean annual runoff equations 2 and 3 in Table 9. In this region MAR is expressed as a function of catchment shape, area, relief and mean annual precipitation and the more significant peak discharge equations express the dependent variable as a function of catchment area alone.

In developing equations to calculate runoff from catchment physiographical and rainfall variables, the results of this study have revealed differences in the variables controlling the runoff of catchments in the winter and summer rainfall regions of South Africa. This finding may be attributable to a number of factors. Firstly, the two regions have distinctly different climatic and geological characteristics (Table 1). Secondly, the catchments selected for the winter rainfall region were spread over a considerably smaller area than those selected for the summer rainfall region. Thus, greater geologica and climatological homogeneity existed for catchments of the winter rainfall region. Finally, the data obtained for catchments in the summer rainfall region were less reliable than those of the winter rainfall region, particularly for peak discharges. This latter point may account for the poor equations

Table 8 : Regression equations and associated statistics for catchments of the winter rainfall region (N = 23)

Regression Equation	r	Significance (%) Sy	β	r
$MAR = (5,395 \times 10^{-4})(A)^{1,916}(S85)^{0,788}$	0,633	0,687	95,00	99,00
$MAF = (1,049 \times 10^{-4})(D)^{1,799}(12,33)^{2,500}$	0,787	0,474	99,50	99,50
$MAF = (1,260 \times 10^{-4})(D)^{1,791}(RE)^{-2,003}(12,33)^{2,781}$	0,839	0,429	97,50	99,50
$MAF = 2,358\ (D)^{2,150}(MAP)^{0,976}$	0,705	0,545	95,00	99,50
$MA10 = (25,745 \times 10^{-4})(D)^{1,546}(I10)^{1,808}$	0,796	0,374	97,50	99,50
$MA10 = (6,440 \times 10^{-4})(D)^{1,555}(RE)^{-1,527}(I10)^{1,945}$	0,843	0,341	97,50	99,50
$MA10 = 1,104\ (D)^{1,817}(A)^{0,579}$	0,718	0,430	95,00	99,50

r = multiple correlation coefficient; Sy = standard error of estimate;

β = regression coefficient

Table 9 : Regression equations and associated statistics for catchments of the summer rainfall region (N = 25)

Regression Equation	r	Significance (%)		
		Sy	β	r
MAR = $0,112\ (RE)^{-3,801}(MAP)^{1,617}$	0,716	0,420	97,50	99,50
MAR = $0,269\ (RE)^{-2,651}(MAP)^{1,668}(A)^{0,561}$	0,782	0,383	97,50	99,50
MAR = $0,024\ (RE)^{-1,998}(MAP)^{1,279}(A)^{0,845}(S85)^{0,502}$	0,831	0,351	96,50	99,50
MAF = $0,091\ (A)^{1,194}(MAP)^{1.137}$	0,707	0,489	90,00	99,50
MAF = $0,093\ (LB)^{1,779}(MAP)^{1,207}$	0,714	0,484	90,00	99,50
MAF = $0,093\ (A)^{1,371}$	0,681	0,496	99,95	99,50
MA10 = $0,120\ (A)^{1,371}$	0,692	0,563	99,95	99,50
MA10 = $0,118\ (A)^{1,395}(MAP)^{1,40}$	0,713	0,560	85,00	99,50
MA10 = $0,127\ (LB)^{2,060}(MAP)^{1,218}$	0,713	0,559	85,00	99,50

r = multiple correlation coefficient; Sy = standard error of estimate;

β = regression coefficient

which were developed to calculate peak discharges in the summer rainfall region.

In both regions equations which did not include rainfall variables were not notably less significant than those equations which included these variables (Table 8 and Table 9). This finding suggests that physiographical variables are mor important than rainfall variables in explaining the variance in runoff variables. The factor analysis also indicated that physiographical variables accounted for most of the variance in the correlation matrix.

CONCLUSIONS

The results of the three related analyses have been presented and discussed, namely, Q-mode analysis, R-mode analysis and regression analysis. From these analyses a number of conclusions may be drawn and summarised as follows :

(i) The catchments of the winter and summer rainfall regions are physiographically and hydromorphometrically distinct.

(ii) In both regions the major physiographical dimensions are catchment shape, relief, length and drainage density.

(iii) Physiographical variables appear to be more important than rainfall variables in explaining the variance in the runoff variables. This observation is true for both the winter and summer rainfall regions and may be taken to indicate that the physiography of the catchments reflects, amongst other things, the climate which acted on them.

(iv) The equations developed to calculate peak discharges (MAF and MA10) in the winter rainfall region are more significant than the equations developed to calculate the mean annual discharge (MAR). The reverse is true in the summer rainfall region.

While this study was conducted using a limited number of catchments and in some cases, poor quality runoff data, the results obtained are considered to be highly satisfactory. The analyses have shown that multivariate techniques may be used beneficially to regionalise catchments and to assist in developing runoff equations. Finally, it may be concluded that the development of equations using physiographical and rainfall data could be a viable undertaking for other regions in South Africa.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to Professor P. Meiring, Head, Department of Agricultural Engineering, University of Natal for placing the facilities of the Department at their disposal. The assistance given by the staff of this Department is also gratefully acknowledged.

This research was funded by the Water Research Commission, Republic of South Africa.

REFERENCES

Harman, H.H. (1968) Modern factor analysis. The University of Chicago press, Chicago, 486 pp.

Hope, A.S. (1979) Calculating catchment runoff : the importance of catchment physiography. In : Hope, A.S. and Mulder, G.J. Hydrological investigations of small catchments in the Natal Coastal Belt and the role of physiography and land use in the rainfall-runoff process. Final report series B, 6, Hydrol. Res. Unit, Univ. Zululand, 110-167.

Horton, R.E. (1945) Erosional development of streams and their drainage basins. Bull. Geol. Soc. Amer., 56, 275-370.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. J. Educational Psychol. 24, 417-441.

Kaiser, H.F. (1958) The varimax criteria for analystic rotation in factor analysis. Psychometrika 23, 187-200.

Maxwell, J.C. (1960) Quantitative geomorphology of the San Dimas experimental forest, California. Office of Naval Res., Geog. Branch, Washington D.C., Tech. Rep., 19.

Miller, V.C. (1953) A quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain area, Virginia and Tennessee. Dept. Geol., Columbia Univ., New York, Tech. Rep. 3.

Morisowa, M.E. (1958) Measurement of drainage-basin outline form. Water Resour. Res. 66, 587-591.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H. (1975) SPSS (Statistical package for the social sciences). McGraw-Hill, New York, 661 pp.

Schulze, R.E. (1982) Agrohydrology and climatology of Natal. Water Res. Commis., Pretoria (in press).

Schumm, S.A. (1956) Evolution of drainage systems and slopes in badlands in Perth Amboy, New Jersey. Bull. Geophys. Soc. Amer., 67, 597-646.

Seyhan, E. (1976) Calculation of runoff from basin physiography (CRBP). Geographical Studies 2, Instit. Geog., State Univ. Utrecht, Utrecht, 176 pp.

Seyhan, E. (1977) Mathematical simulation of watershed hydrologic processes. Instit. Geog., State Univ. Utrecht, Utrecht, 179 pp.

Seyhan, E. (1981a) Introduction to multivariate statistical analysis in hydrology. Instit. Earth Sci., Free Univ., Amsterdam, 204 pp.

Seyhan, E. (1981b) A universal application of the multivariate statistical techniques to hydrology. Proceedings of the International Symposium on Rainfall-Runoff Modelling. Mississippi State Univ., Mississippi (in press).

Seyhan, E. and Keet, B. (1981) Multivariate statistical analysis (Part I): Application to hydromorphometrical data (Case Study : Ahr river basin, Bolzana, Italy). Aqua-vu 8, Instit. Earth Sci., Free Univ., Amsterdam, 114 pp.

Seyhan, E. and Keet, B. (1982) Comparative case studies with multivariate statistical techniques. Geologie en Mijnbouw 61(3), (in press).

Snedecor, G.W. and Cochran, W.G. (1972) Statistical Methods. The Iowa State Univ. Press, Ames, 593 pp.

Strahler, A.N. (1957) Quantitative analysis of watershed geomorphology. Trans. Amer. Geophys. Union, 38, 913-920.

II

ESTIMATION OF SURFACE SOIL MOISTURE IN THE SIAYA CATCHMENT

ESTIMATION OF SURFACE SOIL MOISTURE CONTENT IN THE SIAYA CATCHMENT

A.S. Hope, E. Seyhan and G.J. Mulder

ABSTRACT

Regression equations are developed to calculate surface soil moisture of bare lands in the Siaya catchment which is located on the coastal belt of Zululand. Soil moisture was measured at depths of 25 mm and 75 mm along with 15 soil and terrain variables at each site. Six representative sites were selected and five samples of each depth were collected. At the time of the experiment colour and infra-red aerial photographs of the catchment were taken with the intention of relating, at a later date, digitised images to variations in measured and calculated soil moisture. Factor analysis, correlation analysis and regression analyses were used to develop the equations. These equations were found to be unsatisfactory predictors of soil moisture content, particularly at 75 mm. The poor regression results were attributed to the extremely dry soil moisture conditions on the day of the study.

BACKGROUND

Soil moisture has long been ascribed a prominent role in the hydrological cycle (Henninger, Petersen and Engman, 1976). The major dynamic storage component of a catchment is the soil matrix and generally soil moisture variations need to be considered in both time and space for hydrological modelling. Temporal variations in soil moisture at different depths in the soil horizon are controlled by different mechanisms. Tischendorf (1969) concluded that the deeper soil moisture changes were related to season while researchers such as Carlson, Reinhart and Horton (1956) found soil moisture in the surface 300 mm to depend primarily on the sequence of rainfall amounts, findings which have been substantiated in South Africa by Hope and Schulze (1979). The depth to which daily fluctuations in evapotranspiration affect soil moisture is determined by the rooting depth of the vegetation (e.g. Grindley, 1967 and Jones, 1976). Research findings, particularly in the United States of America, generally attribute spatial differences in soil moisture to one or more of the following : climate, soils, vegetation and topographical position (e.g. Wild and Scholz, 1930; Platt, 1955; Kovner, 1955; Stoeckeler and Curtis, 1960; Whipkey, 1965; Tischendorf, 1969 and Helvey, 1971).

The measurement of catchment soil moisture is time consuming and at best point readings are made at selected sites. In view of the highly variable nature of soil moisture in a catchment estimates of total soil moisture content are usually crude even when numerous point samples are taken (Blyth, 1981). Over the past decade much attention has been given to remote sensing techniques for the estimation of soil moisture (e.g. Schmugge, Meneely, Rango and Neff, 1977; Schmugge, Blanchard, Anderson and Wang, 1978; Price, 1980; Newton, 1981 and Heilman and Moore, 1981). Reflectances in the green-yellow band ($0,5\mu - 0,6\mu$) and near infra-red band ($0,8\mu - 0,9\mu$) have been found to be correlated with surface soil moisture conditions of bare land (Morris, Blyth and Clarke, 1980; Hardy, 1980 and Marcolongo, 1980).

An aerial survey of the Siaya catchment is made annually in order to monitor land-use changes in the catchment. This survey includes both colour and infra-red photography and attempts are to be made at relating digitised multispectral images to the surface soil moisture of bare lands in this area. Thus, surface soil moisture measurements at representative sites in the catchments are required at the time of the photography. This study was designed to provide information on the surface soil moisture of bare soil in the Siaya catchment on August 30, 1982 when aerial photographs of the catchments were being taken.

AIMS AND DATA

In order to relate remotely sensed data to surface soil moisture conditions it is necessary to sample soil moisture at sites which represent the variety of conditions found in the study area. Data from these control sites may then be used to establish a relationship between the appropriate spectral signature (combination of wavelengths) and surface soil moisture. Should such a relationship be found then the soil moisture of other bare fields in the catchment could be classified from the remotely sensed data.

The principal aim of this investigation was to develop regression equations from readily measurable soil and terrain variables which could be used at a later date to calculate surface soil moisture conditions which prevailed at the time of the flight. Estimates of surface soil moisture could then be made in areas of the catchment which were not included in the original sampling and assist in the interpretation of the aerial photographs.

An initial consideration in this study was to select six sites which were representative of the physical characteristics of the Siaya catchment and then to determine the surface soil moisture of each site at the time of the aerial data collection (between 11h00 and 13h00).

The Siaya catchment is located on the northern coastal belt of Zululand and covers an area of $15,3^{2}$ (Figure 1). Altitudes range from 90 m in the west to sea level at the outlet of the catchment. Mean annual rainfall for the area is approximately 1350 mm and most of the catchment is

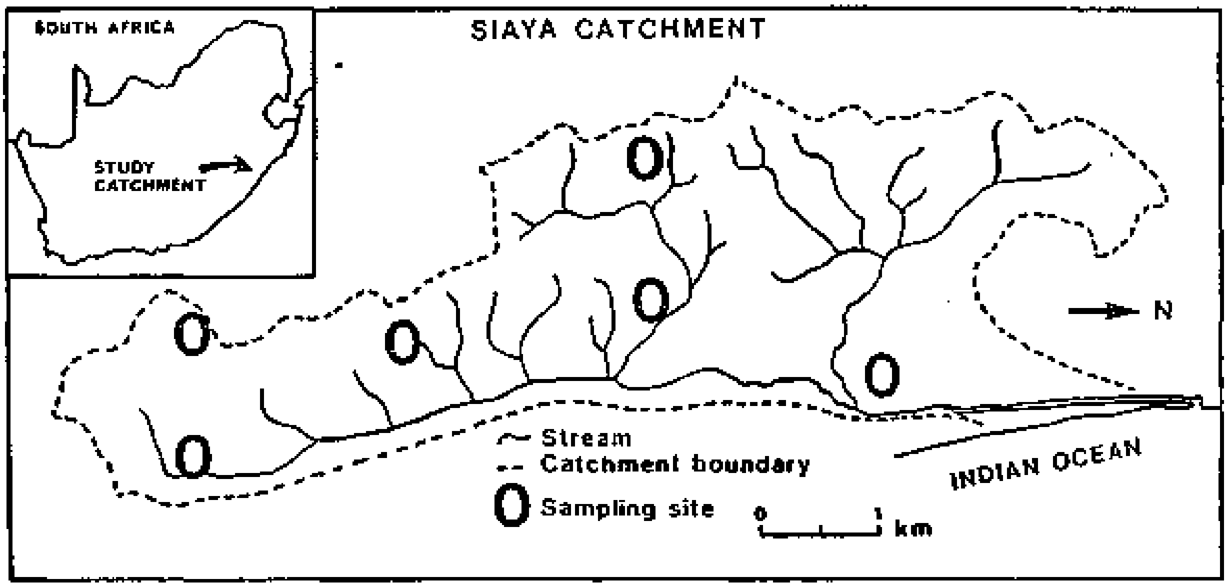


Figure 1 : The Siaya catchment and location of sampling sites

covered by sugar cane. The selection of sites for this study was restricted by the location of fallow fields, however, an attempt was made to select sites which were representative of the soils, slopes and aspects found in the catchment. Furthermore, attention was given to providing a good spatial cover within the catchment which would include sites at various altitudes and distances from the sea. The locations of the selected sites are given in Figure 1.

At each site five sampling points were selected and samples of the soil at depths of 25 mm and 75 mm were taken. The five individual sampling points were located at least 75 m apart so as to represent the local features of each site such as topographical position, soil and altitude. At least 100 g of soil was collected for each sample and place in an air-tight bag and soil moisture (percent by mass) was determined by gravimatric analysis in a laboratory. Besides expressing soil moisture as a percent by mass, the variable was also converted to a relative value as is defined in the following expression :

$$RSM = \frac{SM_{ij} - SM_L}{SM_L} \qquad \text{..........} \quad (1)$$

where

RSM = relative soil moisture (-),

SM_{ij} = soil moisture at the i'th sample point at site j (%) and

SM_L = the lowest soil moisture value for any sample point for the chosen depth (%).

The concept of relative soil moisture was introduced so that the relattionship between catchment variables, soil variables and the relative difference in soil moisture on the selected day could be evaluated. Independent variables were correlated with the variations in soil moisture rather than absolute soil moisture values.

Fifteen independent variables were measured at each sampling point for possible inclusion in the regression equations to calculate soil moisture at the two selected depths. Although not included in the regression analysis, soil temperature was measured at 25 mm (TEMP1) and 75 mm (TEMP2) to determine whether aspect and slope affected the temperatures at these depths.

Antecedent rainfall was not included as an independent variable in this study since no rain had been recorded for fifteen days prior to the day of the flight. Furthermore, Hope and Mulder (1979) have established that rainfall amounts in this region are correlated highly with altitude and distance from the sea, both these variables being recorded at each sample point. Since the six selected sites were all located in ploughed fields, soil texture and organic matter were found to be very similar. Textural analyses were thus conducted using the combined samples from the two depths. A summary of all the variables used in this study is given in Table 1 while the mean, standard deviation and coefficient of variation of each variable is presented in Table 2.

Table 1 : Selected independent and dependent variables

a) Independent Variables

Variable	Description	
CSAC	coarse sand (%)	TEXTURE
CSAM	medium sand (%)	
CSAF	fine sand (%)	
CSI	silt (%)	
CC	clay (%)	
CSSI	sand + clay (%)	
CSIC	silt + clay (%)	
CO	organic matter (%)	
SB	soil surface slope (deg.)	TERRAIN
H	height above nearest drainage channel (m)	
L	distance to nearest drainage channel (m)	
AL	altitude read from topographical map (m)	
ALB	altitude determined by barometer (m)	
DS	distance from sea (km)	
AZ	azimuth (°N)	

b) Dependent Variables

Variable	Description
SM1	soil moisture at 25 mm (%)
SM2	soil moisture at 75 mm (%)
RSM1	relative soil moisture at 25 mm (-)
RSM2	relative soil moisture at 75 mm (-)

Table 2 : Means ($\bar{x}$), standard deviations (s) and coefficients of variation (CV) of selected independent and dependent variables

Variable	$\bar{x}$	s	CV(%)
CSAC (%)	34,945	11,754	33,6
CSAM (%)	47,225	11,018	23,3
CSAF (%)	9,314	3,613	38,8
CSL (%)	3,271	3,218	98,4
CC (%)	5,304	4,271	80,5
CSSI (%)	94,755	4,319	4,6
CSIC (%)	8,575	7,139	83,3
CO (%)	2,010	1,633	81,4
SB (deg.)	5,068	2,312	45,6
H (m)	5,603	6,149	109,7
L (m)	109,883	62,866	59,9
AL (m)	45,057	16,243	36,0
ALB (m)	55,087	18,977	34,4
DS (km)	2,226	0,957	43,0
AZ (°N)	164,967	91,848	55,7
SM1 (%)	1,055	0,808	76,6
SM2 (%)	6,188	3,796	61,3
RSM1 (-)	1,681	1,981	117,8
RSM2 (-)	1,242	1,375	110,7

ANALYTICAL PROCEDURES

A fundamental principle of regression analysis is that the independent variables contained in a regression equation should not be intercorrelated. The application of principal components factor analysis to hydrological problems is described by Seyhan (1981) and this procedure was adopted to identify the underlying independent dimensions or factors which explain the variance in the 21 selected variables measured for this study (Table 1).

An examination of the correlation matrix used for the principal components factor analysis may reveal some association between variables included in different factors. This would also be seen if the factor matrix were presented without simplification. However, each factor would contain those variables which are least related to variables in other factors and are, for practical purposes, taken as independent of variables in these factors.

The degree of association between a single variable and the other variables in a factor is given bythe factor loading. This loading may be taken to have a similar meaning to that of a correlation coefficient in a correlatio matrix. Since it is difficult to interpret a factor matrix which contains all the factor loadings, it is necessary to retain only those variables having a loading greater or less than a set value (Seyhan, 1981). The selection of this threshold value is arbitrary and depends largely on the experience of the investigator. However, in many analyses there is a sharp transition between the group of high/low factor loadings and the group of high/low factor loadings and the group of values close to zero which assists in the selection of the critical value. For each factor the cumulative eigenvalue is given which represents the cumulative amount of variance explained by successive factors. The communality, h_i^2, gives the percentage of variance for the <u>i</u>th variable in the matrix of <u>m</u> factors.

The factor analysis in this study was based on a symmetrical correlation matrix (of Pearson's product-moment correlation coefficients) with variance orthogonal rotation criterion (Seyhan, 1981). Following the identification of independent factors, correlation, multiple linear regression and stepwise multiple linear regression analyses were used to develop the regression equations with only a single variable from each factor being permitted to enter each equation. All data were transformed logarithmically since, according to Seyhan (1981), the transformation generally results in a better approximation of the normal distribution, this distribution being assumed in all the statistical procedures adopted for this study.

A flow chart of the data collection procedure, analytical steps and intended remote sensing applications is given in Figure 2.

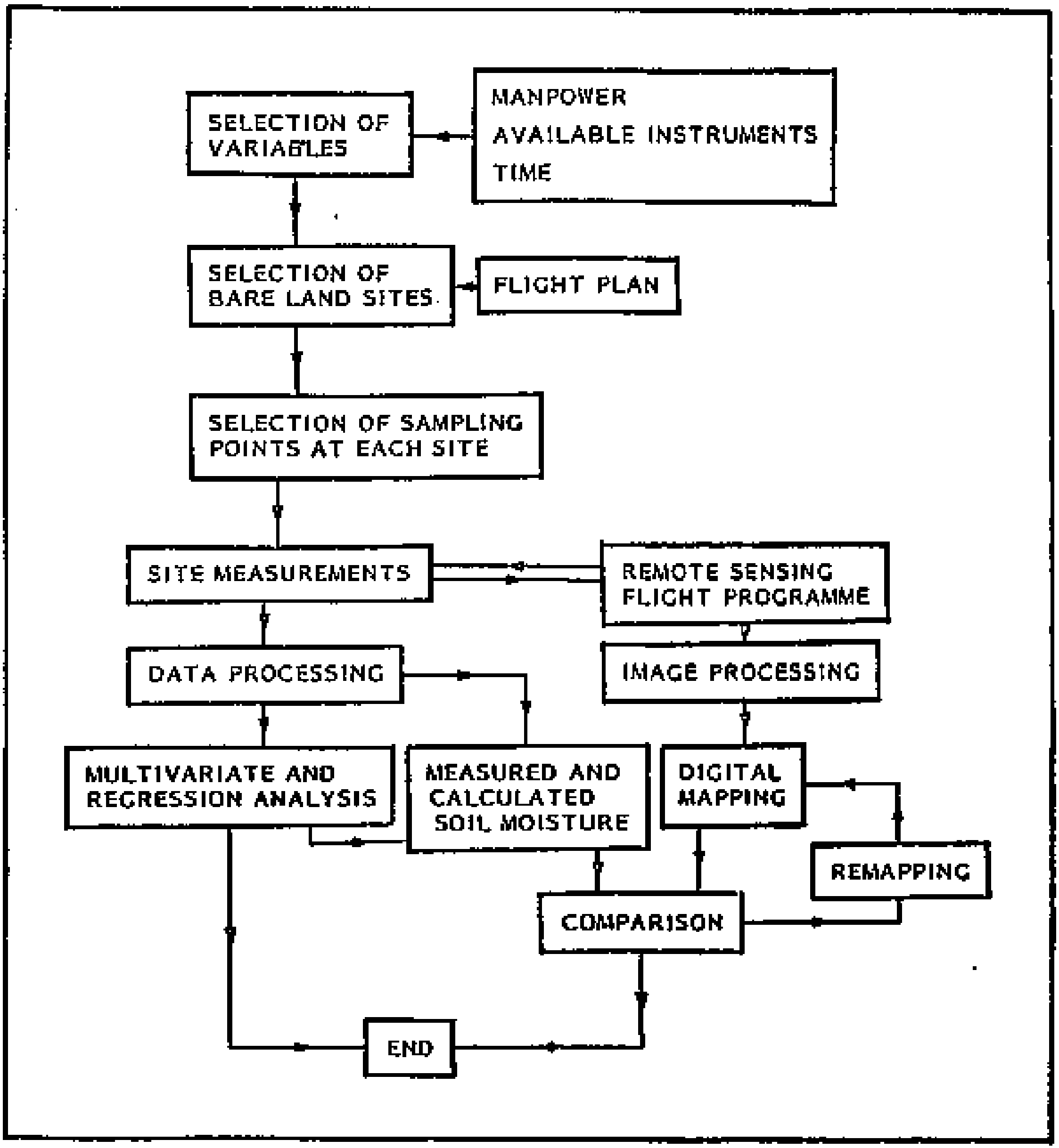


Figure 2 : Flow chart of data collection procedure, analytical steps and intended remote sensing application

The flow chart in Figure 2 gives an indication of the constraints under which such an investigation is invariable conducted, namely, availability of manpower, time and instruments for data collected at the required time.

RESULTS AND DISCUSSION

The surface soils of bare fields in the Siaya catchment were found to be exceptionally dry on the day of the study with the mean soil moisture content at depths of 25 mm and 75 mm being 1,055% and 6,188% respectively

(Table 2). The greater soil moisture content at 75 mm may have been expected since losses occur from above and rainfall had not occurred for 15 days. The coefficient of variation for soil moisture in the upper zone was 76,6% while that for the lower zone was 61,3%. These differences in variability suggest that localised factors, such as microtopography, have a more pronounced effect on soil moisture closer to the surface. Quantification of factors affecting soil moisture could thus be expected to be more difficult for the shallower depth. The two relative soil moisture indices, RSM1 and RSM2, also reflect the greater relative variability in soil moisture at 25 mm. However, the coefficients of variation for both indices are markedly higher than the associated values for soil moisture content (RSM1 : CV = 117,8%; RSM2 : CV = 110,7%).

The simplified rotated factor matrix of the 21 selected variables (Table 1) is given in Table 3. Six underlying dimensions or factors were identified and accounted for 84,8% of the variance in the correlation matrix. Factor I is clearly the most important dimension explaining 31,9% of the variance. This factor represents the strong relationship between soil texture and distance to the nearest drainage channel (L). The negative sign associated with L indicates that there is an inverse relationship between this variable and the silt content (CSI), clay content (CC) and organic matter content (CO) of the soils. Since these fine particles are generally transported downslope by the movement of water, a greater concentration in each of these fractions would be expected closer to the drainage channels.

Factor II is a dimension which may be broadly described as the altitude factor. The variables included in this factor are either direct measures of altitude (AL, ALB) or related to altitude (H, DS). The coarse, medium and fine sand textural components constitute an independent dimension,

Table 3 : Simplified rotated factor matrix of soil, terrain and soil moisture variables (factor loadings $\geq 0,65$ and $\leq -0,65$ retained)

Variables	Factors						h_i^2 (%)
	I	II	III	IV	V	VI	
CSAC			0,945				90,18
CSAM			0,873				84,54
CSAF			-0,930				94,33
CSI	0,650						48,03
CC	0,883						89,59
CSSI	-0,934						95,18
CSIC	0,937						93,60
CO	0,805						93,89
SB						-0,809	85,44
H		0,813					68,69
L	-0,808						87,94
AL		-0,967					93,98
ALB		-0,954					95,93
DS		-0,930					96,23
AZ							74,36
SM1				0,690			82,42
SM2				0,794			79,64
RSM1				0,806			79,31
RSM2				0,871			80,66
TEMP1					-0,635		82,00
TEMP2					-0,799		64,81
Eigenvalue	6,79	3,69	3,50	1,40	1,39	1,18	
Explained Variance (Cu. %)	31,90	49,48	66,15	72,82	79,18	84,80	

factor III. Factor IV is the independent dimension of soil moisture containing soil moisture content and relative soil moisture for the two depths. The high positive factor loadings in factor IV indicate that soil moisture at the two depths is highly related. Soil temperatures, TEMP1 and TEMP2, were included in a single factor (V) and did not include azimuth (AZ) or slope SB as was expected. Finally, the soil surface slope (SB) was found to be an independent factor (VI) which accounted for 10,61% of the observed variance in the correlation matrix.

Following the correlation, multiple linear regression and stepwise multiple linear regression analyses two equations were developed for calculating soil moisture content at a depth of 25 mm (SM1), viz.

$$SM1 = 0,534\,(CO)^{0,886};\quad r^2 = 0,563 \qquad \ldots\ldots\ldots\ldots \quad (2)$$

and

$$SM1 = 0,038\,(CO)^{0,834}(CSAF)^{0,759};\quad r^2 = 0,631 \qquad \ldots\ldots\ldots\ldots \quad (3)$$

These equations indicate that organic content (CO) has the greatest association with moisture content in the upper zone accounting for 56,3% of the observed variance in SM1 (Equation 2). By including the fine sand content (CSAF) in the equation the explained variance increases by less than seven percent to 63,1 percent. Since the standard error of estimate for both equations was greater than 30 percent and the explained variances were low, the equations were not regarded as satisfactory models for calculating SM1.

A single equation was developed for soil moisture content at the depth of 75 mm (SM2) and took the following form :

$$SM2 = 0,232\,(CO)^{0,497};\quad r^2 = 0,430 \qquad \ldots\ldots\ldots\ldots \quad (4)$$

The standard error of estimate for this equation also exceeded 30 percent and a mere 43 percent of the observed variance in SM2 was explained by the model (Equation 4). These statistics were not improved by including additional independent variables and Equation 4 was also considered unsatisfactory for calculating soil moisture content.

The importance of organic matter (CO) in retaining moisture in the soil under dry conditions was illustrated in this regression analysis, being important at both depths. Generally, adsorptive forces are important in retaining soil moisture under dry conditions and despite the greater specific surface associated with silt and clay, fine sand was the only textural component included in the regression equations. Contrary to expectations, the equations developed for SM1 were better than the equation for SM2.

Attempts to develop regression equations for relative soil moisture at the two depths (RSM1, RSM2) were not successful. The best equation was for RSM1 and was of the form :

$$RSM1 = (0,13 \times 10^{-4})(CO)^{3,221}(ALB)^{2,116};\ r^2 = 0,429 \quad \ldots\ldots\ldots \quad (5)$$

Besides the low coefficient of determination ($r^2 = 0,429$) the standard error exceeded 30 percent. It may thus be concluded that the relative soil moisture as defined by these indices had a limited relationship with the independent variables measured. However, organic matter (CO) was also a significant independent variable in this equation at the 95% level.

CONCLUSIONS

Soil moisture contents determined for the 30 sample points on bare land and at two depths in the Siaya catchment revealed that the soil moisture conditions for the selected day were particularly dry. The major findings of this investigation may be summarised as follows :

(i) Soil moisture contents at the depth of 75 mm were markedly higher than those at the depth of 25 mm and exhibited less variability.

(ii) The 21 variables measured in this study were reduced, successfully, to six independent dimensions using principal components factor analysis.

(iii) Organic matter was found to be the most important independent variable affecting soil moisture contents at both depths investigated.

(iv) Regression equations for calculating surface soil moisture were not satisfactory, particularly for estimates at a depth of 75 mm. These equations were characterised by poor coefficients of determination (r^2) and high values of the standard error of estimate.

(v) Expressing soil moisture as a quantity relative to the lowest recorded value resulted in weaker regression equations and could not be related successfully to the selected independent variables.

The results and conclusions of this investigation were dominated by the dry soil moisture conditions on the day of the study. Different results may be found under wetter conditions where factors such as antecedent rainfall could be included in the analysis. Furthermore, independent variables which were found to be unimportant in this study could be significant under wetter conditions and may result in more accurate predictive equations. The results of the regression analysis for these dry conditions are inadequate for the proposed application to remote sensing. However, the measured values of the 30 sample points could be used as a training sample for digitised images. The procedures adopted in this investigation were considered to be satisfactory and suitable for repeated analyses of this nature.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to Professor P. Meiring, Head, Department of Agricultural Engineering, University of Natal, for placing the facilities of the Department at their disposal. The assistance given by the staff of this Department is also gratefully acknowledged. A particular word of thanks goes to Messrs. S. Hlatswayo, E. Ntuli and C. Depraetere for their assistance in collecting the data and to Miss R. Sands for the data processing. Thanks are also due to Professor D.A. Scogings of the Department of Land Surveying, University of Natal for his assistance in the planning of the project.

This research was funded by the Water Research Commission, Republic of South Africa.

REFERENCES

Blyth, K. (1981) Remote sensing in hydrology. Rep. No. 74, Institute of Hydrology, Wallingford, 169 pp.

Carlson, C.A., Reinhart, K.G. and Horton, J.S. (1956) Predicting moisture in the surface foot of soil. Soil Sci. Soc. Am. Proc., 20, 412-415.

Grindley, J. (1967) The estimation of moisture deficits. Met. Magazine, 96, 97-108.

Hardy, J.R. (1980) Survey of methods for the determination of soil moisture content by remote sensing methods. In : G. Fraysse (Ed.), Remote Sensing Application in Agriculture and Hydrology. Balkema, Rotterdam, 153-213.

Heilman, J.L. and Moore, D.G. (1981) Soil moisture applications of the Heat Capacity Mapping Mission. In : M. Deutsch, D.R. Wiesnet and A. Rango (Eds.), Satellite Hydrology, American Water Resour. Assoc., Minneapolis, Minnesota, 371-376.

Helvey, J.D. (1971) Predicting soil moisture in the Southern Appalachians. Unpubl. Ph.D. Dissertation, Univ. Georgia, Athens, Georgia, 79 pp.

Henninger, D.L., Petersen, G.W. and Engman, E.T. (1976) Surface soil moisture within a watershed - variations, factors influencing, and relationship to surface runoff. J. Soil Sci. Soc. Am., 40, 773-776.

Hope, A.S. (1979) Calculating catchment runoff : the importance of catchment physiography. In : A.S. Hope and G.J. Mulder: Hydrological investigations of small catchments in the Natal Coastal Belt and the rôle of physiography and land-use in the rainfall-runoff process. Rep. series B., No. 6, Hydrol. Res. Unit, Univ. Zululand, Kwa Dlangezwa, 110-167.

Hope, A.S. and Mulder, G.J. (1979) Hydrological investigations of small catchments in the Natal Coastal Belt and the role of physiography and land-use in the rainfall-runoff process. Rep. series B, No. 6, Hydrol. Res. Unit, Univ. Zululand, Kwa Dlangezwa, 282 pp.

Hope, A.S. and Schulze, R.E. (1979) Catchment soil moisture distribution at Cedara. In : R.E. Schulze (Ed.), Field Studies, Data Processing, Techniques and Models for Applied Hydrological Research. Agric. Catchm. Res. Unit (ACRU), Dep. Agric. Eng., Univ. Natal, Pietermaritzburg, ACRU Rep. 7(1), 275-297.

Jones, J.R. (1976) Physical data for catchment models. Nordic Hydrol., 7, 245-264.

Kovner, J.L. (1955) Changes in streamflow and vegetation characteristics of a Southern Appalachian mountain watershed brought about by forest cutting and subsequent regrowth. Unpub. Ph.D. Dissertation, State University of New York, Syracuse, New York.

Marcolongo, B. (1980) Photointerpretation applied to hydrological problems : soil moisture content, soil permeability, hydrogeological structures. In : G. Fraysse (Ed.), Remote Sensing Application in Agriculture and Hydrology. Balkema, Rotterdam, 329-338.

Morris, E.M., Blyth, K. and Clarke, R.T. (1980) Watershed and river channel characteristics and their use in a mathematical model to predict flood hydrographs. In : G. Fraysse (Ed.), Remote Sensing Application in Agriculture and Hydrology. Balkema, Rotterdam, 431–446.

Newton, R.W. (1981) Characteristics of microwave emission of significance to satellite remote sensing of soil water. In : M. Deutsch, D.R. Wiesnet and A. Rango (Eds.), Satellite Hydrology, American Water Resour. Assoc., Minneapolis, Minnesota, 353-362.

Platt, R.B. (1955) An ecological study of the mid-Appalachian shale barrens and of the plants endemic to them. Ecological Monograms, 21, 268-300.

Price, J.C. (1980) The potential of remotely sensed thermal infrared data to infer surface soil moisture and evaporation. Water Resour. Res., 16, 787-795.

Schmugge, T.J., Blanchard, B., Anderson, A. and Wang, J. (1978) Soil moisture sensing with aircraft observations of the diurnal range of surface temperature. Water Resour. Bull., 14, 169-178.

Schmugge, T.J., Meneely, J.M., Rango, A. and Neff, R. (1977) Satellite microwave observations of soil moisture variations. Water Resour. Bull., 13, 265-281.

Seyhan, E. (1981) Introduction to multivariate statistical analysis in hydrology. Institute of Earth Sciences, Free University, Amsterdam, 204 pp.

Stoeckeler, J.H. and Curtis, W.R. (1960) Soil moisture regime in southwestern Wisconsin as affected by aspect and forest type. J. Forestry, 58, 892-896.

Tischendorf, W.G. (1969) Tracing stormflow to varying source areas in a small forested watershed in the southern Piedmont. Unpub. Ph.D. Dissertation, Univ. Georgia, Athens, Georgia, 113 pp.

Whipkey, R.Z. (1965) Subsurface stormflow from forested slopes. Bull. Int. Assoc. Sci. Hydrol., 2, 74-85.

Wild, S.A. and Scholz, H.F. (1930) The basis for subdividing mountainous forests for the purpose of management. J. Forestry, 28, 1104-1109.

III

ESTIMATION OF STREAMFLOW LOSS BY EVAPOTRANSPIRATION FROM THE RIPARIAN ZONE

ESTIMATION OF STREAMFLOW LOSS BY EVAPOTRANSPIRATION FROM THE RIPARIAN ZONE

E. Seyhan, A.S. Hope and R.E. Schulze

ABSTRACT

Recession hydrographs from a small research catchment ($0,67 \; km^2$) in Zululand exhibit diurnal fluctuations in streamflow which are superimposed on the general recession trend. The daily additional reduction in flow is assumed to be the result of greater evapotranspirational losses in the riparian zone during the day time. This reduction in flow is calculated for 76 days and an attempt is made to relate these values to hydrometeorological variables which could be expected to correlate with evapotranspirational losses. Factor analysis, correlation analysis and regression analysis are used to develop equations to estimate the daily additional reduction in streamflow. These equations were found to be satisfactory with hydrograph stage being the most important independent variables.

INTRODUCTION

Five nested research catchments on the northern coastal belt of Natal are monitored by the Hydrological Research Unit of the University of Zululand. Rainfall, runoff and other hydrometeorological variables are measured continuously in these catchments which range in area from 0,67 km^2 to 85 km^2. In dry periods, when the rate of discharge recedes over a number of days, consistent diurnal fluctuations in streamflow, superimposed on the long-term trend of the hydrograph recession, have been observed. These fluctuations are most evident in the smallest of the five catchments, namely, catchment W1M17.

Such diurnal fluctuations in the recession hydrograph of small catchments have been reported by numerous researchers, for example, Troxell (1936), Croft (1948), Todd (1959), Meyboom (1966), Nänni (1972), Weisman (1977) and Burt (1979). These authors generally attribute such fluctuations to the diurnal variations in evaporative water losses in the riparian zone of the stream. Vapour losses are at a maximum when the solar energy load is at its greatest and tend to zero at night time. Evapotranspiration in the riparian zone is assumed to interrupt the supply of soil water and water from the groundwater aquifer to the stream.

Additional daily water loss from the riparian zone may be expected to be related to daily pan evaporation, temperature and the water available for evapotranspiration. This hypothesis is examined using data from catchment W1M17.

STUDY CATCHMENT, AIMS AND DATA

Catchment W1M17 has an area of 0,67 km² of which 10 percent is marshland (Figure 1). These marshlands are adjacent to the river channels and are covered by phreatophytic grasses (Hope and Mulder, 1979). Catchment W1M17 is located at 31°46'E and 28°50'S, is underlain by biotite and granite-gneiss and most of the area is covered by grassland. Mean annual precipitation for the region is 1450 mm and mean annual runoff is approximately 31% of this total (Hope and Mulder, 1979).

Two specific aims are embodied in this study, namely,

(i) to determine whether the daily excess water loss from the recession hydrographs of catchment W1M17 can be related to daily pan evaporation, temperature and hydrograph stage and

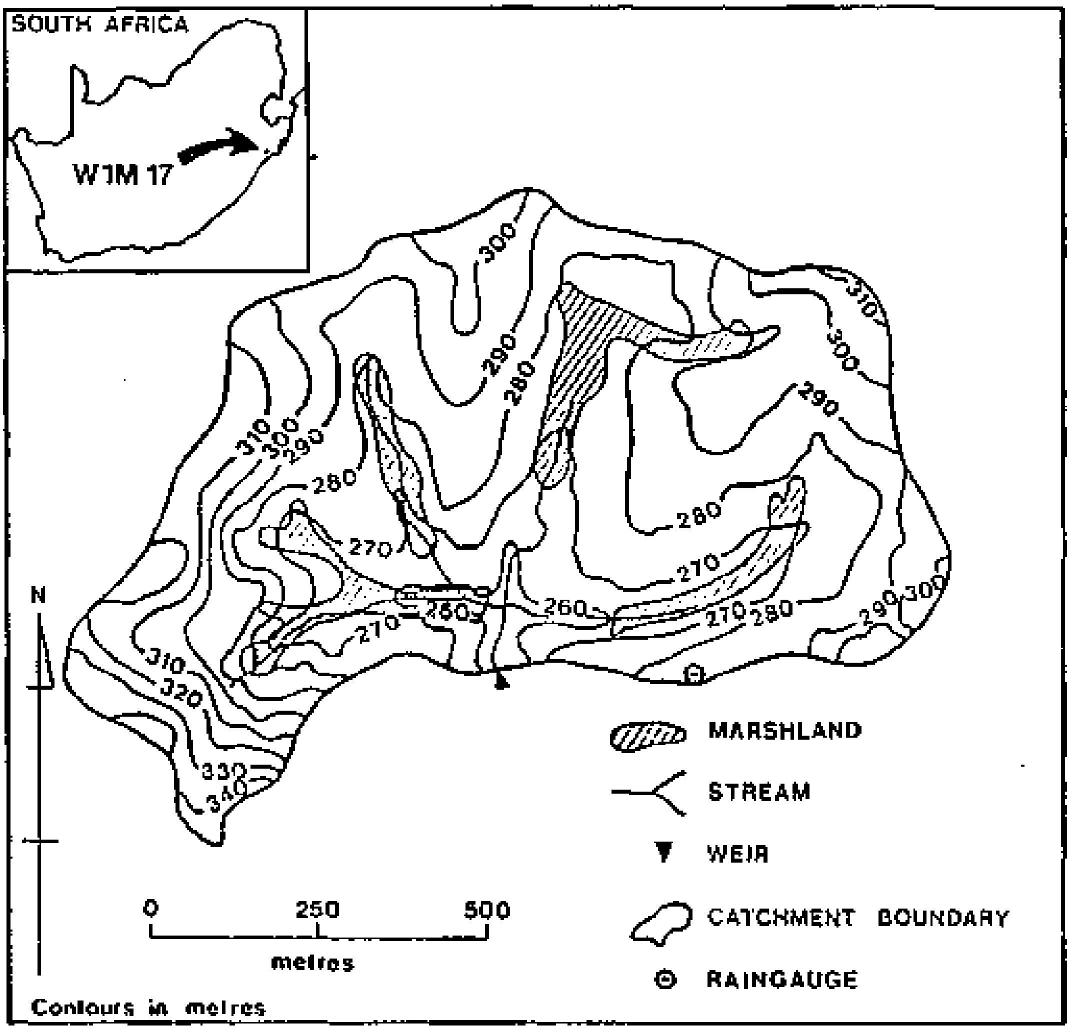


Figure 1 : The selected study catchment, W1M17

(ii) to develop regression equations from the above variables in order to establish daily additional water loss, ER ($m^3 day^{-1}$).

Seven independent variables were defined and included in the analyses, namely,

HS = hydrograph stage (m),

EA = A-pan evaporation (mm day^{-1}),

EA1 = A-pan evaporation for the preceding day (mm day^{-1}),

ES = S-pan evaporation (mm day^{-1}),

ES1 = S-pan evaporation for the preceding day (mm day^{-1}),

AVT = daily average temperature (°C) and

AVT1 = average temperature for the preceding day (°C).

Evaporation and temperature variables for the preceding day were included to represent any lag phenomena or residual effect which may affect the losses of the next day. Daily evaporation and temperature data were measured at a meteorological station 13 km from the study catchment. This station is also 180 m lower in elevation than the outlet of the catchment.

Stream discharge records for the period February 1980 to March 1982 were analysed for this study. A typical recession hydrograph illustrating the diurnal fluctuations superimposed on the general recession trend of the hydrograph is shown in Figure 2.

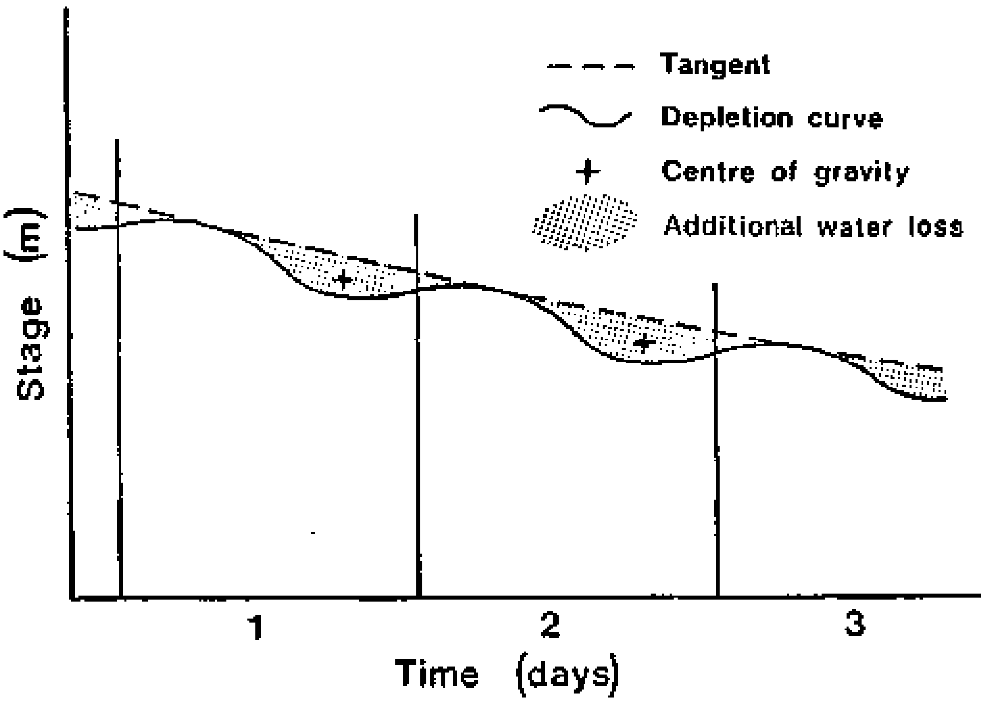


Figure 2 : Flow depletion curve and additional water loss

The rate of discharge which may be assumed to occur without these daily fluctuations was taken to be a straight line drawn tangential to the points where additional losses begin. Total additional water loss for any day is the difference in discharge between this line and the actual recession curve. Since there is clearly a time lag in the onset and termination of the daily reduction in flow, calculated loss for any one day ends at the start of additional loss for the next day. Hydrograph stage, HS, was taken to be the centre of gravity of the area enclosed by the tangential line and recorded recession curve (Figure 2).

Seventy-six observations were recorded for the period under investigation. The means and standard deviations for the dependent and independent variables are given in Table 1.

Table 1 : Means and standard deviations of the independent and dependent variables using linear and logarithmic data (N = 76)

Variable	Mean		Standard Deviation	
	Linear	Log	Linear	Log
ER : Additional water loss (m^3)	73,30	1,792	45,23	0,256
HS : Hydrograph state (m)	11,71	1,055	3,03	0,110
EA : A-Pan evaporation (mm day^{-1})	5,79	0,719	2,29	0,213
EA1 : EA-preceding day (mm day^{-1})	5,85	0,724	2,30	0,213
ES : S-pan evaporation (mm day^{-1})	4,80	0,605	2,10	0,375
ES1 : ES-preceding day (mm day^{-1})	4,86	0,611	2,11	0,375
AVT : Average temperature (°C)	22,97	1,357	3,31	0,065
AVT1 : Average temperature/day (°C)	23,08	1,359	3,24	0,063

ANALYTICAL PROCEDURES

Since the observations for this investigation were collected over a period of two years, the first analytical step was to test whether the collected data exhibited any climatic non-homogeneity or whether they could be considered part of the same population. This initial test was undertaken using linear discriminant analysis described in detail by Seyhan (1981). Data for this and subsequent analyses were transformed logarithmically to approximate the normal distribution better. The distribution is a fundamental assumption in all the statistical procedures adopted.

In order to develop regression equations to calculate additional diurnal water losses (ER) it is necessary to select 'truly' independent variables for any one equation. Principal components factor analysis was used to identify independent dimensions or factors which would explain the variance in the observed data. Highly interrelated variables are grouped into a single factor and one variable from each factor may be chosen for the regression equation. The factor analysis was based on a symmetrical correlation matrix (of Pearson's product-moment correlation coefficients) with variance orthogonal rotation criterion (Seyhan, 1981).

The degree of association between a variable and other variables in a factor is given by the factor loading of that variable. These loadings have a similar meaning to that of a correlation coefficient in a correlation matrix. Since it is difficult to interpret a factor matrix which contains all the factor loadings, it is necessary to retain only those variables having a loading greater or less than a set value (Seyhan,1981). The selection of this threshold value is arbitrary and depends largely on the experience of the investigator. However, in many analyses there is a sharp transition between a group of high/low factor loadings and a group of values which are close to zero, thus assisting in the selection of the critical value. For each factor the cumulative eigenvalue is given which represents the cumulative amount of variance explained by successive factors. The communality, h_i^2, gives the percentage of variance of the ith variable in the matrix of m factors.

Step-wise multiple linear regression was used to develop regression equations to calculate ER. The selection of truly independent variables was based on the results of the principal components factor analysis.

RESULTS AND DISCUSSION

The null hypothesis that the data for the 76 observations was drawn from different populations was rejected at the 0,1% level of significance. Thus it was concluded that there was no climatic non-homogeneity in the data and that stratification of the sample was not necessary.

The simplified rotated factor matrix produced by the principal components factor analysis of the six independent and one dependent variable is presented in Table 2.

Table 2 : Simplified rotated factor matrix of independent and dependent variables (factor loads $\geq$ 0,6 and $\leq$ -0,6)

Variable	Factor loads			h_i^2(%)
	I	II	III	
Additional water load (ER)		0,940		90,5
Hydrograph stage (HS)		0,904		89,1
A-Pan Evaporation (EA)	0,838			73,5
A-Pan Evaporation/day-1(EA1)			0,744	74,0
S-Pan Evaporation (ES)	0,801			69,9
S-Pan Evaporation/day-1(ES1)			0,894	80,3
Average temperature (AVT)	0,685			66,6
Average temperature/day-1 (AVT1)	0,602			67,7
Eigenvalue	3,35	1,66	1,06	
Cumulative % of total explained variance	41,84	62,62	75,82	

Three independent factors were identified and which together explained 75,82% of the variance in the original data. The first factor, which explained 41,84% of the variance, is clearly the most important dimension.

This factor defines the intercorrelation between pan evaporation and temperature and unexpectedly includes the average temperature of the preceding day. Factor II gives a clear indication of the strong relationship between ER and HS (stage) while factor III reflects the association of the A-pan and S-pan evaporation depths for the preceding day.

Seven significant regression equations to calculate ER were developed and are presented along with selected statistics in Table 3.

Table 3 : Regression equations of additional water loss (ER) with the multiple correlation coefficient (r), standard error of estimate (S_y) and significance of the regression coefficient (c)

Equation	r	$S_y(m^3)$	c(%)
$ER = 0,983\ (HS)^{1,706}$	0,729	1,500	99,95
$ER = 0,018\ (HS)^{1,915}(AVT)^{1,125}$	0,778	1,455	99,95
$ER = 0,767\ (HS)^{1,748}(ES)^{0,105}$	0,745	1,493	97,50
$ER = 0,451\ (HS)^{1,844}(EA)^{0,267}$	0,760	1,476	99,50
$ER = 0,702\ (HS)^{1,766}(ES1)^{0,137}$	0,756	1,479	99,00
$ER = 0,372\ (HS)^{1,889}(EA1)^{0,316}$	0,771	1,462	99,95
$ER = 0,063\ (HS)^{1,856}(AVT1)^{0,761}$	0,750	1,486	99,00

Each of the regression equations includes HS since, as was pointed out previously, this variable has a strong association with ER. The equation which was developed using HS alone has a correlation coefficient (r) of 0,729, thus accounting for 53,1% of the variance. However, by including any of the other variables in the equation variances increased slightly (Table 3) to a maximum of 60,5% ($r = 0,778$).

The strong relationship between HS and ER in this analysis may be explained in terms of the water which was available for evapotranspiration in the riparian zone. When discharge from catchment W1M17 is high (i.e. high value of HS) the groundwater aquifer and soil moisture of the catchment and riparian zone may be assumed to be greater than at times when HS is low. Under such wet conditions evapotranspiration is likely to be at the potential rate and not restricted by the availability of water. When evapotranspiration is limited by the availability of water (low HS) then losses are presumed at a rate less than the potential rate and ER is reduced proportionately. However, in an attempt to improve the regression equations, actual evapotranspiration in the riparian zone was estimated for each day using a soil moisture budgeting procedure described by Schulze (1982). These values were used in place of pan evaporation, but failed to improve the results.

CONCLUSIONS

The regression equations developed to calculate ER may, on the basis of the minimal standard errors of estimate (< 10 percent), be regarded as satisfactory. The low percentage of variance accounted for by these equations may be attributed to one or more factors. First, the temperature and evaporation measurements were made at a point some 13 km from catchment W1M17 and at an altitude 180 m below that of the catchment. Secondly, the assumption that the uninterrupted recession curve could be approximated by straight tangential lines may have introduced slight additional error. Finally, the processes controlling the magnitude of ER losses may be different to those hypothesised in this study. Factors such as air pressure may have an effect on fluctuations in the recession hydrographs.

This preliminary study has indicated that the diurnal fluctuations in the recession hydrograph from a small catchment may be related successfully to selected variables. However, a better understanding of the processes involved and variables which need to be measured is required. An extension of this study to include additional catchments, more observations and additional independent variables may help towards this understanding.

ACKNOWLEDGEMENTS

The authors wish to thank Professor P. Meiring, Head, Department of Agricultural Engineering, University of Natal, for placing the facilities of the Department at their disposal. Thanks are also due to the staff of the Department of Agricultural Engineering, in particular, Miss R. Sands, for their assistance in this study.

This research was funded by the Water Research Commission, Republic of South Africa.

REFERENCES

Burt, T.P. (1979) Diurnal variation in stream recharge and throughflow during a period of low flow. J. Hydrol., 41, 291-301.

Croft, A.R. (1948) Water loss by stream surface evaporation and transpiration by riparian vegetation. Trans. Am. Geophys.Unión, 29, 235-239.

Hope, A.S. and Mulder, G.J. (1979) Hydrological investigations of small catchments in the Natal Coastal Belt and the rôle of physiography and land-use in the rainfall-runoff process. Rep. series B, (6), Hydrol. Res. Unit, Univ. Zululand, Kwa Dlangezwa, 283 pp.

Meyboom, P. (1966) Three observations on streamflow depletion by phreatophytes. J. Hydrol., 2, 248-261.

Nänni, U. (1972) Water use by riparian vegetation at Cathedral Peak. S. Afr. For. J., 80, 1-10.

Schulze, R.E. (1982) Agrohydrology and climatology of Natal. Water Research Commission, Pretoria, 176 pp (In press)

Seyhan, E. (1981) Introduction to multivariate statistical analysis in hydrology. Institute of Earth Sciences, Free Univ., Amsterdam.

Todd, D.K. (1959) Groundwater Hydrology. Wiley, New York, pp.

Troxell, M.C. (1936) The diurnal fluctuation in groundwater and flow of the Santa Ana river and its meaning. Trans. Am. Geophys. Union, 17, 496-504.

Weisman, R.N. (1977) The effect of evapotranspiration on streamflow recession. Bull. Hydrol. Sci., 22, 371-377.