International Seminar on Hydrology of Extremes

1-3 December, 1988
Roorkee (India)

PRELIMINARY ANALYSES OF LOW FLOW CHARACTERISTICS OF UNGAUGED CATCHMENTS

R. J. Nathan K. C. Gan N. Jayasuriya

T. A. McMahon I. C. O'Neill

Department of Civil and Agricultural Engineering, University of Melbourne, Parkville, Victoria, Australia-3052.

SYNOPSIS

The ability to assess the low flow yield of rural catchments is fundamental to the efficient utilisation of Australia's scarce water resources. This paper describes a systems approach concerned with the investigation of low flow hydrology of small ungauged rural catchments. The study involves three main areas of research: estimation of yield from both catchments and reservoirs, and rainfall-runoff modelling.

1.0 INTRODUCTION

Much of the development of Australia's agricultural and energy-related industries is dependent on the efficient utilisation of our scarce water resources. Not only is Australia the driest continent, but its streams are among the most variable in the world (McMahon, 1982). Consequently, there is a fundamental need to develop methods that can readily be used to assess the hydrology during times of low flow. Where streamflow data exist, a variety of analytical methods can be used to analyse the low flow characteristics; the main problem faced by the hydrologist in this case is the selection of the technique most appropriate to the nature of the data and the task at hand. Ungauged catchments pose a more difficult problem. A lack of suitable data often means that design or environmental decisions are based on little or no hydrology. In the case of farm water supplies for instance, it is likely that application of a rational design method could result in smaller storages sizes, and hence savings in construction cost.

An extensive survey was carried out to determine the requirements for low flow data of such organisations such as water authorities, engineering consultants and environmental bodies (McMahon, 1984). The survey clearly showed the need for low flow and yield data in such fields as water supply design, licensing of stream abstractions, groundwater and environmental studies. To meet this need, a systems approach is currently being developed for quantitatively estimating low flows and reservoir sizes for small ungauged rural catchments. A team of up to five has been assembled for the project, and it is envisaged that the study will take about 12 person-years to complete. This paper describes the investigations currently being undertaken, and briefly presents the final results for the section of the study concerned with storage yield estimation.

2.0 OVERVIEW OF SYSTEMS APPROACH

The main objective of the study is the quantitative assessment of low flow and yield characteristics of small ungauged rural catchments. Given the broad scope of this project it is convenient to describe its various components within the overall framework of a systems approach.

Figure 1 illustrates schematically a number of possible approaches to the analysis of low flow in gauged catchments. Reservoir capacity-yield analyses can be conveniently divided into three main groups represented by critical period methods, probability matrix methods and stochastic data generation. These methods may be used to determine the relationship between inflow characteristics, reservoir capacity, controlled release and reliability. The low flow characteristics of the catchment can be described in a

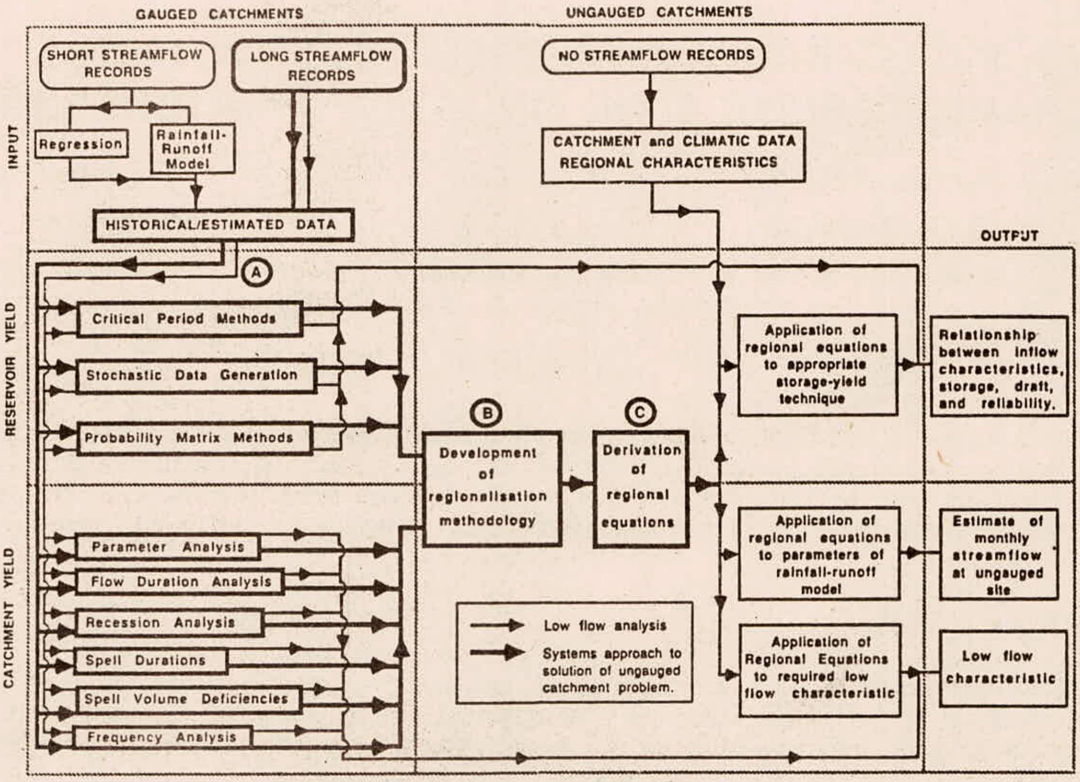

FIGURE 1. Schematic methodology of low flow analysis, showing systems approach to solution of ungauged catchment problem.

number of ways, ranging from a simple statement concerning the annual minimum n-day flow, to frequency analyses which allow estimation of the probability of occurrence of a drought of a specified magnitude. All these analyses require streamflow records of sufficient length to be representative of the long term variability of flow at the gauging station. If the historic record is too short, then the record length may be extended by regression with nearby station records or, if suitable rainfall data are available, by use of a rainfall-runoff model.

The analytical techniques mentioned above are obviously unsuited to catchments for which there are no available streamflow records. There are at least three possible solutions to the problem of ungauged catchments: (i) if rainfall records are available undertake short term stream gauging and extend the record using a rainfall-runoff model; (ii) directly transpose results from an adjacent catchment with similar characteristics; and (iii) use regional prediction equations. The size of projects normally associated with small catchments does not usually justify the time and costs involved with short term stream gauging. It is also unusual for there to be a nearby gauged catchment from which it is possible to confidently transpose low flow and yield results. Figure 1 illustrates the steps necessary in a regional approach, in which catchment and climatic data are used in lieu of streamflow records as inputs to regional prediction equations for the quantitative assessment of low flow and yield characteristics.

The problem thus arises as to the derivation of suitable regional prediction equations. Within the scope of the ungauged catchment problem defined by Figure 1, there are three main areas that require investigation: storage yield analysis, assessment of low flow characteristics, and the estimation of streamflow. To this end, a systems approach is being developed to derive the necessary prediction equations.

A systems approach to a problem requires three steps: (i) description of the system; (ii) definition of an objective function, and (iii) optimisation. The first step is defined in Figure 1 by the flow path A to B.

The analytical methods developed for investigating the low flow and yield characteristics of catchments with adequate streamflow data are applied to a large number of gauged catchments within the study area. The results of these analyses are represented in some non-dimensional form, either as low flow or model parameters, or else as some relationship between the low flow parameters and storage-yield characteristics of a regulated catchment. The objective function is defined by the development of a suitable regionalisation methodology (box B in Figure 1), in which it is desired to minimise the errors associated with low flow and yield predictions in ungauged catchments. The third step, optimisation, is represented by box C in Figure 1. Here the form and parameters of the prediction models within the various regions are adjusted until prediction errors are minimised.

The first two steps of the systems approach are briefly described in the following sections, and the results of optimising the storage-yield techniques are also presented. Development of prediction models for the estimation of streamflow and low flow characteristics remain the object of further research, and are consequently omitted from discussion.

3.0 REGIONALISATION METHODOLOGY

3.1 Data Selection

The study area chosen for the project is located in south-eastern Australia, encompassing 17 drainage divisions and two federal states (Figure 2). The two northern-most regions not contiguous with the rest of the study area were chosen so as to overlap with other regionalisation and landscape classification procedures currently under investigation (Laut et al, 1984). The prediction equations for storage-yield analysis were developed using monthly streamflow data derived from 9 coastal drainage divisions in the south of Victoria, though ultimately these equations will be tested on the whole of the study area.

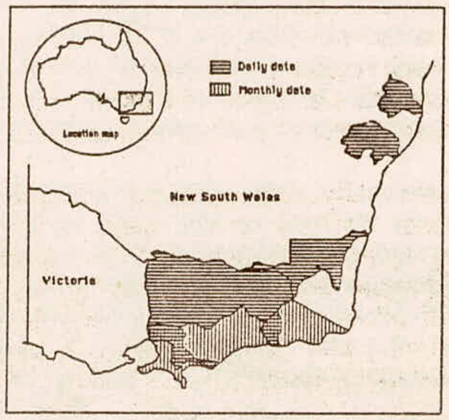

FIGURE 2. Location of study areas.

Daily streamflow data were obtained for a total of 222 gauging stations distributed throughout the drainage divisions indicated in Figure 2, and monthly streamflow data for 81 stations. Each of the selected stations fulfilled a number of selection criteria related to catchment size, data accessibility, length and quality of record, and effect of artificial influences. The streamflow records were then subject to detailed quality control which resulted in the rejection of some station records - the final data set represents a total of about 4000 years of daily streamflow data, and 930 years of monthly data, or in other words, a total of around 1.5 million flow values. Another data base of similar size has also been established for the processing of rainfall and climate data.

In developing the regional prediction equations it was decided to impose the constraint that all input data must be easily obtainable from published maps, tables etc. - thus low flow and yield estimates may be obtained without the need for field reconnaissance. To this end, a further data base has been compiled from 1:100 000 topographic maps consisting of a range of physiographic characteristics, including information on catchment area, slope, elevation, fraction of area covered by forest, or grassland, etc., and it is intended that geological indices will be abstracted from 1:250000 geological maps.

3.2 Determination of Homogeneous Regions

Before developing prediction equations it is desirable to divide the study area into homogeneous sub-regions that can be considered to behave in a similar fashion. Application of prediction equations to homogeneous sub-regions is likely to aid model development and improve the predictive ability of the final equations. Mosley (1981) discusses the problems associated with regionalisation, and notes that the theory and practice of its application has received the attention of many workers in the geographical sciences, but no general methodology for identifying regions is available. Statistical methods have been used to test for homogeneity of hydrologic response within regional groupings, though the initial choice of the regional groupings has been generally either highly subjective or based on irrelevant considerations, such as administrative boundaries; for example, the Natural Environment Research Council (NERC, 1975) commenced its regionalisation with groupings of hydrometric areas.

Many investigators have defined homogeneous subregions using residuals from an overall regression equation (Tasker, 1982a; NERC, 1975), though the large subjective element in such an approach means that different hydrologists would obtain different groupings, or that a different data set (in time or space) would have yielded different regions.

Cluster analysis can be used to objectively identify homogeneous hydrological regions based on catchment characteristics (Hawley and McCuen, 1982; Acreman and Sinclair, 1986) or hydrological characteristics (Hughes, 1987; Haines et al, 1988). Assuming that the distribution of basin characteristics for each region is multivariate normal, an ungauged site can be allocated to a particular grouping using discriminant analysis, or alternatively, a site may be classified by developing a decision tree based on the characteristics of each grouping (Haines et al, 1988).

Traditionally, regionalisation groupings encompassed geographically contiguous areas, however, subregions defined on the basis of similarity of hydrologic or basin characteristics may not have geographical significance. With classification techniques such as cluster analysis, it is not necessary for catchments within a given grouping to be geographically contiguous. Such regions would exist in a multi-dimensional predictor variable data space rather than in geographical space. Tasker (1982b) and Acreman and Sinclair (1986) have successfully applied cluster analyses to both catchment and hydrological characteristics without assuming the groupings to be geographically contiguous.

Other techniques have been used to identify homogeneous regions such as principal components (Blake et al, 1970), principal coordinate analysis (Laut et al, 1984), and the optimisation of statistics that describe the efficiency of the groupings (Wiltshire, 1986). All these techniques have computational disadvantages and consequently have been rejected in favour of cluster analysis.

3.3 Prediction Model

Selection of an appropriate prediction model for application within each homogeneous region also requires careful consideration. Methods successfully used in regionalisation studies can be loosely classified as either multiple regression or multivariate techniques.

Multiple regression has been used for many years as a predictive tool in hydrology. It permits evaluation of the combined effects of many parameters on a dependent variable, and is relatively straightforward to apply. The parameters of a multiple regression model are estimated by the method of least squares. However, the least squares estimate is based on a number of assumptions which are often violated in hydrology. For instance it is assumed that the independent variables are fixed variates and are measured without error, that they are random events exhibiting negligible serial correlation, that they are not correlated with one another, and that they are homoscedastic. Sharp et al (1960), Snyder (1962), and Wallis (1965) discuss different aspects of these assumptions when applying multiple regression techniques to hydrologic data. In the application of tests of significance, it is further assumed that the residuals are normally distributed with zero mean and homoscedastic variance. These errors tend to bias the parameter estimates and may result in high values of the correlation coefficient, or misleading levels of significance. In practice, such errors tend to hamper attempts to identify important variables or correct model structure.

Stepwise regression techniques are commonly used in hydrology for the selection of regressors in the model. These procedures should however be used with caution (Kendall, 1980; Weisberg, 1985) - while stepwise methods are easily understood and widely used, they may not give the same answer and, if they do, that answer may not be the optimum. Stepwise regression techniques tend to pick variables that confound several independent effects and to build models that are hard to interpret in terms of the real world. Sharp et al (1960) illustrate the possible inconsistencies in stepwise regression techniques in a study of parameters affecting water yields. The use of residual or partial residual plots are especially useful in determining the correct variable specification or transformation, diagnostic examples of which may be found in Weisberg (1985).

Biased regression techniques such as ridge regression, shrunken estimators, regression on principle components and latent root estimators are gaining increasing popularity as a means to avoid the problems associated with inter-correlated data. A number of comparative studies of biased estimators have been undertaken (for example see De Coursey and Deal, 1974, and Gunst and Mason, 1977), though there appears to be a certain amount of controversy surrounding their use, especially those techniques used less widely than ridge regression or regression of principal components.

While prediction is the primary purpose of a regression equation, multivariate analysis can also be used to investigate the structure of dependence among variates. Multivariate analysis includes such techniques as cluster analysis, principal component analysis (PCA), principal co-ordinate analysis and factor analysis. The objective of principal components is to transform a set of correlated variables into a new set of uncorrelated, or orthogonal components. These new components are a linear combination of the original variables and are derived in decreasing order of importance so that, for example, the first principal component accounts for as much as possible of the variation in the original data. The main benefits of PCA lie in its ability to reduce the dimensionality of a problem and to more clearly reveal underlying relations and influences between the selected variables (Shelton and Sewell, 1969). It is also possible to use the principal components in regression equations, which has several desirable properties over those of conventional multiple regression techniques, including more stable regression coefficients that clearly reflect the independent contribution of each term in the equation, and the equations tend to yield more reliable results when used with different populations (Wallis, 1965). It is worth noting that there is little advantage in using PCA if the original data are already nearly uncorrelated, or else if the initial components are not rotated (optimised).

Although factor analysis has been purported to have been used in a number of hydrological studies (for example Abrahams, 1972; White, 1975; Dawdy and Feth, 1967) the writers have in fact been unable to find a single application of its use - confusion arises as factor analysis terminology is often applied to principal component analysis, even though the two statistical models are quite distinct. There are doubts as to the practical usefulness of factor analysis (Matalas and Reiher, 1967) and there seems little to recommend its use above that of PCA.

Principal coordinate analysis is a technique commonly used in numerical taxonomy as an alternative to cluster analysis, though recently it has found applications in hydrology for both regionalisation (Hughes, 1987) and prediction (Laut et al, 1980). Again, there appears little point in using the method of principal coordinate analysis in preference to PCA as the latter requires significantly less computational effort, is as efficacious in classification, and has the added advantage of being able to be interpreted physically in terms of the independent predictor variables.

At this stage it appears that PCA is best suited for examination of the underlying structure and relative importance of the variates, and perhaps also for prediction purposes in ungauged catchments. If, however, the data are near-orthogonal and the variables require transformation to allow for non-linear or heteroscedastic effects, then stepwise regression procedures will be adopted, and techniques such as residual and partial-residual plots will be used to investigate the required structure of the model.

4.0 LOW FLOW CHARACTERISTICS

In low flow analysis it is important to consider both the magnitude and the duration of the event. Many low flow analyses have been developed over the years for investigating different aspects of the low flow regime. The low flow analyses selected for this study include straightforward parameter analysis, flow duration analysis, low flow spells, frequency analysis, base flow and recession analyses. Particular levels or parameters are chosen from these analyses in order to relate them to catchment characteristics - these are termed "low flow indices", as used by the Institute of Hydrology (1980). The following sections briefly describe the low flow analyses evaluated to date - in all cases the procedures were developed for automated processing on a Macintosh SE microcomputer.

4.1 Parameter Analysis

A number of statistical parameters were calculated for each station in the data base. The parameters selected provide information on the central tendency, dispersion and serial correlation of the streamflow data, and were evaluated for daily, monthly, and annual time steps. The descriptive statistics used in the analysis are: the minimum, maximum, mean, standard error of the mean, standard deviation, standard error of the standard deviation, coefficient of variation, coefficient of skewness, and the lag-one serial correlation. These parameters are all commonly used and do not warrant further discussion.

4.2 Flow Duration Analysis

Flow duration curves show the relationship between any given discharge and the percentage of time that the discharge is exceeded. The curve is a very simple, but useful device for illustrating the flow characteristics of a stream throughout the range of discharge, without regard to the sequence of occurrence. Flow duration analysis was undertaken for all streamflow stations contained in the project data base. Daily streamflow data were input to the program and were summed to provide complete series of monthly and annual data. The daily, monthly and annual data were ranked into descending numerical order and flow values corresponding to exceedence probabilities of 1%, 2%, 5%, 10%, 20%, 30%, ... , 80%, 90%, 95%, 98%, and 99% were then extracted. The flow duration curve was standardised by representing each value as a percentage of the mean annual flow, and a non-dimensional indication of the shape of the curve was obtained by dividing each ordinate by the 90% exceedence value. An example of the graphical output is shown in Figure 3.

4.3 Frequency Analysis

Flow frequency curves allows the estimation of the probability of occurrence of a flow event of specified magnitude. In this study only the annual series is considered, where the curves are based on the minimum flow event in each year of record. The two most commonly used approaches to frequency analysis that are suited to automated processing are data transformation methods and probability distribution methods. Examples of data transformations used in low flow analysis include the SMEMAX geometric transformation (Prakash, 1981) and the Box-Cox transformation (Kumar and Devi, 1982). Although data transformations are attractive in principle, it appears that their use does not always guarantee the desired result, especially if the original data is highly skewed or variable - consequently the more traditional approach of fitting a probability distribution to the data was adopted.

The choice of which particular distribution to use required careful consideration. It is generally accepted that no one distribution will suit all applications, and that the data under consideration should dictate use of a particular distribution. Ideally, the applicability of several distributions should be evaluated using some objective goodness-of-fit criteria, however application of different distributions to different sites is obviously an impractical approach for regional analysis, and it was therefore necessary to select a single distribution for use on all catchments. A number of comparative studies have been undertaken to determine the suitability of various distributions to the analysis of low flows (for example Matalas, 1963;

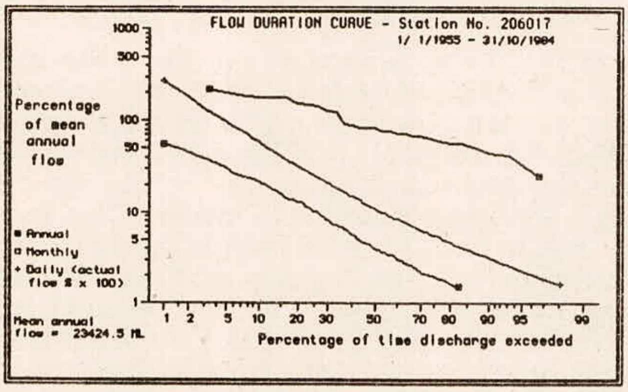

FIGURE 3. Flow duration curves.

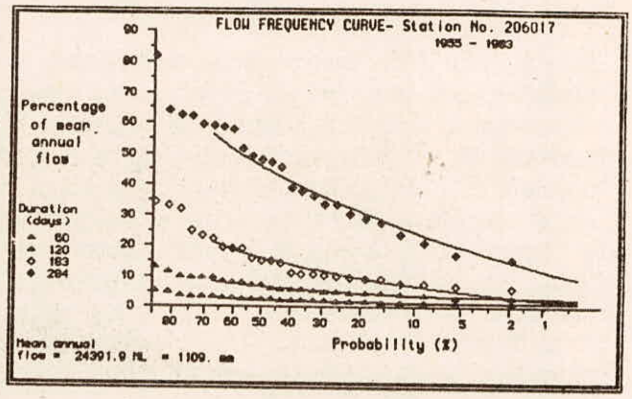

FIGURE 4. Probability plot of annual n-day minima.

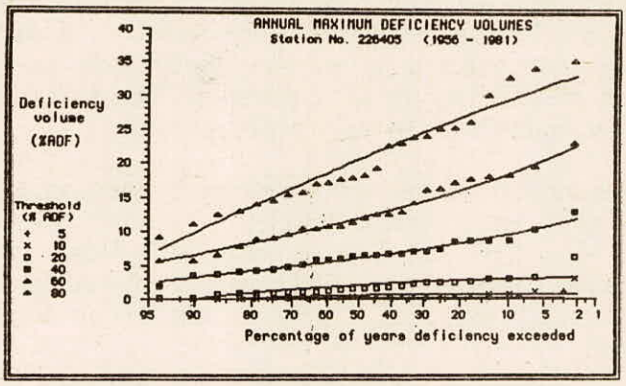

FIGURE 5. Probability plot of annual maximum deficiency volumes.

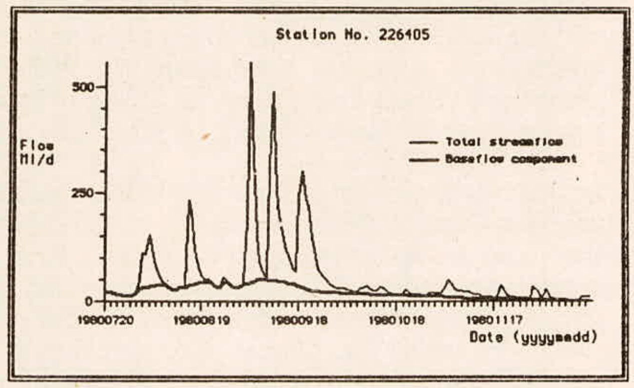

FIGURE 6. Illustration of base flow separation.

Eratakulan, 1970; Zucchini and Adamson, 1984). Unfortunately, the results of these studies are not directly comparable as the authors have applied different forms of the distributions to different types of populations. Inspection of the data set revealed a large range of skew values and non-zero minimum flows - accordingly, on the basis of the characteristics of the potentially suitable distributions, the log-Pearson Type III and Weibull distributions appeared to be equally suitable, though the latter was adopted for it avoids the necessity of transforming the original data into the logarithmic domain.

A two- or three-parameter form of the Weibull distribution was fitted to the 1, 7, 15, 60, 120 and 284 day minima for all stations in the data set that contained an adequate length of record. The distributions were fitted by either the method of maximum likelihood or the method of moments. Maximum likelihood estimates, if they existed, were adopted in preference to the moment estimates because of their greater efficiency (Dubey, 1967). Approximately one third of the data set included streams that are intermittent - in such cases the distribution was fitted to the non-zero data and a joint/conditional probability function was then used to factor the resultant probabilities (Condie, 1979). Also, examination of a number of preliminary results revealed that the first few (most frequent) values often described a much steeper curve. This characteristic has been noted by a other authors (Velz and Gannon, 1953, Tollow, 1987) and may be assumed to be the demarcation between "normal" and "drought" flows. Accordingly, the distributions were only fitted to those flows with exceedence probabilities greater than 80%, and a correction based on a conditional probability function (Jennings and Benson, 1969) was then applied to the resultant probabilities.

In total, over 1000 distributions were fitted. Approximately 20% of the samples indicated a non-zero minimum and required the 3-parameter form of the distribution, and overall a maximum likelihood solution was found in about 80% of cases. Figure 4 illustrates an example of the graphical output produced by the frequency analysis program. In the Weibull distribution, the relationship between the variate and its associated probability is dependent upon the skewness of the sample - thus to obtain a linear probability plot it is necessary to alter the abscissa scale for each individual distribution. Plots such as that indicated in Figure 4 were generated merely in order to check the conformity of the data to the fitted probability distribution; linear plots of single distributions could also be generated if more detailed inspection was warranted. The plotting position formula used in the plots is the approximate form recommended by Cunnane (1978) - the exact formulae presented by Arnell et al (1986) was found to be impractical when applied to sample sizes of greater than about 15.

4.4 Low Flow Spells

While the flow duration curve gives the duration below any flow, it provides no information about the length of consecutive periods below this flow, or how large a deficit has been built up. Two streams with similar flow duration curves may exhibit very different low flow sequences: for instance a sluggish stream may spend the time below the 90% flow duration value in a few long spells, whereas a responsive stream may divide the time into many short spells below the same threshold. The different responses would have a significant effect on say, dilution requirements in water quality control.

Procedures for investigating the low flow spell characteristics of streams were developed by the Institute of Hydrology (1980). The analysis of frequency of spell durations and deficiency volumes below threshold values corresponding to 5%, 10%, 20%, 40%, 60% and 80% of mean daily flows are currently underway. The results are expressed in two ways: (i) frequency per 100 years of an event, and (ii) proportion of years in which a deficit duration or volume is exceeded. The former result is obtained simply by noting the number of occurrences in the available length of record and factoring the total to yield the equivalent frequency per 100 years of record. The latter result is obtained by fitting a probability distribution to the annual duration and deficit volume maxima. Initially, a log-normal distribution was fitted to the annual maxima as undertaken by the Institute of Hydrology (1980). However, for record lengths of around 30 years and over, it was apparent that the cumulative density curve asymptotically approached an upper limit, ie was bounded in the direction of the desired extreme. This characteristic ideally suits application of the 3-parameter Extreme Value Type III distribution, and consequently this distribution was adopted for analysis. If a variate x has an Extreme Value Type III distribution then -x has a Weibull distribution and thus, except for a few sign changes associated with some terms in the estimation equations, estimation of the Extreme Value Type III parameters was the same as that used in the low flow frequency analysis. A non-linear example of the volume deficit probability curves is shown in Figure 5.

4.5 Base Flow Analysis

A base flow index is being calculated for each catchment based on concepts developed by Lvovich (1972) and the Institute of Hydrology (1980). The daily flow record of each station is analysed to determine the annual ratios of baseflow volume to total streamflow volume. This index can thus be thought of as the proportion of runoff that is derived from stored sources, and has been found to be highly correlated with geology and the 95% flow duration value.

The procedure used to separate the base flow component is based upon a recursive digital filter commonly used in signal analysis and processing (Lyne and Hollick, 1979). The filter was of the simple form:

$$f_k = a f_{k-1} + \frac{(1+a)}{2} (y_k - y_{k-1}) \quad (1)$$

where f_k is the filtered quick response at the kth sampling instant, y_k is the original streamflow, and a is the filter parameter. The optimum value of the filter parameter was found to be 0.925, and the filter was passed twice forward and once backward over the data. The output of the filter was constrained so that the separated slow flow was not negative or greater than the original streamflow. An example of the filtered base flow response is given in Figure 6.

5.0 RESERVOIR YIELD

Several methodologies have been investigated with regard to the estimation of yields and storages for small, rural, ungauged catchments. These are reviewed and summarised below.

5.1 Regression of the Storage Matrix

For 71 catchments, storage sizes at various drafts and probabilities of failure were calculated from the available historical streamflow data using a behaviour analysis and Morton's (1983) method to calculate net reservoir evaporation. For each catchment, a storage matrix was constructed, giving storage sizes for 56 combination of drafts and probabilities of failure (see Figure 7). The technique of least squares regression was used to study the relationship between storage on the one hand, and a variety of catchment physiographic and rainfall parameters on the other. The data did not significantly violate the least squares assumptions discussed in Section 3.3, and thus a multivariate analysis was not considered. Several types of regressions were performed, defined as point, column, row or block regression depending on the way in which data were extracted from the storage matrices (see Figure 7). The block regression results are regarded as most practical. In this, the storage matrix was divided into blocks, and separate regression equations developed for each block which covered an unique range of drafts and probabilities of failure (see Figure 8). The regression equations contain as independent variables, the catchment area and one or more rainfall statistical parameters (mean annual rainfall, coefficient of variation and coefficient of skew of monthly rainfall) , as well as the percentage probability of failure and the draft as a percentage of the mean annual flow (see Table 1). The quality of the regressions vary from block to block, being better for blocks associated with higher drafts and lower probabilities of failure. However, in the independent testing of the method for 10 catchments, the prediction errors did not follow the expected pattern as observed for block regression quality, but instead, were randomly organised. The errors, compared to the storage sizes that would be obtained form a behaviour analysis of the historical streamflow data, can be as low as ±10 % and in the majority of cases, do not exceed ±50 %. The mean annual streamflow for a catchment is estimated from the following regression equation:

$$X_m = 5.4954 \times 10^{-6} A^{0.9916} R^{1.5480} \qquad (N=70,\ R^2=0.96,\ SE= -34\%, +54\%) \quad (2)$$

where X_m is the mean annual flow (million m^3), A the catchment area (km^2), R the mean annual rainfall (mm), N the number of data points in the regression, R^2 the coefficient of determination, and SE the standard error of regression. Estimates of mean annual streamflow are used to derive the specified draft in volumetric units.

5.2 Gould Gamma Method

Both the mean annual streamflow (X_m) and the annual coefficient of variation of flows (C_v) for an ungauged catchment are estimated by regression equations relating these to catchment area and rainfall statistical parameters. The two streamflow parameters estimated in this way are then used in the Gould

Gamma technique (Gould, 1964) for storage size (C) estimation to a required draft (D) and probability of failure (p%) as below, where Z_p is the standardised normal variate at p%, and d is the difference between the p percentile flow of the Gamma and Normal distributions.

$$C = \frac{Z_p^2}{4(1-D)} C_v^2 X_m - d\, C_v^2 X_m \tag{3}$$

This is essentially a theoretical formulation derived by a consideration of streamflow statistics, and as only annual streamflow data is used, the technique estimates only the carry-over storage requirement while ignoring the within-year storage requirement. X_m may be estimated by Equation 2, while C_v may be estimated by Equations 4 and 5.

$$C_v = 0.690\ A^{0.1421}\ R_{cv}^{0.4692c} \quad (N= 50,\ R^2= 0.53,\ SE = -29\%, +41\ \%) \quad (A \leq 50\ km^2) \tag{4}$$
$$C_v = 1022\ R_{cv}^{1.1559}\ R^{-0.8240} \quad (N= 25,\ R^2= 0.80,\ SE = -19\%, +24\%) \quad (A > 50\ km^2) \tag{5}$$

where A is the catchment area (km^2), R the mean annual rainfall (mm), and R_{cv} the coefficient of variation of monthly rainfall.

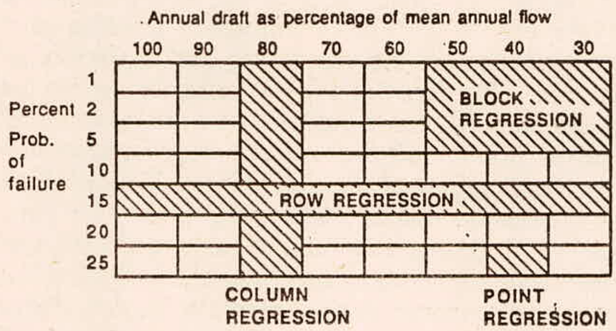

FIGURE 7. The storage matrix and types of regression.

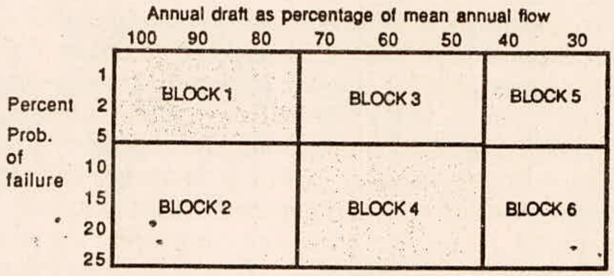

FIGURE 8. Division of storage matrix into blocks.

The errors in the estimated storage from this method depend greatly on the particular combination of errors in the estimated annual mean and coefficient of variation of flows. Positive errors in both (i.e. over-estimation compared to the historical values) result in a larger error in the storage than if both errors are negative. However, where the errors in the annual mean and coefficient of variation are of opposite signs, there is sometimes a compensating effect and the overall error in the storage may be small. Hence, the expected error arising from this method is highly unpredictable and may range from a few percent to over two hundred percent.

	Block Number					
	1	2	3	4	5	6
Draft (%)	100-80	100-80	70-50	70-50	40-30	40-30
P.Fail (%)	1-5	10-25	1-5	10-25	1-5	10-25
Num. Cases	639	852	639	852	426	568
R^2	0.97	0.95	0.93	0.81	0.78	0.51
Std. Error	0.194	0.250	0.307	0.522	0.525	0.811
Std. Err.(%)	-36,+56	-44,+78	-51,+103	-70,+232	-70,+234	-85,+547
F-statistic	3180	2444	1199	591	309	120
Variables	Regression Coefficients					
Log10 (A)	1.1063	1.0890	1.0866	1.0196	0.9104	0.6689
Log10 (PF)	-0.1462	-0.8038	-0.2292	-0.9779	-0.3754	-1.1857
Log10 (D)	3.3100	3.5129	2.4411	3.1886	2.3959	3.1058
Log10 (R_{cv})	1.1412	2.2367	1.9054	2.7287	1.4391	
Log10 (R)	1.2170	1.2872	1.1609	0.9691		-0.8591
Log10 (R_{cs})	0.0525	-0.1220	-0.1891	-0.4031	-0.3160	-0.5567
Constant	-10.5081	-10.2152	-8.4996	-8.3091	-4.7455	-2.6119

List of Symbols: A: catchment area (km^2); PF: probability of failure (%); D: annual draft as % of mean annual streamflow; R_{cv}: monthly rainfall coeff. of variation; R: mean annual catchment rainfall(mm); R_{cs}: monthly rainfall coeff. of skew.

TABLE 1. Block regresion results, with the division of the storage matrix as in Figure 8.

5.3 Transposition of Streamflow Data

The purpose of this exercise was to find a method to transpose monthly streamflow data from a gauged catchment to a neighbouring ungauged catchment so that the estimated historical flows can be used to calculate storage sizes using a conventional technique. The research aimed to find the best regression model to correlate the flow information of two neighbouring catchments with a view to regionalising the regression coefficients, or relating them to basic catchment physiographic and climatic data. Various forms of inter-catchment regressions were attempted to relate the concurrent monthly streamflow depths (X, Y) of two catchments, among which, were: linear regression through the origin Y=BX; linear regression with intercept Y=AX+C; linear regression through the origin using log-transformed data Log Y = B Log X; and linear regression through the origin using a two season division of the year. Finally, a Fourier series model was adopted, based on the assumption that the monthly streamflows of two neighbouring catchments can be described by a linear regression model through the origin, Y=BX. Thus, there is only one regression coefficient, B which is presumed to vary with the time of the year in a 12-month cycle for which the Fourier series of one harmonic gives an adequate fit. The three parameters of the Fourier series are related by regression to the catchment area and rainfall statistical parameters of an ungauged catchment. However, the regression coefficients tended to vary greatly from one catchment pair to another so that they were difficult to generalize for the region as a whole. No great success was found with any model, as all gave large errors in the individually transposed monthly flows. However, the Fourier model is recommended because of its sounder theoretical basis. Flows transposed by the Fourier model tend to be in error by about 40 - 50%, while monthly streamflow statistical parameters are reasonably well reproduced. Storage capacity calculated from the transposed flows are generally in error by not more than ±60 %, which is relatively small, probably because of compensating errors in the transposed flows.

5.4 Data Generation for Ungauged Catchments

The annual Markov model of lag 1 is used to generate annual flows for an ungauged catchment. The required inputs to the model are, the mean annual flow, the standard deviation of annual flows, the autocorrelation coefficient of annual flows, and the coefficient of skew of annual flows. The mean and the standard deviation can be estimated from regression equations relating these to the catchment area and the rainfall statistical parameters of an ungauged catchment, while for the autocorrelation coefficient and coefficient of skew, regional average values were used because of insufficient data available for regression. The annual flows thus generated are then disaggregated into monthly flows using sets of fragments varied according to the magnitude of the generated annual flow. The fragments were derived from a study of and averaging of observed fragments for real streams. Individual replicates of generated flows, each of a certain length, are then used to estimate the storage size for a particular draft and degree of reliability by which an average storage size can be calculated. However, the errors depend greatly on the accuracy of inputs to the generating model. It may be expected that good results can be obtained only with good inputs, however, storage errors are generally less than the sum of errors in the input parameters. For the catchments tested, they were generally not more than about ±60 %.

5.5 Comparison of Methodologies Investigated

All the methodologies investigated require some basic data, in particular, the catchment area and two or three catchment rainfall statistical parameters such as the mean annual rainfall, the standard deviation, coefficient of variation or coefficient of skew of monthly rainfall.

The use of block regression equations is a simple and rapid method of obtaining an estimate of the storage size for an ungauged catchment, requiring only a hand-held calculator. Likewise, the use of regression estimated annual mean and coefficient of variation of streamflow in the Gould Gamma formula for calculating storage sizes is quick, simple and requires only a hand-held calculator. However, the latter technique estimates only the carry-over storage as only annual flows are used, and this may result in significant under-estimation of storage capacity. The storage errors arising from the use of the Gould Gamma formula are highly unpredictable, while those arising from the use of block regression equations can be known on an average statistical basis. Thus, in the latter case, the designer is at liberty to add an amount to the estimated storage capacity as a safety factor.

The transposition of streamflow data assumes that a sufficiently long record of monthly streamflow exists in a gauged catchment near to a hydrologically similar ungauged catchment where storage is required to be estimated. The parameters required in the transposition exercise are easily calculated from a set of equations, but for a long record of streamflow, the transposition itself is tedious if not done on a computer. Large errors are to be expected in the transposed flows. The transposed flows will then be used to calculate storages using one of the many available techniques. If a behaviour analysis is used, the errors in estimated storage sizes have been shown to vary much as they do for the results arising from the use of block regression equations, however, the difference in the input effort between both techniques is considerable.

The generation of monthly streamflow data for an ungauged catchment requires the use of a computer. The method is attractive because the distribution of the storage as a random variable can be investigated using any number of generated replicates, and hence the mean and the variance of the storage can be derived. The analyst thus has a better feel of the possible accuracy of selecting a particular storage size. However, good results can be expected only with good parameter estimation. The use of regression equations to estimate these parameters for an ungauged catchment means that these errors can only be known on an average, statistical basis. It is possible for the designer to increase the estimated storage capacity based on the worst expected errors. However, the high level of uncertainty may not justify the sophistication of the approach to the problem.

Based on these considerations, the best regional method of estimating storages for small ungauged catchments is deemed to be that of the storage block regressions.

6 RAINFALL-RUNOFF MODELLING

Development of a suitable rainfall runoff model for application to ungauged catchments involves consideration of three main problems. Firstly, it is necessary to select or develop a rainfall runoff model that is potentially suited to regional analysis; in order to relate model parameters to catchment characteristics the model should be physically based and of simple structure. Secondly, in order to develop regionalisation equations for the model parameters it is necessary to select an optimisation technique that is able to efficiently estimate model parameters for a large number of gauged catchments. The third problem relates to the estimation of evapotranspiration, a model input that is often characterised by its unreliability and extensive data requirements. The following sections review preliminary results of an inter-comparison of three Australian rainfall-runoff models coupled with four optimisation techniques applied to Yallourn Creek catchment (40 ha) in south east Victoria. The results of this preliminary investigation should not be considered conclusive, but are presented to illustrate the salient aspects of the investigation being undertaken.

6.1 Rainfall-Runoff Models and Optimisation Techniques Considered

The rainfall-runoff models selected for comparison are the Boughton (Boughton, 1966), SFB (Boughton, 1984) and the semi-arid zone (Sukvanachaikul and Laurenson, 1983) models. The Boughton model simulates daily runoff from daily rainfall and potential evaporation - it was modified by McMahon and Mein (1973) to incorporate baseflow and consists of a total of 10 parameters. The SFB rainfall-runoff model (Boughton, 1984) is a simplified version of the Boughton's original model; it incorporates a baseflow response and requires only 3 parameters to be fitted. The daily model presented by Sukvanachaikul and Laurenson (1983) was developed for application to semi-arid catchments; this model was also modified by Jayasuriya et al (1988) to include baseflow and consists of a total of 6 parameters. All these models possess a relatively parsimonious structure and attempt to reflect catchment-wide processes by simple conceptual storages. A more detailed description of these models may be found in Jayasuriya et al (1988).

The four optimisation techniques chosen for investigation were the Simplex (Nelder and Mead; 1965), pattern-search (Hooke and Jeeves; 1961), steepest ascent and the Gauss-Marquardt (Kuczera, 1983) methods. The first two belong to the direct search category and the latter two to the gradient category.

The most commonly used objective function is the sum of squares of the differences between predicted and observed streamflows. Minimisation of the least squares objective function provides reliable parameter estimates provided the least squares assumptions are satisfied, i.e. provided the residuals have zero mean, are normally distributed and are time-independent. Kuczera (1983) incorporated the Gauss-Marquardt algorithm in an interactive optimisation package called NLFIT that includes various data transformation subroutines for removing errors associated with heteroscedasticity and time-dependency of residuals. This feature was not available with the other optimisation packages used.

6.2 Results and Discussion of Rainfall-Runoff Modelling

Table 2 gives the values of the goodness of fit statistical measures obtained by applying the three rainfall-runoff models, each coupled in turn with the four optimisation techniques. It can be seen that a similar level of performance was achieved by all models. Importantly, however, the study showed that the modified Boughton model is over parameterised with only 5 parameters being sensitive; this large number of redundant parameters contributes substantially to modelling uncertainty and thus the model is probably not suited to regionalisation.

Examination of the differences between observed monthly flow volumes and those predicted by the three models revealed that the residuals violated the least squares assumptions of heteroscedasticity and time-dependency - consequently the NLFIT program was used to transform the data to rectify the residuals, a step not possible with the other optimisation packages. While it would be expected that NLFIT would out-perform the other optimisation packages the results indicate that if anything the reverse is true. However, as NLFIT is the only package not to violate the least squares assumptions, it is likely that its use will yield more reliable parameter estimates that are better suited to regionalisation, where conceptual model parameters are related to physical catchment characteristics. This aspect will be investigated further by independent tests on catchments not used in development of the regionalisation equations.

6.3 Evaporation estimates in Ungauged Catchments

Potential evaporation is an input factor in most rainfall-runoff models - in Australia, 90% of precipitation is lost to the atmosphere by evapotranspiration (Watts and Hancock, 1984), and it is thus important to model this loss accurately. Frequently, researchers use open pan evaporation readings to evaluate potential evaporation as the data are easily available; Watts and Hancock (1984), however, illustrate the inaccuracies inherent in this approach and indicate that pan coefficients are highly variable and depend upon the environment.

The complementary equation initially proposed by Bouchet (1963), completely ignores the availability of water, soil moisture and plant complexities, and provides estimates of areal evapotranspiration based on the premise that for a large uniform and closed system, actual and potential evapotranspiration are complementary quantities. Brutsaert and Stricker (1979) and Morton (1983) re-assessed the evaluation of some of the terms in Bouchet's original complementary equation to derive two separate approaches to the estimation of actual areal evapotranspiration; Morton's method requires only wet and dry bulb temperature, and sunshine duration data, and Brutsaert and Stricker's method requires wet and dry bulb temperatures, as well as wind run and observed net radiation data. The information requirements of

Morton's method are generally easily satisfied, and hence it is ideally suited to the estimation of evapotranspiration from ungauged catchments.

Preliminary applications of the models to data from forested areas confirm the potential of these models for computing actual evapotranspiration, and further research is currently underway to compare the estimates with lysimeter observations.

TABLE 2. Goodness-of-fit parameters obtained from the 3 rainfall-runoff models coupled with the steepest ascent (SA), Simplex (SX), pattern search (PS) and NLFIT (NL) optimisation techniques.

Goodness-of-fit parameters	Modified Boughton model				SFB model				Modified semi-arid-zone model			
	SA	SX	PS	NL	SA	SX	PS	NL	SA	SX	PS	NL
Mean estimated flow	19.19	20.00	19.78	16.38	20.33	20.40	20.20	17.60	20.79	20.33	20.49	19.56
Mean observed flow	20.70	20.70	20.70	20.70	20.70	20.70	20.70	20.70	20.70	20.70	20.70	20.70
Estimated std. deviation	20.06	20.56	20.33	18.76	20.12	18.24	18.32	18.17	21.38	21.87	23.49	22.81
Observed std. deviation	26.05	26.05	26.05	26.05	26.05	26.05	26.05	26.05	26.05	26.05	26.05	26.05
Coeff. of determination	0.84	0.83	0.83	0.84	0.79	0.73	0.74	0.75	0.85	0.84	0.85	0.85
Coeff. of efficiency	0.81	0.82	0.82	0.78	0.77	0.71	0.72	0.71	0.85	0.84	0.85	0.84
Residual mass curve coeff.	0.73	0.77	0.80	0.63	0.94	0.94	0.94	0.95	0.77	0.78	0.78	0.77

7.0 CONCLUSIONS

A systems approach has been presented for the estimation of low flow and yield parameters from ungauged catchments. A brief review of regionalisation methodologies concluded that cluster analysis should be used to define regions of hydrological homogeneity, and that either stepwise regression or regression on principal components should be used for both prediction of hydrological characteristics and for examination of the underlying structure and relative importance of the variates. After a comparative investigation of a number of storage yield estimation techniques, the best estimation procedure was found to be based upon statistical regression analyses that relate storage size to rainfall and catchment characteristics. The directions of current research associated with the estimation of low flow parameters and rainfall-runoff modelling techniques have also been presented.

REFERENCES

Abrahams, A.D. (1972): Factor analysis of drainage basin properties: evidence for stream abstraction accompanying the degradation of relief. *Water Res. Res.*, 8(3): 624-633.

Acreman, M.C. and Sinclair, C.D. (1986): Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. *J. of Hydrol.*, 84: 365-380.

Arnell, N. W., Beran, M. and Hosking, J. R. M. (1986): Unbiased plotting positions for the general extreme value distribution. *J. of Hydrol.*, 86: 59-69.

Blake, G.J., Cook, A.D. and Greenall, D.H. (1970): The use of principal component factor analysis to establish the uniformity of a hydrological region in Northland, New Zealand. *Results of Research on Representative and Experimental Basins*, Proc. of the Wellington Symp., IAHS-AISH: 183-197.

Bouchet, R.J., (1963), Evapotranspiration reelle et potentielle, signification climatique. *Int. Assoc. Sci. Hdrol., Publ.* No. 62: 134-142.

Boughton, W.C., (1966), A Mathematical Model for Relating Runoff to Rainfall with Daily Data. *Civil Eng. Trans.*, Inst. of Eng. Aust.: 83-93.

Boughton, W.C., (1984), A Simple Model for Estimating the Water Yield of Ungauged Catchments. *Civil Engineering Trans, I.E. Aust.*: 83-88.

Brutsaert,W. and Stricker, H., (1979), An Advection-Aridity Approach to Estimate Actual Regional Evapotranspiration. *Water Res. Res.*, 15 (2): 443-449.

Condie, R. (1979): Drought frequency analysis for intermittent streams. *Hydrological Aspects of Droughts. International Symposium*, New Delhi: 242-248.

Cunnane, C. (1978) Unbiased plotting positions - a review. *J. of Hydrol.*, 37: 205-222.

Dawdy, D.R. and Feth, J.H. (1967): Applications of factor analysis in study of chemistry of groundwater quality, Mojave River Valley, California. *Water Res.Res.*, 3(2): 505-510.

DeCoursey, D.G. and Deal, R.B. (1974): General aspects of multivariate analysis with applications to some problems in hydrology. U.S. Department of Agriculture, *Misc. Publ. No. 1275*, Proc. Symp. on Statistical Hydrology: 47-68.

Dubey, S.D. (1967): Revised tables for asymptotic efficiency of the moment estimators for the parameters of the Weibull laws. *Naval Res. Log.Quart.*, 14: 261-267.

Eratakulan, S. J. (1970): Probability distribution of annual droughts. *J. Irrig. & Drainage Div.*, 96 (IR4): 461-474.

Gould, B.W. (1964): Discussion of paper by Alexander, 1964. In *Water Resources Use and Management*, Melbourne University Press, Melbourne.

Gunst, R.F. and Mason, R.L. (1977): Biased estimation in regression: an evaluation using mean squared error. *J. Amer.Stat. Assoc.*, 72: 616-628.

Haines, A.T., Finlayson, B.L. and McMahon, T.A. (1988): A global classification of river regimes. *Appl. Geog*. (in press).

Hawley, M.E. and McCuen, R.H. (1982): Water yield estimation in western United States. *A.S.C.E. J. Irrig.and Drainage Div.*, 108: 25-34.

Hooke, R. and Jeeves, T.A. (1961), 'Direct Search' Solution of Numerical and Statistical Problems, *J. of the Assoc. for Comp.Machi*. 8(2): 212-292

Hughes, J.M.R., (1987): Hydrological characteristics and classification of Tasmanian rivers. *Aust. Geog. Studies*, 25(1): 61-82.

Institute of Hydrology (1980): *Low Flow Studies*., Institute of Hydrology, Wallingford, U.K.

Jayasuriya, L.N.N., O'Neill, I.C., McMahon, T.A. and Nathan, R.J., (1988), Parameter Uncertainty in Rainfall-Runoff modelling. In Proc. *Conf. on Agric. Engin.*, I.E. Aust., Sept. 25-29, Sydney.

Jennings, M. E. and Benson, M. A. (1969): Frequency curves for annual flood series with some zero events or incomplete data. *Water Res. Res.* 5(1): 276-281.

Kendall, M.G. (1980): *Multivariate Analysis*. MacMillan Publishing Co., Inc., New York, 210 pp.

Kuczera, G, (1983), Improved Parameter Inference in Catchment Models: Parts I and II, *Water Res. Res.* 19(5): 1151-1172.

Kumar, A. and Devi, R. (1982): Statistical determination of design low flows - a comment. *J. Hydrol.*, 58: 175-177.

Laut, P., Austin, M.P., Goodspeed, M.J. Body, D.N. and Faith, D.P. (1983): Hydrologic classification of sub-basins in the Macleay Valley, New South Wales: Parts I - III, *Civ. Eng. Trans. I.E. Aus.*, CE26(3): 218-235.

Lvovich, M.I. (1972): Hydrologic budget of continents and estimate of the balance of global fresh water resources. *Soviet Hydrol: Selected Papers*, 4: 349-363.

Lyne, V. and Hollick, M. (1979): Stochastic Time-Variable Rainfall-Runoff Modelling, Hydrol. and Water Res. Symp., Perth, *I.E. Aust., National Conf. Publ.* No. 79/10: 89-93.

Matalas, N. C. (1963): Probability distribution of low flows. *Statistical Studies in Hydrology, Geological Survey Professional paper* 434-A, 27 pp.

Matalas, N.C. and Reiher, B.J. (1967): Some comments on the use of factor analysis. *Water Res. Res.* 3(1): 213-223.

McMahon, T.A. (1982): World Hydrology: does Australia fit? Hydrology and Water Resources Symposium, *I.E. Aust . Nat. Conf. Publ.* No. 82/3, pp 1-7.

McMahon, T.A. (1984): Low flow and yield data requirements. In: Proceedings of Workshop on Surface Water Resources data. *Aust.Water Res. Council Conf. Series* No. 10: 80-124.

McMahon, T.A.and Mein, R.G., (1973), Storage-Yield Estimates with inadequate streamflow data *Symp. on the Design of Water Resources Projects with Inadequate Data*, Madrid, Vol. 2: 355-367.

Morton, F.I., (1983), Operational Estimates of Areal Evapotranspiration and their Significance to the Science and Practice of Hydrology. *J. of Hydrol.*, 66: 1-77.
Mosley, M.P. (1981): Delimitation of New Zealand hydrologic regions. *J. of Hydrol.*, 49: 173-192.
Natural Environment Research Council (1975): *Flood Studies Report*, NERC, London (5 vols).
Nelder, J.A. and Mead, R., (1965), A simple method for function minimization. *Comp J.*, 7: 308-313.
Prakash, A. (1981): Statistical determination of design low flows. *J. of Hydrol.*, 51: 109-118.
Sharp, A., Gibbs, A., Owen, W. and Harris, B. (1960): Application of the multiple regression approach in evaluating parameters affecting water yields of river basins. *J. Geophys. Res.* 65(1): 1273-1286.
Shelton, C.H. and Sewell, J.I. (1969): Parameter screening for watershed analysis. *Trans. Amer. Soc. Agric Eng.*, 12(4) 533-539.
Snyder W.M. (1962): Some possibilities for multivariate analysis in hydrologic studies. *J. Geophys. Res.* 67(2): 721-729.
Sukvanachaikul, Y. and Laurenson, E.M, (1984): An Improved Rainfall-runoff Model for Semi-arid Regions. *Hydrol. & Water Res. Symposium, I.E.Aus.*: 108-138.
Tasker, G.D. (1982a): Comparing methods of hydrological regionalisation. *Water Res.Bull.*, Vol. 18, (6): 965-970.
Tasker, G.D. (1982b): Simplified testing of hydrologic regression regions. *J. Hydraul. Div., Proc. of Am. Soc. Civil Eng.* 108(HY10): 128-1222.
Tollow A.J. (1987): Methods used in assessing low-flow probabilities and their application. *National Hydrol. Symp. Univ.of Hull, 14-16 September,*: 15.1-15.5.
Velz, C.J. and Gannon, J.J (1953): Low flow characteristsics of streams. *Ohio State University Engin. Series*, XXII(4): 138-154.
Wallis, J.R. (1965): Multivariate statistical methods in hydrology - a comparison using data of known functional relationship. *Water Res. Res.* 1(4): 447-461.
Watts, P.J. and Hancock, N.H. (1984); Evaporation and Potential Evaporation - A practical Approach for Agricultural Engineers. *In Proc Conf. on Agric. Eng. I.E. Aust.*: 290-297.
Weisberg, S. (1985): *Applied Linear Regression*. John Wiley & Sons, 324 pp.
White, E.L. (1975): Factor analysis of drainage basin properties: classification of flood behaviour in terms of basin geomorphology. *Water Res. Bull.* , 11(4): 676-687.
Wiltshire, S.E. (1986): Identification of homogeneous regions for flood frequency analysis, *J. Hydrol.*, 84: 287-302.
Zucchini, W. and Adamson, P.T. (1984): Assessing the risk of deficiencies in streamflow. *WRC Report No.* 91/2/84, University of Stellenbosch.