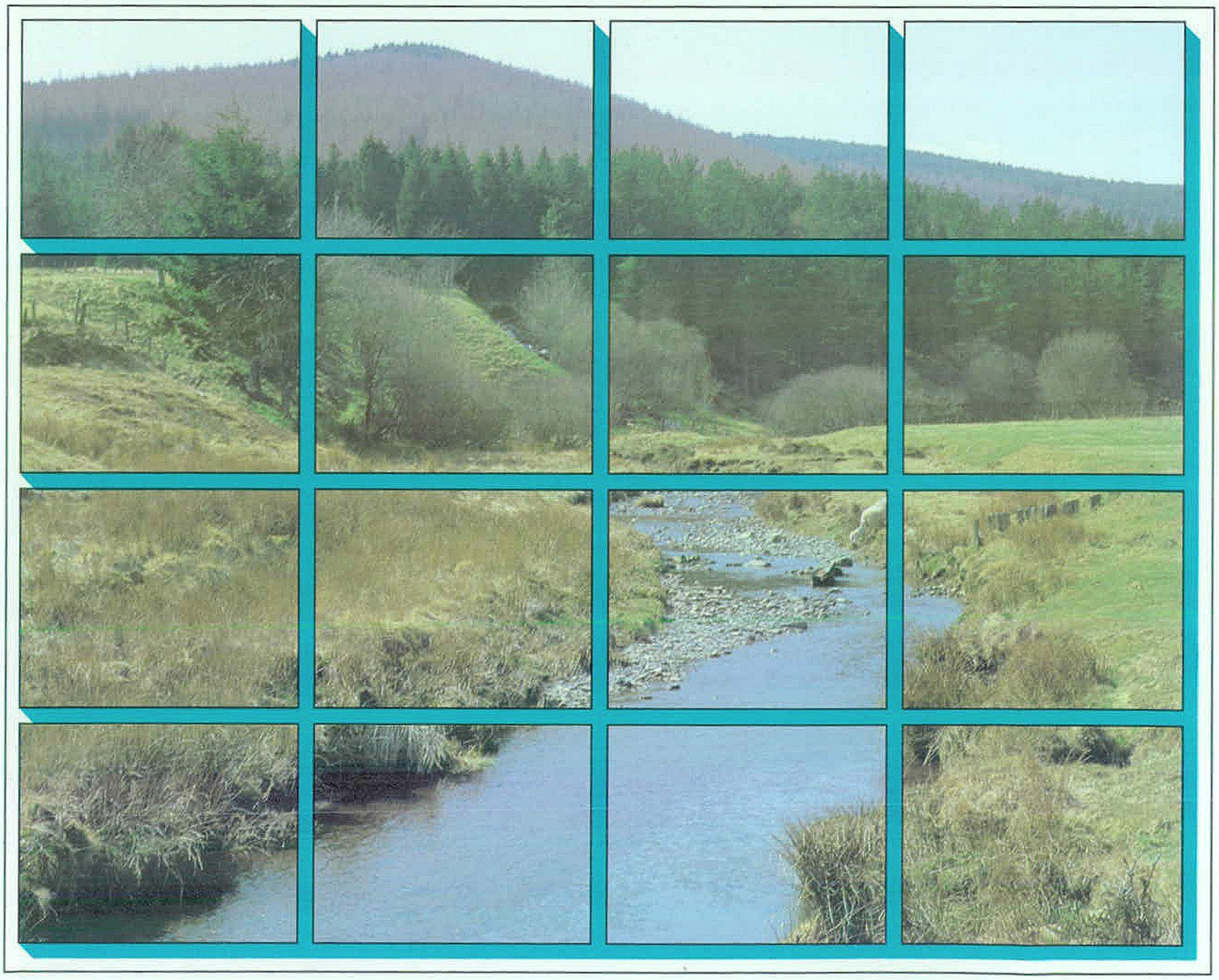
Institute of Hydrology

Report No. 118

Catchment classification applied to the estimation of hydrological parameters at ungauged catchments



Natural Environment Research Council

Report No. 118

Catchment Classification Applied to the Estimation of Hydrological Parameters at Ungauged Catchments

Donald H. Burn and David B. Boorman

April 1992

Institute of Hydrology
Maclean Building
Crowmarsh Gifford
Wallingford
Oxfordshire OX10 8BB

ISBN 0 948540 40 0

IH Report No. 118

Published by the Institute of Hydrology

April 1992

A catalogue record for this book is available from the British Library

Abstract

Estimates of hydrological parameters at ungauged sites have traditionally been obtained from regression equations. This study investigates alternative methods based on the classification of catchments according to their flow regime, the assignment of ungauged catchments to a class based on physical characteristics of the catchment, and the use of similarity measures to transfer parameters from gauged to ungauged catchments. The report considers the methods that can be adopted in this type of approach, and the many variations that must be considered in their implementation. The methods are examined using a set of 99 catchments from the UK, and are seen to be efficacious in estimating the unit hydrograph time to peak and standard percentage runoff, as defined by the UK Flood Studies Report. A step-by-step guide and worked example show how the method can be applied in practice.

Contents

Page

EXECUTIVE SUMMARY

NOTATION

INTRODUCTION

APPROACH

2.1 Grouping of catchments 3
2.2 Assignment of ungauged catchments 6
2.3 Estimation of rainfall-runoff model parameters 8

AVAILABLE DATA 12

ANALYSIS 15

4.1 Grouping of catchments 15
4.2 Assignment of ungauged catchments 17
4.3 Estimation of rainfall-runoff model parameters 19

CATCHMENTS WITH SOME FLOW DATA 40

5.1 Results 40
5.2 Summary 42

CONCLUSIONS AND RECOMMENDATIONS 46

ACKNOWLEDGEMENTS 47

8 REFERENCES 48

APPENDICES

A Data listings 49

B Guide to application of methods 64

Executive Summary

A much used method of estimating hydrological parameters at ungauged sites has been to use empirically derived regression equations. The UK Flood Studies Report adopts such an approach to derive rainfall-runoff model parameters (unit hydrograph time-to-peak and percentage runoff) for use in a design flood estimation procedure. However, users of this procedure are strongly recommended to examine parameters derived on similar nearby gauged catchments to refine or replace the regression estimates at the ungauged site. This recommendation is backed up by experience in using the methods rather than an analysis of data. If, in practice, the best parameter estimate is obtained using so-called *local data* then it seems reasonable to attempt to develop an estimation procedure based on this approach rather than regression analysis. This report examines a variety of methods that adopt this type of approach, and the best of these are found to be as good, or better, than the regression based method.

A three part approach was used in the study. Firstly, catchments were classified (clustered) according to observed flow indices to give a broad distinction between hydrological regimes. Secondly, a method of assigning catchments to the classes based on physical properties of the basins was developed. Thirdly, the catchment properties and cluster membership were used in five different ways to estimate the hydrological parameters. The five methods included averaging the observed parameter values from sites within the same cluster that are most similar to the ungauged sites, and developing separate regression equations for catchments in each cluster. Variations on the former of these allowed the similar catchments to be from more than one cluster where there was some ambiguity about to which cluster the ungauged site belonged, and also to add a geographical distance limit to the search for similar catchments.

Many issues need to be resolved in adopting this approach. The clustering process can be performed using different procedures (eg. agglomerative or hierarchical) and different flow variables (or transformations of variables), and since the hydrological data form a continuum rather than a clumped data set, a decision on the number of clusters is also needed. The assignment process uses discriminant analysis but, as in the clustering process, requires a review of which catchment data are most useful, although this time the data are the physical properties rather than the flow indices.

The data available for this study come from 99 catchments mainly in England and Wales, but with a few from southern Scotland. Seven variables derived from observed data were available to describe various aspects of the flow regime, and a set of nine catchment characteristics abstracted from maps described the physical nature of the drainage basin.

Clustering of the catchments used the K-means algorithm applied to principal components derived from all seven flow variables. Assignment of catchments to clusters used discriminant analysis applied to canonical variables obtained from all of the available characteristics. An examination of the parameter estimation methods showed that good results were being obtained with very few (ie. two or three) clusters. In both of these cases the catchments were divided into groups that contained roughly the same number of members (two clusters 51 and 46 members: three clusters 34, 29 and 36 members). Some geographic trends could be seen in the catchments but location was not a strong distinguishing feature of the groups. The assignment process correctly placed 91 and 76 (of 99) catchments in the two and three cluster cases respectively. The poorer performance of the assignment process in the three cluster case demonstrates why moving to a greater number of clusters, in which the type of catchment becomes better defined, is not necessarily the best option.

The clustering and assignment process described thus far considers all flow variables equally and could be used as an aid in studies of, for example, low flows or floods. The cluster membership information has been taken and used in the specific example of estimating the Flood Studies Report variables time-to-peak and standard percentage runoff.

Of the methods investigated the best for the estimation of standard percentage runoff was found to be from the two cluster scheme using the arithmetic mean of values observed at the two most similar catchments within the cluster to which the ungauged catchment is assigned. The similarity of the catchments should be assessed using the particular definition given in this report, which is based on the five Winter Rainfall Acceptance Potential classes. This estimation method represents a considerable improvement over the regression equation presented in Flood Studies Supplementary Report No.16 both in terms of bias and RMS error. This result reinforces the advice to use local data to refine regression equation estimates. The method can be considered a formalization of this general advice into a defined procedure.

The results for the estimation of time-to-peak are less clear cut. The equation presented in Flood Studies Supplementary Report No.16 is robust and reliable in most applications, and errors in its estimation have a less dramatic effect than do errors in percentage runoff. The best alternative to the regression approach was found to be from a weighted mean of regression estimates obtained from individual equations for each cluster. This method works well in both the two and three cluster cases, giving slightly lower RMS error than the regression equation but with slightly higher bias. Overall there is not sufficient gain with the new method to justify the complexity of the new method.

Further refinements are considered in which one flow variable is available at the site of interest. This information can be used both in the assignment process and the estimation process. The Base Flow Index is beneficial in improving percentage runoff estimation compared with both the method developed for the transfer of local parameter values, and the regression equation using Base Flow Index contained in Supplementary Report No.16.

NOTATION

C^k_m	centroid coordinate for feature m of cluster k
d_{ij}	distance between catchment i and j
E_{ik}	regression estimate of the rainfall-runoff model parameter for catchment i from cluster k
I_k	number of catchments in cluster k
J	number of catchment characteristics
j	catchment characteristic index
K	number of clusters
k	cluster number
L_i	set of n nearest neighbours for catchment i
M	number of attributes (flow response measures)
m	attribute number
N	number of catchments
n	number of nearest neighbours used in the estimation of parameters
nc	number of catchments used in the similarity comparison
nn_{ik}	number of nearest neighbours to catchment i that are from cluster k
nm_k	number of catchments in cluster k
p_{ik}	probability that catchment i is a member of cluster k
pr_k	prior probability of membership in cluster k
q	number of canonical variables, min (J, K-1)
R_l	rainfall-runoff model parameter for catchment l
S	sample covariance matrix for catchment characteristics
V_i	vector of canonical variables for catchment i
vw	matrix of canonical variable weights
W_m	weight for feature m
X^i_m	value of feature m on catchment i
Y_{ij}	catchment characteristic j for catchment i
$Y_{.j}$	mean of catchment characteristic j
YS_j	standard deviation of catchment characteristic j
y_i	vector of standardized catchment characteristics for catchment i
Δ	total within group variability
δ_k	variability measure for cluster k
Ω_l	weight applied to the rainfall-runoff model parameter for catchment l
ω_t	weight applied to catchment characteristic t in the distance metric

1 Introduction

An approach to the estimation of extreme flow probabilities for ungauged catchments, or catchments for which only limited flow data are available, is to use a rainfall-runoff model to transform a design rainfall event into a peak flow value with a specified probability of exceedence. For catchments in the UK, the Flood Studies Report (NERC, 1975) describes a methodology for conducting flood frequency analysis for ungauged catchments based on this type of procedure. When applying this method to ungauged catchments, it is necessary to determine model parameters using data describing the physical characteristics of the catchment. The two most important model parameters are the unit hydrograph time to peak, Tp, and the standard percentage runoff, SPR, which is a measure of the percentage of rainfall that generates runoff. A third rainfall-runoff model parameter, the peak flow of the one hour unit hydrograph, Qp, can be estimated from the value of Tp. The Flood Studies Report (FSR) presents regression equations for obtaining estimates of these model parameters. Later work, Flood Studies Supplementary Report No. 16 (IH, 1985), used slightly redefined definitions of the model parameters, and presented new regression equations for use at ungauged sites.

This regression based approach to the estimation of hydrological parameters is by no means unique to the UK or to the studies mentioned above. The time of concentration has been used in design flood studies for many years; the Bransby-Williams (1922) equation is one of the best known for estimating this characteristic time at an ungauged site. Even restricting a survey to the estimation of unit hydrograph model parameters yields a considerable literature. In a British context Nash (1960) presents regression equations for the estimation of unit hydrograph parameters, but offers no advice on how to estimate rainfall loss model parameters. Heerdegen & Reich (1974), in their study of catchments in Pennsylvania, USA, and Cordery & Pilgrim (1983) in a study of Australian catchments also fail to relate parameters describing a rainfall loss model to characteristics of the catchment or rainfall event.

The FSR recognised that there would be considerable uncertainty associated with the model parameter estimates obtained from the regression equations, and suggested that this might be reduced by incorporating local data (i.e., results from the analysis of flood event data recorded on nearby catchments). The first formalization of this recommendation is to be found in Lowing & Reed (1980). For time to peak, the recommendation is a simple scaling of the regression equation estimate by a local scaling factor (i.e., the observed value divided by the estimated value averaged over all nearby sites), subject to conditions concerning the similarity of the site of interest and its neighbour. These conditions are that the catchments are of about the same size and that they are both on the same river channel. The suggested size constraint was that the larger catchment should be no more than five times the area of the smaller. For SPR, it was suggested that a value should be transferred from a nearby site if the neighbouring catchment is broadly similarly in terms of geology, topography, and land use. These recommendations were based on "experience of using the model" rather than an analysis of gauged data.

Two issues are raised by these recommendations. Firstly, is it possible to demonstrate that this use of local data is beneficial, and if so can the benefit be quantified? Secondly, if it is better to use local values of SPR rather than the regression equation, should a method of estimation based solely on the use of local data be recommended in place of the regression equation?

This report examines alternative methodologies for estimating the requisite rainfall-runoff model parameters based on the use of local data, and compares such estimates with those derived from the FSSR16 regression equations. However, an important feature of this study was to move away from the idea of using data only from catchments in the proximity of the study catchment, towards the use of information from sites that are similar, but not necessarily close, to the ungauged catchment. The only available method by which an ungauged catchment can be judged similar to, or different from, others is by comparing catchment characteristics.

The problem described above has some parallels with work in regional flood frequency analysis based on estimating extreme flow probabilities at gauged or ungauged sites using annual flow data from a collection of gauged locations. Wiltshire (1986) employed a combination of cluster analysis and discriminant analysis to estimate extreme flows at ungauged sites. Acreman (1987) and Acreman & Wiltshire (1989) suggested a framework for regional flood frequency analysis that dispenses with the need for unique regions. Burn (1990) presented an evaluation of such a technique referred to as the Region of Influence approach wherein flood frequency analysis for a selected site uses information from all other gauged sites that are sufficiently similar to the site of interest.

In the present study, preliminary work showed that some catchments with similar physical characteristics had different flow regimes, and that an initial division of catchments into a number of groups may prove necessary, or beneficial, in estimating hydrological parameters at the ungauged site. The fact that this seemed necessary indicates that the guidelines for assessing the similarity of catchments prior to transfer of model parameters were an essential consideration in the use of data from nearby sites.

In addition to comparing methods of model parameter estimation using catchment characteristics, this study considers how the derived flow parameter, base flow index (BFI), can be used to improve estimation. This parallels the use of BFI in a regression equation to estimate SPR that is recommended in preference to the catchment characteristics regression by FSSR16.

2 Approach

The FSSR16 approach to model parameter estimation was to consider all catchments together and to find the best, physically acceptable, regression model. Local data could then be used according to the suggested guidelines and at the users' discretion to refine the estimates. In this study, we start by grouping the catchments into a small number of classes according to their hydrological similarity. While it may have been possible to make this hydrological classification based only on the variables of interest (i.e., Tp and SPR), it was considered better to include a greater number of flow variables as this would give a more general view of the catchment's flow regime, and lead to a classification that may be appropriate beyond the present application. The information about class membership is an extra item of data that may then be used in building a parameter estimation model. Various models are considered for parameter estimation, including a regression approach identical to that used in FSSR16, and a "transfer value" approach similar to the one built into the "local data" recommendations. Thus in the (trivial) case of grouping the catchments into one class, estimation using the traditional regression equation, and local data approaches may be compared. Where a greater number of classes are used, then an extra component must be added to consider how ungauged catchments can be assigned to one of the groups defined by hydrological similarity. The approach therefore comprises three parts.

1 Classification of catchments based on hydrologic similarity.
2 Assignment of catchments to classes based on their catchment characteristics.
3 Development of a method of model parameter estimation based on available information, including probable class membership.

2.1 GROUPING OF CATCHMENTS

The intent within the grouping process is to subdivide the entire set of catchments based on similarity in the flow response of the catchments. Thus, for example, catchments that have a very quick runoff response should be distinguished from those that have a slower and more sustained response.

2.1.1 Clustering Methods

To carry out the grouping of the catchments, cluster analysis (see, for example, Anderberg 1973) can be employed. There are two generic types of cluster analysis; the agglomerative or hierarchical approach and methods based on partitioning the data set. The former approach is based on assembling individual objects into larger groups whereas the latter approach is based on dividing a data continuum into distinct groups. The nature of our problem and the intent of the catchment grouping process conform more closely to the second approach and therefore in this work, the version of the K-means clustering algorithm described by Burn (1989) was used.

2.1.2 The K-Means Algorithm

This algorithm divides the entire set of objects (catchments) into K clusters (groups) based on the values of M features, or attributes, of the objects. For the work described herein, the features comprised either a subset of the available flow response measures, or variables derived as a linear combination of the flow response measures. The objective of the clustering process is to minimise:

$$\sum_{k=1}^{K} \sum_{i \in I_k} \sum_{m=1}^{M} W_m (X^i_m - C^k_m)^2 \tag{1}$$

where W_m is the weight applied to feature m in the Euclidean distance measure; X^i_m is the value of feature m for object i; C^k_m is the centroid coordinate for feature m of cluster k; K is the number of clusters; I_k is the set of objects in cluster k; and M is the total number of features.

The K-means algorithm involves selecting K objects to function as seed points, or initial cluster centroids. Each of the objects is then assigned to the cluster corresponding to the centroid that it is nearest to where proximity is measured in terms of a weighted Euclidean distance in the M-dimensional space defined by the features selected (see Equation 1). After all of the objects have been assigned to a cluster, the centroid for each cluster is recalculated based on the membership of the cluster. Each object is then again assigned to the cluster corresponding to the centroid that it is now nearest to, and after all objects have been assigned, the centroids are recalculated. This process is repeated until no object changes cluster membership in successive applications of the assignment step. Within the clustering process, there are several issues to be resolved, including:

1 The identification of a global optima.
2 The determination of an appropriate number of clusters.
3 The selection of clustering variables and associated weightings.

Identification of a global optima

This issue arises from the nature of the K-means algorithm. The algorithm, as outlined above, will identify the optimal partitioning of the catchments for the particular combination of K catchments selected as seed points. Although the results tend to be reasonably robust with respect to the starting seed points, there is no guarantee that a global, as opposed to a local, optima has been achieved. To compensate for this, the catchments that act as the seed points were randomly selected and the algorithm was repeatedly solved with different combinations of catchments functioning as seed points. It was found that 200 repetitions of the process provided a reasonable assurance that a global optima had been identified.

Determination of number of clusters

The selection of the number of clusters (groups) to divide the catchments into involves choosing a particular value for K. Formal approaches for selecting the preferred number of clusters have been outlined in the literature (Galeatti *et al.*, 1986) where the intent is to obtain homogeneous groupings of the objects while still retaining a reasonable number of objects in each group. The employment of such a procedure to this application would have resulted in a fairly large number of groups (of the order of ten clusters) with a corresponding small number of members in each group. Since the intent with the clustering component of the estimation process is to obtain a broad classification of the catchments, in terms of flow response, a more subjective process was used to select K. Results were examined with the number of groups ranging from two to five and a subset of these results were retained for consideration within the remaining stages of the procedure.

Selection of variables and weights

The clustering variables selected are used as the features, or attributes, in the proximity measure which defines catchment similarity. Since the ultimate intent is to estimate the value of Tp and SPR for the ungauged catchments, one option would be to use only these two variables as the attributes in the distance metric. While this would result in clusters of catchments that are similar in terms of the two variables that are of primary interest, this is not necessarily the best approach to partitioning the data. Since it will subsequently be necessary to assign catchments to a cluster based on data describing the catchment characteristics, it is apparent that adopting a broader definition of catchment similarity will be likely to improve the assignment of catchments to the appropriate group using catchment characteristic data. As such, the approach ultimately taken herein was to consider all available flow response variables within the clustering process. However, the question of the relative importance, or weight, to assign to the various variables still remains. Related to the question of appropriate weights is the issue of correlation between the flow response variables. If two variables are correlated, they are, to some extent, measures of the same attribute of catchment flow response. It could, therefore, be argued that a reduced weight in the distance metric should be assigned to such variables to reflect this. However, the subjective assignment of weights will potentially impact the resulting partitioning of the catchments, implying that an objective weighting scheme is desirable.

Alternative approaches to clustering, in terms of a combination of the available flow response variables, are based on calculating principal components from the entire set of flow response variables and using the significant principal components as attributes in the proximity measure. Nathan & McMahon (1990) describe a number of clustering options that incorporate principal components. One result of adopting the principal components approach is to reduce the dimensionality of the problem while still retaining the effects of all of the variables in the analysis. The principal component approach will also provide an objective assignment of weightings to the original variables in that the weightings are determined directly from the principal components analysis. In addition, the correlation between variables is explicitly accounted for in that variables with a high correlation will tend to be associated with the same principal component. Within the principal component approach to clustering, several alternatives exist, including:

1 Clustering of unweighted principal components;
2 Clustering of principal components weighted by eigenvalue;
3 Clustering of rotated principal components, either weighted or unweighted.

The first option assumes that each principal component is of equal importance in defining catchment similarity whereas the second option assumes that the importance of the principal components is related to the fraction of the variance of the original data explained by the component. The third option entails a rotation of the principal components which results in a clearer distinction of the variables associated with each of the principal components.

2.2 ASSIGNMENT OF UNGAUGED CATCHMENTS

The assignment of ungauged catchments to one of the clusters identified in the first stage of the process can be accomplished using discriminant analysis. However, not all of the variables describing the physical characteristics of a catchment are necessarily of use in discriminating between the clusters. Furthermore, the preferred variables for use in the discrimination process may change as a function of the number of groups into which the original set of catchments has been divided. The first step in the assignment of ungauged catchments to a cluster was thus the identification of a suitable set of catchment characteristic data for use within a discriminant analysis technique.

2.2.1 Selection of catchment characteristics

Formal statistical procedures exist for the identification of relevant variables (SAS, 1985). These include step-wise discriminant analysis, which attempts to determine the best combination of variables to distinguish between a given set of clusters, and canonical discriminant analysis, which defines a set of canonical variables that are linear combinations of all of the variables considered. A total of q canonical variables can be calculated where $q = \min(J, K-1)$ where J is the number of catchment characteristic variables and K is the number of clusters, as before. The canonical discriminant analysis procedure calculates a set of weights and the canonical variables can then be calculated using these weights through:

$$V_i = vw^T y_i \tag{2}$$

where V_i is the (K-1) dimensional vector of canonical variables for catchment i; vw is the matrix of weights, of dimension J by (K-1), determined from the canonical discriminant analysis; and y_i is a J-dimensional vector of standardized catchment characteristics. The elements of the vector y_i are calculated through:

$$y_{ij} = \frac{Y_{ij} - Y_{\bullet j}}{YS_j} \tag{3}$$

where y_{ij} is the jth element of the standardized catchment characteristic vector for catchment i; Y_{ij} is the jth element of the catchment characteristic vector for catchment i; and $Y_{.j}$ and YS_j are the mean and standard deviation for catchment characteristic j calculated from all catchments in the data set.

Both the stepwise and canonical discriminant analysis procedures assume that the set of variables used in the discriminant analysis follow a multivariate Gaussian distribution. Since the nature of the data describing catchment characteristics may result in the validity of this distribution assumption being questionable, additional analysis was used to assist with the determination of an appropriate set of variables. The main additional investigative tool used was a graphical display of various combinations of variables. This was used to subjectively evaluate the capability of the variables to distinguish between the defined clusters. The strategy taken was to use the results of the step-wise discriminant analysis and the canonical discriminant analysis as a starting point for identifying potentially useful variables, or combinations of variables. Graphics plots were then used to confirm the utility of proposed combinations of variables. Finally, the results of actually assigning gauged catchments to the clusters, where the appropriate cluster membership is known a priori, were used to refine the selection of discriminating variables.

2.2.2 Discriminant analysis

Once a set of discriminating variables was selected, the specific discriminant analysis technique adopted was a non-parametric discriminant analysis technique based on the nearest neighbour approach (SAS, 1985). This approach involves assigning a catchment to a cluster in accordance with the cluster membership of the n catchments nearest to the ungauged catchment. The proximity of gauged catchments to the catchment to be classified is defined in terms of a distance metric using the catchment characteristics identified as being of relevance. Two common options for the distance metric are the Euclidean distance and the Mahalanobis distance, where the latter accounts for correlation between the variables. The Euclidean distance between two catchments is defined as:

$$d_{ij}=(Y_i-Y_j)^T(Y_i-Y_j) \tag{4}$$

where d_{ij} is the distance between catchment i and catchment j; and Y_i is the catchment characteristic vector for catchment i. The Mahalanobis distance metric is defined as:

$$d_{ij}=(Y_i-Y_j)^T S^{-1}(Y_i-Y_j) \tag{5}$$

where S is the sample covariance matrix of the catchment characteristic data and all other variables are as previously defined. The output from this procedure includes: 1) the probability of membership for each catchment in each cluster, based on the cluster membership of the n nearest neighbours; 2) the cluster to which each catchment is assigned; and 3) a summary of the performance of the assignment process. In this discriminant analysis procedure, a catchment is assigned to the cluster to which it has the highest probability of membership. The probability of membership for a catchment in a cluster is given by:

$$p_{ik}=\frac{nn_{ik}pr_k}{\sum_{l=1}^{K}\left(nn_{il}pr_l\right)} \tag{6}$$

where p_{ik} is the probability that catchment i is a member of cluster k; nn_{ik} is the number of nearest neighbours to catchment i that are from cluster k; pr_k is the prior probability of membership in cluster k; and the summation is over all clusters. The prior probability of membership in a cluster is defined as:

$$pr_k = \frac{nm_k}{N} \tag{7}$$

where nm_k is the number of members (catchments) in cluster k; and N is the total number of catchments.

2.3 ESTIMATION OF RAINFALL-RUNOFF MODEL PARAMETERS

In Sections 2.1 and 2.2, a methodology has been described that will allow catchments to be classified according to their hydrological similarity, and assigned to classes based on their physical properties. The next stage is to consider how catchment characteristics data, and class membership information, can be used to estimate the rainfall-runoff model parameters. Again many possibilities present themselves for consideration. The initial aim of the study was to use the similarity of the catchments in characteristic space, to determine how variables might be transferred between catchments. Having classified the catchments, this similarity could be assessed by only considering other catchments within the same group, by including a term in the similarity measure that reflects group membership, or by ignoring the groupings altogether.

In any of these scenarios, the first step is to determine a subset of the catchment characteristics to be used in the similarity measure. It is to be expected that the set of characteristics will vary according to the hydrological variable that is to be estimated, and between different classifications of the catchments.

To determine an appropriate set of catchment characteristic variables, the correlation was calculated between the difference in each catchment characteristic, for all combinations of station pairs, and the corresponding difference for each rainfall-runoff model parameter. From this analysis, a set of variables with significant correlations were identified and retained for possible use in the estimation process. The retained variables were evaluated in terms of their capability to identify catchments that are similar in terms of the rainfall-runoff model parameters. Specifically, the agreement between the nc nearest catchments in terms of a rainfall-runoff model parameter and the nc nearest catchments in terms of a combination of catchment characteristics was determined. The catchment proximities were calculated in terms of the Euclidean distance. The analysis was carried out individually for each catchment in the collection of catchments and a network average performance was determined. In this way, it was possible to evaluate potential combinations of variables for each of the rainfall-runoff model parameters of interest.

With the variables identified through the above analysis, several options were developed for estimating the rainfall-runoff model parameters for an ungauged catchment.

Option 1

Option 1 entails identifying the n nearest neighbours to the ungauged catchment from the catchments in the cluster to which the ungauged catchment has been assigned. The proximity of gauged catchments to the catchment of interest is calculated as a weighted Euclidean distance in the T-dimensional space defined by the catchment characteristics selected for the particular runoff parameter of interest. Catchment proximity is thus given as:

$$d_{il} = \sum_{t=1}^{T} \omega_t (y_{it} - y_{lt})^2 \tag{8}$$

where d_{il} is the distance from catchment i to catchment l; ω_t is the weight applied to catchment characteristic t; and y_{it} is the standardized value for catchment characteristic t for catchment i. The estimate for the rainfall-runoff model parameter for the ungauged catchment is then calculated as the weighted average of the rainfall-runoff model parameter values for the n nearest catchments, resulting in:

$$\hat{R}_i = \frac{1}{n} \sum_{l \in L_i} \Omega_l R_l \tag{9}$$

where $\hat{R}_i$ is the estimate of a rainfall-runoff model parameter for catchment i; n is the number of nearest neighbours selected; L_i is the set of catchments that comprise the n nearest neighbours to catchment i; Ω_l is the weight applied to the rainfall-runoff model parameter value for catchment l; and R_l is the rainfall-runoff model parameter value for catchment l.

Option 2

Option 2 is similar to Option 1 except that the search for the n nearest neighbours is restricted to those catchments which are less than a specified geographic distance from the catchment of interest. The logic behind this option is that catchments that are physically close together should be similar in terms of both catchment characteristics that affect runoff production (such as rainfall, soil type, etc.) and also, therefore, in terms of runoff response. The distance constraint imposed on the search domain for similar catchments is intended to avoid spurious similarities where catchments with very different flow response measures happen to have similar catchment characteristics.

Option 3

Option 3 entails identifying the n nearest neighbours where the pool of available catchments corresponds to the set of catchments from all clusters to which the ungauged catchment has a probability of membership that exceeds a specified probability threshold level. This option is considered since the catchment assignment process will result in some catchments being incorrectly classified as to cluster membership. This results from an imprecise relationship between similarity in flow response and similarity in catchment characteristics. Furthermore, not all of the catchments will be unambiguously assigned to one cluster but rather there may be several clusters which can lay a claim to the catchment. This could arise as a result of the inability of the discriminant analysis procedure to properly classify the catchment but could also, however, reflect the fact that the catchment is somewhat similar to catchments from several clusters and does not properly fit entirely in any one cluster. This predicament, which may be referred to as the border effect, invariably occurs when a continuous data space is divided into discrete segments. A catchment of this type will benefit from an expanded pool of catchments such that the n nearest catchments could well comprise a set of catchments coming from more than one cluster. The estimation of the rainfall-runoff model parameters for this option is as described in Equation (9).

Option 4

This option adopts a regression based approach in which separate regression relationships between catchment characteristics and the rainfall-runoff model parameters are developed for each of the clusters. The variables included in the regression relationships are those used in the regression equations developed in FSSR16, and of course the case of a single cluster is essentially equivalent to the regressions of FSSR16. For an ungauged catchment, separate cluster specific estimates of the runoff parameter of interest are obtained from each of the regression relationships. The final estimate is obtained as a weighted combination of the individual estimates where the weights are the probability of membership of the catchment in each of the clusters, resulting in:

$$\hat{R}_i = \sum_{k=1}^{K} p_{ik} \hat{E}_{ik} \tag{10}$$

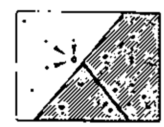
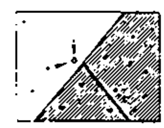
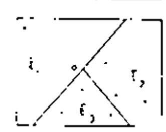
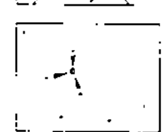
where p_{ik} is the probability of membership in cluster k for catchment i; $\hat{E}_{ik}$ is the regression estimate of the rainfall-runoff model parameter for catchment i from cluster k; and all other symbols are as previously defined.

Option 5

The final approach entails considering all catchments when determining the n nearest neighbours, regardless of the cluster membership of the catchments. The estimation of rainfall-runoff model parameters then proceeds as in Equation (9). Option 5 corresponds to Option 3 with a probability of membership threshold of zero. The purpose of this option was to evaluate the merits of the initial classification of the catchments into groups. If this option were to be the preferred approach, this would imply that there is nothing to be gained from the clustering process.

Summary and comparison of options

For all of the above options, except Option 4, the value of n is a parameter which can be varied to identify the value which provides the best estimate for the particular rainfall-runoff model parameter. Similarly, the distance threshold in Option 2 and the probability of membership threshold in Option 3 are parameters that can be modified in order to improve the estimation. Figure 1 presents a schematic comparison of the estimation options considered. None of these options exactly represents the FSR recommended procedure for using local data. However, the distance threshold in Option 2 restricts the search space to those catchments that are physically close to the catchment of interest, and Option 5 represents an unrestricted search for catchments that are similar in terms of catchment characteristics. Both of these options have parallels with the FSR recommended procedure.

Option	*Description*	*Parameters*	*Illustration*	
OPTION 1	Nearest neighbours in cluster.	n - nearest neighbours		Proximity from equation (8) Estimate from equation (9)
OPTION 2	Nearest neighbours in cluster plus distance threshold.	n - nearest neighbours d - distance threshold		As above, but catchment must be within distance d.
OPTION 3	Nearest neighbours in any cluster with membership probability greater than threshold.	n - nearest neighbours t - probability threshold		As for Option 1, but using catchments in possible clusters.
OPTION 4	Regression in each cluster. Use weighted mean of regression estimates	regression coefficients		Estimate using equation (10)
OPTION 5	Nearest neighbours without clustering.	nearest neighbours		Estimate as for Option 1

Options 1,2,3 & 5 also require weights to define nearest neighbours.

Illustrations are for 2-dimensional parameter space with 3 clusters.
- ∘ - site of interest
- · - gauged site
- ▨ - excluded region

Figure 1 ***Parameter estimation options***

3 Available Data

The data required for the approaches outlined above consisted of both measures of catchment flow response and data describing the catchment characteristics. In addition to the data needed to derive the rainfall-runoff model parameters (Tp and SPR), other measures of the flow response of a catchment that are of potential use for defining similarity in catchment response were available. The additional flow response measures considered included a seasonality index (RBAR) which describes the regularity of the timing of peak flows (RBAR can vary from 0, for a catchment where every day of the year is equally likely to correspond to the peak flow, to 1 for a catchment where the peak flow always occurs on the same day); the average daily flow (ADF) in m^3s^{-1}; a low flow measure (Q95) which is the flow value which is exceeded 95% of the time, expressed as a percentage of the average daily flow; a base flow index (BFI) which expresses the base flow component of runoff as a fraction of the total runoff; the mean value of the annual flood series, normalized by dividing by the drainage area, giving QBAR in $m^3s^{-1}km^{-2}$; and the coefficient of variation of the annual flood series (CV) which is dimensionless.

The flow response measures outlined above have been estimated from the available streamflow record for each catchment. The reliability, or representativeness, of the estimates is a function of both the amount of streamflow data available (i.e., the length of the data record) and the quality of the gauging station. To be included in the data base used herein, the data record at a catchment had to contain at least ten years of annual flow data to provide reasonable estimates for the mean and the coefficient of variation of the annual flood series. For the estimation of the seasonality index, a minimum of five years of peaks over threshold data were required and data from at least five storm events were required for the estimation of the rainfall-runoff model parameters. Finally, each gauging station has a quality rating associated with the estimation of the low flow regime. A gauging station for a catchment had to be rated as "good" or "very good" for the catchment to be included. The total number of catchments with data available for all of the above-noted flow response measures was quite small. However, if Q95 and ADF were to be excluded, a total of 99 catchments, satisfying the above-noted quality criteria, would have data for all of the remaining variables. Since Q95 is correlated with BFI (ϱ = .69) and ADF divided by the catchment drainage area is correlated with QBAR (ϱ = .89), it was felt that excluding these two measures of flow response would be justified. A listing of this data, which includes the catchment identification numbers and the flow response measures, is presented in Appendix A.

Prior to clustering the catchments, the time to peak, Tp, was transformed to define a new variable, LTp, where:

$$LTp = \log(Tp) \tag{11}$$

This new variable was used in the clustering process in place of time to peak due to the substantially greater skewness of the time to peak relative to the remaining flow response measures. This transformation of variable thus results in greater uniformity in the skewness of the flow response measures used as attributes in the cluster analysis.

In addition to the flow response measures, data describing the catchment characteristics of the catchments were also required since estimates of the various flow response measures are generally not available at catchments that are ungauged. The available data base for each catchment included the drainage area (AREA) in km^2; the main stream length (MSL) in km; the dry valley factor (DVF) which is the length of dry valley from the catchment divide to the head of the mainstream divided by the distance from the catchment outlet to the divide; a measure of channel slope (SL1085) calculated as the difference in elevation, in metres, between two points along the main channel corresponding to 10% and 85% of channel length upstream of the catchment outlet divided by the distance (in km) along the channel between these two points; the stream frequency (STMFRQ) which is the number of stream junctions in the channel network divided by the catchment area; the standard annual average rainfall (SAAR) for the period from 1941 to 1970 in mm/year; the fraction of the catchment draining through lakes (LAKE); the fraction of the catchment that is urbanised (URBAN); and soil classification variables (SOIL1, SOIL2, SOIL3, SOIL4, SOIL5) which give the fraction of the catchment corresponding to each of five soil classes. The soil classes are based on the winter rainfall acceptance potential (WRAP) of the soil. An additional variable, SOIL, is defined as a linear combination of the five individual soil variables. A listing of the values for all of the catchment characteristic data for the catchments used in this study is contained in Appendix A.

Table 1 contains the mean and standard deviation as well as the minimum and maximum values for all of the variables included in the data set. As can be inferred from this table, the catchments used in this study were predominantly of a moderate size (i.e., maximum drainage area of 544 km^2) from largely non-urbanised areas. Many of the physiographic features exhibit a substantial range in value (e.g., slopes that range from less than 1 to in excess of 60 m/km) implying that the data set includes catchments with diverse characteristics.

For all clustering options considered, the set of variables used as catchment attributes were standardized prior to clustering. The standardization was accomplished by dividing each variable by its standard deviation where the standard deviation of a variable is calculated from the entire set of catchments. The purpose of standardization is to account for differing amounts of variability in the various catchment attributes used and to remove the effects of the arbitrary selection of the units of measure for a catchment attribute.

Table 1 Summary statistics for catchment data

Variable	Units	Mean	Standard Deviation	Minimum	Maximum
Flow Response Measures					
Tp	hours	9.19	6.77	2.18	42.8
Qp	$m^3s^{-1}100km^{-2}$	27.6	13.9	6.50	68.7
SPR	%	38.5	13.4	7.49	74.3
RBAR		0.52	0.12	0.19	0.82
BFI		0.49	0.14	0.15	0.86
QBAR	$m^3s^{-1}km^{-2}$	0.44	0.39	0.02	1.95
CV		0.42	0.17	0.15	1.16
Catchment Characteristics					
AREA	km^2	165.	135.	8.0	544.
MSL	km	25.0	15.6	3.90	84.6
DVF		0.05	0.08	0.00	0.53
SL1085	m/km	9.30	10.8	0.92	63.7
STMFRQ	km^{-2}	1.34	1 12	0.10	6.28
SAAR	mm	1155	505	559	3030
LAKE		.025	.055	.000	.260
URBAN		.052	.112	.000	.810
SOIL1		115	.211	.000	1.00
SOIL2		.171	.294	.000	1.00
SOIL3		.148	272	.000	1.00
SOIL4		296	354	.000	1.00
SOIL5		.270	.360	.000	1.00
SOIL		.396	.079	.150	.500

4 Analysis

4.1 GROUPING OF CATCHMENTS

4.1.1 Application of K-means algorithm

As noted in Section 2.1.1, it is possible to apply the K-means algorithm in a number of different ways. The strategy ultimately selected was clustering using the principal components approach. Principal components were calculated from the data set comprising all seven flow response measures. The clustering variables then corresponded to all principal components with an eigenvalue greater than unity, which resulted in the retention of three principal components. The weighting on each of the principal components for each flow response measure is shown in Table 2. The values for the three principal components for each of the catchments are presented in Appendix A. The variables with the largest weight on the first principal component are LTp, Qp, and QBAR which implies that this principal component represents a measure of the magnitude of the flood response. The variables associated with the second principal component are SPR and BFI implying that this component can be viewed as a measure of runoff production. The third principal component has the variables RBAR and CV associated with it suggesting that this principal component reflects the annual variability in flood response, both in terms of the timing (RBAR) and the magnitude (CV) of the response.

Each of the three significant principal components identified were weighted by the associated eigenvalue resulting in weightings of 3.13, 1.44, and 1.05, respectively. The weighted principal components clustering option was selected since the eigenvalues, which indicate the amount of the variability of the data explained by the associated principal component, provide a useful way of quantifying the relative importance of each of the principal components. There appeared to be no intrinsic advantage to the two rotated principal components options considered so these were not pursued in depth.

The number of clusters delineated using the three weighted principal components was varied from two to five. After examining the results of these partitionings, it was decided to focus on the results for the two and three groups cases. The pertinent results are summarized below.

4.1.2 Resulting classifications

Two Clusters

Table 3 provides a summary of the characteristics of the groups identified for the two cluster case. The number of catchments in each group is roughly the same and each group contributes essentially the same amount to the total within group variability of the data. The variability for a cluster is calculated as:

$$\delta_k = \sum_{i \in I_k} \sum_{m=1}^{M} W_m (X^i_m - C^k_m)^2 \tag{12}$$

where δ_k is the variability measure for cluster k, and all other symbols are as previously defined. The variability contribution is shown in Table 3 as a fraction and is calculated as δ_k divided by the total within group variability, which is given by:

$$\Delta = \sum_{k=1}^{K} \delta_k \tag{13}$$

where Δ is the total within group variability. The cluster membership for each catchment is given in Appendix A.

An examination of the geographic location of the catchments reveals that group 1 consists of catchments primarily from the western portion of the UK while group 2 contains catchments from the eastern and central areas (see Figure 2). It is apparent, however, that geographic location is not a strong distinguishing feature of the clusters. Figures 3 to 5 contain scatter plots of the first three principal components for the catchments. From these scatter plots it can be seen that the first principal component strongly distinguishes between the two groups with group 1 associated with high values on this component and group 2 associated with low values. However, the second and third principal components do not differentiate between the catchments in the two clusters. This latter result can also be inferred from the centroid coordinates presented in Table 3.

In terms of the flow response measures, group 1 is characterised by high values for Qp, SPR, and QBAR and low values for LTp and BFI. Catchments in this group thus tend to exhibit a rapid, flashy, runoff response with a high runoff production. Group 2 catchments have high values for LTp and BFI and low values for Qp, SPR, and QBAR. These catchments therefore tend to exhibit a slow and sustained runoff response with comparatively low runoff production. Figures 6 to 9 contain scatter plots of the flow response variables for the catchments. The drainage area for each catchment, although not a variable used in the clustering process, is also included to ascertain if there are any scaling factors, related to catchment size, associated with the partitioning of the catchments. Figure 6 reveals that the catchments are well separated in terms of LTp and SPR. A similar result for Qp and BFI is noted from Figure 7. Figure 8 reveals that CV is not a distinguishing variable for the clusters but there are patterns in terms of QBAR. Figure 9 implies that neither RBAR nor AREA is capable of distinguishing between the clusters.

Three Clusters

Table 4 summarizes the characteristics of the groups for the case of three clusters. From this table, it can be seen that there are roughly the same number of catchments in each of the three groups. The variability contribution of the three groups is not, however, as evenly dispersed as for the case of two clusters. The somewhat larger fraction of the variability contributed by group 1 implies that the catchments in this group are not as cohesive, or tightly grouped, as the catchments in the other two groups. The cluster membership for each catchment is given in Appendix A.

Figure 10 reveals that the catchments in group 1 are primarily from the west of the UK, group 2 contains catchments that are not located in the north, while group 3 has a slight southern bias but is basically quite evenly dispersed. As for the case of two clusters, the geographic location of a catchment is not a strong distinguishing feature of the clusters. Figures 11 to 13 contain scatter plots of the three principal components for the catchments from which it can be seen that group 1 catchments have high values for principal component 1, group 2 catchments have low values for both of the first two principal components, and group 3 has low values for principal component 1 and high values for principal component 2. The groups are well separated in terms of the first two principal components but the third principal component does not differentiate between members of the various groups. This latter result, also apparent from the centroid coordinates in Table 4, undoubtedly follows from the comparatively low magnitude of the eigenvalue associated with the third principal component. A value of only slightly above one for the eigenvalue implies that this component is not as important as the other two components for explaining the variance of the original variables. The fact that this component does not substantially differentiate between the clusters is therefore consistent with the characteristics of the data.

In terms of the flow response measures, group 1 contains catchments with high values for Qp and QBAR and low values for LTp. The catchments in this group therefore have a flashy runoff response that produces comparatively large floods. Group 2 consists of catchments with high values for LTp and BFI and low values for Qp, SPR, and QBAR. These catchments thus have a slow and sustained runoff response with comparatively low runoff production. Group 3 catchments have high values for LTp and SPR and low values for Qp, QBAR, and BFI. The catchments in this group also have a slow and sustained runoff response but have a comparatively high runoff production. The three groups are well defined in terms of LTp and SPR (see Figure 14) and also in terms of Qp and BFI (see Figure 15). QBAR is able to distinguish between group 1 and the other two groups (see Figure 16) but there is no discernable pattern in terms of CV (Figure 16) or RBAR (Figure 17). As with the case of two clusters, there are no noticeable catchment size patterns.

4.2 ASSIGNMENT OF UNGAUGED CATCHMENTS

4.2.1 Selection of variables for discriminant analysis

The first stage in the catchment assignment process involves the identification of relevant catchment characteristics for use as discriminating variables. Step-wise discriminant analysis was first applied to the entire set of variables describing catchment characteristics to determine a preferred set of variables for each of the partitionings of the catchments considered. For the case of two clusters, the variables which exhibited a capability for discriminating between clusters were, in order of importance, SOIL5, SAAR, SOIL, STMFRQ, SOIL1, and SL1085. Since several of these variables are correlated, the variables ultimately selected by the step-wise discriminant procedure were SOIL5, SOIL1, URBAN, SAAR, and SOIL4, where the variables are listed in the order in which they entered as discriminating variables. It is noteworthy that the variable URBAN is selected even though this variable was not one of the variables identified as having a discriminating capability for the entire data set. This variable is entered because it is capable of discriminating between the clusters after other variables are included in the discriminant function.

For the case of three clusters, the initial variables, in order of importance, were SOIL5, SOIL, SAAR, SL1085, STMFRQ, SOIL4, and SOIL1. The variables ultimately selected by the step-wise discriminant analysis procedure were SOIL5, SOIL, URBAN, and SL1085. Although there were more variables that exhibited a capability for discriminating between the clusters for this case as opposed to the two cluster case, there were fewer variables ultimately selected.

As indicated earlier, the step-wise discriminant analysis procedure assumes that the variables considered follow the multivariate Gaussian distribution. Since this distributional assumption is suspect at best, the variables identified by the step-wise discriminant analysis process were merely used as a starting point in the variable selection procedure. For example, Figure 18 shows a scatter plot of SOIL5 and SOIL for the three cluster data set. It is clear from Figure 18 that these two variables alone are unlikely to provide satisfactory results within the catchment assignment process, although some patterns do exist in the data. Note that SOIL5 and SOIL were the first two variables selected by the step-wise discriminant analysis procedure. Figure 19 shows a scatter plot of SOIL5 and SOIL1 for the two cluster data set. This plot reveals a more promising configuration in that the catchments from group 1 tend to have a low fraction of soil type 1 whereas catchments from group 2 tend to have a low fraction of soil type 5.

Canonical discriminant analysis was next used to identify a set of canonical variables that are capable of discriminating between the clusters. Tables 5 and 6 present the weighting matrices for the canonical variables for the case of two and three clusters, respectively. Note that one canonical variable can be calculated for the two cluster case while two canonical variables can be calculated for the case of three clusters. The canonical variables calculated using the weighting coefficients from Tables 5 and 6 and the relationship given in Equation (2) were considered as potential discriminating variables. The values for the canonical variables for each catchment are presented in Appendix A.

Since both of the above approaches to identifying variables for use in the catchment assignment process involve a distributional assumption, the variables suggested by the above procedures were subjected to further scrutiny before a final set of variables was selected. Scatter plots of the type shown in Figures 18 and 19 were examined for different combinations of pairs of variables. From plots of this type, it was possible to identify a limited number of promising combinations of variables which were then tested in the nearest neighbour discriminant analysis procedure. From this process, the option involving the use of canonical variables was selected for both the two and the three cluster case. Figure 20 shows a scatter plot of the first and second canonical variables for the three cluster case. This plot reveals that the two canonical variables separate the catchments, as to cluster membership, reasonably well. The first canonical variable tends to separate cluster 1 from the remaining two clusters while the second variable distinguishes between members of cluster 2 and cluster 3. The single canonical variable for the two cluster case was also observed to provide a satisfactory separation of the catchments.

4.2.2 Results of assignment process

The nearest neighbour discriminant analysis procedure was then applied with the canonical variables used as the discriminating variables. Table 7 summarizes the performance of the discriminant analysis procedure for the two cluster case and Table 8 gives the corresponding results for the case of three clusters. From Table 7 it is seen that the procedure has correctly classified 91 out of 99 catchments with 5 catchments from cluster 1 and 3 from cluster 2 being incorrectly classified. The results in Table 7 are based on catchment assignment using the nearest five neighbours. The results in Table 8 indicate a considerably lower overall success rate, with 76 out of 99 catchments correctly classified. The nearest three neighbours were used to obtain these results. It is to be expected that as the number of clusters is increased, (e.g., with a transition from two to three groups) the number of catchments correctly classified will decrease. However, increasing the number of clusters will also result in greater homogeneity of the catchments in each group. A trade-off clearly exists between the capability to correctly classify a catchment and the similarity of the collection of catchments in the group to which the catchment is assigned.

The classification performance, as summarized in Tables 7 and 8, does not constitute a completely satisfactory evaluation of the effectiveness of the catchment assignment process. The results presented are based on a comparison of the actual cluster membership of a catchment with the cluster to which a catchment has been assigned based on the largest probability of membership. There is thus no distinction made between the case where a catchment is unambiguously assigned to a cluster (i.e., a probability of membership of unity) and the case where the probability of membership is slightly larger for one cluster than it is for any of the other clusters. In either case, a success or failure is determined by the agreement or disagreement between the actual cluster and the assigned cluster for the catchment. An incorrect assignment could result when the probability of membership for an incorrect cluster is slightly higher than the probability of membership for the true cluster. Such a situation would not necessarily be a cause for concern, particularly if the catchment was on the border between the two clusters in question and its actual cluster membership was therefore somewhat ambiguous. Conversely, if a cluster has a probability of membership of near zero for its true cluster then this is clearly an unsatisfactory classification performance. Two of the parameter estimation options (Options 3 and 4) incorporate the probability of membership in the various clusters in an attempt to account for the former type of behaviour.

4.3 ESTIMATION OF RAINFALL-RUNOFF MODEL PARAMETERS

4.3.1 A baseline for comparison

Estimates of rainfall-runoff model parameters obtained using the procedure developed herein were compared with estimates obtained from the regression equations given in FSSR16. For the time to peak, the regression estimate is from:

$$Tp_r = 283\,SL1085^{-.33}(1+URBAN)^{-2.2}SAAR^{-.54}MSL^{.23}+0.5 \tag{14}$$

where Tp_r is the regression estimate for the time to peak of the one hour unit hydrograph; and all other terms are as previously defined. The regression estimate for the standard percentage runoff is from:

$$SPR_r = 10SOIL1+30SOIL2+37SOIL3+47SOIL4+53SOIL5 \tag{15}$$

where SPR_r is the regression estimate for the standard percentage runoff; and all other variables are as previously defined. Parameters for regression equations of the same form as Equations (14) and (15) were estimated using only the data from this study and were found to be in good agreement with the results presented above. The coefficients obtained from FSSR16 were used in preference to the coefficients obtained from our (smaller) data set since the intent is to develop an estimation system which can be applied to ungauged catchments in the UK. As such, the regression relationships developed from the larger data set available for the FSSR16 study represent the appropriate basis for comparison.

To evaluate the relative merit of the estimation framework developed herein, each of the catchments was sequentially considered to be ungauged. The catchment presumed to be ungauged was assigned to a cluster based on the classification process described above. Estimates of the runoff parameters were then obtained using the appropriate regression equation and the parameter estimation options developed herein. For the latter approach, the pool of available catchments consisted of all other catchments in the data base. This process was then repeated with each of the remaining catchments individually considered to be ungauged. It was thus possible to compare estimates of rainfall-runoff model parameters from catchments for which the true values could be assumed known. Values for three performance measures were calculated for each of the estimation options developed herein as well as for the FSR regression approach. The performance of the estimators was evaluated in terms of measures of the bias, the imprecision, and the worst performance for the estimator. The bias measure is calculated as:

$$BIAS = \frac{1}{N}\sum_{i=1}^{N}\frac{\hat{R}_i - R_i}{R_i} \tag{16}$$

where BIAS is the normalized bias; N is the number of catchments in the data base; $\hat{R}_i$ is the estimate for a rainfall-runoff model parameter (i.e., either Tp or SPR) for catchment i; and R_i is the actual value for the rainfall-runoff model parameter for catchment i. The imprecision is calculated as:

$$RMSE = \left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{R}_i - R_i}{R_i}\right)^2\right]^{\frac{1}{2}} \tag{17}$$

where RMSE is the normalized root mean squared error, which is used as a measure of imprecision. The measure of the worst performance is calculated as:

$$D_{max} = \max_{i} \left[\frac{\hat{R}_i - R_i}{R_i} \right] \tag{18}$$

where D_{max} is the normalized measure of the worst estimate for a rainfall-runoff model parameter, for a given estimator.

4.3.2 Determination of catchment proximity

The nearest neighbour based estimation options (Options 1, 2, 3, and 5) require a measure of catchment proximity that is based on a subset of the variables in the catchment characteristic data base. From an examination of the correlation structure of differences in variable values for pairs of catchments, a reduced set of potential variables was identified for each of the rainfall-runoff model parameters of interest. Further examination of the agreement between catchment pair similarity, as measured by catchment characteristic variables, with catchment similarity in terms of rainfall-runoff model parameters, resulted in the selection of a final set of variables for defining catchment proximity for each rainfall-runoff model parameter. The variables selected for time to peak were SL1085, SAAR, and MSL. For the standard percentage runoff, the variables selected were the five soil class variables, SOIL1, SOIL2, SOIL3, SOIL4, and SOIL5. It is interesting to note that the variables ultimately selected are all included in the FSSR16 regression equations, although URBAN, which is included in the Tp regression equation, was not found to be beneficial for determining catchment proximity for Tp estimation.

4.3.3 Assessment of estimation options

Two clusters

Table 9 presents a summary of the estimation performance for the two cluster case for Tp and SPR estimated using the five options developed herein and the FSSR16 regression equation. The cluster specific regression equations for Tp, included in option 4, are given as:

$$Tp_{r1} = 95 SL1085^{-.24}(1+URBAN)^{-1.6} SAAR^{-.36} MSL^{.13} + 0.5 \tag{19}$$

where Tp_{r1} is the regression estimate for Tp from cluster 1, and:

$$Tp_{r2} = 155 SL1085^{-.34}(1+URBAN)^{-1.3} SAAR^{-.44} MSL^{.24} + 0.5 \tag{20}$$

where Tp_{r2} is the regression estimate for Tp from cluster 2. The two nearest neighbours, based on weighted Euclidean distance (Equation 8), form the basis for the estimation of SPR with options 1 to 3 while the five nearest neighbours are used for option 5. The number of nearest neighbours selected for each case corresponded to the number which gave the best estimation performance. A simple search routine was used to select the weights for the five soil class variables (0.50, 0.95, 0.90, 1.75, and 1.25, respectively). The preferred weights were found for the option which gave the best performance with equal weights. The results in Table 9 then present the performance of each option when this same set of weights was used for each option, except for option 4. The values of the weights reflect the relative importance within the clusters of differences in each of the soil class variables for determining catchments with similar SPR values.

The estimation of Tp using option 4 gives a small negative bias but also results in lower RMSE and lower maximum error values than for the FSSR16 regression estimate. The other options examined for estimating Tp did not perform as well as option 4. It should be noted that the distance threshold for option 2 was forced to be sufficiently small so that the catchments considered for information transfer could be deemed to be in close physical proximity. This explains why option 2 results can be worse than the results for option 1. Without imposing a constraint on the distance threshold, option 2 could, of course, be equivalent to option 1. For SPR, estimation using option 1 results in substantially lower bias, relative to the FSR approach, as well as improved performance in terms of both RMSE and D_{max}. Option 3 resulted in estimates nearly as good as those for option 1 while the remaining options were inferior. Option 5, which is the no clustering option, was noted to be a particularly poor choice which implies that the initial clustering of catchments is meritorious. Option 4, involving cluster specific regression relationships, was not implemented for the estimation of SPR since the individual regression relationships obtained were not physically realistic.

Three clusters

Table 10 summarizes the performance of the estimators for the case of three clusters. The cluster specific regression equations, required for option 4, are given as:

$$Tp_{r1} = 32 SL1085^{-.11}(1+URBAN)^{-1.0} SAAR^{-.33} MSL^{.24} + 0.5 \qquad (21)$$

for cluster 1, and:

$$Tp_{r2} = 122 SL1085^{-.25}(1+URBAN)^{-1.6} SAAR^{-.43} MSL^{.25} + 0.5 \qquad (22)$$

for cluster 2, and:

$$Tp_{r3} = 570 SL1085^{-.53}(1+URBAN)^{-2.4} SAAR^{-.58} MSL^{.21} + 0.5 \qquad (23)$$

for cluster 3. The preferred values for the weights applied to the variables in the Euclidean distance measure were found to be 0.85, 1.20, 0.95, 1.70, and 1.20, respectively, for the five soil class variables. The weights for the three cluster case, although similar to those for the two cluster case, do differ from the earlier results reflecting differences in the range of catchment characteristics for the catchments in each cluster. The estimation of Tp with option 4 results in a slight negative bias and a lower RMSE than the FSSR16 regression estimate. As with the results for two clusters, the other options examined tended to perform worse than option 4. For the estimation of SPR, option 1 results in improved performance in terms of BIAS, RMSE, and D_{max}, relative to the FSSR16 alternative. The remaining options resulted in estimates that were worse than the results from option 1, with Option 5 again noted to be a particularly poor choice.

4.3.4 The preferred approach

The preferred approach for estimating SPR for an ungauged catchment is to use option 1 and two clusters. The estimates obtained using this option are better, in terms of all three performance measures, than results from the FSSR16 approach and the best of the results from using three clusters. For the estimation of Tp, the selection of a preferred approach is more ambiguous since no single estimator is best for all three performance measures. In fact, the best value of each of the performance measures occurs for a different estimator. The FSSR16 approach results in the lowest bias with the three cluster option 4 providing the next best result. The RMSE values are essentially equivalent for option 4 with two and three clusters. Both of these alternatives provide better RMSE performance than is obtained with the FSSR16 approach. In terms of the maximum error criteria, the two cluster option 4 results are somewhat better than both of the other alternatives, although the difference is not large. The selection of a preferred estimation approach for Tp could entail assigning a relative importance to the three evaluation criteria so that the performance measure values are combined into a single number. This process would, however, require subjective inputs that may not be easy to obtain. A perusal of the results indicates that one would be likely to select either option 4 with three clusters or the FSSR16 approach, with perhaps preference for the former. However, considering the ease of implementing each alternative, the FSSR16 regression approach is best retained.

Table 2 Weightings of the variables on the three principal components

Variable	Principal Component 1	Principal Component 2	Principal Component 3
LTp	-.455	0.425	0.035
Qp	0.483	-.367	-.008
SPR	0.306	0.642	-.004
RBAR	-.202	-.016	0.673
BFI	-.381	-.509	0.107
QBAR	0.500	-.074	0.155
CV	-.169	-.086	-.714

Table 3 Summary of characteristics for two clusters

Cluster Number	Number of Members	Variability Contribution	Centroid Coordinates		
			PC 1	PC 2	PC 3
1	51	.520	-.20	0.37	1.73
2	48	.480	-1.78	0.25	1.59

Table 4 *Summary of characteristics for three clusters*

Cluster Number	Number of Members	Variability Contribution	Centroid Coordinates		
			PC 1	PC 2	PC 3
1	34	.412	0.09	0.07	1.62
	29	.285	1.87	-.44	1.81
3	36	.302	-1.24	1.13	1.58

Table 5 *Weights on the catchment characteristics for the canonical variable for two clusters*

Catchment Characteristic	Canonical Variable
	1
MSL	-.1922
DVF	.0774
SL1085	-.2422
STMFRQ	-.1292
SAAR	0.9152
LAKE	-.0761
URBAN	.3104
SOIL1	-44.226
SOIL2	-105.60
SOIL3	-125.19
SOIL4	-179.87
SOIL5	-200.60
SOIL	79.241

Table 6 *Weights on the catchment characteristics for the canonical variables for three clusters*

Catchment Characteristic	Canonical Variable	
	1	2
MSL	-.0697	-.4442
DVF	.1237	-.2576
SL1085	.2550	.1507
STMFRQ	.1389	-.3938
SAAR	.3066	-.1749
LAKE	-.1500	.0253
URBAN	.4058	.1661
SOIL1	-8.9814	11.265
SOIL2	.3231	53.656
SOIL3	7.947	72.926
SOIL4	15.365	109.66
SOIL5	21.668	128.01
SOIL	-22.460	-68.866

Table 7 *Results of catchment assignment for two clusters*

Cluster	Number of Members	Number of Members Classified into Clusters		Percent Correctly Classified
		1	2	
1	51	46	5	90.2
2	48	3	45	93.8
Total	99	49	50	

Table 8 *Results of catchment assignment for three clusters*

Cluster	Number of Members	Number of Members Classified into Clusters			Percent Correctly Classified
		1	2	3	
1	34	25	3	6	73.5
	29		24		82.8
3	36	3	6	27	75.0
Total	99	32	33	34	

Table 9 *Runoff parameter estimation performance for two cluster case*

Runoff Event Parameter	Option	n	BIAS	RMSE	D_{max}
Tp	1	6	.156	.443	1.923
			.171	.516	2.376
			.153	.446	1.923
			-.044	.277	.645
			.146	.554	2.447
	FSSR16		.000	.284	.787
SPR	1	2	.040	.340	1.277
			.085	.399	1.623
			.052	.344	1.277
			108	.433	1.799
	FSSR16		120	.390	1.703

† This option was not pursued for SPR.

Table 10 ***Runoff parameter estimation performance for three cluster case***

Runoff Event Parameter	Option	n	BIAS	RMSE	D_{max}
Tp	1	3	.169	.499	1.923
			.116	.492	2.098
			.156	.486	1.985
			-.026	.276	.787
			.146	.554	2.447
	FSSR16		.000	.284	.787
SPR	1	3	.060	.372	1.602
			.057	.392	1.681
			.078	.384	1.590
			.120	.440	1.799
	FSSR16		.120	.390	1.703

† This option was not pursued for SPR.

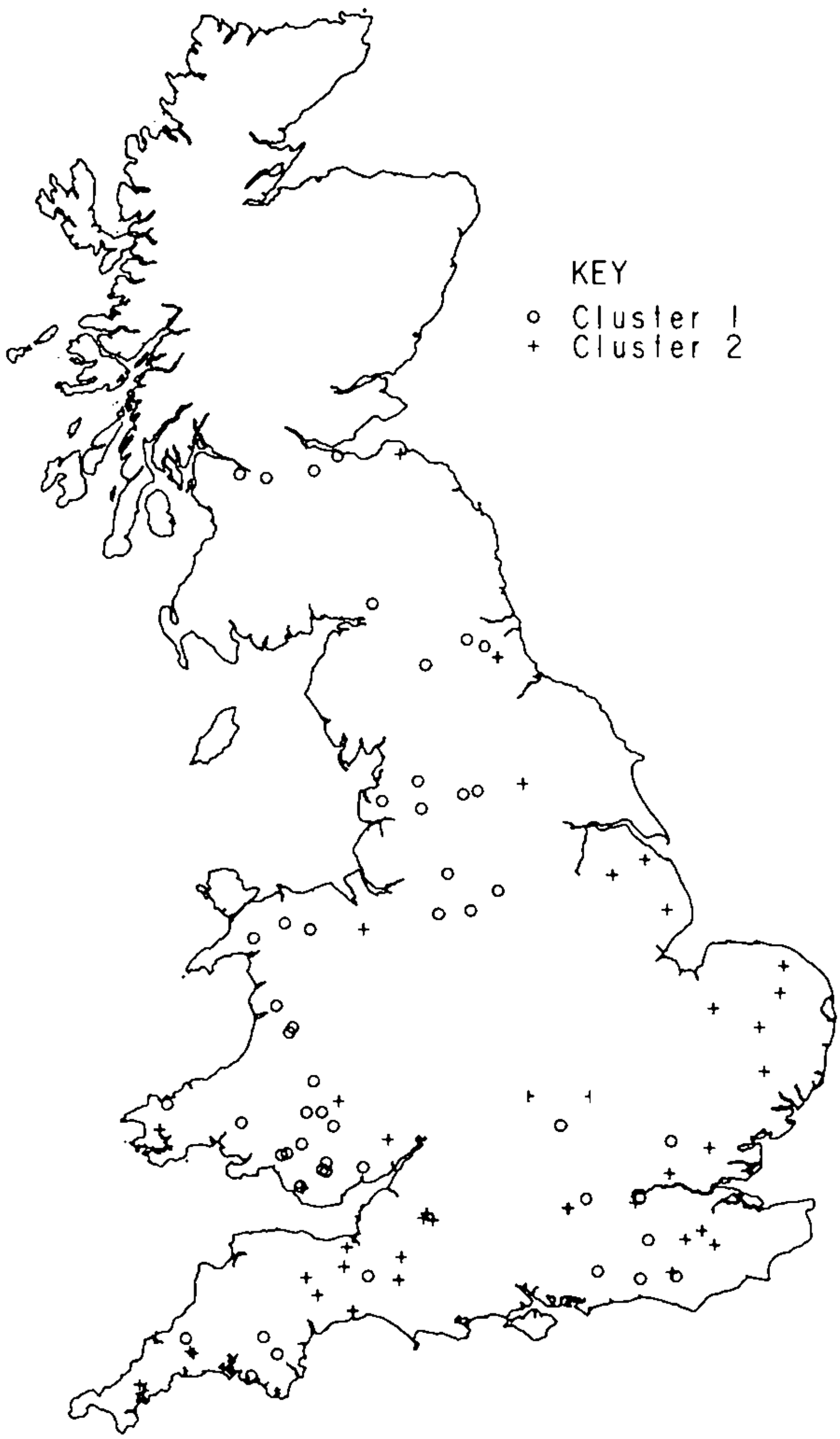


Figure 2 Two cluster case: geographical distribution of catchments

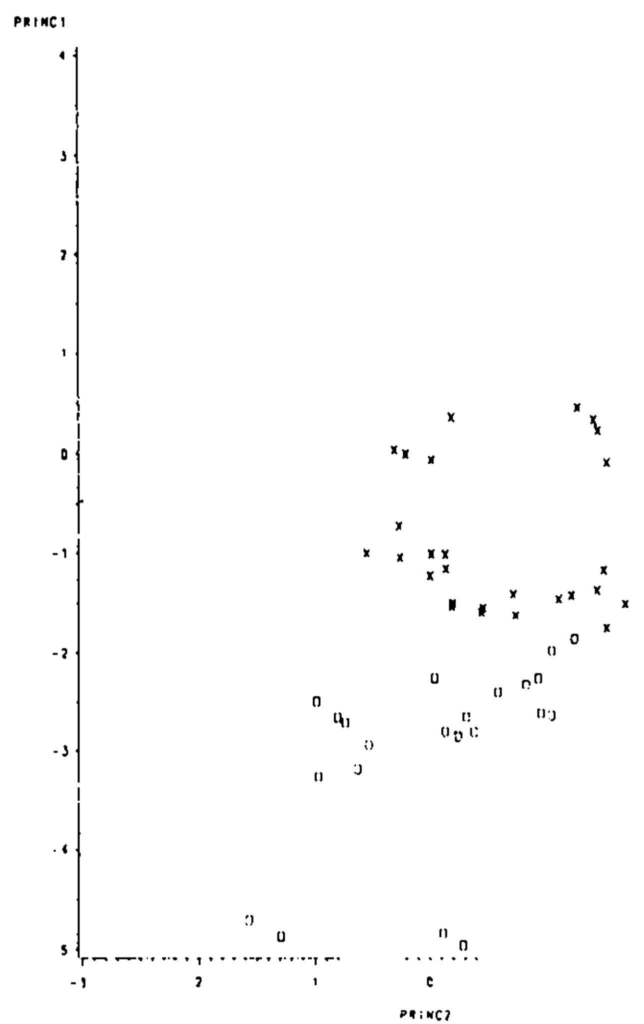


Figure 3 *Two cluster case: principal component 1 against principal component 2 with cluster membership*

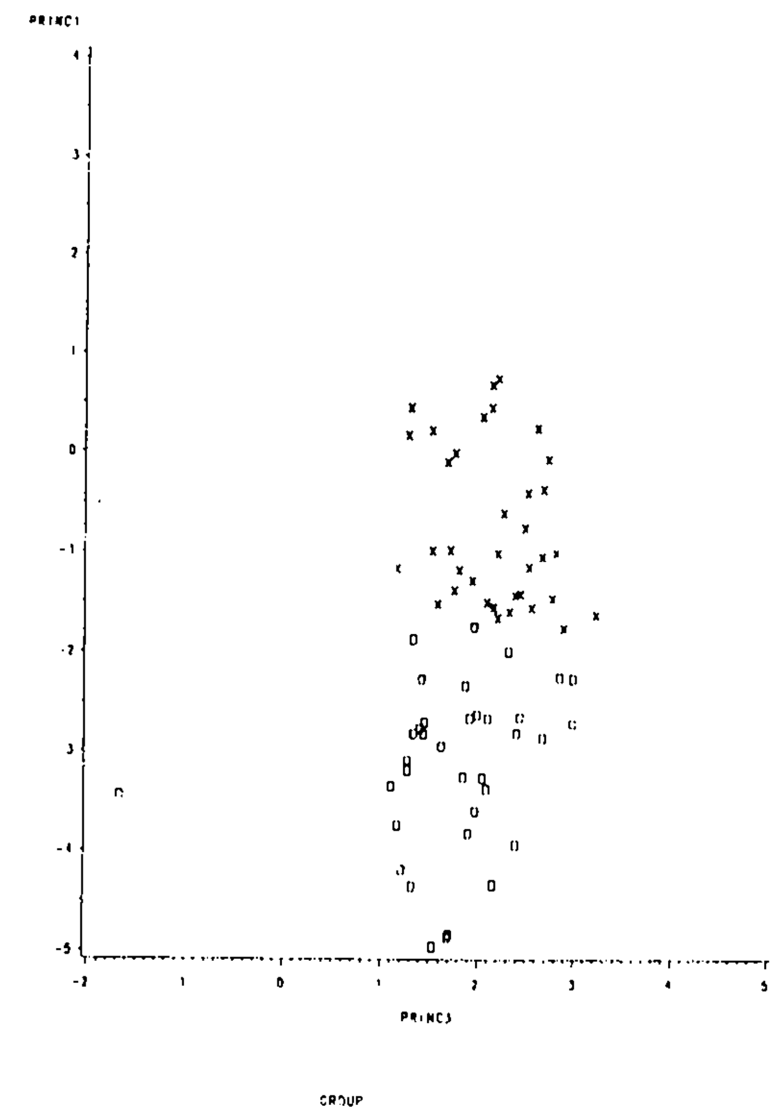


Figure 4 *Two cluster case: principal component 1 against principal component 3 with cluster membership*

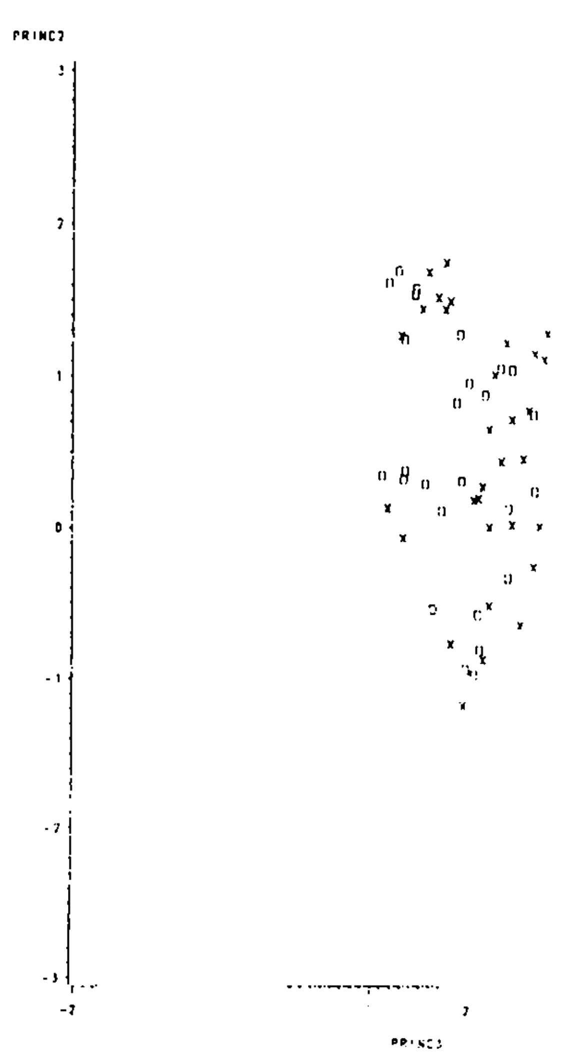


Figure 5 *Two cluster case: principal component 2 against principal component 3 with cluster membership*

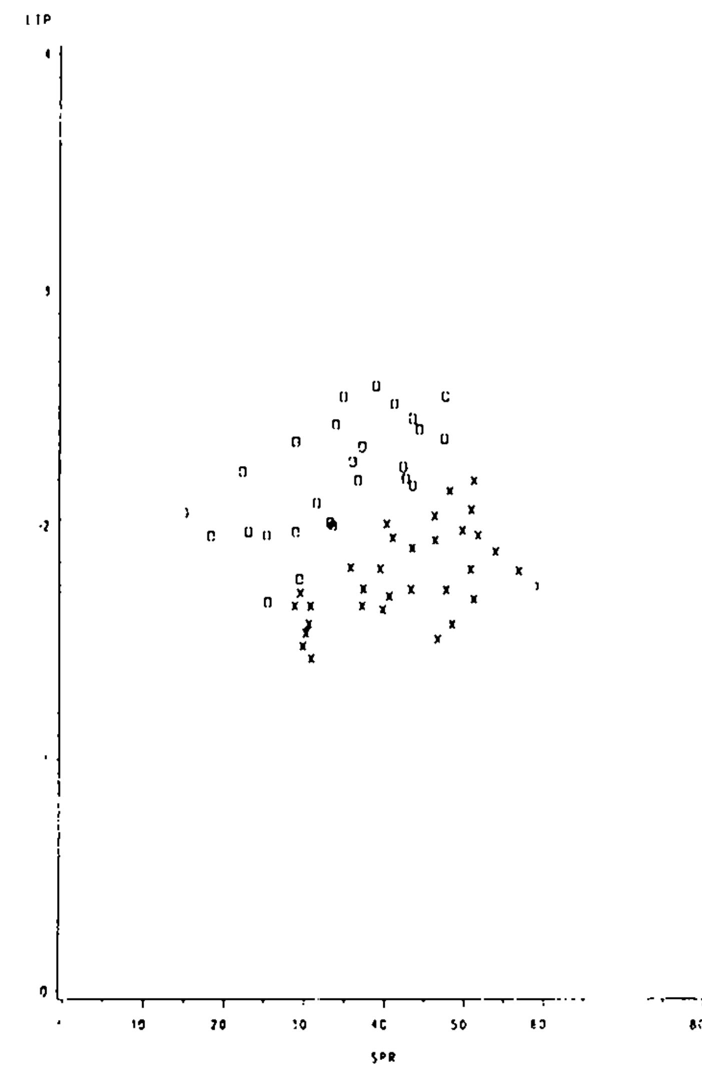


Figure 6 *Two cluster case: LTp against SPR with cluster membership*

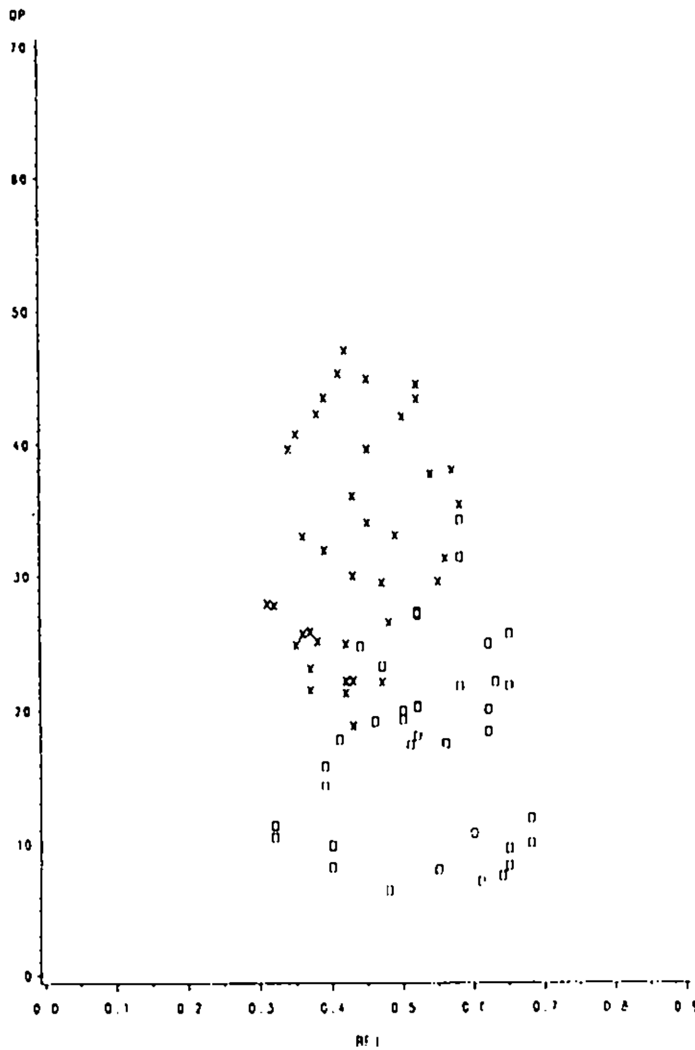


Figure 7 *Two cluster case: Qp against BFI with cluster membership*

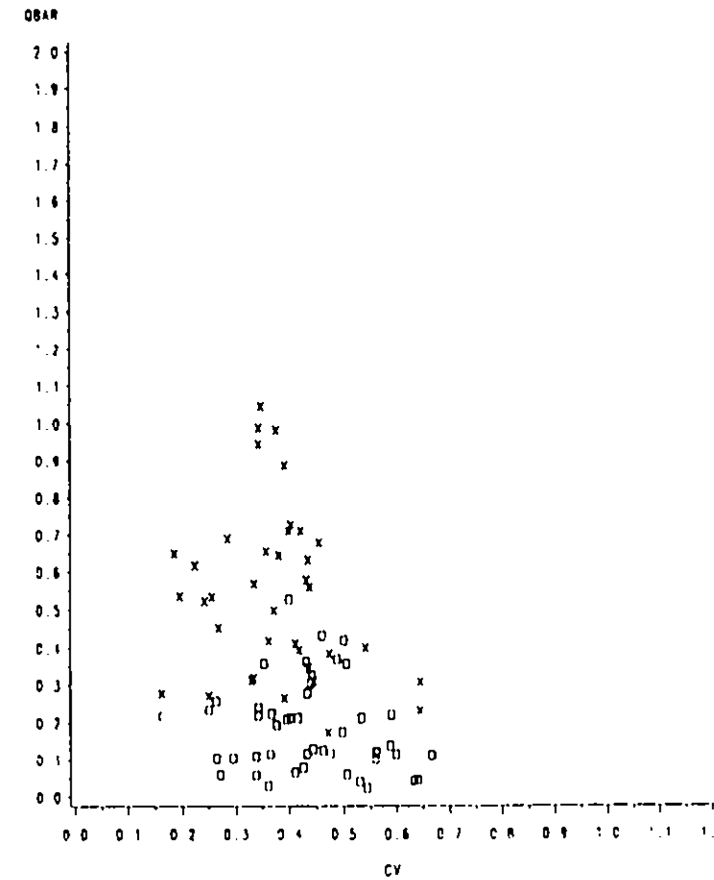


Figure 8 *Two cluster case: QBAR against CV with cluster membership*

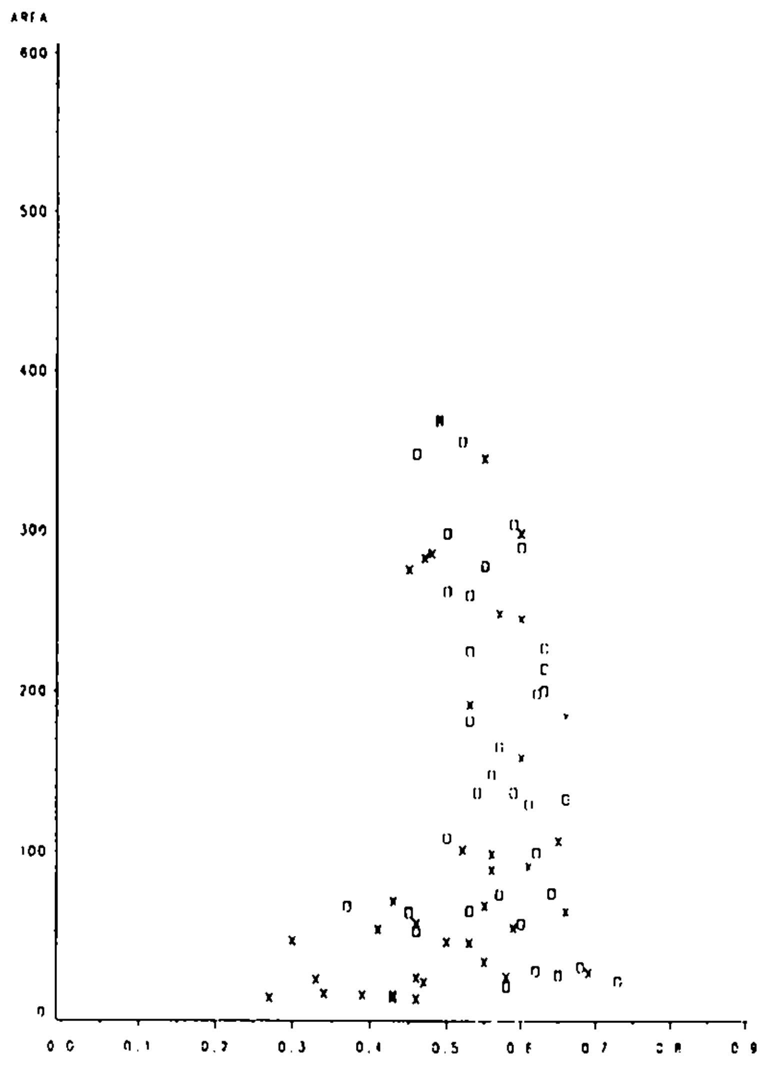


Figure 9 *Two cluster case: AREA against RBAR with cluster membership*

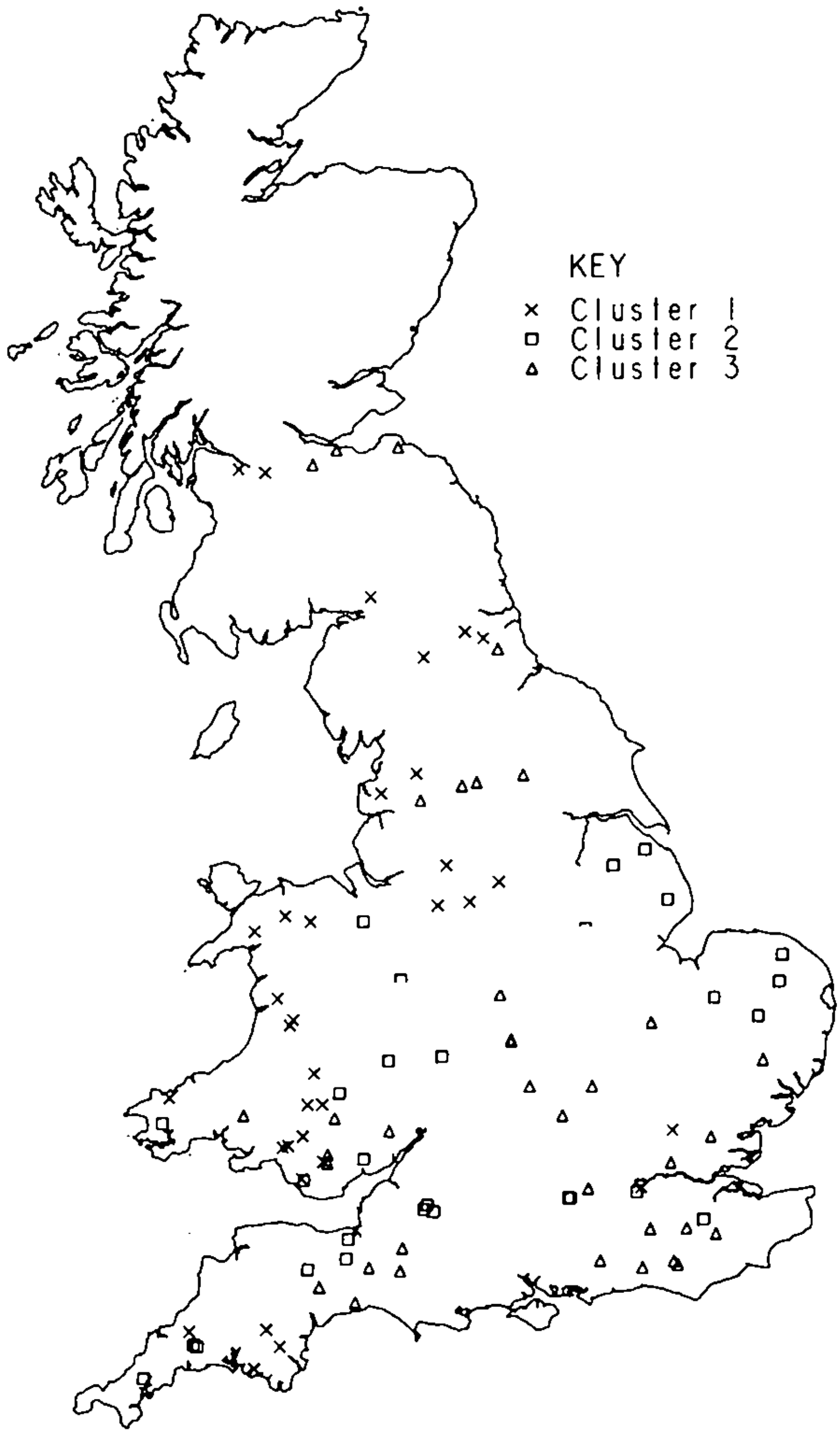


Figure 10 *Three cluster case: geographical distribution of catchments*

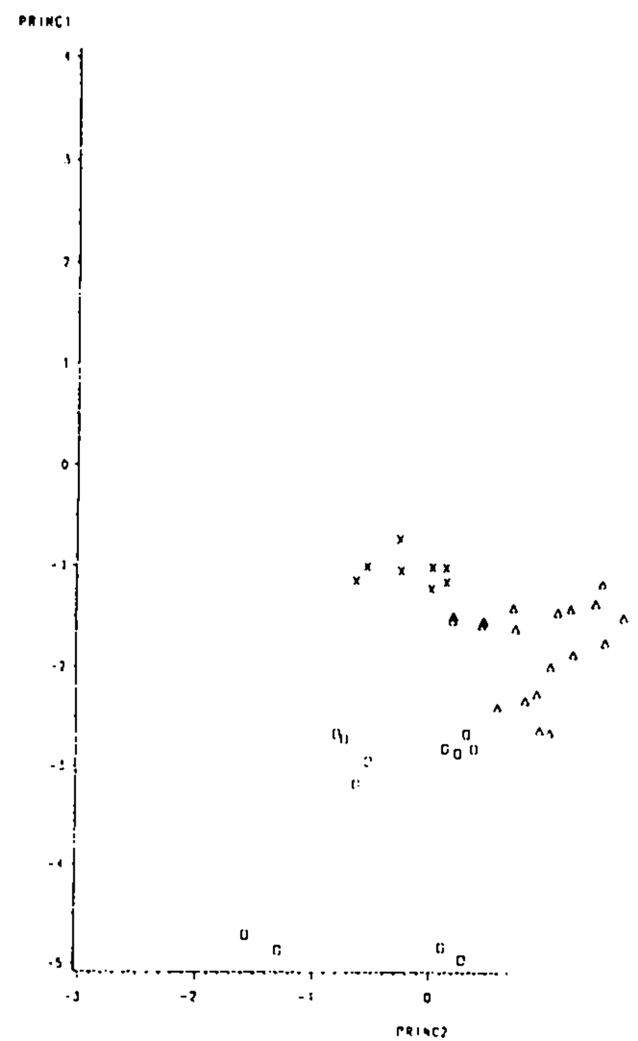


Figure 11 *Three cluster case: principal component 1 against principal component 2 with cluster membership*

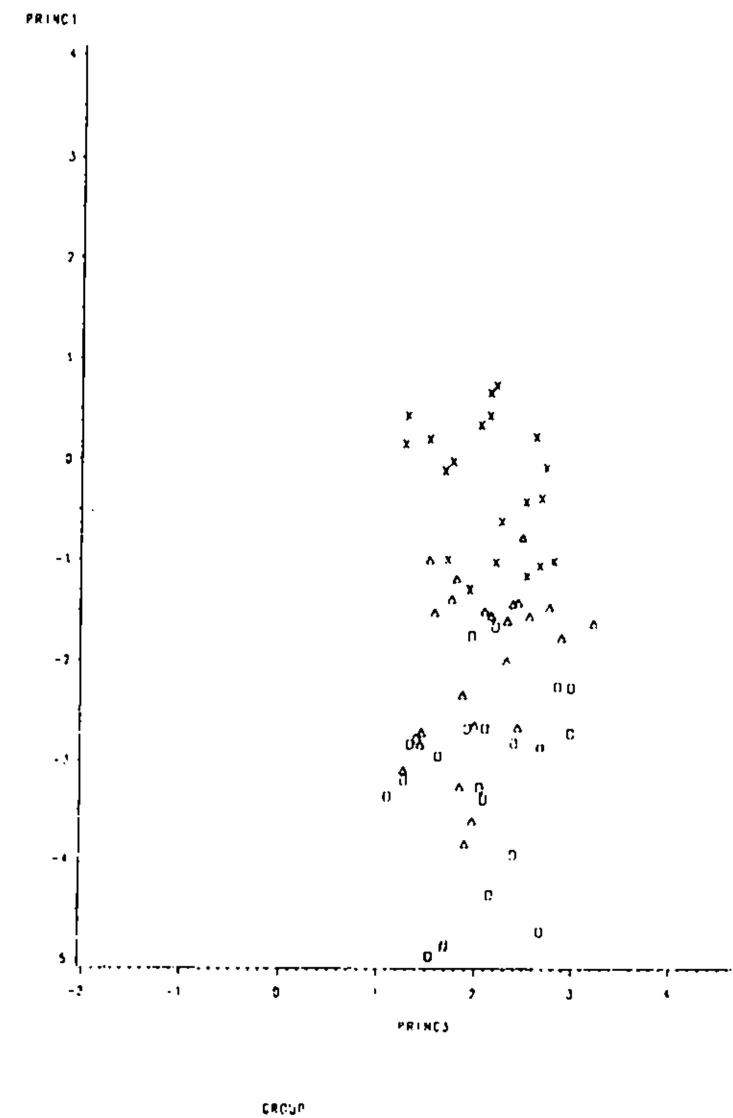


Figure 12 *Three cluster case: principal component 1 against principal component 3 with cluster membership*

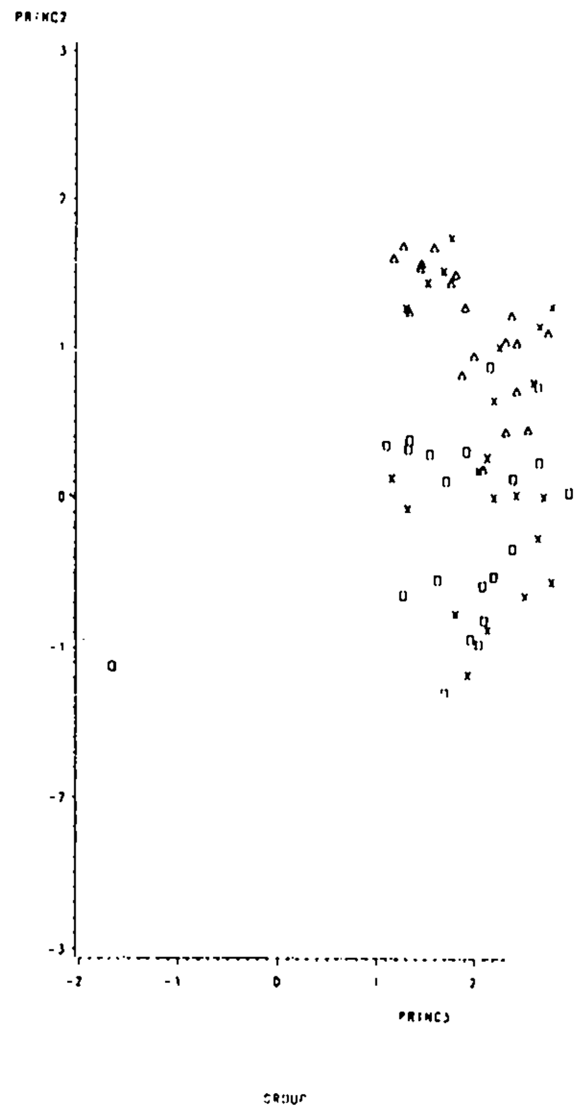


Figure 13 *Three cluster case: principal component 2 against principal component 3 with cluster membership*

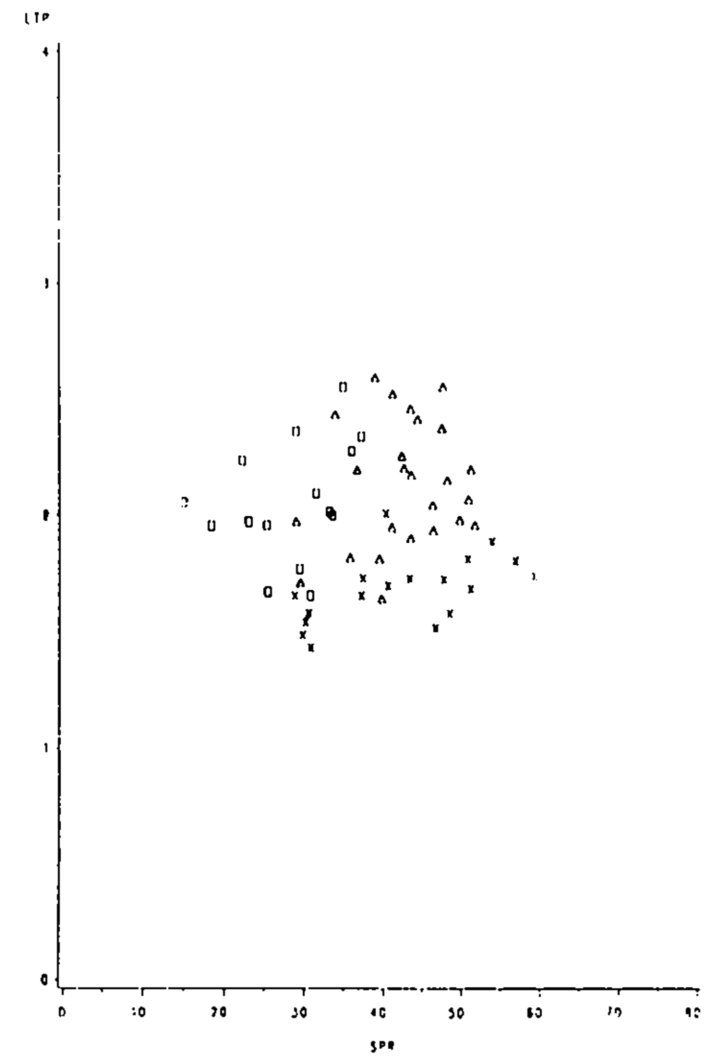


Figure 14 *Three cluster case: LTp against SPR with cluster membership*

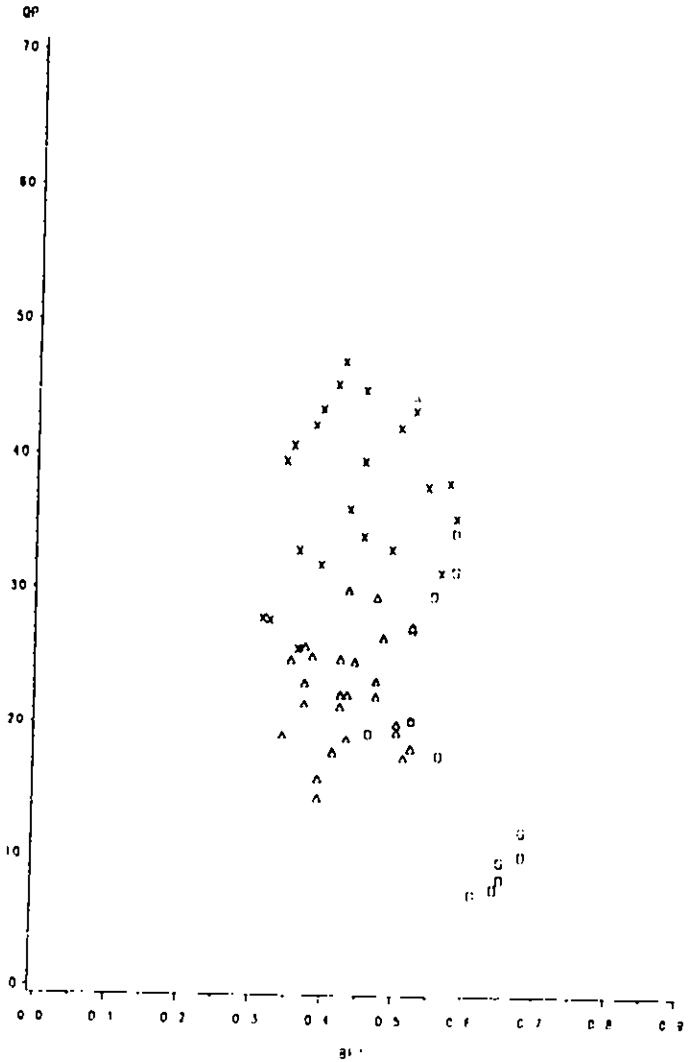


Figure 15 *Three cluster case: Qp against BFI with cluster membership*

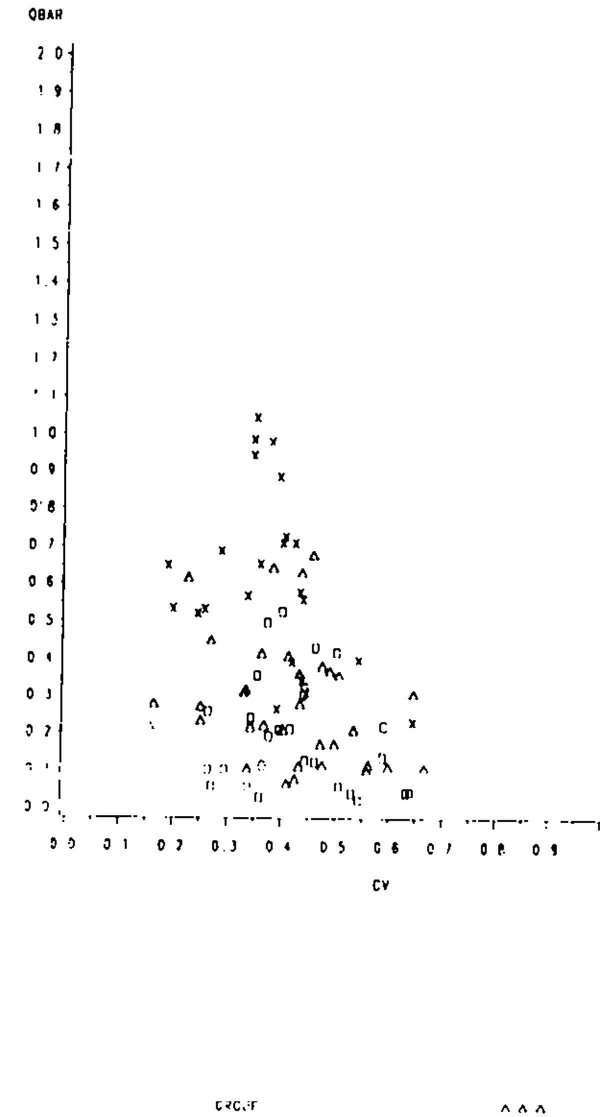


Figure 16 *Three cluster case: QBAR against CV with cluster membership*

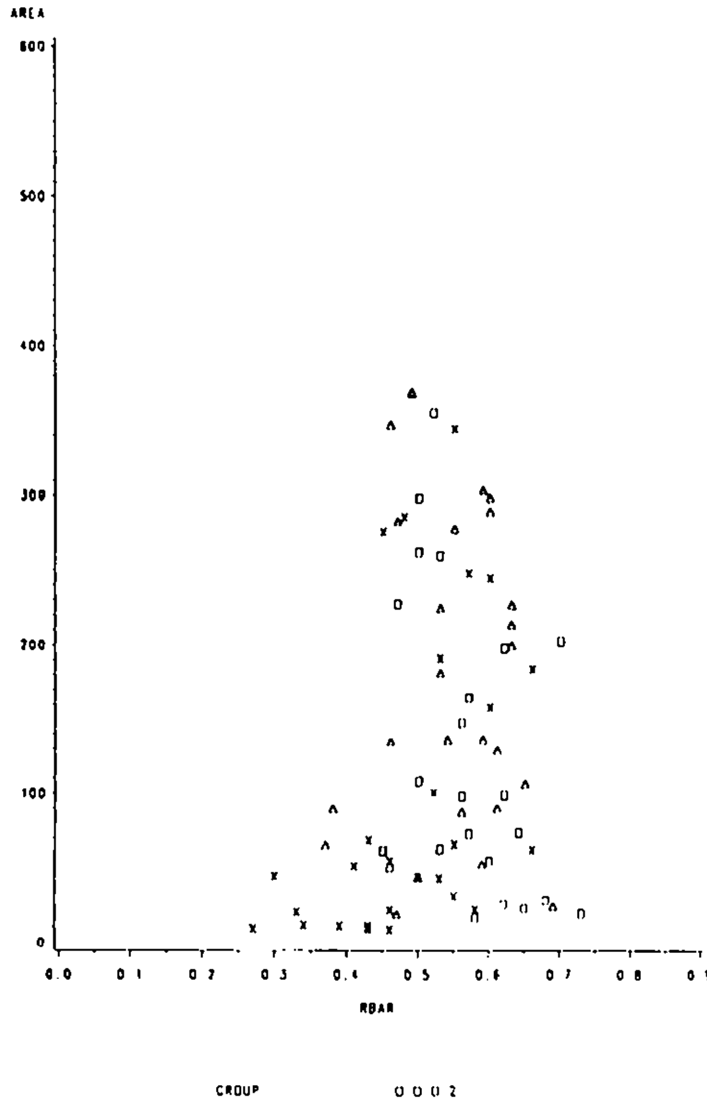


Figure 17 *Three cluster case: AREA against RBAR with cluster membership*

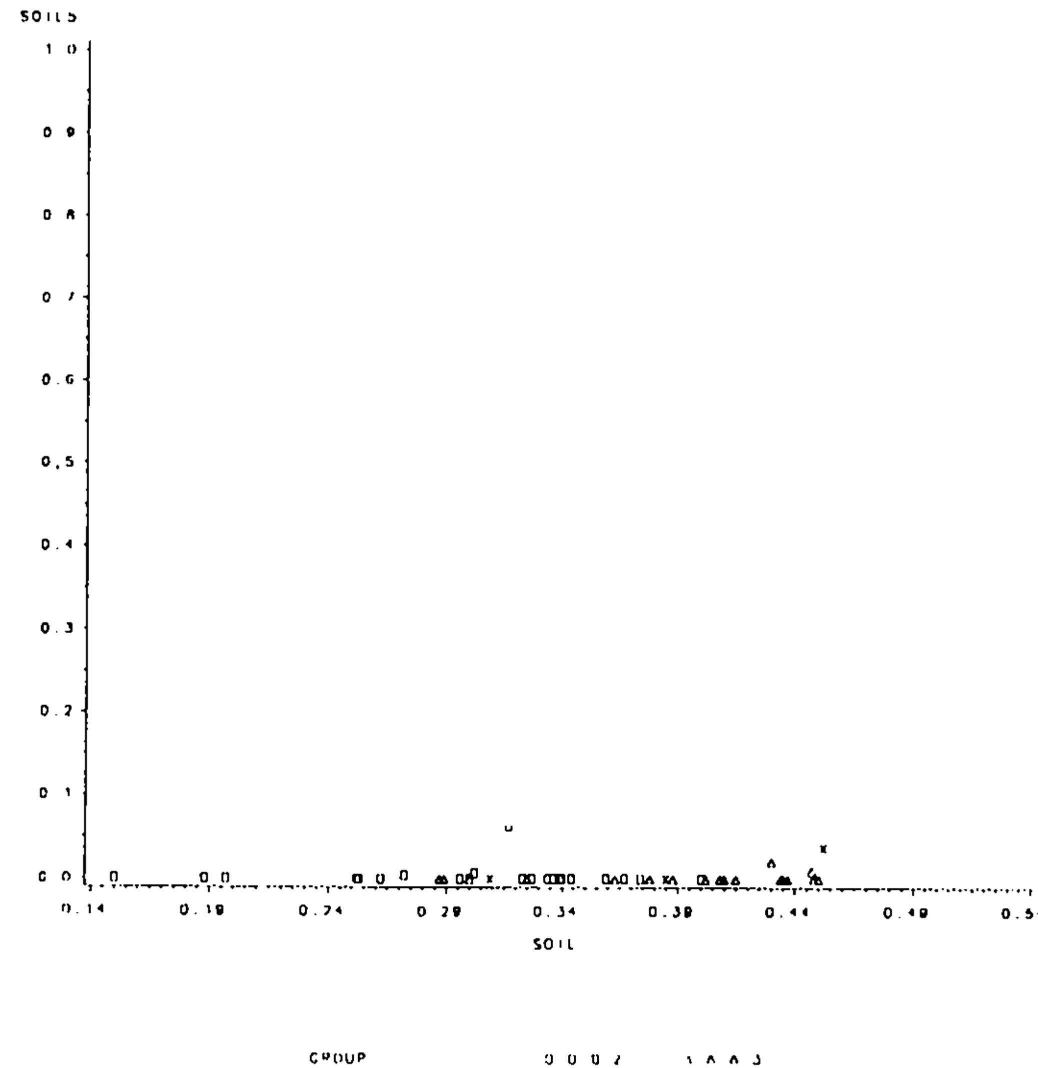


Figure 18 *Three cluster case: discrimination of catchments using SOILS and SOIL*

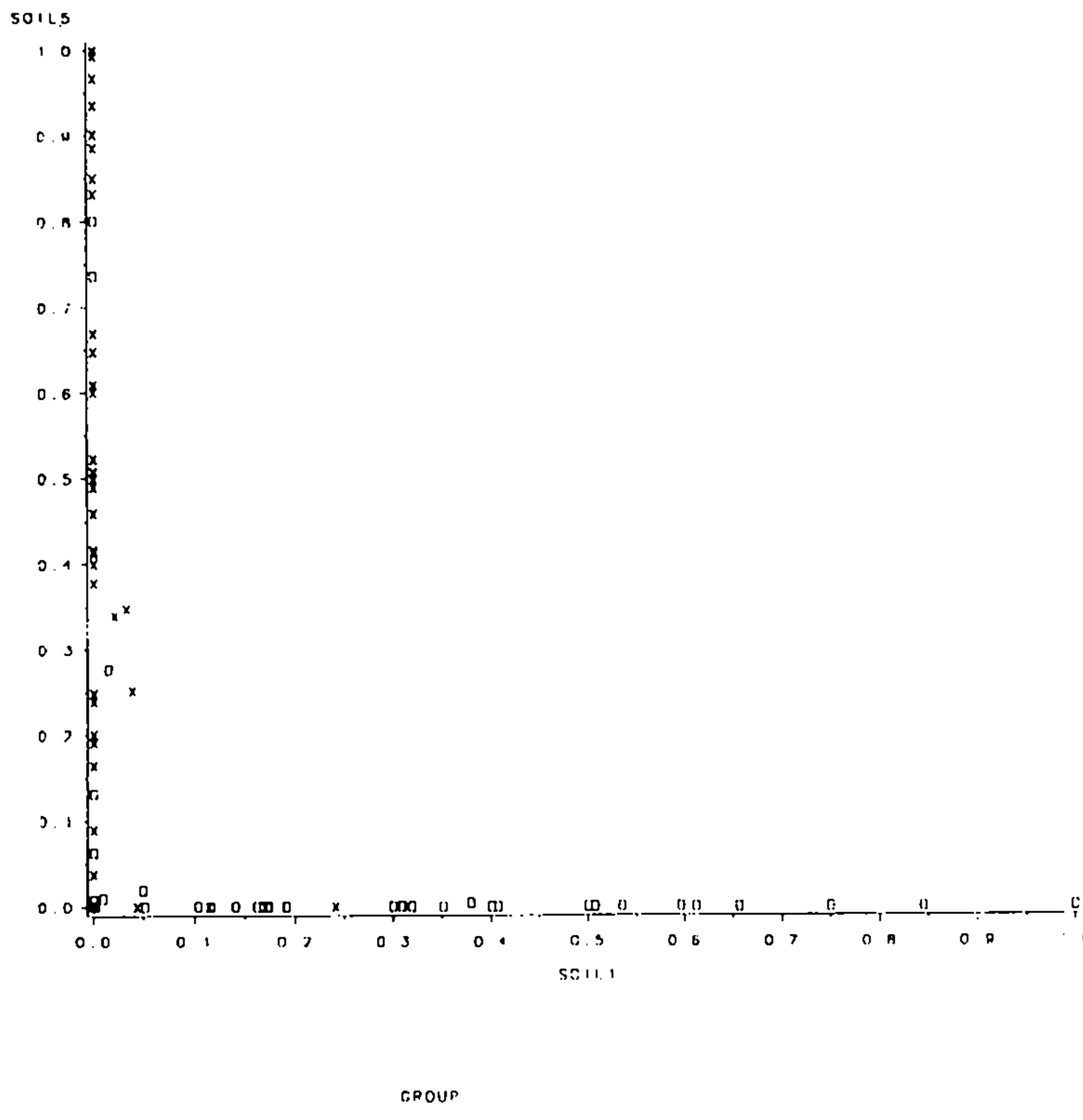


Figure 19 *Two cluster case: discrimination of catchments using SOIL5 and SOIL*

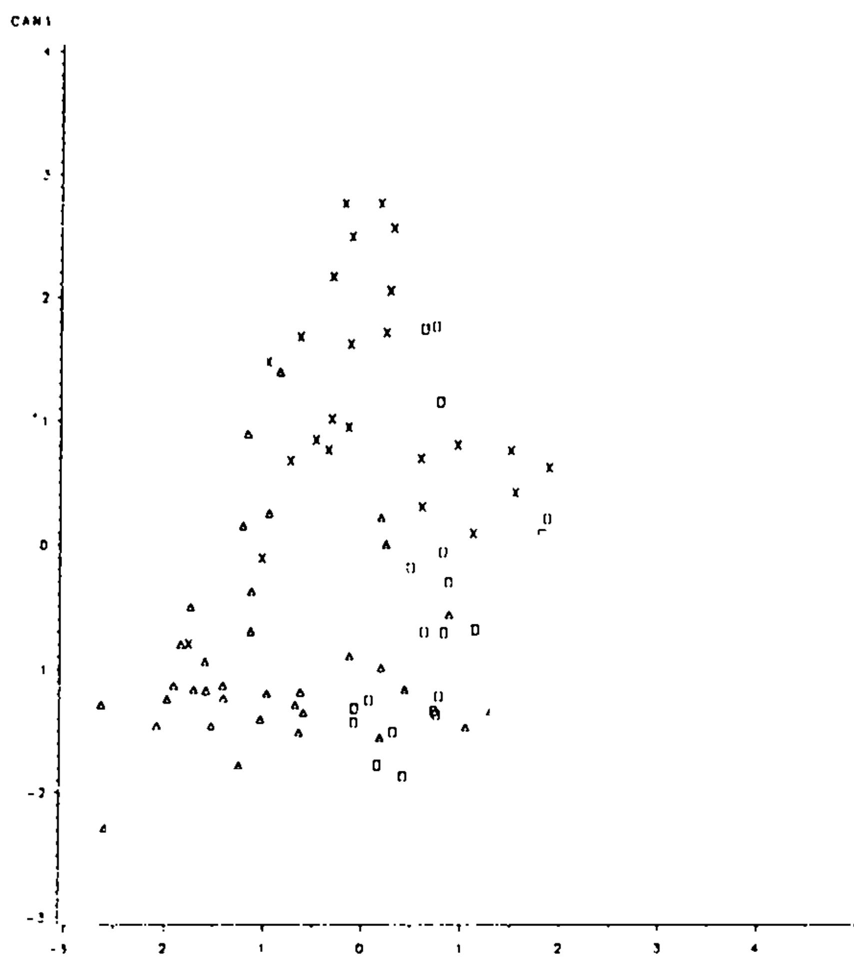


Figure 20 *Three cluster case: discrimination of catchments using the two canonical variables*

5 Catchments with Some Flow Data

The analysis to this stage has considered the estimation of rainfall-runoff model parameters for catchments for which no flow data are available. This section explores the possibility of using a limited amount of flow data, that may be available for a catchment, to enhance the parameter estimation. This situation, which is referred to as the partially gauged case, could arise when the catchment of interest has a very short gauging record resulting in insufficient storm event data from which to obtain estimates of the rainfall-runoff model parameters. However, the estimation of values for some of the other flow response measures may still be possible with the available flow record. A flow response measure from a catchment with some flow data could be used in the catchment assignment process, in order to (i) enhance the classification of a catchment, (ii) as a variable in the similarity measure for determining the n nearest neighbours to the catchment of interest, and (iii) to estimate the model parameter through a regression relationship.

To explore the utility of additional information, portions of the latter two stages of the investigation (i.e., catchment assignment and parameter estimation) were repeated considering the catchments to have data for one of the flow response measures. BFI and QBAR are the two flow response measures that are most likely to be available for a catchment that has only a limited flow record. The utility of each of these measures for enhancing the estimation of rainfall-runoff model parameters was examined separately. To simplify this stage of the analysis, the same basic framework has been adopted as was used for the case of ungauged catchments. In particular, the canonical variable approach was again used in the discriminant analysis procedure although the availability of a flow response measure will result in new canonical variables that should lead to enhanced classification of the catchments. In addition, the same parameter estimation options were considered with the available flow response measure also considered as a potential variable for defining catchment proximity.

5.1 RESULTS

5.1.1 Catchment assignment

A new set of canonical variables was calculated including the additional variable (i.e., BFI or QBAR) in the procedure. Table 11 gives the new weights for the calculation of canonical variables for the two cluster case and Table 12 contains the corresponding three cluster weightings. The new canonical variables for each catchment are presented in Appendix A. Using the new canonical variables in the catchment assignment process resulted in the same overall success rate for the two cluster results when BFI was the additional variable (based on the nearest seven neighbours). When QBAR was used in the formation of the canonical variable, the result was one fewer catchment correctly classified (based on the nearest five neighbours). Given that the original results of the catchment assignment process for the two cluster case were quite good, it is perhaps not surprising that the additional information has not improved the overall success rate of the process. It should be noted that, although the overall success rate was basically unchanged, the probability of membership in the two clusters for a given catchment will frequently have changed, thus potentially altering the estimation performance for the options that utilize this information.

For the case of three clusters, the somewhat less satisfactory original catchment classification results were improved as a result of including either of the additional variables. When BFI was available, the overall success rate was 83 out of 99 catchments, an improvement of seven in the number of catchments correctly classified. These results, which are summarized in Table 13, are based on using the nearest three neighbours. Using QBAR as the additional variable resulted in an overall success rate of 78 out of 99 catchments correctly classified, again based on the nearest three neighbours. These results are presented in Table 14. BFI is clearly the more useful variable, of the two flow response measures considered, for the catchment assignment process.

5.1.2 Parameter estimation

Revised procedures for estimating rainfall-runoff model parameters were based on the availability of either BFI or QBAR for a catchment. BFI was found to be most useful for the estimation of SPR while QBAR was found to be of greater benefit for estimating Tp. Revised SPR estimation procedures, using BFI, were therefore examined, and revised procedures, using QBAR, were examined for estimating Tp.

For Tp, the addition of QBAR as a proximity measure resulted in only marginal improvement in the estimation results. The improvement in estimates of Tp thus arises primarily from the enhanced classification of the catchments. Table 15 summarizes the results for the estimation of Tp using QBAR within the catchment assignment process. Note that only the best estimation option for each parameter and cluster combination is presented in Table 15. It is interesting to note that the results for the two cluster case are marginally better than the previous results even though the overall success rate of the catchment assignment process was slightly worse. This illustrates the inadequacy of the overall success rate as a measure of the utility of a classification of the catchments for the purposes of parameter estimation. Table 15 reveals that the improvement in estimation for both the two and three cluster case is small, although there is a marginally stronger case for selecting option 4 with three clusters as the preferred alternative.

BFI is beneficial for defining the set of catchments that are the nearest neighbours to the catchment of interest for estimating SPR. In addition, it was found that the use of the SOIL variable in combination with BFI was preferred to using BFI with the five individual soil class variables that were used previously. The utility of BFI as a measure of catchment proximity is not surprising since BFI and SPR are related. The FSSR16 presents a regression equation for estimating SPR when BFI is known. This equation results in improved estimates of SPR, relative to the estimates obtained using the relationship given in Equation (15), and thus provides the new reference for comparison with the other options. The FSSR16 regression equation is given as:

$$SPR_r = 72.0 - 66.5BFI \qquad (24)$$

where all symbols are as previously defined. The results for the best options for estimating SPR using BFI are presented in Table 15. This table reveals substantially improved estimates relative to the results previously obtained for the ungauged catchment case. Both the two and three cluster cases result in essentially unbiased estimates of SPR with a slightly lower RMSE for the two cluster case relative to the three cluster case. Both of the nearest neighbour based approaches give substantially better results than the FSSR16 approach, especially in terms of the bias measure. For the two cluster case, option 1 is the preferred approach, as was previously the case. The weights selected for the variables in the catchment proximity measure were 1.6 for SOIL and 1.0 for BFI. For the three cluster case, option 3 gives the best results. The probability threshold for this option was set to 0.25. The weights applied to the SOIL and BFI variables were 1.5 and 1.0, respectively.

5.2 SUMMARY

The results presented in this section indicate that the availability of the BFI can greatly improve the estimation of SPR for an ungauged catchment. The resulting estimates are essentially unbiased and have improved values for RMSE and D_{max} relative to those obtainable when an estimate of BFI is not available. The improvement in the results comes from a combination of enhanced catchment classification, which is most notable for the three cluster case, and through a better identification of catchments that are similar to the catchment of interest. The availability of QBAR improves the estimation of Tp, but the performance gain is not substantial.

It should be noted that if QBAR is available, it would be inappropriate to use this information only to refine the estimate of Tp. The user should compare the rainfall-runoff model estimate of the flow event with a return period corresponding to QBAR with the value of QBAR and consider how the model parameters (especially SPR) might be modified to reconcile any difference. In an application in which relatively frequent flood magnitudes are being estimated, serious consideration should be given to a rejection of the rainfall-runoff method in favour of the statistical method described in the FSR. When used in these ways the QBAR data will, undoubtedly, improve the required flood estimate.

Table 11 *Weights on the catchment characteristics and additional data for the canonical variable for two clusters*

Catchment Characteristic	Canonical Variable	
	BFI	QBAR
	1	1
MSL	-.3140	-.1903
DVF	.0483	.0683
SL1085	-.2903	-.3767
STMFRQ	-.3374	-.1050
SAAR	.7428	.4855
LAKE	-.0085	-.0332
URBAN	.3834	.2753
SOIL1	-32.766	-42.846
SOIL2	-81.670	-102.94
SOIL3	-98.218	-122.35
SOIL4	-141.66	-175.94
SOIL5	-158.13	-196.34
SOIL	64.582	77.978
BFI	-.8233	
QBAR		.5940

Table 12 Weights on the catchment characteristics and additional data for the canonical variable for three clusters

Catchment Characteristic	Canonical Variable			
	BFI		QBAR	
	1	2	1	2
MSL	-.1260	-.2815	.0010	-.4508
DVF	-.2371	-.0698	.1494	-.2256
SL1085	-.0173	.3132	.0519	.3283
STMFRQ	-.1213	-.0394	.2307	-.3938
SAAR	-.0837	.2300	-.2504	.3344
LAKE	.0337	-.1356	-.0952	-.0432
URBAN	-.2500	.3628	.3304	.2623
SOIL1	-2.1578	-7.0660	-8.5478	8.1154
SOIL2	3.5966	19.493	-3.6119	49.638
SOIL3	7.5911	35.895	1.4239	69.481
SOIL4	12.415	58.135	5.0528	105.37
SOIL5	15.033	72.136	9.1138	123.95
SOIL	-12.318	-52.506	-14.127	-69.495
BFI	1.0828	.5439		
QBAR			.7979	-.6324

Table 13 *Results of catchment assignment for three clusters with BFI*

Cluster	Number of Members	Number of Members Classified into Clusters			Percent Correctly Classified
		1	2	3	
1	34	28	1	5	82.4
	29		25		86.2
3	36	4	2	30	83.3
Total	99	34	28	37	

Table 14 *Results of catchment assignment for three clusters with QBAR*

Cluster	Number of Members	Number of Members Classified into Clusters			Percent Correctly Classified
		1	2	3	
1	34	25	4	5	73.5
	29		27		93.1
3	36		7	26	72.2
Total	99	28	38	33	

Table 15 *Runoff parameter estimation performance with additional data*

Runoff Event Parameter	Number of Clusters	Option	n	BIAS	RMSE	D_{max}
Tp	2	4		-.045	.271	.645
				-.029	.275	.705
		FSSR16		.000	.284	.787
SPR	2	1	1	-.001	.287	.987
				.002	.298	.987
		FSSR16 (BFI)		.104	.333	1.163

6 Conclusions and Recommendations

This report has presented a methodology for estimating rainfall-runoff model parameters for catchments that are either ungauged, or for which only a limited flow record is available. The procedure involves:

An initial grouping of the available set of gauged catchments, in terms of measures of runoff response;

2 The determination of the group membership for an ungauged catchment;

3 The estimation of rainfall-runoff model parameters based on the results of the catchment assignment process.

The methodology developed gives estimates of the standard percentage runoff that are substantially improved relative to estimates obtained using a regression approach. If an estimate of the base flow index for the catchment is available, the estimate of the standard percentage runoff can be improved even further. An estimate of the base flow index may be obtainable if a fairly short flow record is available for the catchment.

The application of the methodology to the estimation of the time to peak of the unit hydrograph resulted in estimates that were not definitively superior to the estimates from regression. This may be at least partially due to the comparatively good estimates that are obtained with the regression relationship. If an estimate of the mean annual flood for the catchment is available, the estimates can be improved, but the benefits of this additional information are fairly small.

A step-by-step guide for users wishing to apply the methods contained in this report is presented in Appendix B. That appendix also contains a worked example.

Future work should consider the application of the estimation framework outlined herein for the prediction of other hydrologic parameters for ungauged catchments. This could include additional flood event characteristics or involve the estimation of low flow characteristics for ungauged catchments. Consideration should also be given to refining the estimation procedure for the time to peak. Since the regression based approach is superior for some catchments and the methodology developed herein is superior for others, it would be interesting to determine if there are catchment characteristics that can be used to determine which approach should be taken for a particular ungauged catchment. It is thus possible that a composite estimation procedure could be developed involving a combination of the alternatives examined herein.

7 Acknowledgements

The authors gratefully acknowledge the financial support for this research from the Ministry of Agriculture, Fisheries and Food. Data have been made available from the UK Surface Water Archive and the Representative Basin and Flood Event Archive. This research was conducted while the first author was on sabbatical leave from the Department of Civil Engineering, University of Manitoba.

8 References

Acreman, M.C. 1987. Regional flood frequency analysis in the UK: Recent research - new ideas. Institute of Hydrology, Wallingford, UK.

Acreman, M.C. & Wiltshire, S.E. 1989. The regions are dead; long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis in the UK. In L. Roald, K. Nordseth, & K.A. Hassel (eds) *FRIENDS in Hydrology*, IAHS Publ. No. 187, 175-188.

Anderberg, M.R. 1973. *Cluster Analysis for Applications*. Academic Press, N.Y.

Bransby-Williams, G. 1922. Flood discharge and the dimensions of spillways in India. *Engineer*, 34, 321.

Burn, D.H. 1989. Cluster analysis as applied to regional flood frequency analysis. *J. Wat. Resour. Planning & Management*, ASCE, 115, 567-582.

Burn, D.H. 1990. Evaluation of regional flood frequency analysis with a Region of Influence approach. *Wat. Resour. Res.*, 26, 2257-2265.

Cordery, I. & Pilgrim, D.H. 1983. On the lack of dependence of losses from flood runoff on soil and cover characteristics. Hydrology of Humid Regions, Proc. Hamburg Symposium, *IAHS Publ. No. 140*, 187-195.

Flood Studies Supplementary Report No. 16 1985. The FSR rainfall-runoff model parameter estimation equations updated. Institute of Hydrology.

Galeatti, G., Rossi, G., Pini, G. & Zilli, G. 1986. Optimization of a snow network by multivariate statistical analysis. *Hydrol. Sci. J.* 31, 93-108.

Heerdegen, R.G. & Reich, B.M. 1974. Unit hydrographs for catchments of different sizes and dissimilar regions. *J. Hydrol.* 22, 143-153.

Lowing, M.J. & Reed, D.W. 1980. Recent advances in flood estimation techniques based on rainfall-runoff. Flood Studies Report-Five Years On, Manchester Conference, Institute of Civil Engineers, 49-56.

Nash, J.E. 1960. A unit hydrograph study with particular reference to British catchments. Proc. ICE, 17, 249-282.

Nathan, R.J. & McMahon, T.A. 1990. Identification of homogeneous regions for the purposes of regionalisation. *J. Hydrol.*, 121, 217-238.

NERC 1975. Flood Studies Report. 5 Vols. Natural Environment Research Council.

SAS 1985. SAS User's Guide: Statistics, Version 5 Edn, SAS Institute Inc., Cary, N.C.

Wiltshire, S.E. 1986. Regional flood frequency analysis II: Multivariate classification of drainage basins in Britain. *Hydrol. Sci. J.* 31, 335-346.

Appendix A Data listings

VARIABLE	CATCHMENT NUMBER						
	19001	19002	20001	23002	23005	24005	24007
Flow Response Measures							
Tp	7.08	7.01	8.98	5.25	6.60	7.18	5.64
Qp	25.1	19.0	18.1	36.0	40.4	27.3	34.0
SPR	51.6	62.6	36.6	37.1	53.8	28.9	37.3
RBAR	0.49	0.50	0.33	0.35	0.48	0.31	0.30
BFI	0.38	0.34	0.52	0.43	0.27	0.52	0.45
QBAR	0.311	0.411	0.173	0.231	0.709	0.210	0.311
CV	0.33	0.41	0.50	0.64	0.40	0.40	0.44
Catchment Characteristics							
AREA	369.0	43.8	307.0	118.0	284.9	178.5	44.6
MSL	42.0	17.9	31.9	21.4	36.3	31.7	11.9
DVF	0.01	0.04	0.03	0.02	0.01	0.00	0.04
SL1085	5.8	5.1	6.1	11.8	4.9	6.4	14.9
STMFRQ	0.75	1.07	0.64	2.14	2.20	1.32	1.64
SAAR	909.	1024	736.	962.	1322.	752.	797.
LAKE	0.04	0.00	0.02	0.00	0.00	0.00	0.00
URBAN	0.11	0.07	0.02	0.00	0.00	0.05	0.00
SOIL1	0.000	0.000	0.050	0.000	0.000	0.010	0.000
SOIL2	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL3	0.000	0.000	0.220	0.000	0.000	0.000	0.000
SOIL4	0.800	1.000	0.720	0.033	0.000	0.980	0.962
SOIL5	0.200	0.000	0.020	0.967	1.000	0.010	0.038
SOIL	0.460	0.450	0.430	0.498	0.500	0.447	0.452
Principal Components							
PC1	-1.18	-0.99	-2.41	-1.23	0.22	-1.92	-1.02
PC2	1.50	2.27	0.58	0.00	1.44	-0.15	0.13
PC3	1.80	1.52	0.29	-0.30	1.51	0.58	0.32
Cluster Membership							
CLUSTER2	1						
CLUSTER3	3	3	3				
Canonical Variables for Two Clusters							
Base Case							
CAN1	0.413	0.672	-0.733	0.645	1.638	-0.674	-0.006
With BFI							
CAN1	0.624	0.976	-1.080	0.941	2.626	-1.379	-0.415
With QBAR							
CAN1	0.363	0.683	-0.764	0.325	1.820	-0.224	-0.047
Canonical Variables for Three Clusters							
Base Case							
CAN1	-0.502	-0.802	-2.278	1.716	1.681	-0.937	-0.797
CAN2	-1.750	-1.845	-2.596	0.228	-0.640	-1.597	-1.765
With BFI							
CAN1	-1.195	-1.163	0.133	-1.419	-2.904	0.389	-0.199
CAN2	-1.703	-2.040	-3.213	1.200	0.135	-1.324	-1.524
With QBAR							
CAN1	-0.305	-0.510	-1.915	1.242	2.006	-0.756	-0.582
CAN2	-1.744	-1.946	-2.866	0.821	-0.569	-1.661	-1.814

VARIABLE	CATCHMENT NUMBER						
	25003	27001	27027	27035	28026	28033	28070
Flow Response Measures							
Tp	3.98	9.52	7.90	7.01	24.94	2.56	2.65
Qp	51.7	19.3	25.8	21.5	8.8	58.0	39.6
SPR	64.8	42.3	50.7	41.0	48.6	24.4	42.5
RBAR	0.34	0.48	0.51	0.47	0.49	0.46	0.27
BFI	0.15	0.50	0.37	0.37	0.47	0.45	0.45
QBAR	1.429	0.278	0.618	0.216	0.108	0.560	0.586
CV	0.19	0.43	0.22	0.16	0.34	0.44	0.90
Catchment Characteristics							
AREA	11.4	484.3	443.0	282 3	368.0	8.0	9.1
MSL	5.1	84.6	55.1	31.7	34.1	3.9	4.2
DVF	0 11	0.00	0.01	0.08	0.07	0.19	0.07
SL1085	35.8	2.5	4.5	4.5	1.4	33.4	35.9
STMFRQ	3.51	1.23	1.67	1.80	0.49	1.99	1.54
SAAR	2027.	975.	1381.	1134.	679.	1363.	985.
LAKE	0.00	0.25	0.09	0.13	0.00	0.00	0.00
URBAN	0.00	0.02	0.00	0.02	0.07	0.00	0.00
SOIL1	0.000	0.015	0.188	0.082	0.000	0.000	0 000
SOIL2	0.000	0.000	0.000	0.000	0.012	0.000	0 169
SOIL3	0.000	0.000	0.000	0.000	0.000	0.000	0 000
SOIL4	0.000	0.708	0.152	0.565	0 988	0.000	0.000
SOIL5	1.000	0.277	0.660	0.352	0.000	1 000	0.831
SOIL	0.500	0.459	0.427	0.443	0.448	0.500	0.466
Principal Components							
PC1	2.92	-2.27	-0.76	-1.43	-3.27	0 17	-0 19
PC2	1.74	0.93	1.54	1.22	2 37	-1 74	-0.58
PC3	1.75	1.41	2.47	2.37	1.83	1.26	-1.69
Cluster Membership							
CLUSTER2	1						
CLUSTER3	1	3	3				
Canonical Variables for Two Clusters							
Base Case							
CAN1	2.555	-0 853	0.850	0.919	-0.195	1.671	0.240
With BFI							
CAN1	3.842	-1.425	1.384	1 224	-0.505	1.680	0.615
With QBAR							
CAN1	2.859	-0.759	0.976	0.658	-0.310	1 210	0 184
Canonical Variables for Three Clusters							
Base Case							
CAN1	3.292	-1.290	0.252	-0.374	-1.135	2.770	2.069
CAN2	-0 340	-2.640	-0.951	-1 709	-1.913	0.169	1 278
With BFI							
CAN1	-4.187	-0.312	-1.550	-1.359	-0.204	-1.794	-1.016
CAN2	1.314	-2.392	-0.815	-1.636	-1.936	2.122	2 120
With QBAR							
CAN1	3 724	-0.756	0.559	-0.472	-0.995	2 105	1 791
CAN2	-0.162	-2.896	-1.030	-1.457	-1.931	1.074	1.626

VARIABLE	CATCHMENT NUMBER						
	29001	29004	30001	30004	33015	33029	33045
Flow Response Measures							
Tp	7.17	8.12	19.94	9.34	19.31	10.46	18.12
Qp	21.1	19.2	10.0	21.8	8.0	13.3	7.5
SPR	7.5	31.4	27.5	22.2	36.0	11.7	21.9
RBAR	0.50	0.60	0.50	0.45	0.55	0.62	0.68
BFI	0.84	0.46	0.68	0.65	0.55	0.86	0.64
QBAR	0.023	0.137	0.060	0.127	0.067	0.029	0.045
CV	0.54	0.59	0.51	0.44	0.41	0.36	0.64
Catchment Characteristics							
AREA	108.3	54.7	297.9	61 6	277.1	98.8	28.3
MSL	20.2	12.1	46.8	15.1	39.1	7.0	7.8
DVF	0.04	0.14	0.01	0.02	0.07	0.53	0.13
SL1085	3.3	2.0	2.1	3.3	1.0	1.6	3.3
STMFRQ	0.10	0.22	0.25	0.89	0.40	0.51	0.14
SAAR	729.	635.	631.	697	656.	634.	627.
LAKE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
URBAN	0.00	0.00	0.02	0.00	0.05	0.00	0.00
SOIL1	0.845	0.319	0.401	0.000	0.174	0.750	0.000
SOIL2	0.000	0.101	0.046	0.000	0.000	0.250	0.000
SOIL3	0.000	0.000	0.000	1.000	0.196	0.000	1.000
SOIL4	0.155	0.580	0.553	0.000	0.631	0.000	0.000
SOIL5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL	0.197	0.339	0.323	0.400	0.388	0.188	0.400
Principal Components							
PC1	-4.20	-2.83	-4.36	-3.20	-3.83	-4.71	-4.84
PC2	-2.25	0.38	0.32	-0.64	1.28	-1.57	0.11
PC3	1.21	1.32	1.31	1.26	1.89	2.67	1.70
Cluster Membership							
CLUSTER2	2						
CLUSTER3	2	2	2				
Canonical Variables for Two Clusters							
Base Case							
CAN1	-2.447	-1.333	-1.703	-1.643	-1.561	-2.031	-1.488
With BFI							
CAN1	-3.045	-0.587	-2.426	-2.681	-1.848	-2.754	-2 286
With QBAR							
CAN1	-2.554	-1.264	-1.740	-1 647	-1.645	-2.016	-1.586
Canonical Variables for Three Clusters							
Base Case							
CAN1	-1.912	-1.337	-1.775	-1.314	-1.352	-0.992	-1 249
CAN2	3.079	0 719	0.152	-0.082	-0.598	2.102	0.066
With BFI							
CAN1	4.325	0.448	2.260	2.175	0.836	3 128	1.928
CAN2	1.119	-0.975	-0.718	-0.342	-1.222	1.188	-0.349
With QBAR							
CAN1	-2.497	-1.340	-1.830	-1.293	-1.361	-1.273	-1.377
CAN2	2.855	0.432	-0.081	-0.277	-0.706	1.900	-0.022

VARIABLE	CATCHMENT NUMBER						
	33809	34003	34005	35008	37001	37007	38007
Flow Response Measure							
Tp	18.65	13.32	23.90	11.14	32.82	13.38	3.67
Qp	11.3	12.8	8.3	15.8	8.2	14.3	45.3
SPR	55.5	13.1	22.6	44.3	47.9	38.9	37.2
RBAR	0.37	0.57	0.64	0.61	0.59	0.54	0.19
BFI	0.32	0.83	0.65	0.39	0.40	0.39	0.41
QBAR	0.113	0.041	0.044	0.118	0.079	0.116	0.346
CV	0.60	0.53	0.63	0.56	0.43	0.48	0.44
Catchment Characteristics							
AREA	65.3	164.7	73.2	128.9	303.3	136.3	21.4
MSL	19.0	22.4	22.7	14.6	62.6	26.9	5.6
DVF	0.05	0.06	0.03	0.07	0.02	0.01	0.24
SL1085	1.6	2.1	1.7	3.4	1.2	1.9	7.5
STMFRQ	0.66	0.33	0.40	0.40	1.17	1.10	0.80
SAAR	559.	686.	647.	606.	610.	606.	611.
LAKE	0.00	0.03	0.00	0.00	0.00	0.00	0.00
URBAN	0.00	0.00	0.00	0.02	0.10	0.13	0.29
SOIL1	0.000	1.000	0.104	0.000	0.000	0.000	0.000
SOIL2	0.000	0.000	0.000	0.118	0.010	0.026	0.161
SOIL3	0.845	0.000	0.896	0.882	0.763	0.877	0.811
SOIL4	0.155	0.000	0.000	0.000	0.227	0.097	0.028
SOIL5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL	0.408	0.150	0.374	0.388	0.410	0.402	0.385
Principal Components							
PC1	-2.45	-4.87	-4.96	-2.71	-3.60	-2.82	0.04
PC2	2.86	-1.30	0.29	1.58	2.74	1.54	-0.32
PC3	-0.05	1.69	1.53	1.44	1.96	1.43	-0.29
Cluster Membership							
CLUSTER2	2						
CLUSTER3	3	2	2				
Canonical Variables for Two Clusters							
Base Case							
CAN1	-1.349	-3.629	-1.931	-1.832	-2.084	-1.688	-0.990
With BFI							
CAN1	-0.478	-3.731	-2.717	-1.063	-1.967	-1.156	-0.199
With QBAR							
CAN1	-1.264	-3.584	-1.994	-1.766	-2.094	-1.647	-0.719
Canonical Variables for Three Clusters							
Base Case							
CAN1	-1.511	-1.924	-1.507	-1.164	-1.132	-0.892	0.308
CAN2	-0.644	4.259	0.304	0.430	-1.415	-0.135	0.595
With BFI							
CAN1	-0.599	4.504	2.255	0.062	-0.477	-0.233	-0.860
CAN2	-2.128	1.577	-0.384	-1.078	-1.909	-1.117	0.090
With QBAR							
CAN1	-1.285	-2.477	-1.620	-1.125	-0.916	-0.804	0.586
CAN2	-0.955	3.849	0.136	0.173	-1.560	-0.316	0.340

VARIABLE	CATCHMENT NUMBER						
	39005	39007	39012	39017	39022	39026	39052
Flow Response Measure							
Tp	3.21	12.60	4.15	9.21	20.54	20.34	5.53
Qp	57.7	11.9	37.1	22.7	11.4	9.8	22.2
SPR	18.2	19.1	19.1	57.4	40.1	34.8	29.4
RBAR	0.50	0.52	0.28	0.47	0.57	0.63	0.19
BFI	0.61	0.68	0.73	0.15	0.75	0.40	0.42
QBAR	0.266	0.060	0.191	0.308	0.103	0.111	0.172
CV	0.39	0.27	0.38	0.64	0.30	0.67	0.47
Catchment Characteristics							
AREA	43.6	354.8	69.1	18.6	164.5	199.4	50.2
MSL	7.4	32.3	11.8	7.1	22.1	27.9	11.0
DVF	0.52	0.08	0.02	0.14	0.15	0.01	0.06
SL1085	2.3	1.0	3.7	4.8	1.6	2.1	3.5
STMFRQ	0.20	0.38	0.19	0.38	0.36	0.64	0.78
SAAR	633.	710	679.	650.	751.	700.	687.
LAKE	0.00	0.00	0.00	0.00	0.00	0 09	0.00
URBAN	0.81	0.33	0.46	0.00	0.02	0 02	0.18
SOIL1	0.115	0.173	0.310	0.000	0.408	0.141	0.000
SOIL2	0.712	0.116	0.153	0.000	0.001	0.000	0.000
SOIL3	0.000	0.418	0.000	0.000	0.000	0.000	0.306
SOIL4	0.172	0.292	0.537	1.000	0.591	0.859	0.694
SOIL5	0.000	0.000	0 000	0.000	0.000	0.000	0.000
SOIL	0 309	0.359	0.334	0.450	0.327	0.408	0.435
Principal Components							
PC1	-0.98	-3.95	-1.92	-0.98	-4.09	-3.76	-1.54
PC2	-2.39	-0.34	-2.01	2.71	0.74	1.61	0.19
PC3	1.70	2.38	0.65	0 20	2.65	1.16	-0.47
Cluster Membership							
CLUSTER2	1						
CLUSTER3	1	2	2				
Canonical Variables for Two Clusters							
Base Case							
CAN1	2.106	-0.905	-0.071	0.113	-1.661	-0.415	0.349
With BFI							
CAN1	2 518	-1.586	-0.570	1.804	-2.635	0.012	0.527
With QBAR							
CAN1	2.202	-1.085	-0.097	0.284	-1.747	-0.437	0.284
Canonical Variables for Three Clusters							
Base Case							
CAN1	2.409	-0.061	0.211	-1.173	-1.215	-1.783	-0.698
CAN2	1.151	0.813	1.858	-1.589	0.774	-1.256	-1.133
With BFI							
CAN1	-1 324	1.503	1.875	-2.556	2.652	-0.235	-0.559
CAN2	2.288	0.953	1.768	-3.054	0.280	-2.435	-1.395
With QBAR							
CAN1	2.343	-0.421	-0.104	-0.700	-1.433	-1.613	-0.613
CAN2	1.397	0.981	1.889	-1.921	0.668	-1.480	-1.150

VARIABLE	CATCHMENT NUMBER						
	39053	40006	40009	40010	41005	41006	41022
Flow Response Measure							
Tp	8.98	7.19	8.77	16.64	17.32	12.85	7.24
Qp	18.8	24.9	24.7	10.4	12.6	24.9	24.8
SPR	51.1	23.0	43.5	48.6	45.2	60.4	49.7
RBAR	0.38	0.46	0.59	0.53	0.53	0.56	0.59
BFI	0.43	0.62	0.44	0.32	0.48	0.42	0.35
QBAR	0.271	0.183	0.217	0.154	0.210	0.418	0.384
CV	0.25	1.16	0.34	1.12	0.53	0.36	0.47
Catchment Characteristics							
AREA	89.9	50.3	136.2	224.3	180.9	87.8	52.0
MSL	14.6	13.5	19.4	30.9	26.7	16.4	16.5
DVF	0.00	0.04	0.02	0.03	0.02	0.01	0.02
SL1085	2.3	6.2	3.2	1.6	2.1	4.0	4.9
STMFRQ	1.23	0.54	1.43	0.79	1.32	1.03	1.33
SAAR	825.	733.	808	764.	835.	837.	887
LAKE	0.19	0.00	0.02	0.04	0.12	0.00	0.00
URBAN	0.09	0.03	0.01	0.03	0.04	0.02	0.01
SOIL1	0.000	0.596	0.000	0.117	0.000	0.000	0.241
SOIL2	0.000	0.117	0.000	0.000	0.000	0.000	0.000
SOIL3	0.000	0.012	0.000	0.135	0.000	0.000	0.000
SOIL4	1.000	0.276	1.000	0.748	1.000	1.000	0.759
SOIL5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL	0.450	0.253	0.450	0.408	0.450	0.450	0.378
Principal Components							
PC1	-1.52	-3.44	-2.00	-3.27	-3.10	-1.55	-1.38
PC2	1.68	-1.13	1.05	2.19	1.69	2.15	1.44
PC3	1.57	-1.68	2.31	-1.33	1.26	2.15	1.74
Cluster Membership							
CLUSTER2	1						
CLUSTER3	3	2	3				
Canonical Variables for Two Clusters							
Base Case							
CAN1	0.152	-2.756	0.048	-0.811	-0.010	0.213	-0.329
With BFI							
CAN1	0.257	-2.278	-0.245	0.036	-0.439	0.109	0.388
With QBAR							
CAN1	0.304	-2.660	0.016	-0.841	0.006	0.425	-0.160
Canonical Variables for Three Clusters							
Base Case							
CAN1	-1.458	-1.285	-1.240	-1.408	-1.459	-1.166	-1.288
CAN2	-1.535	2.929	-1.981	-1.035	-2.086	-1.716	-0.682
With BFI							
CAN1	-0.227	2.339	-0.192	-0.907	0.064	-0.291	-0.542
CAN2	-2.172	0.745	-2.065	-2.298	-2.205	-1.950	-1.872
With QBAR							
CAN1	-1.017	-1.577	-0.980	-1.282	-1.118	-0.618	-0.951
CAN2	-1.889	2.583	-2.102	-1.199	-2.288	-2.088	-1.043

VARIABLE	CATCHMENT NUMBER						
	41028	45002	45003	45004	46003	46005	47007
Flow Response Measures							
Tp	8.58	7.49	11.68	9.05	4.68	3.72	5.25
Qp	23.1	27.1	20.3	20.0	44.5	59.9	37.7
SPR	48.0	33.1	43.4	42.5	30.1	58.3	28.7
RBAR	0.69	0.70	0.63	0.60	0.57	0.46	0.46
BFI	0.37	0.52	0.52	0.50	0.52	0.42	0.54
QBAR	0.320	0.358	0.359	0.370	0.943	1.953	0.391
CV	0.33	0.35	0.50	0.49	0.34	0.29	0.15
Catchment Characteristics							
AREA	24.0	421.7	226.1	288.5	247.6	21.5	54.9
MSL	10.0	48.1	26.4	33.6	35.2	11.8	16.6
DVF	0.02	0.04	0.01	0.01	0.03	0.08	0.05
SL1085	4.9	5.7	6.1	3.6	16.5	22.6	17.8
STMFRQ	0.84	0.85	0.59	0.79	0.78	0.93	0.75
SAAR	847.	1420.	996.	1052	1696.	1987.	1477.
LAKE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
URBAN	0.01	0.00	0.00	0.01	0.00	0.00	0.00
SOIL1	0.044	0.000	0.535	0.508	0.277	0.000	0.039
SOIL2	0.000	0.862	0.017	0.000	0.254	0.000	0.708
SOIL3	0.000	0.000	0.000	0.168	0.000	0.000	0.000
SOIL4	0.956	0.006	0.447	0.324	0.000	0.000	0.000
SOIL5	0.000	0.131	0.000	0.000	0.469	1.000	0.252
SOIL	0.437	0.327	0.287	0.289	0.352	0.500	0.344
Principal Components							
PC1	-1.77	-2.27	-2.63	-2.34	-0.41	2.75	-1.15
PC2	1.52	0.03	0.95	0.82	-0.99	0.02	-0.65
PC3	2.88	2.97	1.98	1.86	2.50	2.43	2.51
Cluster Membership							
CLUSTER2	1						
CLUSTER3	3	2	3				
Canonical Variables for Two Clusters							
Base Case							
CAN1	0.339	-0.056	-1.052	-1.885	0.103	2.963	-0.056
With BFI							
CAN1	0.662	-0.059	-0.697	-1.374	0.296	3.158	0.079
With QBAR							
CAN1	0.387	-0.188	-0.970	-1.781	0.341	4.207	-0.387
Canonical Variables for Three Clusters							
Base Case							
CAN1	-1.231	-0.186	-1.470	-1.345	0.763	2.562	0.623
CAN2	-1.413	0.486	1.040	1.293	1.490	0.300	1.876
With BFI							
CAN1	-0.578	0.492	1.307	1.166	0.307	-1.742	0.761
CAN2	-2.066	0.205	-0.582	-0.438	1.453	1.966	1.656
With QBAR							
CAN1	-0.949	-0.435	-1.505	-1.384	0.855	4.162	-0.106
CAN2	-1.625	0.591	0.717	0.957	1.328	-0.639	2.293

VARIABLE	CATCHMENT NUMBER						
	48004	48005	48009	49003	52004	52005	52006
Flow Response Measures							
Tp	7.39	4.37	10.35	5.61	6.69	9.73	11.37
Qp	28.9	47.0	22.1	35.4	26.5	21.8	17.8
SPR	33.5	12.7	37.2	47.6	43.4	35.9	33.9
RBAR	0.62	0.73	0.65	0.58	0.61	0.70	0.63
BFI	0.72	0.67	0.63	0.58	0.48	0.58	0.41
QBAR	0.420	0.304	0.432	0.709	0.278	0.328	0.222
CV	0.50	0.44	0.46	0.42	0.16	0.44	0.37
Catchment Characteristics							
AREA	25.3	19.1	22.7	21.7	90.1	202.0	213.1
MSL	10.0	7.2	12.2	6.7	14.3	37.3	16.7
DVF	0.08	0.08	0.01	0.16	0.08	0.04	0.05
SL1085	17.5	13.1	18.0	10.8	5.1	5.6	5.5
STMFRQ	1.66	1.77	1.63	1.20	0.70	0.72	1.27
SAAR	1512.	1107.	1622.	1714.	943.	995.	907.
LAKE	0.00	0.00	0.02	0.00	0.00	0.00	0.03
URBAN	0.00	0.06	0.00	0.00	0.02	0.06	0.05
SOIL1	0.000	0.000	0.000	0.000	0.170	0.162	0.192
SOIL2	0.264	1.000	0.200	0.007	0.000	0.406	0.000
SOIL3	0.000	0.000	0.000	0.000	0.226	0.000	0.581
SOIL4	0.000	0.000	0.000	0.000	0.603	0.432	0.226
SOIL5	0.736	0.000	0.800	0.993	0.000	0.000	0.000
SOIL	0.447	0.300	0.460	0.499	0.388	0.340	0.363
Principal Components							
PC1	-2.67	-2.25	-2.82	-1.02	-1.62	-2.87	-2.65
PC2	-0.81	-2.43	0.12	0.00	0.73	0.24	1.04
PC3	2.10	2.84	2.39	2.19	3.20	2.66	2.42
Cluster Membership							
CLUSTER2	2						
CLUSTER3	2	2	2				
Canonical Variables for Two Clusters							
Base Case							
CAN1	1.297	-0.778	1.728	3.225	0.352	-1.304	-1.107
With BFI							
CAN1	-0.048	-1.282	0.845	2.508	0.209	-1.526	-0.583
With QBAR							
CAN1	0.809	-0.764	1.153	2.954	0.280	-1.224	-1.159
Canonical Variables for Three Clusters							
Base Case							
CAN1	1.762	0.078	1.742	2.168	-1.197	-0.705	-0.984
CAN2	0.735	1.806	0.618	-0.313	-0.974	0.626	0.190
With BFI							
CAN1	1.107	1.909	0.447	-0.703	0.337	1.204	0.109
CAN2	2.682	1.626	2.290	1.868	-1.428	0.052	-0.938
With QBAR							
CAN1	0.983	-0.170	0.860	1.824	-1.144	-0.683	-1.073
CAN2	1.509	1.772	1.486	0.314	-1.058	0.422	0.093

VARIABLE	CATCHMENT NUMBER						
	52010	52016	53005	53007	53009	54004	54011
Flow Response Measures							
Tp	10.73	4.31	9.88	10.57	7.80	12.42	12.82
Qp	23.3	34.6	18.4	20.2	20.1	10.7	15.5
SPR	47.4	13.9	18.6	28.9	14.9	41.1	34.9
RBAR	0.46	0.58	0.56	0.50	0.57	0.28	0.43
BFI	0.47	0.71	0.62	0.52	0.62	0.60	0.65
QBAR	0.363	0.218	0.208	0.240	0.211	0.114	0.124
CV	0.43	0.59	0.39	0.34	0.42	0.43	0.46
Catchment Characteristics							
AREA	135.2	15.7	147.4	261.6	72.6	262.0	184.0
MSL	20.4	7.1	24.6	27.7	16.1	28.8	26.9
DVF	0.02	0.02	0.03	0.06	0.02	0.07	0.01
SL1085	4.7	26.8	3.0	2.3	8.1	1.9	4.9
STMFRQ	0.77	0.89	0.73	0.33	0.80	0.72	0.45
SAAR	881.	969.	972.	966.	1018.	691	675.
LAKE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
URBAN	0.00	0.00	0.05	0.02	0.07	0.25	0.03
SOIL1	0.002	0.000	0.612	0.309	0.656	0.052	0.352
SOIL2	0.000	0.553	0.000	0.037	0.000	0.000	0.000
SOIL3	0.686	0.000	0.097	0.321	0.018	0.000	0.000
SOIL4	0.312	0.447	0.292	0.332	0.326	0.948	0.648
SOIL5	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL	0.415	0.367	0.262	0.336	0.252	0.434	0.344
Principal Components							
PC1	-1.88	-2.77	-3.39	-2.67	-3.27	-2.95	-3.35
PC2	1.24	-2.24	-0.58	0.31	-0.98	0.97	0.35
PC3	1.32	1.39	2.08	1.91	2.04	0.36	1.09
Cluster Membership							
CLUSTER2	2						
CLUSTER3	3	2	2				
Canonical Variables for Two Clusters							
Base Case							
CAN1	-0.996	-0.842	-1.880	-0.389	-2.047	-0.018	-1.795
With BFI							
CAN1	-1.144	-2.013	-1.752	-0.395	-1.783	-0.848	-2.301
With QBAR							
CAN1	-0.846	-1.147	-1.945	-0.452	-2.210	-0.180	-1.824
Canonical Variables for Three Clusters							
Base Case							
CAN1	-1.187	-0.303	-1.388	-1.425	-0.985	-0.381	-1.364
CAN2	-0.627	0.867	1.845	-0.084	2.668	-1.126	0.749
With BFI							
CAN1	0.538	2.362	2.107	0.946	2.243	0.531	2.118
CAN2	-1.204	1.142	0.168	-1.079	0.933	-0.506	-0.156
With QBAR							
CAN1	-0.881	-0.836	-1.737	-1.490	-1.591	-0.429	-1.500
CAN2	-0.957	1.130	1.674	-0.228	2.649	-0.993	0.559

VARIABLE	CATCHMENT NUMBER						
	54016	54019	54022	55008	55012	55021	55022
Flow Response Measures							
Tp	25.25	42.75	2.35	2.18	6.12	22.40	12.80
Qp	7.1	6.5	53.9	68.7	31.9	9.6	17.4
SPR	27.1	40.7	36.7	43.7	50.7	32.6	47.6
RBAR	0.53	0.46	0.43	0.39	0.60	0.82	0.77
BFI	0.61	0.48	0.32	0.32	0.39	0.65	0.51
QBAR	0.058	0.102	1.598	1.819	0.726	0.103	0.233
CV	0.34	0.56	0.34	0.52	0.40	0.26	0.25
Catchment Characteristics							
AREA	259.0	347.0	8.7	10.6	244.2	371.0	142.0
MSL	40.2	56.7	4.7	5.4	36.0	48.5	29.8
DVF	0.05	0.04	0.04	0.06	0.01	0.01	0.01
SL1085	0.9	1.4	63.7	47.4	8.0	4.0	3.0
STMFRQ	0.27	0.51	3.60	2.88	1.39	0.64	2.73
SAAR	713.	692.	2249.	2395.	1643.	948	944.
LAKE	0.00	0.00	0.00	0.00	0.00	0.00	0.00
URBAN	0.00	0.04	0.00	0.00	0.00	0.01	0.00
SOIL1	0.500	0.300	0.000	0.000	0.000	0.000	0.000
SOIL2	0.030	0.000	0.000	0.000	0.584	0.891	1.000
SOIL3	0.000	0.700	0.000	0.000	0.000	0.045	0.000
SOIL4	0.470	0.000	0.000	0.000	0.000	0.000	0.000
SOIL5	0.000	0.000	1.000	1.000	0.416	0.063	0.000
SOIL	0.296	0.325	0.500	0.500	0.383	0.317	0.300
Principal Components							
PC1	-4.34	-4.14	2.20	3.11	-0.60	-4.49	-2.82
PC2	0.88	2.28	-0.77	-1.00	1.01	0.84	1.46
PC3	2.15	0.78	1.80	0.92	2.24	4.07	3.76
Cluster Membership							
CLUSTER2	2						
CLUSTER3	2	3					
Canonical Variables for Two Clusters							
Base Case							
CAN1	-1.493	-2.540	2.261	2.983	0.595	-1.040	-1.470
With BFI							
CAN1	-1.653	-2.219	2.352	3.217	1.242	-1.590	-1.295
With QBAR							
CAN1	-1.571	-2.528	2.295	3.412	0.714	-1.122	-1.250
Canonical Variables for Three Clusters							
Base Case							
CAN1	-1.869	-1.555	3.989	3.634	0.701	-0.707	-0.564
CAN2	0.410	0.179	0.172	0.066	0.583	0.823	0.873
With BFI							
CAN1	1.786	0.716	-2.745	-2.729	-0.896	1.792	0.757
CAN2	-0.973	-1.248	2.951	2.543	0.489	0.408	0.058
With QBAR							
CAN1	-2.020	-1.545	3.980	4.168	0.766	-0.932	-0.390
CAN2	0.205	-0.077	0.733	0.158	0.554	0.789	0.536

VARIABLE	CATCHMENT NUMBER						
	55025	56003	56004	56005	56006	57004	57005
Flow Response Measures							
Tp	5.87	3.53	7.70	5.25	4.56	6.17	6.13
Qp	34.2	43.4	29.5	29.6	44.9	21.3	22.1
SPR	29.3	29.8	46.2	30.7	46.6	35.7	39.4
RBAR	0.66	0.66	0.65	0.56	0.66	0.65	0.63
BFI	0.58	0.52	0.47	0.55	0.45	0.42	0.47
QBAR	0.392	0.399	0.632	0.499	0.887	0.679	0.645
CV	0.88	0.54	0.43	0.37	0.39	0.45	0.38
Catchment Characteristics							
AREA	132.0	62.1	543.9	98.1	183.8	106.0	454.8
MSL	18.7	20.2	48.7	25.4	22.4	25.8	42.3
DVF	0.04	0.04	0.01	0.02	0.01	0.04	0.05
SL1085	4.0	9.0	4.6	14.2	8.9	7.3	9.2
STMFRQ	0.98	0.98	1.26	1.17	1.67	2.33	2.17
SAAR	999.	1253.	1488.	1469	1661.	1759.	1838
LAKE	0.17	0.00	0.08	0.00	0.11	0.00	0.15
URBAN	0.00	0.00	0.02	0.16	0.00	0.04	0.05
SOIL1	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL2	1.000	0.751	0.600	0.295	0.478	0.000	0.000
SOIL3	0.000	0.000	0.000	0.327	0.000	0.331	0.400
SOIL4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL5	0.000	0.249	0.400	0.378	0.522	0.669	0.600
SOIL	0.300	0.350	0.380	0.408	0.404	0.467	0.460
Principal Components							
PC1	-2.50	-1.28	-1.41	-1.66	-0.07	-1.60	-1.55
PC2	-0.99	-1.18	0.72	-0.52	0.01	0.44	0.45
PC3	0.59	1.93	2.43	2.19	2.71	2.32	2.54
Cluster Membership							
CLUSTER2	2						
CLUSTER3	2		3				
Canonical Variables for Two Clusters							
Base Case							
CAN1	-1.260	0.124	0.355	0.514	0.655	1.868	1.334
With BFI							
CAN1	-0.915	0.330	0.602	0.207	1.107	1.729	0.941
With QBAR							
CAN1	-0.769	0.113	0.562	0.244	1.040	1.645	1.092
Canonical Variables for Three Clusters							
Base Case							
CAN1	-1.089	0.094	0.219	1.157	0.807	1.393	0.893
CAN2	1.781	1.116	0.185	0.778	0.955	-0.848	-1.170
With BFI							
CAN1	1.587	0.491	-0.330	0.175	-0.438	-1.361	-0.929
CAN2	0.206	0.709	0.129	1.578	0.953	0.535	0.112
With QBAR							
CAN1	-0.685	-0.090	0.466	0.669	1.174	1.206	0.733
CAN2	1.066	1.127	-0.004	1.227	0.653	-0.385	-0.757

VARIABLE	CATCHMENT NUMBER						
	57006	58001	58002	58006	58008	58009	60002
Flow Response Measures							
Tp	2.50	4.43	4.87	3.82	4.32	5.32	6.92
Qp	47.0	33.0	39.6	40.7	43.5	31.4	22.2
SPR	35.3	29.7	30.4	44.5	55.3	25.4	46.2
RBAR	0.52	0.60	0.53	0.55	0.53	0.53	0.60
BFI	0.42	0.49	0.34	0.35	0.39	0.58	0.43
QBAR	0.987	0.688	1.048	0.979	1.046	0.528	0.452
CV	0.34	0.28	0.45	0.38	0.35	0.40	0.27
Catchment Characteristics							
AREA	100.5	158.0	190.9	65.8	43.0	62.5	297.8
MSL	22.9	20.1	28.3	14.7	14.0	13 1	50.0
DVF	0.02	0.02	0 04	0.03	0.01	0.01	0.01
SL1085	7.7	10.3	13.5	25.9	14.9	7.7	4.6
STMFRQ	3.04	2.63	4.04	2.51	6.28	2.10	0.82
SAAR	2200.	1839.	1981.	2107	1756.	1382.	1637.
LAKE	0 06	0.00	0.00	0.12	0.00	0.00	0.00
URBAN	0.13	0.04	0.01	0.00	0.00	0.05	0.00
SOIL1	0.000	0.021	0.000	0.000	0.000	0.380	0.000
SOIL2	0.000	0.169	0.000	0.000	0.000	0.331	0 761
SOIL3	0.510	0.472	0.101	0.000	0.150	0.285	0.000
SOIL4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL5	0.490	0.339	0.885	1.000	0.850	0.004	0.239
SOIL	0.449	0.412	0.483	0.500	0.485	0.272	0.348
Principal Components							
PC1	0.68	-1.01	-0.01	0.45	0.74	-1.75	-1 46
PC2	-0.88	-0.55	-0.22	0 26	0.65	-0.94	1 11
PC3	2.13	2.79	1 75	2 12	2 19	1.95	2 75
Cluster Membership							
CLUSTER2	1						
CLUSTER3	1	1					
Canonical Variables for Two Clusters							
Base Case							
CAN1	2 194	0 772	3.180	2.675	1.613	-1.483	0.516
With BFI							
CAN1	1 880	0.308	2 798	3.078	1.053	-1.439	0 957
With QBAR							
CAN1	2 127	0.574	3.203	2.432	1.924	-1.369	0.320
Canonical Variables for Three Clusters							
Base Case							
CAN1	1.480	0.766	3.502	2.495	2.408	-0.394	0.003
CAN2	-0.970	-0.354	-0.259	-0.120	-1 548	2 136	0.235
With BFI							
CAN1	-1.269	-0.166	-2.056	-2.288	-2.059	1.740	-0.386
CAN2	0.635	0.623	2.635	1.484	1.046	1.143	-0.115
With QBAR							
CAN1	1.518	0.545	3.540	2.161	3.040	-0.553	-0.299
CAN2	-0 656	-0.021	0.252	0.521	-1.493	1.919	0.443

VARIABLE	CATCHMENT NUMBER						
	61001	61003	64001	65001	66011	67003	67008
Flow Response Measures							
Tp	7.09	5.46	4.86	4.19	4.22	5.28	7.06
Qp	25.6	38.0	25.6	27.9	41.0	42.1	17.5
SPR	25.2	40.5	48.3	30.7	57.7	74.3	18.3
RBAR	0.62	0.55	0.52	0.43	0.55	0.33	0.47
BFI	0.65	0.57	0.36	0.31	0.28	0.50	0.56
QBAR	0.257	0.535	0.650	1.228	1 083	0.579	0.114
CV	0.26	0.25	0.19	0 26	0.21	0.43	0.37
Catchment Characteristics							
AREA	197.6	31.3	471.3	68.6	344.5	20.2	227.1
MSL	27.6	9.4	37.5	15.2	29.0	7.1	45.8
DVF	0.00	0.03	0.01	0.01	0.00	0.07	0.00
SL1085	3.2	25.5	5.2	33.6	17.2	13.8	5.0
STMFRQ	0.89	1.85	2.70	5.93	4.18	1.83	0.93
SAAR	1282.	1465.	1836.	3030.	2162.	1300.	901.
LAKE	0.00	0.00	0.00	0.26	0.08	0.00	0.01
URBAN	0.00	0.00	0.00	0.00	0.00	0.00	0.04
SOIL1	0.000	0.000	0.000	0.000	0.000	0.000	0.245
SOIL2	0.992	0.911	0.500	0.099	0.493	0.000	0.594
SOIL3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
SOIL4	0.000	0.000	0.000	0.000	0.000	0.000	0.045
SOIL5	0.008	0.089	0.500	0.901	0.507	1.000	0.116
SOIL	0.302	0.318	0.400	0.480	0.401	0.500	0 293
Principal Components							
PC1	-2.72	-1.05	-0.37	0.36	1 18	0.33	-2.95
PC2	-0.74	-0.27	1.14	0.18	1.28	1 41	-0.54
PC3	2 97	2 65	2.66	2.03	2.79	0.66	1.61
Cluster Membership							
CLUSTER2	2						
CLUSTER3	2	1	1				
Canonical Variables for Two Clusters							
Base Case							
CAN1	-0.211	-0.233	1.248	3.120	0.999	1.858	-1.854
With BFI							
CAN1	-0.698	-0.459	1.683	3.038	1.708	1.727	-1.552
With QBAR							
CAN1	-0.293	-0.340	1.111	2.601	1.191	1.733	-1.920
Canonical Variables for Three Clusters							
Base Case							
CAN1	-0.685	0.420	1 025	3.099	1.623	2.055	-0.685
CAN2	1.138	1.532	-0.318	-1.108	-0.135	0.265	1.621
With BFI							
CAN1	1.961	1.119	-1.488	-2.793	-2.148	-1.021	1.214
CAN2	0.619	1.507	0.248	1.559	0.616	1.774	0.400
With QBAR							
CAN1	-0 962	0.046	0.879	2.539	1.888	1.827	-1 005
CAN2	1.106	1.684	-0.012	-0.068	-0.093	0.707	1.559

VARIABLE	CATCHMENT NUMBER					
	68006	69027	71003	71004	72002	77002
Flow Response Measures						
Tp	5.62	7.44	2.87	5.16	5.69	5.40
Qp	45.6	31.3	59.9	30.0	27.8	42.3
SPR	43.2	40.2	54.1	39.7	59.2	51.1
RBAR	0.37	0.33	0.43	0.72	0.45	0.54
BFI	0.55	0.56	0.35	0.43	0.32	0.38
QBAR	0 393	0 570	1.351	0.558	0.536	0.803
CV	0.42	0.33	0.44	0 59	0.19	0.24
Catchment Characteristics						
AREA	150.0	150.0	10.4	316.0	275.0	495.0
MSL	30.9	41.4	5.2	37 1	34.2	53.4
DVF	0.00	0.00	0.00	0.00	0.00	0.01
SL1085	10.0	5.6	37.8	5.0	7.7	3.7
STMFRQ	1.42	0.72	3.36	0.89	1.00	2.11
SAAR	1053.	1181	1786.	1211.	1251.	1507.
LAKE	0.04	0.17	0.00	0.08	0.00	0.04
URBAN	0.02	0.22	0.00	0.09	0.01	0.00
SOIL1	0.033	0.000	0.000	0.000	0.000	0.000
SOIL2	0.022	0.009	0.000	0.000	0.076	0.000
SOIL3	0.000	0.000	0.000	0.000	0.000	0.388
SOIL4	0.598	0.343	0.065	0.541	0.502	0.000
SOIL5	0.347	0.648	0.935	0.459	0.413	0.609
SOIL	0.454	0.481	0.497	0.473	0.455	0.460
Principal Components						
PC1	-0.74	-1.16	2.16	-1.50	-0.10	0 24
PC2	-0.28	0 13	-0.07	0 19	1.89	0.77
PC3	0.91	1.15	1 31	2.08	2.16	2 60
Cluster Membership						
CLUSTER2	1					
CLUSTER3	'	1				
Canonical Variables for Two Clusters						
Base Case						
CAN1	0.226	1.422	2.170	1 212	1 223	1 198
With BFI						
CAN1	-0.489	1.079	2.315	1 266	1.647	1 121
With QBAR						
CAN1	0.173	1.570	2.546	1 308	1.180	1 462
Canonical Variables for Three Clusters						
Base Case						
CAN1	-0.109	0.850	2.769	0 152	0.951	0.752
CAN2	-1.021	-0.484	-0.196	-1.218	-0.155	-1.436
With BFI						
CAN1	0.273	-0.453	-2.221	-1.029	-1.301	-1.447
CAN2	-0.310	0.526	1.787	-0 747	0.315	-0.386
With QBAR						
CAN1	-0.028	1 109	3.280	0.457	0.905	1 313
CAN2	-0.965	-0.502	-0.175	-1 274	0.039	-1.580

VARIABLE	CATCHMENT NUMBER	
	84008	84012
Flow Response Measures		
Tp	3.87	6.07
Qp	34.6	32.9
SPR	57.8	56.7
RBAR	0.41	0.39
BFI	0.32	0.36
QBAR	0.654	0.523
CV	0.36	0.24
Catchment Characteristics		
AREA	51.3	227.2
MSL	18.9	61.2
DVF	0.04	0.00
SL1085	13.4	6.6
STMFRQ	1.05	1.06
SAAR	1187.	1276.
LAKE	0.00	0.12
URBAN	0.26	0.27
SOIL1	0.000	0.000
SOIL2	0.000	0.000
SOIL3	0.098	0.340
SOIL4	0.738	0.469
SOIL5	0.164	0.191
SOIL	0.453	0.443
Principal Components		
PC1	0.45	-0.10
PC2	1.27	1.52
PC3	1 29	1 67
Cluster Membership		
CLUSTER2	1	
CLUSTER3	1	
Canonical Variables for Two Clusters		
Base Case		
CAN1	1.000	1.010
With BFI		
CAN1	1.665	1.240
With QBAR		
CAN1	1.079	1.008
Canonical Variables for Three Clusters		
Base Case		
CAN1	0.681	-0.103
CAN2	-0.741	-2.095
With BFI		
CAN1	-1.866	-1.734
CAN2	-0.447	-1.659
With QBAR		
CAN1	0.890	0.203
CAN2	-0.709	-2.073

Appendix B Guide to application of methods

This appendix details the procedures to be followed in order to apply the methodology outlined in this report to a catchment for which estimates of Tp and SPR are required. The basis for the estimation of the rainfall-runoff model parameters is the data for the catchments used in this work. The appropriate information for the catchments is contained in Appendix A.

Step 1 *Assemble catchment characteristics*

Assemble catchment characteristic data for the catchment of interest and compare with the range of values for the various catchment characteristics used herein. The maximum and minimum values for each catchment characteristic are summarized in Table 1. Caution should be exercised if the new catchment of interest is substantially different from the catchments in the data base since the methodology involves interpolating between runoff event parameter values for the gauged catchments. Estimates from this procedure may therefore be unreliable for catchments that are substantively different from the catchments in the data base.

Step 2 *Standardize the variables*

Standardize the catchment characteristic data for the new catchment using the relationship in Equation (3), and the mean and standard deviation for the catchment characteristics presented in Table 1.

Estimation of Tp

To estimate Tp in the ungauged case, work through steps 3 to 5. If value of QBAR can be derived from data observed at the site of interest, then jump to step 6.

Step 3 *Calculate canonical variables*

Calculate the two canonical variables for the catchment using the standardized catchment characteristic vector with Equation (2) and the weights summarized in Table 6.

Step 4 *Identify the three nearest catchments*

Identify the three nearest catchments to the new catchment in terms of the two canonical variables. Determine the probability of membership of the catchment in each of the three clusters using Equation (6). The cluster membership of each catchment, which is required for this calculation, is given in Appendix A.

Step 5 *Estimate Tp*

Estimate Tp using the cluster specific estimates of Tp from the relationships in Equations (21) to (23). The final estimate is obtained by combining these three estimates using the probability of membership of the catchment in each of the three clusters and Equation (9).

Step 6 *Estimating Tp with derived QBAR value*

Use the standardized characteristic vector, augmented with the standardized QBAR value, to calculate two canonical variables for the catchment with Equation (2) and the weights in Table 12. Use these two canonical variables in the above procedure starting at step 4.

Estimation of SPR

If an estimate of BFI is available, go to Step 10.

Step 7 *Calculate canonical variable*

Calculate a single canonical variable for the catchment using Equation (2) and the weightings in Table 5.

Step 8 *Identify five nearest catchments*

Using the single canonical variable, identify the nearest five catchments and determine the probability of membership of the catchment in cluster 1 and cluster 2 using Equation (6).

Step 9 *Find nearest two catchments in cluster and calculate mean SPR*

Estimate SPR as the arithmetic mean of the SPR values for the two nearest neighbour catchments from the cluster for which the catchment has the largest probability of membership. Catchment proximity is measured as the weighted Euclidean distance in terms of the five (standardized) soil class variables.

The following steps are for estimating SPR when a value of BFI is available.

Step 10 *Calculate new canonical variable*

Use the standardized characteristic vector, augmented with the standardized BFI value, to calculate the canonical variable for the catchment with Equation (2) and the weights in Table 11.

Step 11 *Identify seven nearest catchments*

Using the single canonical variable, identify the nearest seven catchments and determine the probability of membership of the catchment in cluster 1 and cluster 2 using Equation (6).

Step 12 *Use SPR from nearest neighbour in cluster*

Estimate the SPR value as the SPR for the catchment that is nearest to the catchment from those catchments that are in the cluster for which the new catchment has the largest probability of membership. Catchment proximity, for identifying the nearest neighbour, is measured as the weighted Euclidean distance in terms of the variables SOIL and BFI.

Example: Eden at Kirkby Stephen

Step 1 *Assemble catchment characteristics*

The catchment characteristics are listed below. They are all within the range of the characteristics given in Table 1.

Step 2 *Standardize the variables*

	Value	Mean	Std dev	Standardized
MSL	20.77	24.958	15.555	-0.26924
DVF	0.02	0.0498	0.0811	-0.36745
SL1085	18.47	9.296	10.8408	0.84625
STMFRQ	4.03	1.3439	1.121	2.39616
SAAR	1439	1154.62	504.72	0.56344
LAKE	0	0.0259	0.0554	-0.46751
URBAN	0	0.0516	0.1121	-0.4603
SOIL1	0	0.1154	0.2106	-0.54796
SOIL2	0	0.1706	0.2942	-0.57988
SOIL3	0	0.1475	0.2722	-0.54188
SOIL4	0.16	0.2963	0.3537	-0.3835
SOIL5	0.84	0.2700	0.3595	1.58554
SOIL	0.492	0.3958	0.0787	1.2224

Estimation of Tp

Step 3 *Calculate canonical variables*

	Standardized	Canonical weights	
MSL	-0.26924	-0.0697	-0.4442
DVF	-0.36745	0.1237	-0.2576
SL1085	0.84625	0.2550	0.1507
STMFRQ	2.39616	0.1389	-0.3938
SAAR	0.56344	0.3066	-0.1749

LAKE	-0.46751	-0.1500	0.0253
URBAN	-0.4603	0.4058	0.1661
SOIL1	-0.54796	-8.914	11.2646
SOIL2	-0.57988	0.3231	53.6559
SOIL3	-0.54188	7.9474	72.9260
SOIL4	-0.3835	15.3645	109.6562
SOIL5	1.58554	-21.6675	128.0141
SOIL	1.2224	-22.4602	-68.8663
Canonical variables		1.9479	-1.0576

Step 4 *Identify the three nearest catchments*

The following five catchments have canonical variables fairly close to the values obtained in step 3.

	CV1	CV2	Euclidian distance	Cluster
Target	1.9479	-1.0576		
23005	1.6810	-0.6402	0.495	
49003	2.1678	-0.3134	0.776	
57004	1.3932	-0.8482	0.592	3
57006	1.4804	-0.9696	0.476	1
58008	2.4083	-1.5481	0.673	

The cluster membership for the nearest three is shown in the last column.

Prior probabilities of being in clusters 1 and 3 are 34/99 and 36/99 respectively.

Cluster	Probability of membership		
1	2x(34/99)/[2x(34/99)+1x(36/99)]	=	0.6538
0	0	=	0.
3	1x(36/99)/[2x(34/99)+1x(36/99)]	=	0.3462

Step 5 *Estimate Tp*

The regression equations for the three clusters are:

Cluster	Estimated Tp
1	4.86
2	6.01
3	3.86

Weighting these using the probabilities of cluster membership given above gives a best estimate of 4.525 hours.

Step 6 *Estimating Tp with derived QBAR value*

Standardize QBAR

QBAR value .791 Standardized value 2.064

Calculate canonical variable

	Standardized	Canonical weights	
MSL	-0.26924	0.0010	-0.4508
DVF	-0.36745	0.1494	-0.2256
SL1085	0.84625	0.0519	0.3283
STMFRQ	2.39616	0.2307	0.3938
SAAR	0.56344	-0.2504	0.3344
LAKE	-0.46751	-0.0952	0.0432
URBAN	-0.4603	0.3304	0.2623
SOIL1	-0.54796	-8.5478	8.1154
SOIL2	-0.57988	-3.6119	49.638
SOIL3	-0.54188	1.4239	69.481
SOIL4	-0.3835	5.0528	105.37
SOIL5	1.58554	9.1138	123.95
SOIL	1.2224	-14.127	-69.495
QBAR	2.064	0.7979	-0.324
Canonical variables		3.1814	-1.5862

Find the three nearest neighbours and their cluster membership

Catchment	Canonical variables		Cluster
25003	3.724	-0.162	1
58008	3.04	-1.493	1
71003	3.28	-0.175	1

The catchment is unambiguously assigned to cluster 1.

From Step 5 the Tp estimate for the catchment using the cluster 1 regression equation is 4.86 hours.

Estimation of SPR

Step 7 *Calculate canonical variable*

	Standardized	Canonical weight
MSL	-0.26924	-0.1922
DVF	-0.36745	0.0774
SL1085	0.84625	-0.2422
STMFRQ	2.39616	-0.1292
SAAR	0.56344	0.9152
LAKE	-0.46751	-0.0761
URBAN	-0.4603	0.3104
SOIL1	-0.54796	-44.2259
SOIL2	-0.57988	-105.603
SOIL3	-0.54188	-125.603
SOIL4	-0.3835	-179.871
SOIL5	1.58554	-200.596
SOIL	1.2224	79.2414
Canonical variables		1.3504

Step 8 *Identify five nearest catchments*

	CV1	Euclidian distance	Cluster
Target	1.3504		
48004	1.2967	0.0537	2
57005	1.3340	0.0164	1
64001	1.2475	0.1029	1
69027	1.4223	0.0719	1
72002	1.2230	0.1274	.

The cluster membership for these catchments is given in the last column.

Prior probabilities of being in clusters 1 and 2 are 51/99 and 48/99 respectively.

Cluster	Probability of membership		
1	4x(51/99)/[4x(51/99) + 1x(48/99)]	=	0.8095
2	1x(48/99)/[4x(51/99) + 1x(49/99)]	=	0.1905

The catchment is therefore assigned to cluster 1

Step 9　　*Find nearest two catchments in cluster and calculate mean SPR*

Catchment	SOIL1	SOIL2	SOIL3	SOIL4	SOIL5	Weighted Euclidian distance
Target	0	0	0	0.16	0.84	
71003	0	0	0	0.065	0.935	0.462
23005	0	0	0	0.0	1.0	0.778

Remember that the distance is calculated using the standardized variables and the weighting factors for the five SOIL values are 0.50, 0.95, 0.90, 1.75, and 1.25 respectively. Catchment 71003 is the closest but there are seven other catchments in the data set that are in cluster 1 and have 100% SOIL type 5.

Catchment	SPR
23005	53.8
25003	64.8
28033	24.4
46005	58.3
54022	36.7
55008	43.7
58006	44.5
67003	74.3
71003	54.1

The recommendation is to use the average of the nearest two catchments, and while 71003 must be used there is a dilemma about which other catchment to use. This could be resolved by taking nearby catchments (23005 & 25003) which would give an estimate of 57.6%. The average of all the values is 50.5%.

Refinement of SPR estimate using BFI

Step 10　　*Calculate new canonical variable*

BFI value for catchment 0.24 (Mean 0.49, sd 0.14 hence normalized value -1.786)

Canonical variable is 2.070

Step 11　　*Identify seven nearest catchments*

Catchment	Canonical variable	Distance	Cluster
Target	2.070		
39017	1.804	0.266	

(table continues on page 71)

54022	2.352	0.282
57004	1.729	0.341
57006	1.880	0.190
71003	2.315	0.245
64001	1.683	0.388
72002	1.647	0.423

The catchment is therefore assigned to cluster 1.

Step 12 *Use SPR from nearest neighbour in cluster*

Catchment	BFI	SOIL	Weighted Euclidian distance	SPR
Target	0.24	0.492		
23005	0.27	0.5	.250	53.8
65001	0.31	0.48	.536	30.7
25003	0.15	0.5	.656	64.8
38017	0.34	0.483	.729	30.4

Remember that the distance is calculated using the standardized variables and the weighting factors for BFI and SOIL are 1.0 and 1.6 respectively. Catchment 23005 is the closest and its value of 53.8% is taken as the estimate of SPR. The target catchment has a lower value of BFI than this closest catchment, which would hint at a slightly higher SPR.

Summary

Time to peak

Estimate from FSSR16	4.78 hours
From above	4.53 hours
From above with QBAR	4.86 hours
Derived from event data	3.84 hours

Standard Percentage Runoff

Estimate from FSSR16	52.0 %
From above	55.0 %
FSSR16 estimate with BFI	56.0 %
From above with BFI	53.8 %
Derived from event data	66.8 %