Probability Distribution for “x-day” Daily Mean Flow Events (from Gauged Records)

R&D Technical Report W6-064/TR2

Probability Distributions for "x-day" Daily Mean
Flow Events (from Gauged Records)

Methodology for use with Long Records &
Prototype Methodology for use with Short Records

R&D Technical Report W6-064/TR2

MD Zaidman, V Keller & AR Young

Research Contractor
Centre for Ecology & Hydrology

Publishing Organisation
Environment Agency, Rio House, Waterside Drive, Aztec West, Bristol BS32 4UD
Tel: 01454 624400 Fax: 01454 624409
Website:www.environment-agency.gov.uk

ISBN: 1 85705 997 2

This report is the result of a collaborative study jointly funded by the Environment Agency and the Natural Environment Research Council.

Dissemination Status
Internal: Released to Regions
External: Released to Public Domain

Statement of Use
This report provides probability distributions describing the occurrence of annual minima flow events for 25 British rivers having long, stable and natural flow records. It will be of use to Water Resources staff involved in operational and planning tasks.

Key Words
Low flow, annual minima, flow-duration-frequency models.

Research Contractor
This document was produced under R&D Project W6-064 by:
Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford Oxfordshire, OX10 8BB
Tel: 01491 838800 Fax: 01491 692424

Environment Agency Project Manager
The Agency's Project Manager for R&D Project W6-064 was:
Andy Wall, Environment Agency – Wales

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre c/o
WRc, Frankland Road, Swindon. Wilts, SN5 8YF
Tel: 01793 865000 Fax: 01793 514562 email: publications@wrcplc.co.uk

FOREWORD

The authors would like to acknowledge David Jones (Centre for Ecology and Hydrology, Wallingford) for his considerable input into the Literature Review (Chapter 2). Andrew Wall (Environment Agency, Wales) and Daniel Cadman (Environment Agency, Anglian Region) are thanked for their helpful guidance in regard both to the interpretation of results and in preparation of this manuscript.

The Environment Agency and the Scottish Environment Protection Agency are thanked for confirming observations regarding artificial influences in the study catchments.

EXECUTIVE SUMMARY

Probability distributions describing the occurrence of D-day annual minima flow events have been determined for twenty-five British rivers having long, stable and natural flow records.

In each case a pre-determined set of rules and criteria were applied to the flow record to derive time series of D-day annual minima and to ensure that these were both stationary and independent. So that low flow events of different duration could be examined a range of values were used for D including 1, 7, 30, 60, 90, 180 and 365 days. For each time series, estimates of the non-exceedance probabilities corresponding to the annual minima were derived using the Gringorten Plotting Position Formula. These were then used to build the observed probability curve. It is important to note that the shape of this curve is influenced by the number and rank order of the annual minima derived for each series, as well as the range of annual minima observed.

A representative probability distribution should describe the frequency of occurrence of events beyond the observed range. However, as observed data usually represents only the central portion of the true probability distribution of annual minima, the shape of the tails of the distribution must be inferred from theory or experience. Extreme Value Theory suggests that the frequency behaviour of annual minima will follow that of a Generalised Extreme Value distribution. However Pearson Type III and Generalised Logistic distributions are commonly used for this type of analysis. Using the L-Moment method of parametric estimation, the parameters of each of these three candidate distributions were determined based on the flows and probabilities of the observed data. A fourth distribution, the Generalised Pareto, was also investigated. Although this distribution was not theoretically suitable for describing extreme events such as annual minima, it was included to provide a control.

In order to differentiate the most representative of the four parameterised distributions, their descriptive and prescriptive characteristics were considered. Goodness-of-fit tests and root mean square errors (RMSE) were used to quantify the ability of the modelled curve to match the observed data. The results indicated that there was little difference between the performances of the different distributions, and therefore that with the best goodness-of-fit and lowest RMSE values was considered, in a descriptive sense, the most representative. The analysis showed that no one distribution best represents the frequency behaviour of annual minima. However, for annual minima of short duration (values of D in the range 1 to 30 days) the low flow frequency curves in permeable, high storage catchments tend to be best described using the Generalised Logistic or Generalised Extreme Value distributions. In contrast those for the low storage catchments tend to be best described by the Pearson Type-III or Generalised Extreme

Value distributions. Where the annual minimum is of long duration, with values of D above 90 days, many annual minima series were best described by the Generalised Pareto distribution.

The shape of the flow-return period relationships derived was examined to qualify the predictive ability of each distribution, i.e. to determine whether the tails of the distribution provide sensible estimates of return period for given annual minima and vice versa. In general sensible estimates are obtained for annual minima of short duration (D=1 or D=7 days) where the prescribed flow is less than 10% of the mean flow, only where catchments are impermeable. As annual minima of longer duration are considered, or as the catchment type is permeable sensible estimates are obtained only for higher prescribed flows.

The methodology was also applied to streamflow records shorter than 20 years in length. However, as it is unrealistic to expect to identify a representative probability distribution where there are few observed data points, a slightly different approach was adopted. Properties of the probability distribution for each short-record were inferred from that of an analogue site, having similar low flow behaviour and catchment characteristics but a much longer flow record. Two variations of this approach were considered. Firstly analogue catchments were assumed to have similar probability distributions (provided that the flow values were suitably standardised by, for example, expressing as a ratio of the mean annual minimum value). Hence the probability curve of the long-record analogue was re-scaled by the short-record mean flow to provide an estimate of that for the short record catchment. Although this method proved successful for certain catchments, for others it provided poor results.

In the second approach the true probability of non-exceedance for the annual minima occurring in a particular year within a short record was assumed to be equivalent to the probability of the annual minima occurring in the same year of the long record. This method generally provided fairly accurate predictions of the probability distribution, but requires the flow records for the short-record catchment and its long-record analogue being wholly coincident.

CONTENTS

LIST OF SYMBOLS

k	Shape parameter
p	Probability of non-exceedance
q	Observed annual minimum flow
q_i	Annual minimum flow of rank i (in ascending order)
x	Realisation (observations) of a random variable
y	Year
f(x)	Density function.
x(F)	Quantile function, F denotes a numerical value.
D	Duration of low flow period
N	Number of sampled values
P	probability
Q	Random variable corresponding to annual minimum flow
S	Standard deviation
T	Return period or recurrence interval
X	Random variable
Y	Random variable
Z	Random variable
E(x)	Estimation of x
F_S	Standard distribution function
$F_N(x)$	Empirical distribution function
$F_X(x)$, F(x)	Distribution function
α	Scale parameter
β	Parameter
δ	Parameter / discretization interval
θ	Parameter
?	Location parameter
$?_s$	Spearman Rank correlation coefficient
$?^2$	Chi-square goodness of fit parameter
G	Gamma function

LIST OF ABBREVIATIONS

c.d.f.	Cumulative distribution function
p.d.f.	Probability distribution function
AM	Annual minima
AMS	Annual minima series
MAM(D)	Mean annual minimum flow at duration D.
MF	Mean flow
PDS	Partial duration series
Q95	95th percentile flow
RMSE	Root mean square error

GL	Generalised Logistic
GEV	Generalised Extreme Value
GP	Generalised Pareto
PE3	Pearson Type-III

1. INTRODUCTION

1.1 Project Background, Aims and Specific Objectives

Estimates of the frequency of occurrence of low flow events are important within the context of drought assessment, where drought can be thought of as an extreme and/or persistent low flow event. The Environment Agency, for example, estimate return-periods of low flow events in order to quantify the severity of droughts as they develop and intensify. Similarly in design studies, a return period estimate may be required for a 'design drought' of particular magnitude, or a flow magnitude estimate may be required on the basis of a 'design return period'. Typically, the low flow 'event' of interest is the annual minimum flow; this might be either the smallest daily flow value or the smallest average flow over a period (duration) of 'D' days occurring within a year. Other types of representative annual values might also be used, such as all 'peaks' under a threshold level, as in a partial duration series.

In low-flow frequency estimation statistical procedures are used to formulate a model of the probability distribution of low flow events at a particular site (a similar approach is used in flood frequency estimation). The goal is to produce a model that recreates the behaviour in the extrapolated upper and/or lower tails of the distribution, which are beyond the range of observed events. This is achieved by parameterising, i.e. fitting, a mathematical equation (usually this form is a distribution theoretically suitable for describing the occurrence of extreme events) based on the magnitude and apparent frequencies of the sampled (observed) low flow events at the site.

In the UK there has been little research on this topic since the publication of the Low Flow Studies Report in 1980 (Institute of Hydrology, 1980). Although the Low Flow Studies Report provided guidelines for fitting probability distributions to the annual minima of d-day duration flow events, there remains confusion within a UK context as to which probability distributions are best used for flow events of different duration and for different streamflow regimes. Furthermore, as shall be discussed in Section 1.2, there is currently a lack of a unified approach to low flow frequency analysis in the UK.

Given the large amount of gauged flow data that has been collected since publication of the 1980 report it is timely that the problem of estimating low flow frequency for rivers in the UK should be revisited. The aim of this technical document, therefore, is to establish best practice for the analysis of the frequency low flow events by producing a consistent and scientifically sound method for enumerating probability distributions describing annual minima of D-day duration low flow events. This method is to be robust and objective enough to use in a regulatory manner and to withstand scrutiny at Public Enquiries. The specific objectives are:

1. To formulate a robust methodology for use with “long” gauged flow records (>20 years) taking into account the developments in hydrological frequency analysis over the last 20 years and the increased number of “long” flow records available within the UK.

2. To investigate, by prototyping, the potential for development of a similar methodology for “short” gauged flow records (i.e. those with less than 20 years of data). If the prototyping exercise shows this to be theoretically and practically viable, the methodology for short records is to be developed in a second stage, *phase 2*, of the project.

This technical report describes the researches undertaken based on long record data sets, the assumptions made, and the methodologies developed and their strengths and weaknesses. The results of the prototyping exercise for short records are also presented.

1.2 Estimation of Low Flow Probability in the UK

With its mild maritime climate the UK rarely experiences the hydro-meteorological extremes, such as multi-year droughts, that are common elsewhere in Europe. However, there have been several notable drought events in the UK over the past 30 years, including the 1976, 1984, 1990 and 1995 droughts, where rainfall totals have been well below average. Droughts affect all aspects of the hydrological cycle: reservoir storage, groundwater storage and streamflow levels are all depleted during droughts. As this research deals with low flow events, it is useful to define the term *streamflow drought*, which can be thought of as a prolonged period of low-flow, culminating in the depletion of flow beyond the usual low-flow level. The strain on surface water resources during drought periods is often exacerbated by increased public and agricultural demand for water. During these crucial periods the needs of abstractors have to be balanced carefully against the need to protect the in-stream environment: mismanagement of water resources that are already over-stretched may have serious environmental and socio-economic impacts.

Streamflow depletion is usually expressed in terms of the return period of the event, which is estimated by applying frequency analysis procedures to estimate the flow-return period relationship at the site of interest. Operationally, such methods are important for considering the severity of individual drought events in the context of the long-term behaviour of a site (i.e. determining the return period of a particular flow event) and for determining the severity of a particular ‘design drought’ (i.e. predicting the flow associated with a drought of given probability). The former is particularly important for assessing whether the criteria for Drought Orders have been met in cases of exceptional water shortage.

To-date the water resources industry has taken a rather *ad-hoc* approach to low flow frequency analysis. Although consistent methods for frequency analysis of flood events are widely disseminated with the hydrology and engineering communities in the UK, for instance the approach documented in The Flood Estimation Handbook (Institute of Hydrology, 1999) has been widely adopted, the same cannot be said about low-flow frequency analysis. Many agencies and consultants use their own in-house procedures or arbitrary guidelines for setting flows levels for a particular ‘design drought’. No industry standards have been established and there has been no general agreement regarding which of the various theoretical distributions are preferable or to what fitting techniques are more appropriate. This irregularity is in part due to the fact a whole range of river flow conditions can broadly defined as ‘low’, and as drought analysis in the past has been governed by the operational needs of users, a large number of highly subjective definitions and criteria have resulted. The effect of artificial influences and the problem of zero flows have also been handled differently. However, recent drought events have highlighted the need for a consistent methodology to be applied in the UK. For instance, the Environment Agency concluded that during the 1995 ‘drought’, water resources across most of England and Wales were in a generally satisfactory condition, despite abstractors having successfully argued the case for Drought Orders to be put into place (NRA, 1995).

1.3 Layout of Report

Following this introductory chapter, a review of recent literature on the subject of low flow frequency analysis is presented in Chapter 2. The literature review outlines the theoretical principles of low flow frequency analysis and assesses the commonly used methods of parameterising model distributions; candidate mathematical distributions are identified and a methodology that is appropriate for use with fairly long UK flow records is outlined. Section 2.1 firstly gives a brief overview of low-flow frequency analysis, and describes some of the significant issues that have to be addressed. The theoretical background to statistical frequency analysis is discussed in much more detail in Section 2.2. Issues relating to the data, including homogeneity and independence, and the effect of sample size are discussed in 2.3. Non-parametric methods, including both empirical methods and graphical (plotting position) techniques are discussed in Section

2.4, whilst the use of parametric estimation techniques is described in 2.5. Various aspects of choosing an appropriate family of distributions are discussed in Sections 2.5.1 (candidate families of distributions are reviewed later in 2.6) and methods for fitting distributions (i.e. estimating the parameters of the distribution) are discussed in Section 2.5.2. Formal statistical procedures for seeing whether there is enough evidence in the limited sample of data to rule against use of a particular family or distribution are then outlined: model evaluation procedures such as screening and re-sampling are described in Section 2.5.3. A review of the distributions that might be useful in low-flow analyses is given in Section 2.6, whilst a general assessment of the appropriateness of low-flow frequency analyses reported in the literature is given in 2.7. Finally, recommendations for best practice are made (analysis of long records), and possible techniques for dealing with short records are outlined.

The proposed methodology is discussed and developed in more detail in Chapters 3 to 5 and is illustrated with 25 high-quality flow records from catchments in the UK representing a range of flow regime types. The procedure (e.g. missing data criteria) for deriving the time series of annual minima for low flow events of different duration and the statistical procedures to test the homogeneity and statistical independence of these are described in Chapter 3. The methodology for enumerating distribution models is presented in Chapter 4, and includes consideration of different fitting techniques, the effect of hydrometric errors and the ability of the methods to deal with flow records that are discretized (i.e. contain rounding errors) or contain zero flows. Evaluation of the proposed method, based on the 25 flow records, is reported in Chapter 5. Derived models are assessed in terms of the fit of the flow frequency curve (i.e. descriptive ability) and the uncertainty on the resulting flow estimates at high return periods (i.e. prescriptive ability).

The results of a scoping study to investigate the feasibility of determining probability distributions where the record length are short is described in Chapter 6. In the methodology presented an analogue rather than regionalisation approach is used, i.e. the flow – return-period relationship for a short record is derived from that of a single long-record catchment known to have a similar flow regime. Two different systems of transferring information from the probability distribution of the analogue catchment are presented and evaluated.

A final discussion of the results and conclusions of the research is given in Chapter 7. This includes a final recommendation as to the methodology most appropriate for the kinds of river regimes typically found in the UK. This methodology has been incorporated into a guidance note ‘Guidelines for Best Practice’, which is included as a technical annex.

2. LITERATURE REVIEW

2.1 Overview of Low Flow Frequency Analysis

2.1.1 Introduction

Low flow frequency analysis is based upon the use of a statistical distribution to describe the probability of occurrence low flow events (usually the annual flow minima). As only a very small fraction of the true probability distribution of low-flow events is realised in any flow record, the exact form and parameters of the true distribution can never be known. In practice, therefore, a distribution function that describes the low flow behaviour of the observed data *reasonably well* is used. The form of this distribution can then be used to make inferences about the probabilities of events beyond the observed range (i.e. by extrapolating into the tail of the distribution), allowing the flows associated with long return periods to be determined and vice versa.

Section 2.1 introduces the general methodology for low flow frequency analysis. The procedure for choosing a reasonable distribution involves fitting a number of candidate distributions to the data (i.e. quantifying the distribution parameters using the sample data) and testing each for goodness of fit and bias. These individual points are discussed in more detail in subsequent sections.

2.1.2 Applicability of flood frequency estimation methods

Stochastic methods, such as simulation techniques and frequency analyses, are often used to characterise extreme hydrological events such as floods and droughts. Frequency analysis, in particular, is used widely in flood risk estimation and several detailed and ‘robust’ methodologies on this topic have been developed. In fact much of the research within the context of flood frequency (an excellent review is given by Cunnane, 1989) has contributed to improved understanding and procedures for frequency analysis in general. UK research into flood frequency is also highly advanced, for instance the Flood Estimation Handbook, a nationally recommended methodology, was recently published as a result of a five year study on flood flows in UK Rivers (Institute of Hydrology, 1999).

The use of frequency techniques for characterisation of low flow or drought events is, in contrast, much less advanced, although studies have been recently published in the USA and Europe (a short review of recent research is given by Tallaksen, 2000). There remains a lack of a unified approach to drought frequency estimation both from an UK (the last major work in the UK was that of the Institute of Hydrology (1980)) and world perspective.

Although, low flow frequency analysis shares much with its sister topic of flood frequency, there are several key differences between the two, and therefore, although the concepts and general techniques may be transferred between flood and drought analysis, the 'nuts and bolts' of the methods are, by necessity, different. The main differences can be summarised as:

- The need to represent the possibility of observing a zero flow.
- A series of low flow observations might be considered as a discrete distribution rather than a continuous distribution due to a lack of precision of recording variables at low flows.
- A drought or low-flow event is not instantaneous, rather it may last for several days, weeks or months, and in extreme cases for one or more years. The data therefore may be serially correlated especially where flow events of relatively long duration are considered.

2.1.3 Annual minima and partial duration approaches

Low flow frequency analysis uses sample properties of observed drought events to find the most suitable probability distribution to describe those events. The sample properties can be some kind of representative annual value, such as in an annual minima series (AMS), or all 'peaks' under a threshold level, as in a partial duration series (PDS).

The annual minima (AM) method avoids the need to consider all the complicated day-to-day variations of flow recorded at a given site over a number of years, instead attention is concentrated on a single derived value for each year (the annual minima). The single value for each year is usually chosen to reflect the minimum flow averaged over a period of particular duration. For instance, the annual minimum 1-day flow is the lowest daily flow in the year, the annual minimum 7-day flow is the smallest average flow over any 7 consecutive days in the year, the annual minimum 10-day flow is the smallest average flow over any 10 consecutive days and so on. Using flows averaged over a period of 'D' days allows the duration of low flow events to be considered, in addition to magnitude or severity. Typically short periods, such as 1, 7 10 or 30 days, are used. The annual minimum values for each year are used to build up an AMS to which frequency analysis can be applied. The average value of a particular annual minima series, the mean annual minimum flow, MAM(D), represents the lowest D-day flow on average within the record period. The return period of the MAM(D) flow is always fixed for a given distribution. For example, the MAM(7) flow can be thought of as that which occurs during the driest week (7 days) in the average summer. For many of the distributions assumed in frequency analysis the MAM(D), flow has a return period of around 2 years.

The main disadvantage of using a series of annual minima is that the minimum flows in wet years might not belong to an extreme population, whereas if two random droughts happened to occur in the same year, the smaller of the two would be ignored despite being an extreme event. The use of an annual minima series is recommended only where the flow record at a site is sufficiently long. For shorter record periods it is more applicable to use a partial duration series or employ some means of ‘borrowing’ data from similar sites, such as ‘regionalisation’ or pooling techniques.

A partial duration series consists of all daily flows below a base magnitude or threshold level defining the onset of a streamflow drought, regardless of where they occur in the time series. Although more data from the flow record is incorporated in the analysis, the PDS method has several disadvantages. Firstly the threshold can strongly influence the results of the analysis, yet is difficult to set objectively. On an at-site basis a threshold may be set according to ecological, navigational or recreational constraints. Alternatively thresholds may be set according to the flow regime, and are usually either fixed (e.g. in the UK the 95th percentile flow (Q95) is often used as a low-flow threshold) or variable, in which case they take into account the seasonal variability of flow levels. Dracup *et al*. (1980) and Hisdal & Tallaksen (2000) discuss the advantages and disadvantages of some of the methods available for setting threshold levels. The second major disadvantage of the PDS method is that, as a result of fluctuation of the flow around the drought threshold, drought events are often split into a correlated sequence of shorter droughts. The user has to be subjective regarding independence of adjacent events, and it is likely that a procedure for pooling dependent droughts would have to be adopted.

As this review focuses on at-site analysis, where the period of record is long, say not shorter than the target return period (in terms of a design drought), the use of the annual minima series for characterising the frequency of low-flow events is described. For shorter records, some kind of regionalisation or pooling procedure is required to augment the amount of data available for analysis. Such approaches are discussed briefly where appropriate.

2.1.4 Analysis of annual minima series

In its basic form, low flow frequency analysis can be characterised as dealing with a set of N values, one for each year, i, of record, $\{q_i, …., q_N\}$, where q_i denotes the smallest flow in year i. In practice "smallest flow" is defined as the smallest average flow over a period of given duration, for example 1 day, 7 days, 30 days or even 365 days. The essential problem is to make use of these data to provide information about flows in future years which, by definition, have not yet been observed: clearly, these data can only be expected to provide information about the corresponding low flow statistic for the not-yet-observed years. Consider a single future year, in which the smallest flow will take the unknown value q. In order to make any sort of inference about q, some assumptions are required which in effect say that the collection of data-values $\{q_i, ..., q_N\}$ is somehow representative of the values that q might take. This means that any considerations of change over time in the catchment, or in the climatic regime in which it lies, must have been fully considered and dealt with before attempting the analysis.

The sort of information about q provided by $\{q_i, …., q_N\}$ can be summarised as follows:

(i) The average value, or the median value, of $\{q_i, …., q_N\}$ provides an estimate of the typical size that q might be in any single year.

(ii) The variation between the values in the set $\{q_i, …., q_N\}$ provides a guide as to the likely variation of q between different future years.

(iii) The proportion of the values in the set $\{q_i, ….,q_N\}$ which are below a certain threshold provides an estimate both of the proportion of future values of q in different years which will be below the same threshold and of the probability that q for a single year will be below the threshold.

Other possible types of information could be sought: for example, if there were longish runs of high values and runs of low values in $\{q_{i,} …., q_N\}$ when considered for consecutive years, then the same sort of behaviour could be expected to continue into the future. This last consideration is outside of the central scope of low flow frequency analysis which, as the name implies, deals with estimating how often in future years low flows of a given severity can be expected to occur: this may seem to cover only point (iii) above but in fact the other two points are included as well. Frequency analysis is based on the assumption that the observed data-values and any relevant future values are all outcomes or realisations of random quantities which have the same statistical properties. Thus q and $\{q_i, …., q_N\}$ are assumed to be realisations of random variables Q and $\{Q_i, ..., Q_N\}$ which share the same statistical distribution function.

2.1.5 Methods for determining the statistical distribution function

There are several ways in which the statistical distribution, and the parameters that define its exact form, can be determined. This review focuses on parametric estimation techniques (Section 2.1.6), which involve matching the sample data to a hypothetical distribution and using objective statistical tests to assess the appropriateness of that match.

Empirical distribution functions and simple graphical techniques based on plotting positions are also common ways of estimating low-flows of a given frequency, and are also included in this review. As these do not depend on the having to make correct assumptions about the form of the distribution from which the data arise, they are classed as non-parametric estimation techniques. The graphical method is particularly useful, and, as it very straightforward, it is nearly always used as a preliminary first step in frequency analysis. A plotting position formula is used to estimate the exceedance probability for each discharge and a curve is fitted graphically through these probability points: no prior assumptions are made about the distribution. If the set of data points plots as a straight line on probability paper then its underlying distribution is taken as that corresponding to the probability paper. Graphical procedures are therefore particularly useful for identifying which type or family of distribution function a particular data set belongs to, and are discussed in this context later. The graphical procedure can also be used to assess the goodness of fit of distributions.

Other approaches include transformation (e.g. to a normal distribution) and physically based probability modelling. Transformations are often applied to sample data in order to simplify the search for a suitable probability distribution, and for this transformations to the normal are widely used. The physically based approach relies upon characterising hydrological behaviour during the low flow period and using this to derive theoretical expressions for low flow distribution functions. For example Gottschalk *et al*. (1996) derived expressions for distribution functions based upon low flow recession behaviour. These two methods are not central to frequency analysis, and are therefore beyond the scope of this review.

2.1.6 Fitting hypothetical distributions with parametric methods

Principles of parametric estimation

Parametric estimation is the usual method of determining probability distribution from observed flow data. Because the amount of data available for low flow frequency analyses is usually small (typically 10-30 years of data and rarely more than 60 years) it is usually assumed that the distribution function concerned is a member of some particular family of statistical distributions. Use of a family of distributions involves consideration of the parameters of the family and any inferences about the actual distribution, from which the recorded data set has arisen, are made via inferences about the parameter values (i.e. the distribution parameters are quantified using the sample data). Provided that the true distribution can be closely matched by one of the chosen family of distributions, the uncertainty in the parameter estimates is also much reduced. Despite there being a fairly large body of low-flow literature, there is no clear guidance as to which distribution families are best for low flow frequency analysis. Theory suggests that the Generalised Extreme Value (GEV) family of distributions is good for describing extreme data sets, and for low flow data the reverse EVIII distribution, or Weibull distribution, seems to be most appropriate. In practice a wide range of distribution families have been 'successfully' applied to describe the distribution of low flows; including the Pearson Type-III, the EVI and Log-Normal distributions, in addition to the Weibull/EVIII. In fact no universally applicable distribution for low flows has ever been shown to exist. A more detailed discussion regarding the different candidate distribution families is given in Section 2.6.

Methodology for parametric estimation

The procedures for selecting the particular member of a family of distributions on the basis of the sample of data involve estimating the parameters of the distribution via a fitting technique, such as the method of moments or the method of maximum likelihood, and assessing the goodness of fit using objective statistical or graphical tests. Finally, resampling techniques are employed to assess the variability of the quantile estimates and determine confidence intervals. The procedure should follow the following outline (after Takara and Stedinger, 1994):

(i) Evaluate the homogeneity and statistical independence of the data
(ii) Enumerate several candidate distribution families.
(iii) Estimate parameters for each distribution.
(iv) Screen the distribution to assess goodness of fit (i.e. exclude those distributions for which the agreement between the estimated distribution and the data is poor).
(v) Analyse the bias of estimates for distributions that have not been excluded in step 4 (e.g. by using resampling methods, such as bootstrapping or jack-knifing).
(vi) Select a distribution that fits data well, exhibits the smallest bias and also provides the least biased quantile estimates.

Statistical theory relating to distribution fitting assumes that the sample data represent random occurrences of the quantity of interest and therefore depends upon the data being independent and from the same statistical population. Ensuring that the sample is not serially correlated, that the data is stationary (i.e. shows no trends over time) and homogeneous, and that there are no ‘outliers’ amongst the sample data are therefore important prerequisites for frequency analysis. Unfortunately most hydrological time series are temporally correlated (dependent) in some way. Low flow events are particularly likely to be serially correlated if the duration considered is large. It is therefore very important to evaluate the homogeneity and statistical independence of the data prior to attempting a frequency analysis of low flow data and to be aware of the influence of this on any flow-return period relationships that are derived. Methods for evaluating time series are discussed further in Section 2.3.

2.2 Theoretical background

2.2.1 Distribution functions

As described in Section 2.1 the fundamental concept in frequency estimation is that of the frequency distributions. Statistical distribution functions provide the mechanism whereby the probability that a low-flow event will occur is determined. In fact this is essentially the definition of a distribution function. The frequency distribution of a quantity, X, shows the frequency at which possible values of X occur (for this X is assumed to occur randomly). The cumulative frequency distribution (c.d.f.), known also as the probability distribution (p.d.f.) denoted by $F_X(x)$, gives for any value, x, the probability that X will be below or equal to x.

$$F_X(x) = \Pr\{ X \leq x \} \tag{2.1}$$

Quite often the subscript X is omitted when it is clear to which random variable the distribution function relates (i.e. to give F(x)). For the purposes of this report, capital letters X, Y, Z, etc. are used to denote general types of random variables while Q is reserved for the random variable corresponding to annual minimum flow. Lowercase letters x, y, z, etc. are used to represent particular outcomes (i.e. observations) of the random variables.
The interpretation of F(x) as the probability that X does not exceed x means that distribution functions must be non-decreasing and must take values in the range 0 to 1. Furthermore, in order to be a proper distribution function (i.e. to assign zero probability to infinite values) it must have limiting values of zero for large negative values and one for large positive values of x. When a random variable cannot take negative values, the corresponding distribution function must be such that F(0-)=0, where F(0-) denotes the limiting value of F(x) as x approaches zero from below. If F(0-)=0, then F(0) is the probability that the random variable is exactly equal to zero.

A family of distribution functions is simply a collection of distributions indexed by a set of parameters $\alpha_1, \alpha_2, \ldots, \alpha_n$: all this means is that, for any particular set of parameter values $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$, the function

$$F(x) = F(x; \alpha_1, \alpha_2, \ldots, \alpha_n) \tag{2.2}$$

is a distribution function. Formally, the definition of a family of distributions should include the range of permissible values of its parameters. Here, for brevity, a single symbol is used to denote a collection of parameter values. Thus equation (2.1.1) might be rewritten as

$$F(x) = F(x; \alpha) \tag{2.3}$$

Evans *et al*. (1993) provide a simple guide to the families of distributions most often used in applications of statistics. Families of distributions that have been used for low flow frequency estimation and for the closely connected topic of flood estimation are reviewed in Smakhtin (2001) and Cunnane (1989) respectively.

2.2.2 Quantile (inverse) functions

Clearly, if the true distribution function of annual minimum flows were known, this would immediately provide the answer to questions of how often flows below any given critical value will occur. Similarly, if the question is what value of flow is such that there is a given probability of an annual minimum flow falling below that value, the inverse distribution function provides the answer. Within statistical theory, the standard notation for the inverse distribution function is $F^{-1}(p)$. Recently a new notation, x(F), has been used to describe the same idea, which is then referred to as the **quantile function** for the distribution. The quantile function gives the magnitude of an event corresponding to a particular probability, *p*, of non-exceedance of x. In this case F denotes a numerical value rather than a distribution function. For example, if p is a probability value (where $0 < p < 1$), the quantile function x(p) is defined as follows:

$$x(p) = F^{-1}(p) \tag{2.4}$$

i.e. x(F) is the value satisfying F(x) = *p*.

Where distributions have discrete components the quantile function is defined to take the value x*, where x^* is the largest value of x for which $F(x) \leq p$, hence

$$x(p) = F^{-1}(p) = x^* \tag{2.5}$$

If there is no discrete component then x^* is the value satisfying $F(x^*) = p$. In the same way as for distribution functions, the quantile functions corresponding to a family of distributions with parameter α will be denoted by $x(p; \alpha)$.

2.2.3 Density functions

An alternative to the distribution function F(x) is the **density function** f(x), where

$$f(x) = dF(x)/dx, \qquad F(x) = \int f(z)\,dz. \tag{2.6}$$

Thus the distribution function and density function are essentially equivalent to one another. If the distribution is either discretely-valued or has a discrete component, rather than being continuously-valued, it can still be defined by using special functions that allow for this.

Families of statistical distributions are usually defined by giving a functional form for either their distribution or density functions, or for their quantile functions. There are useful families of distributions for which the quantile function can be written in a simple explicit form, while the distribution and density functions cannot, and vice versa.

2.2.4 Multi-parameter distributions

Some common families of distribution have as many as four parameters. The quantile functions corresponding to a family of distributions with *n* parameters are denoted by $x(p; \alpha_1, \alpha_2, \ldots \alpha_n)$. Inclusion of more parameters will typically imply a wider range of possible shapes for the distributions in the family, and thus an enhanced ability to match the shape of the unknown distribution. It will also generally lead to increased sampling variability in the estimates of the distribution function, for a given sample size, as the estimates attempt to reflect the patterns in the specific sample of data. Two-parameter distributions usually include parameters representing the location and the scale of the distribution. Further parameters usually reflect the level of skew and kurtosis in the shape of the distribution.

2.2.5 Return period

Estimating the aspects of the distribution function by fitting methods allows us to estimate the required quantity relating to a particular probability of occurrence or the probability that the quantity will be below a given value. It also allows an assessment of the uncertainty of that estimate to be made, possibly in the form of an estimated sampling variance or a confidence interval, or as a sampling covariance matrix. The probability of an event is often expressed in terms of its recurrence interval, or return period. The return period is the interval between two consecutive events, and, for the general case, is related to the probability of non-exceedance as follows:

$$p(X \leq X_T) = \frac{1}{T} \qquad (2.7)$$

where T represents the return period or recurrence interval of interest, and X_T is the event associated with that interval.

The return period is also related to the distribution function by

$$F\,(X_T, \alpha_1, \alpha_2, \ldots, \alpha_p\,) = \frac{1}{T} \qquad (2.8)$$

The quantile x(F) can also be determined where p is given as a return period (by substituting the inverse distribution or quantile function into equation 2.8.) and in this case is termed x_T. This can then be used to generate a frequency curve, otherwise referred to as the x_T-T relationship.

$$F\,(x_T) = p = \frac{1}{T} \qquad (2.9)$$

Note that the probability of an event being exceeded, termed F '(x), is given by 1-*p*. The quantile function of that event is the value satisfying Pr[X>x] = F'(x) =1- *p*, so that , in this case, x_T and T are related as follows

$$F'(x_T) = 1 - p = 1 - \frac{1}{T} \qquad (2.10)$$

The goal of frequency analysis is to obtain useful estimates of x_T for return period of scientific interest. To be useful an estimate should not only be close to the true quantile, but should also come with an assessment of how accurate it is likely to be. Usually it is assumed that x_T can only be reliably estimated where T is less than or equal to the number of years in the sample.

2.2.6 Confidence intervals for quantiles

The accuracy of the quantile estimators can be quantified using the variance, standard error or using confidence intervals (for large sample sizes quantile estimators tend to be normally distributed). Confidence intervals are usually based upon the standard error of the quantile estimator and it is possible to calculate confidence intervals relating to quantile values for some of the simpler distributions. Where the statistical distribution is very complex, confidence intervals can be estimated by resampling methods (Section 5.4).

2.2.7 Probability relationships for annual minima distributions

Equations 2.1 to 2.10, which describe the features of the probability distribution for the general case, can also be stated for the case where the distribution of annual minima values is of interest. In this case the p.d.f., $F_Q(q)$, gives for any value, q, the probability that Q will be below or equal to q .

$$F_Q(q) = \Pr\{ Q \leq q \} \qquad (2.11)$$

Note that as Q cannot take negative values, the corresponding distribution function must be such that F(0-)=0, where F(0-) denotes the limiting value of F(q) as q approaches zero from below. If F(0-)=0, then F(0) is the probability that the random variable is exactly equal to zero.

Similarly the density function is given by f(q), whilst the return period represents the interval between two consecutive low-flow events of size q_T or smaller. The quantile function now gives the magnitude of the annual minima corresponding to a particular probability of non-exceedance, and is denoted by q(F).

2.3 Evaluating Observed Low Flow Series

2.3.1 Introduction

A number of assumptions are central to expositions of statistical theory relating to distribution fitting. Most importantly the data must be independent and must originate from the same statistical population. The later assumption implies that the observed data must be stationary (i.e. show no trends over time) and homogeneous, and there must be no outliers amongst the sample data. This section discusses methods by which the applicability of these assumptions to annual minima series can be evaluated.

2.3.2 Statistical independence

Validity of the assumption of statistical independence

The assumption of statistical independence requires that the observed data-values $\{q_1, q_2, \ldots, q_N\}$ are realisations of random variables $\{Q_1, Q_2, \ldots, Q_N\}$. Sample data possessing serial elements cannot be thought of representing a population of random variables. Whereas flood events can essentially be thought of as stochastically independent, in the case of low flow frequency analysis the assumption of statistical independence seems untenable because of the serial correlation between data: low flows tend to follow low flows. Catchment storage processes increase the degree of dependence, or auto-covariance, of low-flow data; hydrographs for catchments with high storage capacity usually show only one recession period each year. Carry-over of base flow conditions from one year to the next might also be expected to occur in some years (Round & Young, 2001). The auto-covariance will be more pronounced for longer duration minima where adjacent samples will share d-1 common components, where d is the number of days in the duration.

Statistical dependence has an important effect on the interpretation of the results from low flow frequency analyses. Suppose, for example, that the analysis suggests that an annual minimum flow will be below a given level with a probability, p. Then, under the assumption of independence between years, the probability that minimum flows for two adjacent years will both be below the chosen level is given by p^2. In contrast, if there were strong positive dependence, the probability would be close to p. It is clear that ignoring statistical dependence may lead to substantial underestimation of the risks of sequences of years with low flows.

Testing for dependency within time series

A pair of random variables are only truly stochastically independent if their joint density is equal to the product of their marginal densities. This and other criteria can be used to test for auto-correlation in data. Provided the data is stationary the degree of independence can be characterised by determining the autocorrelation function of the observed data, or by constructing an auto-correlogram from the data. There are a number of methods and statistics for testing for the presence of autocorrelation. Some of the more widely used include the Durbin-Watson statistic, Anderson's Test and the Wald-Wolfowitz Test; these are described in most statistical texts (e.g. Shahin *et al*., 1993). However, as low flow frequency analysis usually has to be performed on a very limited data, these tests cannot always be expected to perform well and it may be difficult to quantify the level of statistical dependence in the data.

Where a time series is found to be statistically dependent, there are three possible actions: filter the time series to remove autocorrelation effects (requires expert judgement), allow for dependence in the distribution function models or take account of independence when quantifying the model uncertainty. It is difficult to formulate and to fit statistical models incorporating dependence so as to create new procedures for estimating the distribution function of annual minimum flow. Therefore unless linear stochastic models are used (Sen, 1980; Chung & Salas, 2000) it can be more useful simply to be aware of the effects of dependence on the usual procedures that happen to have been derived on the assumption of statistical independence between sample values. While parameter estimation will be correct (i.e. then the estimates will converge to the correct answer: to the true value of the distribution function or to the true value of the parameter of the distribution), serial persistence will effect:

i) Assessment of the uncertainty in parameter estimation
ii) Goodness of fit tests.

Where the serial persistence is strong, the assessment of uncertainty will probably be incorrect. In fact, it is likely that the true uncertainty with which the parameters of the distribution can be estimated will be underestimated. Thus in this situation an estimated sampling variance for an estimate provides only a lower bound for the true variance of the estimate (i.e. the quantile estimates will become more biased and there will be larger standard error in the quantile estimates).

Non-independence can also influence the results of a formal test of fit of a statistical distribution. Again it is difficult to account for statistical dependence explicitly in revised testing procedures. The effect of using tests devised for statistically independent data in the presence of positive statistical dependence is that the test statistics employed will be more variable than the underlying theory suggests. This means that the null hypothesis that the sample data arise from the given family of distributions will be rejected too frequently when it is actually true. Thus the results of tests of fit would need to be treated with some distrust. While it seems possible that a procedure that looks at the relative values of test-of-fit statistics across a number of trial families of distributions would be less affected by statistical dependence, this is at best speculative.

2.3.3 Homogeneity, stationarity and outliers

Homogeneity

As discussed in Section 1.2.2, low flow frequency analysis is based on the assumption that the all the observed data-values derive from a single population that can be described by a particular distribution function. This restriction of homogeneity ensures that the distribution function is also able to describe all future values of Q. Low flow data might not be homogeneous if the flow regime changes over time (i.e. the time series is non-stationary) or if low-flow events have different origins. An example of the latter case occurs in snow and ice regions, where rainfall being stored as snow in the winter period results in low-flow events unrelated to those that occur in summer as a result of dry weather. It is essential to test for non-homogeneity prior to frequency analysis.

Non-stationary behaviour

A data series is said to be non-stationary of the underlying properties of the data change over time (i.e. displays a trend or fluctuations). Causes of non-stationarity include problems with the recording process, such as changes in rating equations, relocation of stations or changes in recording methods, changes in the catchment, such as changes in land use, and climatic variability or climate change. Short records are particularly vulnerable to the effects of short-term trends, which might average out over the long-term. A number of statistical tests can be used to test for non-stationarity, such as Spearman's Rank correlation test and the Mann-Kendall test. These are usually based on finding a relationship between the flow variable and time, and are described in detail in most statistical texts (e.g. Shahin *et al.*, 1993).

Outliers

Outliers maybe caused by artificial influences in the catchment, such as the effects of over-abstraction, and in this case may be removed by naturalising the data. Also, where the period of record is short, and the range of magnitudes of low flow events small, extremely rare flow events may appear as outliers. The subject of outliers in the analysis of extreme events is tricky, as random samples drawn from statistical populations can produce outliers in hydrological samples (Cunnane, 1989). If apparent outliers are removed from the sample could this influence the resulting parameter estimation? One method for accounting for outliers is to use discordancy tests specially adapted to each distribution. Alternatively outlying events may be justified by seeing if they occurred on a regional basis, for example by comparing records obtained at nearby sites, or by using local knowledge or records.

2.3.4 Record length

It is very important to consider record length in frequency analysis. It will affect not only the sample size, but also the performance of tests for stationarity, and homogeneity. Most importantly, it is very unlikely that the frequency distribution will be determined adequately where the record length is short. Different techniques need to be adopted where the record length is inadequate, or it is possible to augment a record with data from a donor or analogue station. A scoping study for short records is presented in Chapter 6. This includes some discussion of the criteria for determining the suitability of analogue stations and the methods by which data or information about the frequency curve may be borrowed. Other approaches include regionalisation methods or the adoption of partial duration data. However these are beyond the scope of this review.

2.4 Non Parametric Methods

2.4.1 Empirical distribution functions

If a sample of data contains a large number of data-points it is not necessary to make use of a parametric family of distributions in order to provide an estimate of the underlying distribution function. Instead, one could use the empirical distribution function of the data. This is defined as the function whose value, for any x, is the proportion of the values $\{x_1, x_2, \ldots, x_N\}$ which satisfy $x_i \leq x$: this is the direct analogy of the underlying distribution function. Formally, the empirical distribution function, denoted $F_N(x)$, is defined by

$$F_N(x) = N^{-1} \Sigma_i \, I(x_i \leq x) \qquad (2.12)$$

where I is an indicator function which has the value 1 if the condition in brackets is true, and 0 otherwise, and N is the number of data points.

There is a considerable amount of statistical theory relating to the empirical distribution function, and in fact many formal tests of the fit of a given family of distributions are based upon it. Empirical distribution functions have two advantageous features. Firstly $F_N(x)$ is always an unbiased estimator of F(x), regardless of statistical dependence amongst the random variables. Secondly, given the assumption of statistical independence, for any fixed x, the distribution of the estimator $F_N(x)$ is completely known in a simple form: in particular $N \times F_N(x)$ is a Binomial random variable corresponding to a sample size N and a "probability of success" p = F(x).

Unfortunately the empirical distribution function is always a step-function, as is the corresponding estimated quantile function. Thus, the estimated flows of a given probability, p, take sudden jumps in value as p varies. This feature is unrealistic and best avoided (this is one reason why parametric methods are preferred over empirical techniques). In order to avoid the step-wise jumps in the empirical distribution function, some form of smoothing could be contemplated. Alternatively, the problem could be viewed as one of finding a reasonably smooth curve passing near the plotting positions. While a range of techniques based on smoothing are available for estimating density functions (Silverman, 1986), and hence distribution functions, these are not often used in the analysis of hydrological extremes. This is because estimates are usually required for values in the tails of the distributions, often beyond the range of the observed data. Kernel-based smoothing techniques, however, have the advantage of not requiring the specification of a particular family of distributions to be fitted, but would not be expected to perform well beyond the range of the data, unless the kernel chosen is specifically adapted to the underlying `true' distribution.

2.4.2 Probability and plotting positions of sample data

In hydrology, the empirical distribution method it is not often applied for graphical purposes. Instead the idea of a set of plotting positions is used to produce probability plots. A probability plot depicts the variation in the annual minimum flow with probability. However as the observed data is only a sample from the true population of flows, the true probability of each data point is unknown. The probability at which each data point $x_{(i)}$ should be plotted may be estimated; typically estimation follows the procedure outline below, and in this case the estimated value is called the plotting position p_i.

Plotting positions are calculated according to the position or rank of the data point within the data set. For this, the set of data-values is reordered according to size, with the largest value being assigned a rank of 1 and smallest value being assigned a rank of N. The new set is denoted by $\{q_{(1)}, q_{(2)}, \ldots, q_{(N)}\}$ in which $q_{(i)} \geq q_{(i+1)}$, thus $q_{(i)}$ denotes the i'th largest observation. An estimated value of p_i is then assigned to each rank, i. Various formulae for assigning values of p_i have been proposed, each evaluating the distribution function at $q_{(i)}$ in one sense or another. Cunnane (1978) and Stedinger *et al*. (1992) have discussed the rationale behind the use of different plotting position formulas. The plotting position should be a distribution free estimate of the probability: most plotting positions formulas try to ensure the resulting quantile estimates are unbiased for different distributions. Thus plotting positions of i/N and i-1/N, which were used in early frequency plots, have been replaced by more robust formulas. Some of most widely applied plotting positions are the Gringorten, Hazen, Weibull and Blom formulae as shown in equations 2.13 to 2.16.

$$p_i = (i - 0.44)/(N + 0.12) \qquad \text{(Gringorten plotting positions)} \qquad (2.13)$$

$$p_i = (i - ½)/N \qquad \text{(Hazen plotting positions)} \qquad (2.14)$$

$p_i = i/(N+1)$ (Weibull plotting positions) (2.15)

$p_i = (i – 3/8)/(N + 1/4)$ (Blom plotting positions) (2.16)

Certain formulae are optimised for particular distributions. For instance the Gringorten formula is optimised for the EV1 distribution, but is also a good choice for many others including the EVIII (Weibull) and exponential distributions. The Hazen formula is a traditional choice, and assumes that the probability scale is divided into N equal intervals, with p_i the midpoint of each interval. The Weibull formula provides unbiased estimates of the exceedance probability associated with each data point, rather than for the quantiles of the distribution, and for this reason, it is best used with a uniform probability scale. The Blom formula works best with data sets that are normally or log-normally distributed.

Typically, probability plots consist of displaying $q_{(i)}$ vertically against $[q_S(p_i)]$ for some standard distribution function $F(q)_S$ with quantile function qF_S. By using a probability scale along the ordinate axis the data points can be made to plot approximately along a straight line. The probability scale may be expressed in terms of the reduced variate for any particular distribution (for example, the general case reduced variate for the EVI or Gumbel distribution is (-ln(-lnF))). If the probability scale used corresponds to the true distribution underlying the sampled data, then the data is more likely to plot linearly. If the standard distribution used matches the true distribution then

$$q_{(i)} = q_S(p_i) \quad (2.17)$$

so that a line of unit slope should arise in the plot. If the true distribution is such that it is a shifted and scaled version of standard one, then

$$q_{(i)} \approx a + b\, q_S(p_i) \quad (2.18)$$

and hence the plot of $q_{(i)}$ against $q_S(p_i)$ will still be a straight line. Relationship (2.18) may then be used to estimate the parameters a and b.

By employing a variety of reduced variate scales, this simple method can be used to indicate which of well-known distribution families best represents the sampled data. Probability plots are therefore extremely useful for visually revealing the character of a data set, without the laborious task of fitting a distribution to the data.

Plotting positions are also the usually means of providing estimates of the probabilities for each of the sample flows, for use in parametric estimation of the probability distribution. It is therefore important to note that the plotting positions are predominantly influenced by the number of observations in the sample set. This means that plotting positions will be equally spaced along the probability axis, regardless of how the *magnitudes* of the observed flows are distributed.

2.5 Parametric Methods

2.5.1 Choice of family of distributions

General considerations

Parametric methods require an appropriate family of distributions to be identified. The choice of family can be based on a number of different considerations: some of the general points to be considered are described in sections a to e below. Although the same kinds of considerations are made in flood frequency analysis, the general conclusions about distribution fitting from floods may not be appropriate when transferred to low-flows because the shapes of distributions likely to arise are different. In particular, for droughts the possibility of the distribution having a discrete component at zero needs to be considered. The role of discretization and zero flows are discussed in sections 2.5.1.2 and 2.5.1.3 respectively.

a) The specific set of data

Formal and informal tests of fit can be applied to decide which of a number of alternative families of distributions is most appropriate. Certain pre-fitting analyses may also be useful. These include (i) graphical techniques, in which plots of the data-values against certain transformations of the plotting positions would be expected to produce straight-lines if the corresponding family of distributions was "correct", and (ii) examination of the sample moments of the data (e.g. judging kurtosis and skewness, or corresponding L-moments, in comparison with the theoretical moments for various families of distributions).

b) Use of similar data-sets

A specific data-set, particularly if it contains few data values, provides little information to discriminate between families of distributions. It can be reasonable to assume that, if there are other similar sets of data of the same type these can be used to help to find an appropriate family of distributions. If there are longer series of annual minimum flows for other nearby sites for which particular families of distributions have been found useful: then the same families might be expected to work well for the target site. Typically, a large number of sites might be analysed simultaneously and, for general use, those distributions that are judged to provide acceptable of goodness-of-fits statistics for the most number of sites are chosen. The question of what other sites provide data that are 'similar' to a given site is an ill-defined one: although droughts are. by definition are strongly spatially correlated, geographical considerations other than just distance have to be accounted for (e.g. see Acreman & Sinclair, 1986).

c) Applicability of Extreme Value Theory

"Extreme Value Theory" is a subset of statistical theory relating specifically to the statistical behaviour of maxima and minima of large collections of (not necessarily independent) random variables. This theory strongly suggests that the Generalised Extreme Value (GEV) family of distributions is appropriate for describing such data sets. Some statistical aspects of extreme value theory are outlined in Appendix 1.1: a full treatise can be found in most textbooks on statistical frequency analysis. In practice, however, a relatively short series of annual minima flows is unlikely to behave like a "true" extreme value data set. Therefore there are doubts about the applicability of extreme value theory to low flow frequency analysis. Furthermore, while this theory may be relevant to the independent data, it is not effective where data is serially correlated. In addition, the theory requires that the probability of a zero flow occurring is not positive, a condition that cannot be guaranteed for many sites.

d) Empirical considerations (optimising the number of parameters)

Particular families of distributions are "more flexible" than others in the sense that they are able to provide matches for a wide range of shapes of the underlying distribution. However, it is unlikely that the "true" distribution of annual minimum flows will coincide exactly with any member of a particular distribution family. To increase their flexibility, families of distributions can, in principle, be constructed in such a way that any possible distribution can be matched by some family member or other. This essentially involves using increasing the number of parameters that index the family. A large number of parameters may be required in a family of distributions in order for it to adequately represent the underlying distribution. Unfortunately the more parameters that are estimated in fitting the distribution, the more variability there is in the final estimates. Therefore the need for flexibility in the chosen family of distributions has to be balanced carefully against the need to restrict the number of parameters fitted. Such considerations obviously need to take into account the amount of data available for fitting the distribution and also the range of return periods of interest. Furthermore, any conclusions will also depend on the methods of fitting assumed to be used, especially with regard to the statistical efficiency of the method, its robustness to outliers and, more generally, its behaviour when the underlying distribution is not that assumed in the fitting procedure.

The evaluation of these effects may best be achieved by stochastic simulation studies where pseudo-random samples representing typical sets of data are generated from known population distributions and a variety of combinations of distributions and fitting procedures are applied to them. Such studies have been done in the case of floods - a summary is given by Cunnane (1989). Cunnane (*op cit.*) concluded that, generally, unless there are more than 20 to 30 years of data for a site, it is best to fit a two parameter distribution rather than one having more parameters. He also emphasised that, even for larger sample sizes, consideration has to be given to balancing the effects of bias, arising from lack of fit to the unknown underlying distribution, with those of sampling variability.

e) Applicability across different events of different duration

Several low flow frequency analyses may be carried out on the same data set, with annual minima derived using a different duration, D, in each case. Although each analysis has to be conducted separately, it might sometimes be preferable to use the same family of distributions to characterise low flow events for all durations of interest. This consideration may help to choose between candidate families if one appears to fit adequately well for a range of durations, while others do not. This, of course, leads on to the question of fitting distributions for different durations as part of the same analysis. Note that the question of treating different durations has less prominence in studies of floods, so that there is no experience of this problem on which to draw.

Modelling the recording process: discretization

In choosing a family of distributions to attempt to represent a set of data, it is important to take account of the way in which the individual data values recorded relate to the flows that actually occurred. One aspect of this relates to measurement error, which may occur, for example, due to a not-quite-correct rating curve or from observer-derived random errors. Such measurement errors should be small and the usual approach would be, effectively, to ignore them, on the understanding that the results of analyses apply to the distribution of measurements likely to arise, rather than to the distribution of the actual flow values. For streams where low flow values are very small or close to zero, the question of discretization, or rounding error, takes on rather more importance. For example if the precision of the gauging process is 0.01, it is readily apparent that all the values recorded will be multiples of 0.01 Cumecs, with the effect that the data-values to be modelled will have a discrete distribution. While it would be possible to attempt to model these data with a discrete distribution selected from one of the standard sets of families of discrete distributions (e.g. Poisson, Binomial, negative Binomial etc.), this is not the approach generally taken.

Instead, as for sites where minimum flows are typically larger, continuously-valued distributions are used. In effect the continuously-valued distribution is assumed to reflect the distribution of the "true" annual minimum flow and the discretization is modelled explicitly. Thus if the minimum flow, Q, has an underlying distribution function of $F_0(q)$, the probability that the observed value Q = q arises by rounding to the nearest multiple of δ is given by

$$\Pr(Q = q) = \Pr(q - \tfrac{1}{2}\delta < X \leq q + \tfrac{1}{2}\delta) = F_0(q + \tfrac{1}{2}\delta) - F_0(q - \tfrac{1}{2}\delta). \quad (2.19)$$

In this way, the discrete distribution for the observations is implicitly defined within Eqn. 2.19. The advantage of using a continuously-valued underlying distribution is that it provides a means of dealing with cases where the rounding interval is different in different portions of the record: for example, on the change from imperial to metric units. However, the algorithmic procedures for fitting the underlying continuous distribution cannot simply be carried over to the discretized case. Where the discretization effect is large, new procedures that are able to account for this effect need to be employed: the method of maximum likelihood applied to the discretized distribution would usually provide a reasonable estimation procedure.

The above approach is appropriate when the discretization effects are such that it is as if the annual minimum values being analysed are the direct outcome of a discretization process. Thus, for daily flows one has that:

discretized minimum daily average flow = minimum discretized daily average flow

However, when longer durations are used, the values extracted for a particular D-day duration are averages of discretized values. In this case the minimum D -day average discretized flow is not identically the same as the discretized value of the minimum of the D -day averages. Hence, although it may still be useful as a way of allowing for the measurement uncertainty arising from discretization, it is not exactly correct to employ relation (2.19) for longer durations. Note that, while the data-values for minima of D-day averages of values (originally recorded at a spacing of δ) would have an effective spacing of δ/D, it seems appropriate to use the uncertainty range of $\pm\tfrac{1}{2}\delta$, as in equation (2.19).
Some consideration should always be given to the question of discretization, even when its effects are not apparent in the data, since neglecting it when it exists will lead to incorrect answers. When the range of variation between values in the annual minimum data-set is considerably larger than the discretization error, it may be possible to ignore discretization entirely. However, it is important to at least consider giving special treatment to the smallest observation of annual minima in a data-set and determining what range of values this observation might reasonably represent. One reason for this is that, when estimating an unknown lower bound to possible values of annual minimum flow, one of the standard

statistical estimation procedures (maximum likelihood) is known to perform poorly if applied to "exact" data. This problem is overcome if the smallest observation is explicitly treated as representing a range of values.

Treatment of zero flows

Zero flows lie on the boundary of admissible values, as all observed flows must take non-negative values. In this sense occurrences of zero flow require some sort of "special" treatment. Even if zero values do not occur in the data-set to be analysed, there is still the question of how the possible occurrence of zero values should be treated in the formal description of the statistical distribution to be fitted. The consequences of treating zero flows in a particular way, therefore, need to be considered in any procedure for fitting the distribution. If zero values do occur in a data-set, some thought needs to be given to whether these are "real" zero flows, whether they represent values which are actually positive but are below some discretization threshold (as in Section 2.5.1), or whether they may represent a mixture of both types. Depending on the care with which the original data were recorded, it may be possible to distinguish between real zero values and small positive flows, in the same way that manually recorded daily rainfalls distinguish between zero and "trace" amounts.

Suppose that a data-set contains no "exact" zeroes and no discretization. It is possible that, if a statistical distribution is selected and fitted without taking into account the important role of zero as a lower bound, the fitted distribution may be such that positive probability is assigned to negative flow values. This would mean that some estimates of flows corresponding to given return periods might turn out to be negative. Some families of distributions do have zero as a fixed lower bound. However it may be necessary to consider more general families with the aim of achieving an improved fit. Suppose that an initial candidate family of distributions is such that the distributions either have no lower bound or the lower bound is non-zero, or negative (the particular location of the bounds depends on the parameters of the particular family). Let $F_0(x; \theta)$ denote one of the initial families of distributions, with parameter vector equal to θ, and let $x(0; \theta)$ be the quantile function evaluated at probability zero: i.e., the lower bound of the distribution.

Four ways of proceeding are as follows.

(i) *Ignore possible negative values.*

It may be the case that fitting the initial family of distributions, without taking any specific action about negative values, yields a fitted distribution for which the probability assigned to negative values is very small, say of the order of 10^{-10}. It is arguable that such small probabilities are irrelevant to the question of estimating quantiles at more moderate return periods, for example in estimating events occurring once in 1000 years or equivalently at a probability point of 10^{-3}. For comparison with other approaches this approach can be described as being to fit the family of distributions $F(q; \theta)$, where

$$F(q; \theta) = F_0(q; \theta). \tag{2.20}$$

While this approach may be tenable where analyses are done in detail on a case-by-case basis, so that the probabilities of negative values can be properly examined, it would be dangerous to adopt it as a fully automatic estimation procedure.

(ii) *Treat as a censored distribution*.

There are two standard ways of modifying the initial family of distributions to create families in which the probability assigned to the occurrence of negative values is zero. In the first of these, the new family is derived by assuming that the recorded data represent values arising from one of the initial family of distributions, but that any negative values would be recorded as zero. These might be thought of, for example, as water levels falling below the river bed level.

This rationale effectively creates a new family of distribution functions $F(q; \theta)$, defined by

$$F(q; \theta) = 0 \qquad \text{if } x < 0,$$
$$F_0(q; \theta) \qquad \text{if } x \geq 0. \tag{2.21}$$

Thus the new model incorporates a discrete component at zero: the model says that the probability of recording zero is $F_0(0; \theta)$. Here, even though the observed data include no "exact" zeroes, the statistical model does allow for their occurrence. Note that, if no zero values are observed, the estimated parameter values for this model may or may not coincide with those obtained for the model in Eqn. 2.20, depending on the method of estimation used. For example, maximum- likelihood estimates would remain unchanged, while estimates derived by the method of moments would change, essentially because the formulae for the population moments of the distribution defined by Eqn. 2.21 differ from those of the distribution defined by Eqn. 2.20. However, the estimated quantiles would never be negative.

(iii) *Treat as a truncated distribution.*

The second standard way of modifying the initial family of distributions is that of truncation, which may be considered rather more artificial in the present context than that in (ii). In the truncation model the final recorded value for a data-item is said to arise from a process in which a sequence of values is generated according to a particular distribution (family), but which stops at the first non-negative value. This effectively creates a new family of distribution functions $F(q; \theta)$, defined by

$$F(q; \theta) = 0 \qquad \text{if } q < 0,$$
$$F(q; \theta) = \{F_0(q; \theta) - F_0(0; \theta)\}/\{1 - F_0(0; \theta)\} \qquad \text{if } q \geq 0. \tag{2.22}$$

Here it is assumed that the initial distribution does not have a discrete component at zero. Then, in contrast to case (ii), the new model also does not incorporate a discrete component at zero. Instead, the unwanted probability of negative values is effectively spread over the positive values by scaling up their probabilities of occurrence by an appropriate factor. The estimated parameter values for this model would generally not coincide with those obtained for either of the two models above. Again, as for case (ii), the estimated quantiles would never be negative.

(iv) *Constraining the lower bound*.

The fourth possible approach is to devise an estimation scheme that will force the parameters estimated to be such that the corresponding lower bound of the fitted distribution will always be non-negative. This of course assumes that the initial family of distributions contains distributions that do have non-negative lower bounds. Thus the estimation problem can be stated as fitting the following constrained family of distributions, which coincides with the original family except that the range of acceptable parameter values is restricted:

$$F(q; \theta) = F_0(q; \theta), \qquad \text{with } \theta \text{ satisfying } q(0; \theta) \geq 0. \tag{2.23}$$

The point here is that estimation procedures for fitting this family of distributions would need to be specifically tailored to this problem and would not be identical to those for the initial family in Eqn. 5.21. Note that a different family of distributions would be constructed by using an alternative constraint specifying that the lower bound should be exactly zero.

Suppose now that a data-set does contain some "exact" zeroes, but that there are no other discretization effects. In the first instance, suppose that a family of distributions for positively-valued quantities has been found useful elsewhere – the next obvious step is to extend it to cope with the discrete component at zero. Then main approach is to construct an augmented family of distributions, which contains distribution functions $F(x; \beta, \theta)$ of the following form

$$\begin{aligned} F(q; \beta, \theta) &= 0 && \text{if } q < 0, \\ &\beta + (1-\beta)F_0(q; \theta) && \text{if } q \geq 0. \end{aligned} \tag{2.24}$$

where it is assumed that the initial distribution functions, F_0, satisfy $F_0(0; \theta) = 0$. Then the new parameter β represents the probability that a zero value will be recorded. If instead, the initial family of distributions does contain members that allow negative values to occur, then one possibility is to treat the observed zero values as arising from a censored version of one of the underlying distributions, as in case (ii) above and Eqn. 2.20. In this case the number of zero values recorded, as well as the other data-values, would be taken into account in procedures to estimate θ. This provides, albeit by a roundabout route, the corresponding value of $F_0(0; \theta)$ as an estimate of the probability that a zero value will be recorded. A second possibility is to augment the truncated version of the underlying distribution with "exact" zeroes (as in Eqn. 2.24 above). This approach may be favoured if the proportion of data-values that are recorded as exact zeroes is large, and gives the revised distribution function $F(q; \beta, \theta)$ of the form

$$F(q; \beta, \theta) = 0 \qquad \text{if } q < 0,$$
$$F(q; \beta, \theta) = \beta + (1-\beta)\{F_0(q; \theta) - F_0(0; \theta)\}/\{1 - F_0(0; \theta)\} \qquad \text{if } q \geq 0. \qquad (2.25)$$

Note that while it is also possible to augment the censored version of the underlying distribution, this yields a similar result to the family of distributions obtained by truncation. However, the probability given to zero values would be constrained to be above $F_0(0; \theta)$ if, as formally required by the derivation, the augmentation probability were forced to be non-negative. In the form Eqn. 2.25, the probability of recording a zero value is treated as a separate parameter, β, which can lie anywhere in the range 0 to 1.

Finally, suppose that a data set is affected by discretization, as in Section 2.5.1. The number of values recorded as zero could be modelled by augmenting an existing family of distributions, where this family formally allows the possibility of negative values. Let δ represent the discretization interval, so that a recorded value of zero represents underlying values of up to ½δ, a recorded value of "δ" represents values between ½δ and 1½δ, and so on. Then, one way of proceeding is to augment the underlying distribution once it has been truncated at ½δ. For the discretely-valued observations: this gives:

$$\Pr(Q = j\delta) = \beta, \qquad j = 0,$$
$$\Pr(Q = j\delta) = (1-\beta)[F_0\{(j+½)\delta; \theta\} - F_0\{(j-½)\delta; \theta\}]/\{1 - F_0(½\delta; \theta)\}, \qquad j = 1,2,3,... \qquad (2.26)$$

This may be compared with Eqn. 2.18, which omits both the augmentation at zero and the truncation.

2.5.2 Methods of fitting distributions

Introduction

Two types of parametric estimation are available for low flow frequency estimation: fixed-estimate substitution and Bayesian estimation. Nearly all of this report is concerned with estimation by fixed-estimate substitution, because this is the approach usually taken in practice. Fixed estimate substitution is described in more detail in Section 2.5.2. Bayesian inference techniques provide an alternative approach, which is able to take full account of all the uncertainties involved in a problem, including those arising from only having a limited amount of data on which to base a decision. In particular, Bayesian inference provides a complete and self-consistent approach to decision making is in contrast to non-Bayesian estimation, where new and modified ways of estimating parameters are often created on an *ad hoc* basis. There are, however, several problems related to applying Bayesian methods computationally. Although a full description of Bayesian inference techniques is beyond the scope of this report, a good introduction to the topic is given by Box and Tiao (1973), whilst the Bayesian approach to decision making is described in DeGroot (1970).

Fixed substitution methods

Assuming that a family of distribution functions $\{F(q; \alpha)\}$ has already been chosen, in fixed-estimate substitution the data are used to construct an estimate $\hat{a}$ of α. The parameter estimate is then used to estimate the form of the distribution function and the quantile function (i.e. $\hat{a}$ is substituted into the formulae for the distribution and quantile functions providing $F(q; \hat{a})$ and $q(p; \hat{a})$ respectively. Estimates of other quantities can then be derived in simple ways from these. For example, under the assumption of independence between yearly minima, the probability that the annual minima in K consecutive years will all be above a value x is

$$\Pr(Q_1 > x, Q_2 > x, \ldots, Q_K > x) = \{1 - F(x; \alpha)\}^K, \qquad (2.27)$$

and the estimated value of this quantity is provided by $\{1 - F(x; \hat{a})\}^K$.

Methods of parameter estimation can be divided broadly into five types, which may be described as follows:

(i) *Graphical techniques*

It may be possible to devise a way of plotting the recorded data-values against values derived from the "plotting positions" (Section 2.4.2) so that a straight line should be seen if the assumption about the form of the distribution is correct. The intercept and slope of this line can then be used to estimate the parameters of the distribution. Fitting of the straight line might be done by eye, but might also be done by formulae, for example, based on a least-squares fit for the line. The use of graphical techniques is described further in Section 2.5.2.3.

(ii) *Likelihood based techniques*

Estimation methods based on the idea of maximising the "likelihood function" of the data are well-established in statistical theory and practice and, under certain conditions, they are known to have certain optimality properties, at least for large samples of data. Bayesian parameter estimation might be thought of as a variant of this approach. The use of likelihood based techniques is described in more detail in Section 2.5.2.4.

(iii) *Exact matching techniques.*

A variety of different methods can be categorised as 'exact matching' techniques, because they are based on the same rationale. In this rationale, the parameter estimates used are those for which theoretical values of certain statistics calculated from the assumed distribution, exactly match the values of the same statistics, as calculated from the data. An example of this approach is the "method of moments". This type of approach has certain advantages over likelihood methods and has been popular within hydrology. However there are also some disadvantages associated with this methodology. For example, although the number of statistics used is always the same as the number of parameters estimated, there might be no set of parameter values for which an exact match is achieved. Further details regarding exact matching techniques are given in Section 2.5.2.

(iv) *Close matching techniques*.

In those methods that can be classed as 'close-matching' techniques, the estimated values for the parameters are specified to be those which provide the best match between certain properties of the theoretical distribution and the corresponding properties as measured for the data-values. This, is in a sense, a simple extension of the set of exact matching techniques but, rather than just allowing the number of different statistics used for fitting to increase, it does allow more direct ideas about measuring the closeness of distributions (of the theoretical population and of the data) to be used in estimating the parameters.

(v) *Optimal estimation techniques*.

Statistical theory can be used to construct "best possible" estimators. This approach typically only leads to practicable estimation procedures in a limited range of cases. It is usual to restrict the search to estimators that are "optimal" in the sense of being unbiased and of minimum variance. The optimality of an estimator produced by these techniques, of course, only applies if the assumptions made (either about the distributional form or about statistical independence of the observations) actually do hold.

There are four criteria that are relevant to the selection of a method for parameter estimation or distribution fitting: these are as follows.

(a) *Reproducibility*.

It is desirable that if different people apply the same estimation method to the same set of data, then they should obtain the same estimates. Similarly, if computer programs are used to execute the estimation method, programs constructed separately should also provide the same estimates. This is clearly not the case for methods relying on graphical procedures and on lines fitted by eye. Other methods may also fall short of fully meeting this criterion. For example, both likelihood-based and close-matching methods of estimation may involve numerical search techniques for locating the maximum or minimum of certain functions of the parameters. The results achieved will depend on the search method used, the initial values used for the parameters and the rules used for stopping the search.

(b) *Simple formulae*.

There may be a preference for estimation methods that involve application of simple formulae for various reasons, including ease of understanding, ease of programming and speed of execution.

(c) *Universality and uniqueness*.

In principle an estimation procedure should always provide a set of estimated values for the parameters. Certain basic estimation methods may sometimes fail to do this. For example, a set of equations in a moment-matching method may not have a solution, or a numerical search procedure may be faced with a function that is continually improving in certain directions of the parameter space. A related difficulty might occur where more than one set of estimates could be produced by certain methods. Thus a set of moment equations might have more than one solution or, in an optimisation problem, there may be two local optima of the same size. These different answers may or may not be such that they correspond to essentially different estimated distribution functions. A basic estimation procedure may need to be modified, possibly by adopting a second procedure as a fall-back, in order to ensure that a single answer is always produced.

(d) *Accuracy and robustness*.

Different estimation methods will have different statistical properties. Methods that provide better performance than others in terms of estimation accuracy, as measured by the sampling variability of the estimates, will be favoured. However there is the competing issue of robustness, which relates to the question of how well the method works in cases where the modelling assumptions do not hold. Measures of performance could be based on the estimation of the parameters, the distribution function or the quantile function.

No one type of estimation technique always does well under all of these criteria. In the context of low flow frequency estimation, it is arguable that the last criterion (accuracy and robustness) is most important, while simplicity and speed of execution are irrelevant. While there are certain theoretical results that suggest that likelihood-based methods are "good" in terms of accuracy, these apply only for large sample sizes. Thus comparisons of estimation methods for reasonable sample sizes, for both accuracy and robustness, must rely on the results of computer simulation experiments. While such computer experiments have been done with some success for flood frequency estimation, this seems not to be the case for low flow frequency estimation. The major question is the construction of suitable sets of "true" distribution functions for the present context: these would have to take account of the considerations of boundedness, "exact" zeroes and discretization which were discussed in Section 2.5.1.

Graphical techniques

Graphical estimation techniques are mainly useful for problems involving only location and scale parameters. Note that graphical procedures for assessing the fit of a distribution are applicable to all methods of estimation and to all families of distributions. In order to obtain a little extra generality for the technique, it is assumed that the set of values of annual minimum flow, $\{q_i\}$, may first be subjected to a known mathematical transformation to create a corresponding set of transformed values, $\{x_i\}$. For example a logarithmic transformation might be used in fitting the two-parameter lognormal distribution.

Let F_S be a standard distribution function (that is, one for which there are no parameters to be estimated) and let x_S be its (known) quantile function. A location-shift model for the observed data, which are represented by the typical random variable X, can be created by assuming that

$$X = \alpha + \beta X_S, \tag{2.28}$$

where X_S represents a random variable from the standard distribution. Here α and β will be parameters the distribution function F of X, and there is an assumption that $\beta > 0$. The model implies that F is given by

$$F(x; \alpha,\beta) = F_S\{ (x-\alpha)/\beta \}, \tag{2.29}$$

and that the quantile function corresponding to F is given by

$$x(p; \alpha,\beta) = \alpha + \beta\, x_S(p). \tag{2.30}$$

The graphical estimation technique is then essentially the same procedure used for plotting data: the transformed annual minimum flow data are ranked in ascending order and a plotting position is assigned to each point. The plotting position, p_i, is defined according to the rank, i, of the data point x_s, using the Hazen or Gringorten formulae as follows:

$p_i = (i - ½)/N$ (Hazen plotting positions) (2.31)

$p_i = (i - 0.44)/(N + 0.12)$ (Gringorten plotting positions) (2.32)

The general rationale for this procedure and for the specification of the plotting positions was discussed in Section 2.4.2. A scatter plot is created by plotting the N pairs of values $\{x_S(p_i), x_{(i)}\}$ for all ranks. The slope and intercept parameters of a relationship of the form

$$x_{(i)} \approx \alpha + \beta\, x_S(p_i) \qquad (2.33)$$

are then estimated by eye. A straightforward way of avoiding the subjective quality of the estimation procedure which is inherent in fitting by eye is to specify that the values of α and β should be determined by a least-squares fit of the "estimated values" $\{\alpha + \beta\, x_S(p_i)\}$ to the "target values" $\{x_{(i)}\}$; the standard formulae for this apply. However, such a procedure would not necessarily reproduce the behaviour of estimates created by fitting by eye. There would be some scope for tailoring the weights in a weighted least-squares procedure to the distribution being fitted, to reflect the idea that the largest and smallest values in the ordered sequence $\{X_{(i)}\}$ are subject to different sampling variability than those closer to the centre. It is also possible to extend the graphical estimation procedure to distributions that have more than two parameters. The methodology for this is briefly summarised in Appendix 1.2.

Likelihood based techniques

This section provides a brief introduction to likelihood-based estimation techniques, focussing on the Method of Maximum Likelihood, its advantages and disadvantages. For a more detailed description of the method, the reader is referred to Appendix 1.3 of this review. Maximum likelihood estimation is also discussed is most statistical textbooks, although this is often in the context of particular standard types of distributions. General discussions are provided by Cox and Hinkley (1974) and Hinkley and Reid (1991).

Likelihood theory deals with the probability of occurrence of a particular sample of data. Likelihood based estimation techniques can possibly best be thought of as providing the main general purpose set of estimation techniques within the framework of classical statistics. When a new estimation problem arises, likelihood theory can usually be relied upon to lead to the construction of a reasonably practicable procedure for parameter estimation. Usually the parameter values for which the probability of occurrence of the sample at hand is maximised (i.e. at the point of maximum likelihood) are used as the parameter estimates. Methods for finding maximum likelihood estimates include graphical or tabulation methods, numerical procedures for optimisation, numerical procedures for root-finding, algebraic solutions. It is also possible to use a mix of these approaches.

The advantages of being able to derive estimators of the parameters of a distribution via likelihood theory are as follows:

(a) The approach provides a single well-defined procedure for estimation on a firm theoretical basis, in contrast to other general procedures for which there may be a range of choices of detail that are either dealt with arbitrarily or demand specific investigation.

(b) It is applicable in a wide range of circumstances, including non-identically distributed observations and dependence between observations.

(c) The estimators are known to have certain optimal properties in the sense that, once the sample size is large enough, no other estimators have better properties.

(d) The encompassing likelihood theory provides methods for formally testing for the inclusion or exclusion of subsets of parameters within a model and hence, implicitly, for testing the fit of a model by seeing whether a more complicated model does better.

(e) Likelihood theory provides ways of assessing uncertainty both for model-parameters and for derived quantities such as quantiles.

Despite these many advantages, for the purposes of flood and drought analysis likelihood methods have often been passed over in favour of methods that use "non-optimal" estimating equations. This is because likelihood theory has a major disadvantage in that it does not always produce an estimation method that “works” in an acceptable sense. Likelihood theory works well provided certain conditions hold: unfortunately these conditions fail to hold for some of the distribution-fitting problems commonly encountered in analysing annual maxima and minima. Various problems can arise when attempting to apply maximum likelihood estimation in practice: the non-existence of a maximum, the existence of several local maxima, the presence of several equal-valued maxima, and the problem of non-distinct maxima. In some cases maximum likelihood estimators can only be derived by iterative numerical solutions. It should also be kept in mind that maximum likelihood estimators have asymptotic properties (for large N), and are not as applicable when the sample size is small.

Exact matching techniques

Well-known examples of exact matching techniques include estimation methods such as the "method of moments" and "probability weighted moments". It is convenient to group these moment or generalised-moment methods together with certain other techniques that would not typically be thought of as estimation via moments. The basis of this general class of procedures is to use the available sample of data to provide estimates of certain properties of the underlying population, and then to select the member of the family of distributions which has exactly the same properties as those estimated from the sample. Suppose that S denotes a vector of sample-derived estimates of statistical properties of the underlying population and $s(\theta)$ denotes the vector-function describing how the values of these properties for the family of distributions vary with θ, the vector of parameters indexing the family of distributions. The estimated parameter values are thus defined to be equal to the solution to the set of estimating equations:

$$S = s(\theta). \quad (2.34)$$

Here, it is understood that the solution sought is one that lies within an allowable set of parameter values. The term "exact matching techniques" refers to the fact that an exact identity is sought in Eqn. 2.34. "Close matching techniques" can be thought of as seeking an approximate solution to equations of the same form as Eqn. 2.34, but where the equations cannot be solved simultaneously.

There are many different ways of arriving at a set of estimating equations appropriate to a particular problem. In some approaches one would start with a set of sample statistics in the vector S, and then deduce a corresponding vector of "population-values", $s(\theta)$. In other approaches one can appear to be moving in the opposite direction, by starting with $s(\theta)$ and then seeking a reasonable set of sample statistics, S, which in some sense measure the same sort of properties as measured by $s(\theta)$. These approaches are discussed in more detail in Appendix 1.4. For certain purposes it can be helpful to slightly extend the notion of estimating equations to encompass estimating functions - this procedure is also outlined in Appendix 1.4.

To a considerable extent the choice of which statistical properties are used to construct the estimating equations is rather arbitrary. The choice for S or $s(\theta)$ represents a set of statistics which will be "reproduced" by the estimation technique: for example, samples generated from the fitted family of distributions would have properties exactly matching those of the observed data sample, in either a large-sample sense, or in an expectation sense. Hence it can be argued that the statistics used should be "important" ones to be reproduced in some practical sense, reflecting what the fitted distribution will be used for. However, in practice, this type of argument does not tend to be helpful in choosing statistics for use in the estimating equations. It is possible to try to set out certain general theoretically-based principles or procedures that can be used to define estimating functions or estimating equations. Some of these are as follows.

Moments

Estimation by the "method of moments" is well-known and is based on using the ordinary moments of a distribution as measures of location, scale and shape. Note that it is possible to use moments of a transformation of the original data (such as the logarithm) and that in various circumstances "mixed-moments" (combinations of moments of original and transformed data) have been employed.

L-moments

Recent developments in flood frequency estimation have led to the suggestion that L-moments should always be used in preference to ordinary moments on the grounds that these provide measures of location, scale and shape which are more reliably estimated than are ordinary moments (Hosking, 1990; Hosking and Wallis, 1997). However, these studies have been principally concerned with floods rather than low flows, and the suitability of L-moments for low flows seems to require further investigation. Note that using L-moments is equivalent to using "probability weighted moments", but with the advantage that they are more easily interpreted as measures of shape.

Distribution function values

A direct approach to constructing estimating equations is to use as the target values, $s(\theta)$, the values of the theoretical distribution function evaluated at a number of points (equal to the number of parameters). While it is better, theoretically, to use a fixed set of points on the data-value scale (which not related to the actually observed values), it is also possible to define these points in terms of the ranked observations in order to ensure that points are always chosen reasonably spaced within the range of the observed values. In this type of approach, the standard empirical distribution function (or some reasonable revision of it) would provide the set of sample statistics, S.

Quantile function values

An alternative direct approach to finding estimating equations can be based on using the theoretical quantile function of the distribution, evaluated at a set of fixed percentage points to form the target values, $s(\theta)$. Interpolating between the ranked observations in some standard way would provide the sample estimates of these quantiles. A different version would allow the percentage points used for evaluating the quantile function to vary slightly with the number of observations in order to avoid the problem of interpolation.

Maximum entropy

According to the principle of maximum entropy, a family of distribution functions can be characterised by a set of constraints, which in the present application are equivalent to the estimating equations, together with the condition that the family is precisely that set of distribution functions for which the entropy is maximised subject to the given constraints. Thus the constraints or estimating equations arise in a supposedly clear-cut way from the set of distributions being fitted. Singh & Guo (1995) provide an example of derivation of estimating equations via this approach. This approach seems to be extremely difficult to apply from scratch, and comparisons with other estimation methods via simulations indicate that it provides improved estimates only in exceptional cases.

Optimal estimating equations

While it is possible to develop a theoretically based approach to constructing an "optimal" set of estimating functions, it turns out (Godambe; 1960, 1976) that these are more or less the same as the likelihood equations. Thus use of optimal estimating equations is therefore essentially identical to maximum likelihood estimation.

Exact matching techniques have many advantages. In particular, it is possible to construct estimating-equation methods that provide estimates with good properties in cases where maximum likelihood estimation fails to provide an estimate at all. Estimating-equation methods can also have considerable advantages over maximum-likelihood estimation in terms of simplicity. Indeed, one possible criterion for the choice of a set of estimating equations is whether, or not, simple explicit expressions for the parameter estimates can be found.

The practical problems involved in implementing an exact-matching or estimating equation approach to parameter estimation are rather similar to those encountered for maximum-likelihood, once a set of equations to be solved has been settled upon. The problem is to find a solution to a set of equations under the condition that the solution must be in some allowable region of the parameter space. However, there may be no such solution, or there may be more than one. It might be possible to solve all of the equations algebraically to provide an explicit solution, or it might be possible to employ a mixed approach in which explicit solutions to a subset of the equations are substituted into the others, leaving a reduced number to be solved by numerical procedures for root-finding. There is clearly scope for alternative sets of estimating equations for the same problem to have rather different degrees of ease in finding a solution. The selection of a set of estimating equations can be rather arbitrary, particularly when a new problem is being dealt with, and also when improvements to existing procedures are being sought. However, guidance in suggesting suitable structures can often be gained from successful methods applied to other similar problems. Here "success" would typically have been judged from the results of simulation experiments. Similarly, it would be good practice to use simulation experiments to check out any new exact-matching estimation procedure and to compare it with any other competitors. For example in attempting to construct an exact-matching estimation technique for a new problem, which may typically involve modelling an observed data-set via a truncated and/or discretized version of a standard family of distributions, it is important to take into account this truncation or discretization in forming the estimation equations.

2.5.3 Assessing parameter and distribution estimates

Introduction

Although the form of the distribution function can be estimated by fitting methods, it is extremely rare that the estimated form is equal to the 'true' form of the distribution. In assessing whether the fitted distribution is acceptable, the bias, variability and accuracy of parameter estimates are taken into account.

The parameter α is estimated by $\hat{a}$, which is itself a random variable having a probability distribution. Ideally $\hat{a}$ should be noted as $\hat{a}$ $(X_1,\ldots..,X_N)$ to emphasise that its value depends on the observed samples $\{ X_1,\ldots..,X_N\}$. The estimate $\hat{a}$ is thus a realisation of the estimator, $\underline{\hat{a}}$. Probability theory can be used to describe the behaviour of the estimator, $\underline{\hat{a}}$. If the estimation procedure produces as systematic error, the estimator is said to be biased. The bias associated with the parameter α, $B(\alpha)$, is defined statically as follows:

$$B(\alpha) = E[\underline{\hat{a}}] - \alpha \qquad (2.35)$$

There will be no bias if $E[\underline{\hat{a}}] = 0$, and in this case the estimator is said to be unbiased: that is, each time the estimation procedure is performed the estimate may differ from the true unknown value, α, but will, on average, produce the correct result.

It is also desirable to have an estimator that has a small sample-to-sample variability, or variance, Var $[\underline{\hat{a}}]$. An estimator is said to be the most efficient if it is unbiased and if its variance is at least as small as that of any other unbiased estimator. The mean square error is one measure of accuracy that combines both bias and variance (Stedinger, 1993). The mean square error, MSE, is usually defined as follows:

$$\text{MSE} = \{B(\alpha)\}^2 + \text{Var}[\underline{\hat{a}}] \qquad (2.36)$$

An unbiased estimator will thus have a mean square error equal to its variance.

It is usually fair to assume that the estimates of the distribution parameters are in the 'neighbourhood' of the correct values. It is possible, then, to define a range, or **confidence interval**, that will nearly always contain the correct or 'true' value of the parameter, even if that value is still unknown. Similarly, when two or more parameters are involved it is possible to describe a confidence region. The construction of confidence intervals is described in most statistical texts. A useful step-by-step guide, with examples based on hydrological data, can be found in Shahin *et al*. (1993).

Screening distributions

Screening techniques are used to identify those distributions that do not fit the data well (rather than choosing the option with the 'best' parameter estimates). Several screening techniques are available including:

a) *Graphical procedures*

Physical inspection of the data on a probability plot is the most basic and traditional method for assessing the goodness-of-fit of a distribution. Confidence intervals can be drawn about the line or curve of the fitted distribution in order to help demonstrate its suitability or otherwise. However graphical procedures should only be used as a first step or in conjunction with other methods.

b) *Goodness of fit tests*

Goodness of Fit tests are the main method of testing that sample data are derived from a particular distribution. There are several different goodness of fit criterion including the Chi-squared test and the Kolmogorov-Smirnov statistics.

c) *Least Squares Criterion*

In the least squares method the sum of squares of the differences between predicted and observed flows at different return periods are determined and used to assess the applicability of the choice of distribution. The least squares method is dependent on the plotting position used to determine the return period, and is also computationally demanding.

d) *Tests based on skew*

Tests based on skew are generally used when the method of moments has been used to determine the parameter estimates. Skew can be identified visually, from moment-ratio diagrams.

e) Split Sample Test

In a split sample tests distributions are fitted first to half of each record. The second half of the record are used to test expected numbers of exceedance, or non-exceedances of specified magnitudes given the distribution in question.

Classical goodness of fit tests

There are two main types of goodness of fit tests suitable for testing the fit of a distribution to a particular sample. The Chi-Squared test assumes the observed data is drawn randomly from the population described by a particular distribution function. The Kolmogorov-Smirnov test is a similar test, but is only valid for continuous distributions. The Kolmogorov- Smirnov test has the added advantage that it can be used for a small sample size, and can be used even where the parameters of the distribution have not been specified. In this case the parameters are estimated from the mean and variance of the sample data set. Note that goodness of fit tests assume that the data is independent, so that where data is serially correlated the results of these tests are less reliable.

Resampling methods

Resampling methods are applied in order to assess or reduce bias in a biased estimator, and are used when statistical models become so complicated that it is impossible to used ‘standard’ statistical techniques to analyse them. Resampling methods, such ‘bootstrapping’ and ‘jack-knifing’ are used for constructing confidence intervals for any complicated situation. Techniques such as Synthetic Population Assessment also fulfil a similar role (Round & Young, 2001). As an alternative to resampling methods Bayesian analysis may also be used to choose the best distribution out of a number of distributions.

Bootstrapping techniques

Bootstrap methods are statistical procedures used for statistical testing or for assessing the variability of a point estimate where the underlying statistical population is unknown. In order to apply this method it must be assumed that the low flow events in each year are independent of the flows observed in the preceding years and that the range of annual low flow events in the observed data is the same as in the underlying population. Bootstrapping would typically involve determining period of record flow statistics for the observed period, randomly generating resamples, and yielding confidence intervals for these flow statistics. The key assumption for deriving bootstrap confidence intervals is that the bootstrap residuals are assumed to be representative of values drawn from the same distribution as the actual unknown residual. Unfortunately whilst the bootstrap method provides an estimate of the confidence intervals for the period of record statistics, full rank re-samples have to be drawn from the sample used to calculate the period of record statistic (Round & Young, 2001).

2.6 Families of Distributions

A number of families of statistical distributions are commonly used to represent hydrological data. This chapter discusses some of the distributions found useful in flood frequency and low-flow analysis. It is important to note that those distribution useful for flood frequency analysis, may not have good properties when "reversed" for low-flow analysis, because of the restriction of flows to be positive. It is also important to note that distributions with two or three parameters are more appropriate than higher-order distributions.

2.6.1 Generalised Extreme Value (GEV)

The GEV distribution is the general three-parameter case of the range of extreme value distributions (including EV1, EVII and EVIII distributions). It is itself a special case of the Kappa distribution. The three parameters are location, ξ, scale α and shape, k.

The distribution and density functions are given respectively by:

$$F(x) = \exp(-e^{-y}) \quad (2.37)$$

$$f(x) = a^{-1} \exp(-(1-k)y - e^{-y}) \quad (2.38)$$

where the quantity *y* is determined as follows:

$$y = \begin{cases} -\mathrm{k}^{-1} \ln\left(1 - \mathrm{k}\dfrac{\mathrm{x} - x}{a}\right) & where\ \ \mathrm{k} \neq 0 \quad (2.39) \\ \dfrac{\mathrm{x} - x}{a} & where\ \ \mathrm{k} = 0 \quad (2.40) \end{cases}$$

The distribution is bounded above where $k > 0$ and is bounded below where $k < 0$, i.e. the range is $-\infty < x \leq x + \frac{a}{k}$ if $k > 0$, $x + \frac{a}{k} \leq x < \infty$ if $k < 0$, and $-\infty < x < \infty$ where $k = 0$.

Therefore the GEV is bounded above where k > 0 and this is a Type II GEV distribution. Where k < 0, the distribution is known as a Type III GEV and is closely related to the Weibull distribution. In the special case where the third parameter, k is zero, GEV distribution reduces to the EV1 (Gumbel) distribution. As suggested by extreme value theory, the GEV should give a reasonable fit for independent data. Indeed the GEV distribution has been found to work well with flood data and has been recommended for flood frequency analysis by many authors (e.g. NERC, 1975). As discussed earlier, the

GEV distribution is less appropriate for low-flow frequency analysis because of the issue of serial correlation in low flow data.

2.6.2 Gumbel (EV1)

The Gumbel, or EV1 distribution is a special case of the Generalised Extreme Value distribution, and is possibly the most widely used distribution for flood frequency estimation (see Cunnane, 1989). The Gumbel, is a two-parameter distribution with scale parameter, α, and location parameter, ξ, and its p.d.f. is given by:

$$F(x) = \exp\left(- \exp\left\{ - \frac{x - x}{a} \right\} \right) \tag{2.41}$$

The Gumbel distribution is unbounded at both ends of the distribution (i.e. - 8 = x = 8), so its use in low-flow analysis is much less common. Krokli (1989) and Pearson (1995) consider fitting EV1 distributions for low flow data.

2.6.3 Weibull (EVIII)

The Weibull distribution is a three-parameter distribution given by

$$F(x) = 1 - \exp\left(- \left\{ \frac{x - z}{b} \right\}^{d} \right) \tag{2.42}$$

where the parameters δ, β and ζ are defined by $k = \frac{1}{d}$, $a = \frac{b}{d}$, $x = z - b$

The Weibull distribution is the equivalent to the Extreme Value Type III (EVIII) for minimum values. As it is bounded by zero on the left hand tail (i.e. $z \le x < \infty$)., the Weibull distribution is a popular choice for low flow frequency estimation (e.g. Deninger *et al*., 1996; Durrans, 1996; Guo *et al*.,1996), especially for those using graphical approaches e.g. Gustard *et al*. (1992).

2.6.4 Lognormal distribution

The lognormal distribution can be described using three parameters, in which case it can be defined by the following probability density functions:

$$\mathrm{f(x)} = \frac{1}{\sqrt{2p\mathrm{a}.(\mathrm{x} - \mathrm{m})}} \exp\left\{ - \frac{1}{2} \left(\frac{\log(\mathrm{x} - \mathrm{m}) - \mathrm{b}}{\mathrm{a}} \right)^2 \right\} \text{ where m < x} \tag{2.43}$$

If defined using two parameters, the parameter m is set equal to zero in Equation 2.40.

The use of log-distributions is difficult when dealing with low flows because a zero in a set of data that is being logarithmically transformed requires special handling. However the log-normal distribution has been a population choice when dealing with lowflow data (e.g. Vogel & Kroll, 1990): an un-reversed log-Normal distribution has no upper bound and a lower bound at zero.

2.6.5 Generalised Logistic

The Generalised Logistic (GL) distribution is a generalisation of the 2-parameter Logistic distribution and is also a special case of the Kappa distribution (Robson, 1999). It has been recommended for use in flood frequency analysis (Institute of Hydrology, 1999) as it is unbounded above unless the L-skewness is negative. The distribution can be defined by

$$F(x) = \frac{1}{1+e^{-y}} \tag{2.44}$$

where

$$y = \begin{cases} -k^{-1} \ln\left(1-k\dfrac{x-?}{a}\right) & \text{where } k \neq 0 \\ \dfrac{x-?}{a} & \text{where } k = 0 \end{cases} \tag{2.45}$$

In the special case where ?=0 the GL distribution reduces to the 2-parameter Logistic distribution. The range of possible values for the GL distribution is $-\infty < x \leq x + \frac{a}{k}$ if $k > 0$ and $\xi + \alpha\ / \mathrm{k} \leq\ \mathrm{x} < \infty$ if $k < 0$.

Thus the GL is bounded above for k > 0 and below for k < 0, and so is fairly good candidate for low-flow analysis. As the generalised logistic distribution is a fairly ‘new’, there are few references to it in flood frequency analysis, and has yet to be applied to low-flow data.

2.6.6 Pearson Type III

The Pearson Type III distribution is a popular distribution for fitting hydrological data. It can be described by the following probability density function

$$f(x) = \frac{\left(\frac{x-m}{a}\right)^{b-1}}{|a|.\Gamma(b)} \exp\left\{-\frac{x-m}{a}\right\} \quad (2.46)$$

Where ? is the Gamma Function as defined by Bobee & Ashkar (1991).

The range of possible values for the Pearson Type III are *m* = x if a>0 and *x* = 0 if a <0. The Gamma distribution is a special case Pearson Type III with *m* equal to zero. The Log Pearson distribution is another variant, but generally is only appropriate for hydrological analyses when its parameters fall within a small range of values.

2.6.7 Generalised Pareto distribution

The Generalised Pareto (GP) distribution is a three parameter distribution that may be described by

$$F(x) = 1 - e^{-y} \quad (2.47)$$

where

$$y = \begin{cases} -k^{-1} \log\left(1 - k\frac{x - ?}{a}\right) & \text{where } k \neq 0 \\ \frac{x - ?}{a} & \text{where } k = 0 \end{cases} \quad (2.48)$$

The GP distribution is useful for describing events that exceed a specified lower bound, such as low-flow events bounded by zero - it is actually bounded at both tails (i.e the range of x is given by $x < x \leq x + \frac{a}{k}$, where $k > 0$ and by $x \leq x < \infty$, where $k \leq 0$.

2.7 Recommendations and Concluding Remarks

2.7.1 Applicability of flood estimation techniques

This review has aimed to provide an introductory overview of the principles and techniques of low flow frequency analysis. The main elements of low flow analysis have been described and the advantages and disadvantages of some of the possible methods and distribution families have been outlined.

Although consistent methods for frequency analysis of flood events are widely disseminated with the hydrology and engineering communities (for instance, a thorough description of the recommended method of flood frequency analysis in the UK is given by the Institute of Hydrology (1999)), the same cannot be said about low-flow frequency analysis. The subject of low-flow frequency might at first appear to be analogous to that of flood frequency. However there are several key differences between the two, and therefore, although the concepts and general techniques may be transferred between flood and drought analysis, the 'nuts and bolts' of the methods have to be different. The main differences can be summarised as:

1) The need for a lower bound at zero (representing the possibility of observing a zero flow).
2) Due to a lack of precision in recording variables at low flows, a series of low flow observations can essentially be considered as a discrete distribution, rather than a continuous one.
3) The data may be serially correlated, especially where longer durations are considered.

There are two main implications of these nuances. Firstly a methodology for dealing with dependent data is required. Secondly there must be adequate treatment of zeros; a distribution with a lower bounded tail is required taking into account that the bound cannot be negative. The data may require treatment as if censored. The problem of low frequency analysis also needs to be approached differently for long and short flow records. Where records are sufficiently long (e.g. at least 20 to 30 years), it is appropriate to use annual minima series, whereas where the data covers a much shorter time period, partial duration techniques must be employed or at-site analysis must be supplemented by some kind of regionalisation method.

2.7.2 General recommendations

Non independence

In the one-day duration case it is fair to assume that the data is independent. However where longer durations are considered it is important to assess the level of independence. Therefore pre-processing of data is advocated with checks being made for correlation, homogeneity non-stationarity. There is no appropriate means for adjusting the fitting procedures to allow for dependent data. Instead it is suggested that a more consideration is given to the uncertainties involved. The uncertainty and bias in the quantile estimators will be higher, while classical goodness-of-fit tests will be less reliable.

Zero flows and discretization

The possibility of the occurrence of zero flows is a central factor in low flow frequency analysis. Ignoring observed zero flows may lead to a positive probabilities being assigned to negative streamflow values (such results have no physical basis), but even if they are unlikely to be observed in practice, some consideration of zero flows is required. Similarly it is important to be aware of the effect of discretization within the observed data. Several possible methods for dealing with zero flows have been outlined in this report, including treating the observed data as if it represents a censored or truncated distribution, and augmenting the distribution to cope with the discrete component at zero. There is little evidence from the low-flow literature to show which of these might, in practice, give the best results. Durrans *et al*. (1999) advocate censorship of all zero values, regardless of whether they are real or result from the effects of discretization. Conversely Stedinger *et al*. (1993) used a conditional probability adjustment procedure, where the probability of encountering a zero flow is represented by including an additional parameter in the distribution so that the non-zero data can be considered as continuously-valued (the adjustment procedure is used to adjust the results to the full sample). Again, taking a different approach Bulu (1997) outlined the possible use of the theorem of total probability to handle the occurrence of zero flows.

Outliers

A subjective approach is advocated for dealing with outliers. As there is a lower bound on low flows, outliers are not as important as in flood estimation, and therefore a rigorous or automatic treatment of outliers is not required. Outliers maybe caused by artificial influences in the catchment, such as the effects of over-abstraction, and in this case may be removed by naturalising the data, but may also occur naturally, representing extremely rare flow events. Attempts should be made to verify natural outliers, for example by comparing flow records obtained at nearby sites, from local knowledge or 'folklore' or from other types of environmental data where available.

Family of distributions

Table 1 gives details of a selection of low-flow studies, including the methods applied and the distributions tested. Despite the fairly large number of studies, there is no clear guidance in the literature as to which distributions (or families of distributions) are most favourable. The frequency estimation literature contains many examples where particular combinations of statistical and estimation method are examined, and 'good' or 'bad groupings suggested. In each case different criteria for assessing suitability of combinations have been used, ranging from convenience of formulae for parameter estimation (computation convenience) to the statistical performance of the estimators. There have been few objective tests where the same treatment has been applied to a range of distributions using the same method, papers usually focus on fitting one specific distribution, or comparing the relative performance of two methods. Also the range of different ways in which the lower bound issue and data quality have been treated also makes it difficult to make an outright comparison between results presented by different authors.

Different forms of the Weibull, EV1, Person Type-III and lognormal distributions are among those commonly referred to in the literature. Flood studies in the UK and US tend to advocated GEV or Generalised logistic distributions, whereas the Log Pearson type III is commonly used in mainland Europe. The considerations of Section 2.5.1 indicate, however, that although the GEV distribution might be useful for particular sites, its acceptance would need to be purely based on the results of tests of fit, rather than relying on extreme-value theory. In a review of low flow frequency analysis, Smakhtin (2001) concluded that a universally distribution for low flows is unlikely to exist or to ever be identified. Here it is concluded that the Weibull and Generalised Logistic distributions are probably amongst the most appropriate choices for low flow data.

Fitting techniques

The two most widely used techniques for parametric estimation are the method of maximum likelihood and the method of moments. The use of the two methods in some reported low-flow analyses is shown in Table 2.1. Likelihood techniques are well-established and theoretically sound, for example, in an evaluation of seven different methods for estimating parameters Arora & Singh (1987) found that the maximum likelihood technique produced the most efficient quantile estimates. However likelihood methods have been found to be difficult to apply in practice and because of this the method of moments, which is simplistic in comparison, is often preferred. The technique of using linear combinations of moments (L-Moments) became widely used during the 1980's and has since become a well-respected method for fitting distributions. L-moments were recently applied in the Flood Estimation Handbook (Institute of Hydrology, 1999) and have been preferred by Nordic researchers (e.g. Tallaksen *et al*. 1994). Furthermore moment-ratio diagrams are useful in assessing which types of families are appropriate and for assessing the goodness of fit of distributions. However it is important to consider the need to fit multiple durations, and to ensure these bear a sensible relationship to one another: this problem is better approached using the maximum likelihood method.

Table 2.1: Recent low-flow frequency studies

Study	Specific Topic	Families/ Methods	Tests of Fit
Caffey *et al.* (1980)	Example of variation of low-flow quantile along a river. Discussion of estimation at ungauged sites.	Graphical (non parametric) fitting procedure.	None
Deninger, *et al.* (1969)	Comparison of distribution fitting methods, focusing on the estimate of the lower bound.	Weibull-3: MOM, MOMC, MOMS, GOS.	None.
Durrans	Suggests fitting to the lowest 25% of observations.	Weibull-2 ML, (MOM, PWM)	none
Gumbel (1963)	Graphical displays based on logarithm of (flow minus bound) versus Gumbel-reduced variate that should be linear if data is from Weibull distribution.	Weibull-2: MOM Weibull-3: MOMS	Graphical assessment of fit.
Guo *et al.* (1996)	Simulation comparison of kernel estimate for fitting Weibull distributions. Kernel estimator uses an EV1 kernel.	Weibull-2:MOM, ML, PWM Weibull-3:MOM, ML, PWM	Graphical assessment of fit.
Gustard, *et al.* (1992)	Regionalisation, based on "curve types" and catchment characteristics. Different durations treated via estimates of a relationship of mean annual minimum flow of a given duration to the duration.	Non-parametric graphical approach, using Weibull scale.	No formal fitting
Gustard & Gross (1989)	Forming catchments into groups on the basis of mean and CV of annual minima	Non-parametric graphical approach, using Weibull scale.	No formal fitting.
Joseph (1970)	Moment-ratio plots to aid choice of distribution. Gamma distribution preferred in results of tests.	Gamma: ML Log-Normal: ML Sqrt-Normal: ML Normal: ML Weibull-2: ML	χ^2
Kobold *& Brilly* (1993)	Relation of low flow characteristics to catchment characteristics.	Non-parametric graphical approach, using Weibull scale.	No formal fitting.
Krokli (1989)	Derive estimated quantiles for use as dependent variable in regression on catchment characteristics.	EV1: not stated.	Mention of χ^2 tests applied in same region.
Kumar & Devi (1982)	Comment on paper by Prakash (1981). For details of BC-G-Normal see Chander et al. (1978).	SMEMAX: unspecified BC-Normal: MOM BC-G-Normal: MOM	Assessment of moments after transformation.
Lawal *et al*	Looks at relationship of estimated lower bounds to sample skewness.	Weibull-3 ML	none

Study	Specific Topic	Families/ Methods	Tests of Fit
Leppäjärvi (1989)	Tests for trend. Note: formulae given imply EV1 was used, not $EV1_{min}$.	EV1: MOM Log-Normal: MOM, ML Gamma: MOM, ML	KS
Loganathan *et al.*(1985)	Treatment of exact zeroes by conditional probability. Treatment of zeroes by adding constant to flows before analysis and subtracting it from results (avoids log of zero). Notes possibility of estimated lower bound being either negative or greater than smallest observation.	BC-Normal:ML, SMEMAX:unspecified Weibull-2: MOM Weibull-3:ML,MOM (S) Log-Pearson III:MOM, MOML Log-Boughton: GOS	KS
Loganathan *et al*. (1986)	Treatment of exact zeroes by conditional probability. Treatment of zeroes by adding constant to flows before analysis and subtracting it from results (avoids log of zero). Also estimation methods based on Partial Duration data and on a serially dependent stochastic model for recession minima.	Log-Pearson III: MOML DB P.D.F.: ML (upper bound fixed)	KS
Matalas (1963)	Theoretically-based comparison of variances of MOM and ML estimates. suggests distributions should have no more than 3 parameters. Graphical relationship of skewness and lower bound. Moment-ratio plots.	Weibull-3: MOM, ML Log-Normal-3: MOM, ML Pearson III: MOM, ML Pearson V: MOM, ML	Assessment via moment-ratio plot.
Nathan & McMahon (1990)	Treatment of minima below a cut-off value only and of exact zeroes by conditional probability. Comparison of use of hydrological and calendar years. Treatment of negative estimated quantiles by fixing estimated lower bound to zero. Failure of methods to provide acceptable estimates.	Weibull-2: ML, MOM, PWM Weibull-3: ML, MOM, PWM	None
Pearson (1995)	Regionalisation. Relation of probability of zero flow to catchment characteristics. L-moment tests for regional homogeneity. Text clearly states that EV1 and GEV are preferred to $EV1_{min}$ and Weibull-3 (GEV_{min}).	Gen-Logistic: PWM GEV: PWM Gen-Normal: PWM Pearson III: PWM Gen-Pareto: PWM Kappa: PWM EV1: PWM Weibull-3: PWM	L-moment tests of fit on regional basis.

Study	Specific Topic	Families/ Methods	Tests of Fit
Perzyna & Gottschalk (1995)	Compound distributions, with no simple final form, based on an exponential decay over a random time period from a random initial flow.	Compound From statistics of dry period lengths and fitted recession constant.	Graphical assessment
Pilon (1990)	Regionalisation based on using the median as the index of low flow. Features a test of homogeneity of a region	Weibull-2: Regional fit based on scaled order statistics	Suggests but does not use L-moment ratio diagrams.
Polarski (1989)	Joint estimation of flow frequency curves for averages of different durations.	Weibull-3: ML	Graphical assessment of fit.
Prakash (1981)	Paper advocates SMEMAX	Log-Normal:unspecified Log-Pearson III: Weibuull-3: unspecified SMEMAX: unspecified	KS
Rao (1980)	Discussion and illustration of shapes of distributions as parameters vary.	Log-Pearson III: MOM	None
Riggs (1965)	Mentions importance of including most critical periods in analysis. Adjustment of estimated return periods (for probability plots) using related longer data series. Example of correlation across years. Effects of several different sources of water contributing to low flows.	Graphical estimation	None
Tasker (1987)	Treatment of zeroes by conditional probability. Uses bootstrap procedure to compare performances of the different estimates of quantiles.	Log-Pearson III: MOM Weibull-3: mixed ML/MOMS/MOM (zeroes not excluded) BC-Normal: EOS Log-Boughton: GOS	None
Vogel & Kroll (1990)	Summary of previous studies assessing fit of distributions to real data.	Log-Normal: MOML Log-Normal-3: MOM, QLB-MOML Weibull-3: MOM, ML, MOMS Log-Pearson III: MOML	PPCC
Vogel & Kroll (1990b)	Regression of estimated parameters and quantiles on catchment characteristics.	Log-Normal: MOML	PPCC
Wang & Singh (1995)	Treatment of exact zeroes by conditional probability	Gamma: ML, MOM, PWM, EOS.	Graphical assessment of fit.

3. ANNUAL MINIMA DATA

3.1 Gauging Station Selection

Only those UK gauging stations with natural flow regimes and good hydrometric quality were selected for inclusion in the study. Firstly, the classification of UK gauging stations undertaken as part of the Low Flow Studies Report (Gustard *et al.*, 1992) was used to identify potentially suitable flow records. The Low Flow Studies classification graded stations according to hydrometric quality and the level of artificial influence. Here only those stations receiving an AA grading were considered, that is, those that met the following criteria:

i) A gauged Q95/MF ratio differing from the estimated natural Q95/MF ratio by less than 20%,
ii) A gauge sensitivity of less than 20%
iii) No obvious deterioration of the gauging station due to siltation, weed growth or vandalism.

Two further criteria were imposed:

iv) The net artificial influence at mean flow should be less than 7% of the naturalised mean flow.
v) The net artificial influence at the Q95 flow should be less than 20% of the naturalised Q95.

A total of 123 stations met the imposed criteria. Details of these are given in Appendix 2. The relevant measuring authorities were asked to comment whether, on the basis of possible artificial influences and hydrometric errors at low flows, the ‘AA’ classification was appropriate. Comments were received for 64 stations, 37 of which were deemed suitably natural for the purposes of study. From this group, 20 gauging stations of minimum record length of 30 years were finally selected for inclusion in the study. Unfortunately this set was dominated by impermeable catchments (many of which located in Northeast Scotland), and therefore five lowland catchments in southern England were also included (two with AA grading, the remaining three suggested by the Environment Agency). A complete list of the selected catchments is given in Table 3.1, a table of catchment characteristics is given in Appendix 2, whilst Figure 3.1 shows the spatial locations of the catchments.

Table 3.1: Flow records selected for use in the study

Station	River	Site	**Years**
9001	Deveron	Avochie	38
9002 †	Deveron	Muiresk	38
14001	Eden	Kemback	31
19002	Almond	Almond Weir	37
19004	North Esk	Dalmore Weir	38
20001	Tyne	East Linton	38
20003	Tyne	Splimersford	34
20005	Birns Water	Saltoun Hall	34
21006 †	Tweed	Boleside	38
21012	Teviot	Hawick	36
21013	Gala Water	Galashiels	35
21015	Leader Water	Earlston	33
21017	Ettrick Water	Brockhoperig	34
28031	Manifold	Iiam	31
34003 *	Bure	Ingworth	39
39016 †	Kennet	Theale	38
39028	Dun	Hungerford	31
43005 *	Upper Avon	Amesbury	33
43006 *	Nadder	Wilton	33
48010 *	Seaton	Trebrown Bridge	31
51001 *	Doniford Stream	Swill Bridge	32
55016	Ithon	Disserth	31
55026 †	Wye	Ddol Farm	61
60002	Cothi	Felin Mynachdy	38
72004 †	Lune	Caton	38

* *Indicates the five 'south of England' catchments*
† *Indicates the five 'example' catchments*

Five gauging stations (9002, 21006, 39016, 55026, and 72004) were selected for investigation in more detail, and are used to illustrate the general findings of the study. These were the stations with the longest records but, with the exception of 55026, all represent flow records from fairly big catchments. The five catchments represent a range of regime types, as indicated in Table 3.1.

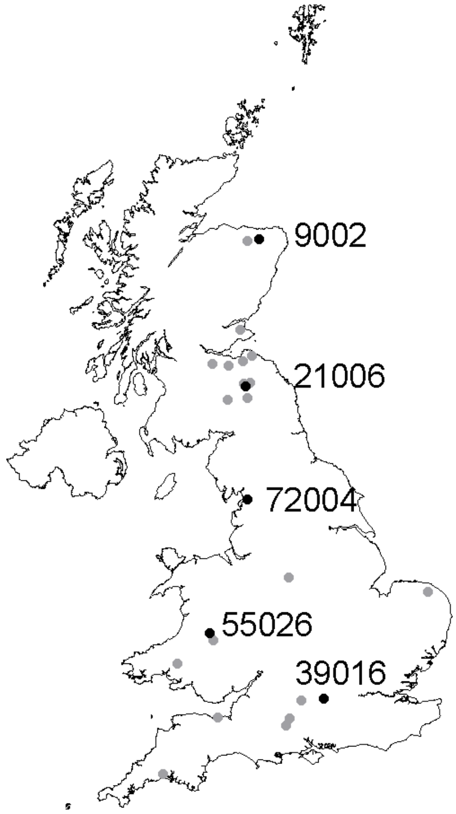


Figure 3.1: Location of selected gauging stations

The subset of five gauging stations (9002, 21006, 39016, 55026, and 72004) were also chosen to represent a range of geographical locations. Their locations in Figure 3.1 are highlighted in black.

3.2 Derivation of Annual Minima

3.2.1 General methodology

Annual minima series were derived for each flow record based on D-day running averages with the duration, D, taking values of 1, 7, 30, 60, 90, 180 and 365 days. Each annual minima was expressed as a percentage of the long-term average daily flow for the entire record, in order to allow direct comparisons between the catchments of different sizes. The annual minima series for the 25 catchments are listed in Appendix 3.

The computation of D-day running averages is based on a window of length D moved sequentially through the entire record using an increment of one-day. For each year, the lowest D-day average occurring within that year is taken as the annual minimum value. Whilst the calculation is straightforward, there are several issues to consider while dealing with data, the most important being how to deal with 'averages' that span two distinct years and the problem of missing data. These issues are discussed in the following sections.

3.2.2 Defining years

Where the duration is large assigning particular 'averages' to particular years becomes more complex, because the daily flow values that are included in each average may come from two distinct calendar years. Calendar years, rather than water years, are used to avoid the change from one year to the next occurring during a period in which the annual minima potentially might occur, and the following convention is applied.

Suppose that the duration, D, is 7 days. The average flow over this 7-day interval is indexed to the middle or central day of the interval, which in this case is the 4th day of the interval. For any interval of length, D, the middle or index day will be the m^{th} day, where *m* is defined as follows:

$$m = \begin{cases} \left(\frac{\mathrm{D}}{2}\right) + 1 & \text{, where D is even} \\ \left(\frac{\mathrm{D}-1}{2}\right) + 1 & \text{, where D is odd} \end{cases} \qquad (3.1)$$

For example, the 7-day average determined using daily flow data from 3rd September 1983 to 9th September 1983 inclusive would be indexed to the 4th day, the 6th September, 1983. As a running average method is utilised with an increment on one day, 365 (366) running average D-day flows can be determined for each calendar year, providing that there is no missing data.

3.2.3 Missing data

Missing data arises when the mean flow for a particular day is not recorded for some reason (e.g. equipment failure, the gauging structure drowned out or damaged, and loss of electronic data). Missing data often occur in blocks, and in extreme cases data for a whole year may be lost. If there are missing data within a particular year the level of uncertainty associated with its annual minima increases. The greatest uncertainty occurs if the missing data is from the time period during which the annual minimum is most likely to occur (for many stations in the UK this will be during the summer). On the other hand, data missing from periods of relatively high-flow are unlikely to have much influence on the calculated annual minima, and need not be treated as rigorously. It is also important to bear in mind that, particularly where the duration considered is large, data missing from one year may also affect the annual minima for the proceeding or following year because some running averages will incorporate daily flow values from two consecutive calendar years.

Generally, missing data is dealt with in one of two ways: missing values may be estimated by interpolation (this is viable only where a short period of data is missing) or years with missing data are excluded from the analysis. Here, to avoid filling in large gaps by interpolation whilst also avoiding rejecting years unnecessarily, an objective criterion for excluding years with too much missing data was applied to each year in turn. The criterion was applied as a 2-step procedure, the first part being to filter out data poor years, the second to ensure that years with too much missing data within the low flow period were excluded as follows:

1) 30-day Criterion

The number of missing data allowed per calendar year was constrained to 30 days, i.e. a year was rejected outright if it contained 30 or more days with missing values. This test was used to filter out years with large amounts of missing data.

2) Missing data criteria for the ‘low flow period’

A second test was applied to all years passing the ‘30-day criterion’. This procedure involved delineating a ‘low flow period’ for the year, which was calculated by ranking the daily flows within the specified year in ascending order, and taking the lowest 30% (110 days) as a subset, and looking where these occurred during the year. The ‘low flow period’ was taken as the longest continuous period covered by this subset of days (if two or more periods of equal length occurred, the one with the lowest rank sum was selected). The year was discarded if the maximum number of consecutive missing values within the ‘low flow period’ exceeded 7 days.. However if there were less than seven consecutive missing days, the year was only discarded if the aggregate number of missing data during the ‘low flow period’ was greater than 10 days, otherwise missing data were filled by interpolation.

Although strictly not missing data, particular problems occur at the beginning and end of the flow record. The first (D-m) days at the start of a record and the last (m-1) values at the end of the flow record cannot be used as index days. For instance to derive a value for 1st January 1968 requires (m-1) extra days from the previous year 1967, yet if 1st January 1968 is the first day in the record, these extra data do not exist. A similar situation arises at the end of the flow record and results in a reduced number of running-averages being computed for years at the end and beginning of the record.

To avoid having to reject these years from the analysis a number of methods for extending the flow record were considered (e.g. modelling/extrapolation). A more simplistic approach of 'borrow' data from elsewhere in the time series was also considered. The beginning of the record was supplemented by data from the last year of the record, whilst the data from the first year was used to extend the end of the record by (D-m). It was found that data for up to 75 days could be 'borrowed' for a year without altering (to within +/-1%) its annual minima estimate. Therefore a special criterion was applied to the first and last years of record, in which the year was not rejected if less than 75 days were borrowed.

3.3 Time Series Analysis

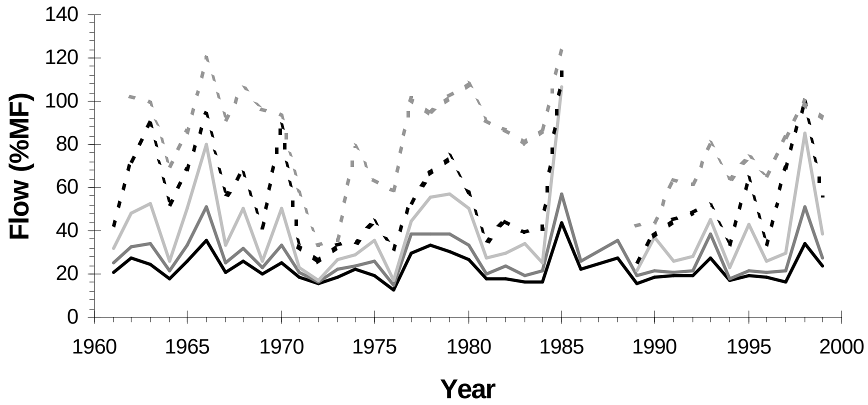
3.3.1 Temporal variation in annual minima

Figure 3.2 contrasts the time series of d-day annual minima for the two different catchment types. The series for the Deveron at Muiresk (9002) is shown in Figure 3.2a, while Figure 3.2b shows the time series for the Kennet at Theale (39016). In both cases the annual minima varies from year to year, and on visual inspection no long-term trends are apparent. The droughts of 1973, 1976, 1989/90 and 1984 (9002 only) appear as well defined troughs in the annual minima series.

Some statistics regarding the D-day annual minima series of the Deveron and Kennet are given in Tables 3.2 and 3.3 respectively. The Deveron is a responsive catchment with little contribution from base flow. When a one-day duration flow is considered the annual minima mostly range between 15 and 30% of the mean flow. For the Kennet, a lowland chalk stream with a high base flow component, the annual minima generally ranges between 30 and 60% of the mean flow at D=1.

As longer durations are considered the contrast between the two time series becomes less apparent. For example, whereas the mean annual minimum for D=1 is 22.6% and 42% of the mean flow for 9002 and 39016 respectively, this changes to 80.9% and 84.2% in the case where D=365 (Tables 3.2 and 3.3). These observations agree with previous work (Institute of Hydrology, 1980) which illustrated a relationship between annual minima and catchment characteristics. The statistical properties of the annual minima series for all 25 stations are summarised in Appendix 4.1.

a)



b)

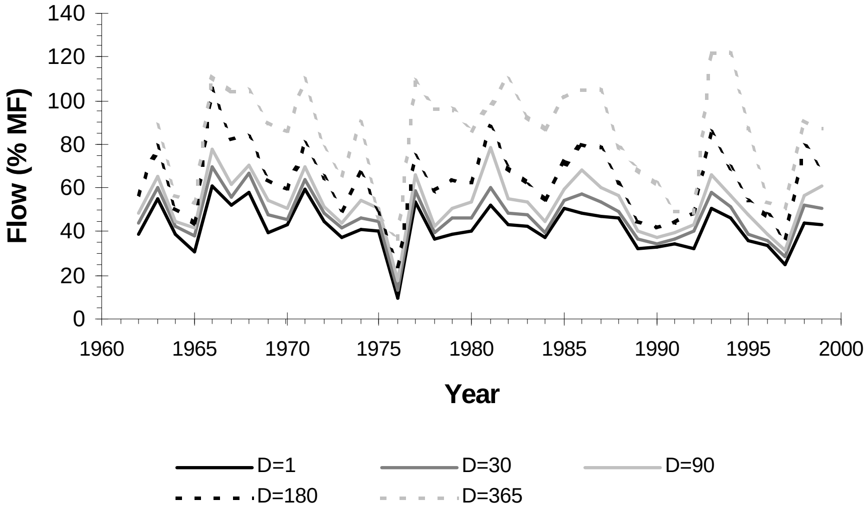


Figure 3.2: Time series of D-day annual minima for a) the Deveron at Muiresk (9002) and b) the Kennet at Theale (39016)

Table 3.2: Properties of annual minima series for the Deveron at Muiresk (9002)

	D=1	D=7	D=30	D=60	D=90	D=180	D=365
N of cases	38	38	38	38	36	36	35
Minimum	12.37	12.80	14.67	15.55	16.92	25.72	33.02
Maximum	43.71	44.82	57.02	92.32	106.90	113.20	122.70
Median	20.20	21.16	24.50	27.53	33.55	49.82	84.85
Mean	22.61	23.77	28.11	33.45	39.67	54.56	80.95
Standard Deviation	6.69	7.19	10.05	14.96	19.54	22.12	22.76
Variance	44.80	51.76	101.02	223.69	381.94	489.35	517.82
C.V.	0.29	0.30	0.36	0.45	0.49	0.40	0.28
Skewness	1.15	1.01	1.23	1.93	1.69	0.93	-0.38
Kurtosis	1.37	0.72	1.16	5.36	3.40	0.24	-0.40

(All values are expressed as a percentage of the mean flow)

Table 3.3: Properties of annual minima series for the Kennet at Theale (39016)

	D=1	D=7	D=30	D=60	D=90	D=180	D=365
N of cases	38	38	38	38	38	38	37
Minimum	9.82	10.84	13.35	15.41	17.43	24.07	36.31
Maximum	61.02	63.11	69.41	73.57	78.10	105.40	122.00
Median	41.80	43.23	46.87	49.85	53.57	63.05	88.94
Mean	42.04	43.73	47.14	50.39	52.79	63.16	84.29
Standard Deviation	9.99	10.21	10.98	11.84	12.66	16.59	23.24
Variance	99.95	104.19	120.62	140.25	160.23	275.51	539.91
C.V.	0.24	0.23	0.23	0.23	0.24	0.26	0.28
Skewness(G1)	-0.59	-0.56	-0.44	-0.36	-0.23	0.08	-0.45
Kurtosis(G2)	1.80	1.73	1.32	0.89	0.54	0.11	-0.84

(All values are expressed as a percentage of the mean flow)

3.3.2 Statistical requirements

For the results of a frequency analysis to be theoretically valid, the data used must meet certain statistical criteria. In particular the data must be stationary, homogeneous and independent. To be stationary the characteristics of the flow series must not change over time (excluding random fluctuations). Non-stationarity may manifest as step changes, which are likely to be associated with anthropogenic or data collection factors, or monotone (gradual) change. Long term cyclic variation is typically related to climatically driven changes. A data series can be thought of as independent if no observation in the data series has any influence on any observations following it, whilst homogeneity implies that all the elements of the data series originate from a single population. For example an annual minimum series where low flows occurred due to over abstraction as well as due to low rainfall, would not be a homogeneous series.

3.3.3 Tests for stationarity

Non-stationarity can be difficult to detect especially where the trend is weak. However, where the trend is strong, statistical tests may be used to identify non-stationary time series. In addition to visual examination of the data, two statistical methods were used to look at monotone or linear trend in the data: linear regression and Spearman's Rank Test.

1) Linear Regression

Linear regression is a commonly used statistical method for describing relationships between two or more variables (in this case 'year' and 'annual minimum flow'). The strength of the regression is given by the coefficient of determination, R^2, which tends to the value of 1 if there is linear relation between the two variables.

Linear regression showed that there were no relationships with annual minimum flow and chronological year, for any of the 25 stations considered, regardless of the duration considered In most cases the R^2 value was, in fact, less than 0.1. The results of the regression tests for the sample set are documented in Appendix 4.2.

2) Spearman's Rank Correlation Test.

Spearman's Rank Correlation Test is a standard non-parametric (distribution-free) test that determines whether the correlation between two variables is significant. The null hypothesis is that there is no association between the rank pairs. Spearman's test is appropriate in this case because it is robust to the effect of extreme values (i.e. to highly skewed hydrological data) and to deviation from a linear relationship.

The test statistic is the correlation coefficient, which is obtained in the same way as the usual sample correlation coefficient but using ranks:

$$\rho_s = \frac{S_{qy}}{\sqrt{S_q S_y}} \qquad (3.2)$$

where

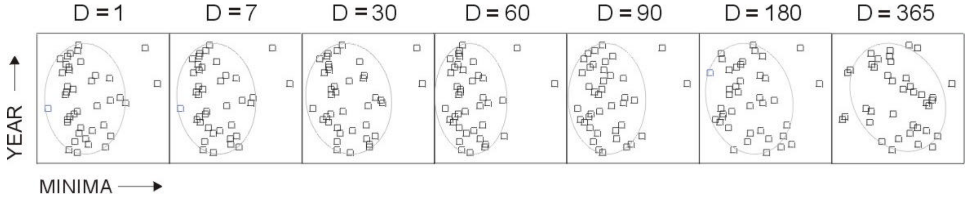
ρ_s is the Spearman Correlation Coefficient, q represents the annual minimum flow, y represents the year variable, S_q is the standard deviation of the sample of annual minima [$= \sum (q_i - q_{mean})^2$], S_y is the sample standard deviation of the years [$= \sum (y_i - y_{mean})^2$], S_{qy} is the sample covariance [$= \sum (q_i - q_{mean})(q_i - q_{mean})$], and q_i, q_{mean}, y_i, y_{mean} refer to the ranks of the i^{th} q value, the mean q, the i^{th} y, the mean y respectively.

When ρ_s takes the value of 1, or –1, the correlation is perfect and a linear relation between the two variables exists, but if $\rho_s = 0$, there is no correlation between the two variables. Significance levels for the test are looked up from tables, such as those given by Neave (1988). For a data set with 38 values the critical value for the test, at 95% confidence level is 0.3209. Table 3.4 gives details of the variation in the Spearman Correlation Coefficient between d-day annual minima and chronological year when different durations are considered, illustrated by the results for stations 9002 and 39016. The coefficients for all 25 stations are given in Appendix 4.3. As shown by Table 3.4 and Figures 3.3a and b, the Spearman's Rank test indicated little correlation with annual minima and time.

Table 3.4: Spearman's Rank Test - correlation coefficients for 9002 and 39016

Station	D=1	D=7	D=30	D=60	D=90	D=180	D=365
9002	-0.185	-0.182	-0.199	-0.164	-0.120	-0.138	-0.217
39016	-0.294	-0.299	-0.261	-0.223	-0.210	-0.200	-0.073

a)



b)

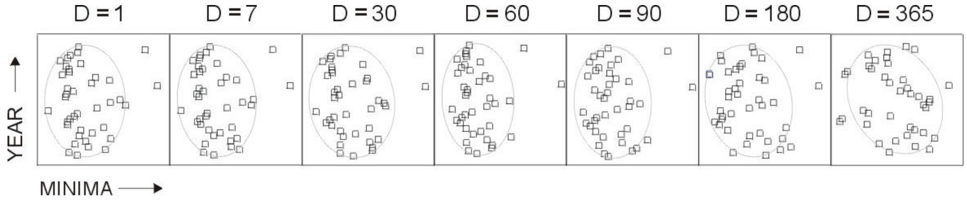


Figure 3.3: Relationship between rank ordered year and minima for a) 9002 and b) 39016.

3.3.4 Tests for homogeneity

Ensuring homogeneity means that the distribution function will be able to describe all future values of flow that might occur. Low flow data might not be homogeneous if the flow regime changes over time (i.e. the time series is non-stationary) or if low-flow events have different origins. An example of the latter case occurs in snow and ice regions, where rainfall being stored as snow in the winter period results in low-flow events unrelated to those that occur in summer as a result of dry weather.

The Mann-Whitney test (Mann-Whitney, 1947) was applied to test the data for homogeneity. In the Mann-Whitney test the sample, N, is arbitrarily split into two sets, p and q (where $p < q$). Here the two sets were defined on the basis of the time of year of occurrence of the minimum i.e. whether the low flow occurred in summer or in winter (for this reason the test was applied only to durations of 90 days and less). The null hypothesis is that the two samples come from the same population. As a trial the Mann-Whitney test was applied to the annual minima series from stations 9002 and 39016. However, the results of these tests were inconclusive, and therefore the test was not applied further.

In the UK short duration annual minima are unlikely to occur outside the summer period, although minima for permeable catchments usually occur later in the season than those for impermeable catchments. This variation is illustrated in Bullock *et al*. (1994). As a result, inhomogeneity can be considered of minor importance where LFFA is to be applied to UK flow data.

3.3.5 Tests for independence

Provided that the data is stationary, the degree of independence can be characterised by determining the autocorrelation function of the observed data, or by constructing an autocorrelogram from the data. The results of the Spearman's Rank Test can also be interpreted in terms of the independence of the data (in which case the results suggest that the data are mainly independent).

The autocorrelation function of each time series was calculated in order to quantify the extent of the correlation (dependence). Autocorrelation measures of the correlation of the series, with itself shifted by a time lag. Autocorrelation can be calculated for a lag of any length, and if autocorrelation is present at one or more lags then the data is not independent. Partial autocorrelation tests were also conducted. Partial autocorrelation plots show the relationship of points in a series to preceding points after 'partialing' out the influence of intervening points, and thus give a more conservative / better perspective of autocorrelation. Partial autocorrelation plots for durations of 1, 7, 30, 60, 90, 180 and 365 days for each of the stations are documented in Appendix 4.4.

Figures 3.4 and 3.5 show the partial autocorrelation plots for stations 9002 and 39016 respectively based on durations of D=1 and D=90, whilst Figures 3.6a and 3.6b show the autocorrelation in the data for stations 9002 and 39016 respectively, in the case where D=365. Values that lie above the dotted (red) line indicate that there is a significant correlation in the data at the lag (in years) indicated. Therefore if most correlation values lie between the dotted lines the data values are to be independent.

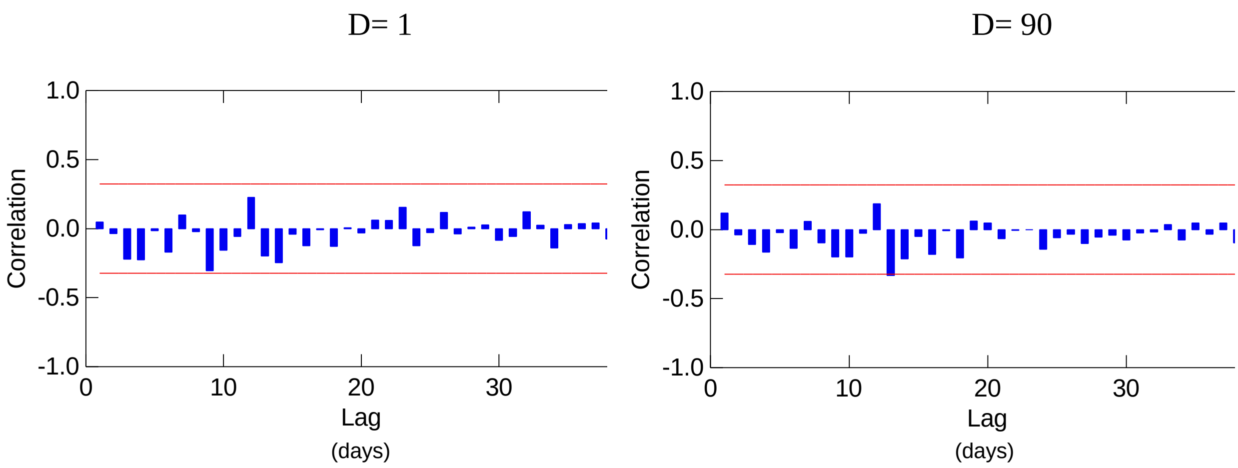


Figure 3.4: Partial autocorrelation functions for durations of 1-day and 90-days for 9002

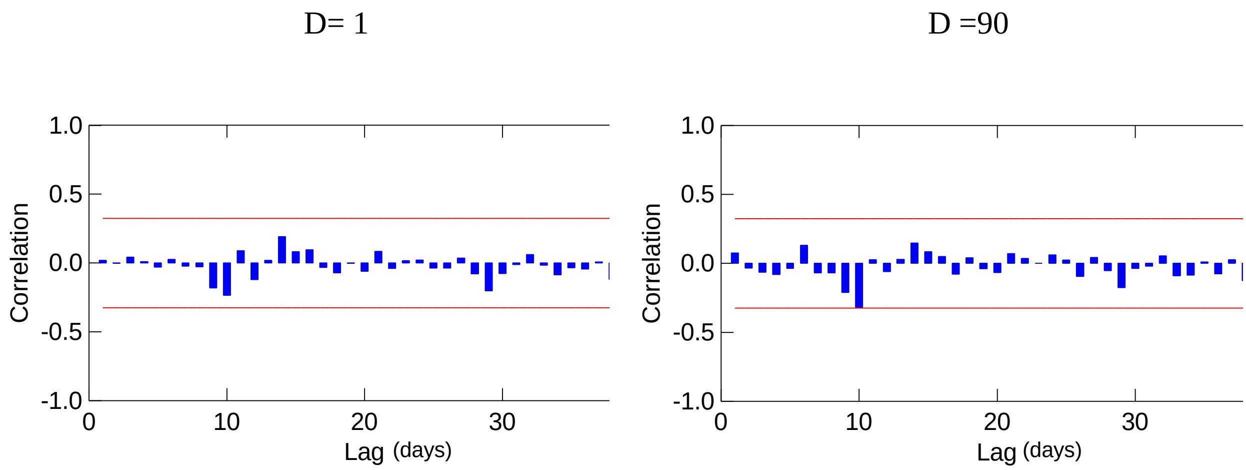


Figure 3.5: Partial autocorrelation functions for durations of 1-day and 90-days for 39016

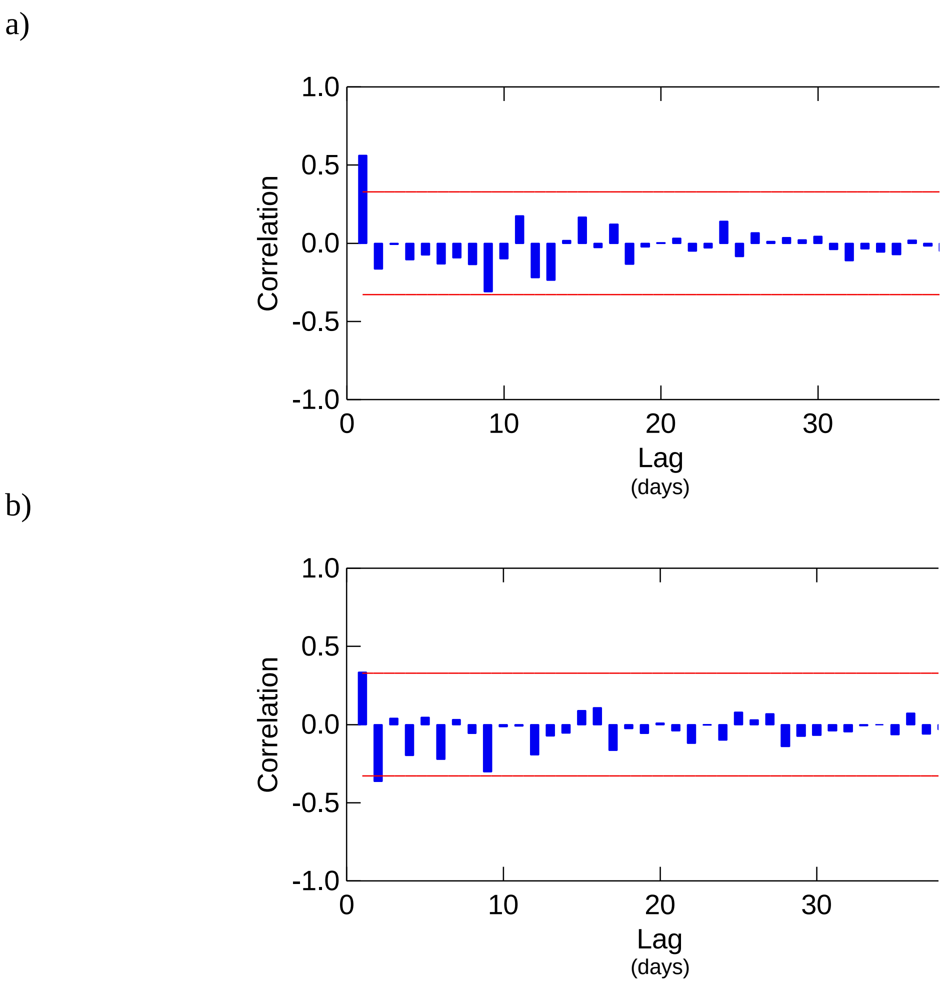


Figure 3.6: Partial autocorrelation for a) 9002 and b) 39016, where D = 365.

In the examples shown in Figure 3.4 and Figure 3.5 there is little or no dependence with the data series. This was typical of most of the time series tested. However Figures 3.6a and 3.6b indicate that where a duration of 365 days is considered the level of autocorrelation and partial auto-correlation is much higher, particularly for lag times of one to three years. This is because consecutive annual minima are each based on average flow over a D-day period, of which ‘x’ days are common to both periods. The number of common days increases as the duration increases, so that dependence is more likely where higher durations are considered.

Although the annual minima for D=365 were found to be dependent no further action was taken. This is because it is beyond the scope of this project to incorporate the problem of dependence into the formulation of the distribution function model to describe each data set. Sen (1980) and Chung & Salas (2000) report it can be more useful simply to be aware of the effects of dependence on the usual procedures that happen to have been derived on the assumption of statistical independence between sample values. While dependence will not influence the process of parameter estimation it will influence the assessment of the uncertainty associated with the resulting quantile (i.e. the quantile estimates will become more biased and there will be larger standard error in the quantile estimates).

4. ENUMERATING DISTRIBUTIONS

4.1 Introduction

This chapter describes the technical details of the methodology employed. Firstly several candidate distribution families are described, and the relationships between F(x), x(F), *p*, x_T and T are given in each case (section 4.2). The use of plotting positions to estimate *p* is discussed in sections 4.3 and different plotting position formulae are evaluated. The two methods of parameter estimation used in the study, Maximum Likelihood Estimation and L-Moments are described and assessed in Section 4.4. The effects caused by hydrometric errors are considered in Section 4.5, whilst in Section 4.6, a method for dealing with zero flows and discretization, that of treating the data as a censored sample, is evaluated.

4.2 Candidate Distributions

4.2.1 Choice of candidate distributions

Due to the increased flexibility of 3-parameter distributions it was decided to concentrate on three parameter distributions including Generalised Extreme Value (GEV) of which the Weibull and Gumbel distributions are special cases, Generalised Logistic (GL), Pearson Type III (PE3) and Generalised Pareto (GPA).

As it has been a fairly popular choice in low flow analysis (e.g. Vogel & Kroll, 1990) the log-normal distribution was also considered, particularly as it has no upper bound and a lower bound at zero. However problems with coding prevented this from being implemented. In addition, a zero in a set of data that is being logarithmically transformed requires special handling. However this could have been avoided by adding an incremental amount (e.g. 1) to all the data.

The following sections give the forms of the distributions for the general case. The equations F(x), f(x) and x(F) are those given by Hosking and Wallis (1997) whilst the equations describing the x_T-T relationships were derived by substituting for F(x) in the x(F) derivation. It should also be noted here that F(x) is the non exceedance distribution and that the exceedance distribution is given by 1-F(x), denoted by F'(x).

4.2.2 Generalised Extreme Value (GEV) distribution

The GEV distribution is the general three-parameter case of the range of extreme value distributions (including EV1, EVII and EVIII distributions). The three parameters are location, ξ, scale α and shape, k.

The distribution is bounded above where $k > 0$ and is bounded below where $k < 0$, i.e. the range is $-\infty < x \le x + \frac{a}{k}$ if $k > 0$, $x + \frac{a}{k} \le x < \infty$ if $k < 0$, and $-\infty < x < \infty$ where $k = 0$.

The distribution and density functions are given respectively by:

$$F(x) = \exp(-e^{-y}) \tag{4.1}$$

$$f(x) = a^{-1}\exp(-(1-k)y - e^{-y}) \tag{4.2}$$

where the quantity y is determined as follows:

$$y = \begin{cases} -\mathrm{k}^{-1}\ln\left(1-\mathrm{k}\dfrac{\mathrm{x}-x}{a}\right) & where\ \ \mathrm{k} \ne 0 \quad (4.3) \\ \dfrac{\mathrm{x}-x}{a} & where\ \ \mathrm{k} = 0 \quad (4.4) \end{cases}$$

Similarly the quantile function is given by

$$x(F) = \begin{cases} x + a\,\dfrac{\left(1-(-\ln F)^{k}\right)}{k} & \text{where } k \ne 0 \quad (4.5) \\ x - a\,\ln(-\ln F) & \text{where } k = 0 \quad (4.6) \end{cases}$$

The x_T-T relationship is given by

$$\mathrm{x_T} = x + a\,\frac{\left(1-(-\ln(1-\frac{1}{\mathrm{T}}))^{k}\right)}{k} \tag{4.7}$$

The GEV is often classified into three types according to the value of *k* as follows:

EVI	$F(x) = e^{-x} \quad -\infty < x < \infty$	where $k = 0$	(4.8)
EVII	$F(x) = \exp(-x^{-1/k}) \quad 0 \le x < \infty$	where $k < 0$	(4.9)
EVIII	$F(x) = \exp(-\lvert x\rvert^{-1/k}) \quad \infty < x \le 0$	where $k > 0$	(4.10)

The special case that *k* = 1 is equivalent to a reverse exponential distribution, whilst if *k* = 0 the GEV distribution reduces to the Gumbel or EVI. The Gumbel, is a two-parameter distribution with scale parameter α and location parameter, ξ, and is unbounded at both ends (i.e. $-\infty \le x \le \infty$). The distribution, density and quantile functions are given respectively by:

$$F(x) = \exp\left(-\exp\left\{-\frac{x-x}{a}\right\}\right) \qquad (4.11)$$

$$f(x) = a^{-1}\exp\left(-\frac{x-x}{a}\right)\exp\left(-\exp\left\{-\frac{x-x}{a}\right\}\right) \qquad (4.12)$$

$$x(F) = x - a\ln(-\ln(F)) \qquad (4.13)$$

The Weibull distribution is a reverse generalised extreme value distribution, and so is equivalent to the EVIII for minimum values. The Weibull distribution has parameters δ, β and ζ where

$$k = \frac{1}{d}, \quad a = \frac{b}{d}, \quad x = z - b$$

and is bounded on the left hand tail, i.e. $z \le x < \infty$).

The distribution function is given by

$$F(x) = 1 - \exp\left(-\left\{\frac{x-z}{b}\right\}^{d}\right) \qquad (4.14)$$

4.2.3 Generalised Logistic

The Generalised Logistic (GL) is 3-parameter Logistic Distribution having location, ξ, scale α and shape, *k* parameters. In the special case where *k*=0 the GL distribution reduces to the 2-parameter Logistic distribution, which is unbound for all values of x.

It is bounded above where $k > 0$ (i.e. the range of x is given by $-\infty < x \leq x + \frac{a}{k}$ if $k > 0$) and below where $k < 0$ (i.e. ξ + α / k ≤ x < ∞ if $k < 0$).

The cumulative frequency and density functions of the Generalised Logistic Distribution

$$F(x) = \frac{1}{1+e^{-y}} \tag{4.15}$$

$$f(x) = a^{-1} \frac{e^{-(1-k)y}}{\left(1-e^{-y}\right)^2} \tag{4.16}$$

$$y = \begin{cases} -\mathrm{k}^{-1} \ln\left(1-\mathrm{k}\frac{\mathrm{x}\ -}{}\right) & \textit{where}\ \mathrm{k} \neq 0 \\ \frac{\mathrm{x}\ -}{} & \textit{where}\ \mathrm{k} = 0 \end{cases} \tag{4.17}$$

$$x(F) = \begin{cases} x + \frac{a}{k}\left(1-\left\{\frac{1-F}{F}\right\}^k\right) & \text{where } k \neq 0 \\ x + a \ln\left(\frac{1-F}{F}\right) & \text{where } k = 0 \end{cases} \tag{4.18}$$

Finally the flow-return period relationship is written as follows.

$$\mathrm{x_T} = x + a \frac{\left(1-(\mathrm{T}-1)^k\right)}{k} \tag{4.19}$$

4.2.4 Generalised Pareto distribution

The Generalised Pareto (GP) distribution is another three-parameter distribution, having location, ξ, scale α and shape, k. It is bounded at both tails where $k > 0$ (i.e. the range of x is given by $x < x \leq x + \frac{a}{k}$, if $k > 0$) and below where k is less than or equal to 0 (i.e. $x \leq x < \infty$, if $k \leq 0$).

$$F(x) = 1 - e^{-y} \tag{4.20}$$

$$f(x) = a^{-1} e^{-(1-k)y} \tag{4.21}$$

$$y = \begin{cases} -k^{-1} \log \left(1 - k \dfrac{x - ?}{a}\right) & \text{where } k \neq 0 \\ \dfrac{x - ?}{a} & \text{where } k = 0 \end{cases} \tag{4.22}$$

$$x(F) = \begin{cases} x + \dfrac{a}{k}\left(1 - \{1 - F\}^{k}\right) & \text{where } k \neq 0 \\ x + a \ln(1 - F) & \text{where } k = 0 \end{cases} \tag{4.23}$$

The x_T-T relationship is given by

$$x_T = x + a \frac{\left(1 - (1 - \frac{1}{T})^{k}\right)}{k} \tag{4.24}$$

In the special case where k = 1 it reduces to the uniform distribution in the interval, ? = x = ? + a and where k = 0 the Generalised Pareto distribution reduces to the exponential distribution, a two parameter distribution having scale, α and location (lower endpoint of the distribution) parameters.

The exponential distribution is bounded below with x having the range, $x \leq x < \infty$

$$F(x) = 1 - \exp\left(1 - \frac{x - x}{a}\right) \tag{4.25}$$

$$f(x) = a^{-1} \exp\left(-\frac{x - x}{a} \right) \tag{4.26}$$

$$x(F) = x - a \ln(1 - F) \tag{4.27}$$

4.2.5 Pearson Type III

The Pearson Type III distribution is a popular distribution for fitting hydrological data. If its three parameters are a (scale), ? (location) and *k* (shape). Where *k*>0, It can be described by the following density and cumulative functions

$$\text{f}(x) = \frac{\left(x - g \right)^{m-1} e^{-(x-g)/b}}{b^{m}\,\Gamma(m)} \tag{4.28}$$

$$\text{F}(x) = \text{G}\left(m, \frac{x-g}{b} \right) / \Gamma(m) \tag{4.29}$$

where $\mu = \frac{4}{k^2}$, $ß = \frac{1}{2} a \left| k \right|$, $? = x - \frac{2a}{k}$ and ? is the Gamma Function as defined by Bobee & Ashkar (1991). The range of *x* is ? = *x* = 8 .

Similarly where *k*<0, the range of *x* is 8 = *x* = ? and the the distribution the density and cumulative functions are as follows:

$$\text{f}(x) = \frac{\left(g - x \right)^{m-1} e^{-(g-x)/b}}{b^{m}\,\Gamma(m)} \tag{4.30}$$

$$\text{F}(x) = \text{G}\left(m, \frac{g-x}{b} \right) / \Gamma(m) \tag{4.31}$$

The quantile function, X(F), has no explicit form for the Pearson Type 3 distribution, but can be approximated by use of the Gamma Function. The X_T-T relationship also, therefore, has no explicit form: Hosking and Wallis (1997) give further details regarding its derivation. The Gamma distribution is a special case Pearson Type 3. The Log Pearson distribution is another variant, but generally is only appropriate for hydrological analyses when its parameters fall within a small range of values.

4.3 Selection of Plotting Position Formula

Plotting positions are empirical formulas for estimating the probabilities (F(q)) associated with the sampled data (as discussed in Section 2.5). Ideally, a plotting position should provide a distribution-free means of estimating probability (i.e. be suitable for use with data sets representing different distribution types). However certain formulae are known to produce biased quantile estimates when used with particular distributions. To ensure that the method applied was as consistent as possible, a single plotting position formula, the Gringorten (Eqn. 2.13), was selected for use in the study. Several authors (e.g. Stedinger, 1992) have advocated the Gringorten formula as it is thought to provide robust results for many different distributions and performs particularly well with the family of extreme value distributions.

To evaluate the uncertainties that might arise from using different plotting position formulae, the results obtained using the Gringorten formula were compared with those obtained using two other popular plotting positions, the Hazen (Eqn. 2.14) and the Weibull (Eqn 2.15). For this test annual minima for durations of 1 day, 30 days and 180 days for gauging stations 9002 and 39016 were used (stations 9002 and 39016 have contrasting flow regimes, the former being flashy, impermeable system, whilst the latter is a permeable catchment). Using the GEV and GL distributions, the flow - return period (Q_T-T) relationship was derived for each time series using the L-moment method to predict the parameters of the distribution.

The results for the GEV distribution are shown in Figures 4.1 and 4.2 for the two stations respectively. In both cases the Hazen and Gringorten formulas produced very similar flow - return period relationships, whilst the Weibull formula predicted lower flows for higher return periods, especially when longer durations were considered. This effect was most pronounced for station 39016. Similar results were observed when using the GL distribution as the parent family.

The results suggest that there is little sensitivity to choice of plotting position for return periods less than or equal to the record length. As longer return periods and/or more permeable systems are considered, the variability between the estimates increases. The greatest difference is between the estimates based on the Weibull formula with those based on the Gringorten and Hazen plotting positions. However as there is no benchmark to which the results may be compared it is not possible to identify which plotting position is providing the correct results. As a result of its reported superiority, the Gringorten plotting position was therefore used as the default for further analyses.

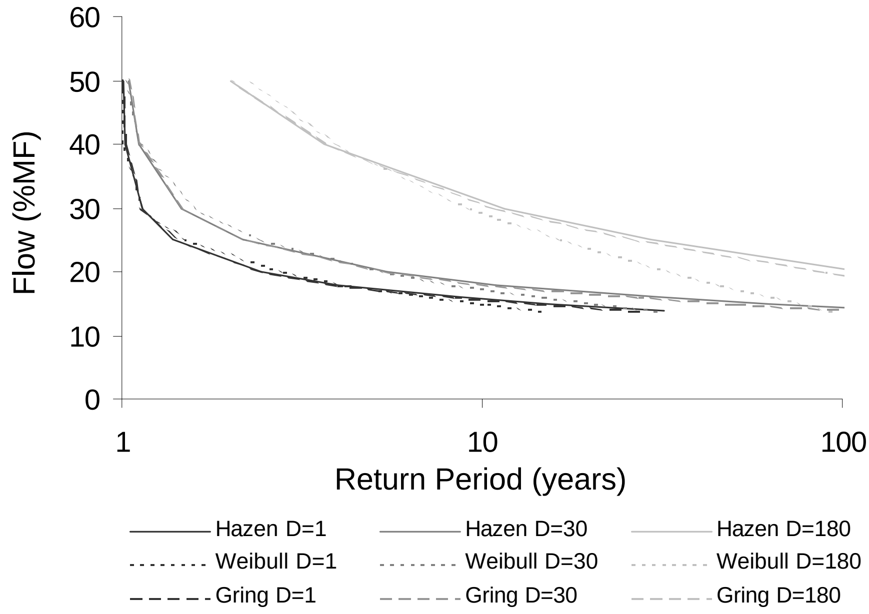


Figure 4.1: Flow-return period relationships derived using different plotting positions to estimate the probabilities of the sample data points (Station 9002)

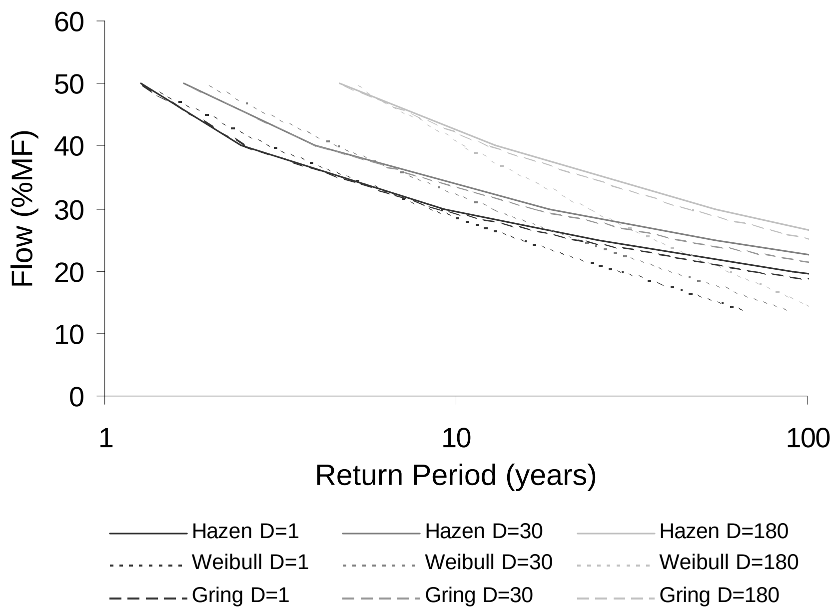


Figure 4.2: Flow-return period relationships derived using different plotting positions to estimate the probabilities of the sample data points (Station 39016).

4.4 Evaluation of Fitting Techniques

4.4.1 Introduction

Two methods of parameter estimation were considered for the study; L-Moments and Maximum Likelihood. As the equations for deriving L-moments for several distribution families are well known, it was thought that employing this estimation technique would allow for greater versatility the analysis (e.g. no restrictions on the distributions considered). In contrast, the equations for Maximum Likelihood estimation are only readily available for the GEV distribution (they could be developed for other distribution, but this would be very time consuming). However, it is important to consider the need to fit different durations simultaneously, making sure that all the solutions are within a particular tolerance of one another. In this case the term 'tolerance' can mean simply that all durations should be fitted with the same distribution, or that, in addition, the parameter values must all fall within a particular range: this problem is probably better approached using the Maximum Likelihood method. After some consideration, the L-Moments method was chosen as the main method for fitting distributions as it can be used with several different distributions. However for the GEV distribution L-Moments and Maximum Likelihood were both used for parameter estimation, to allow a comparison of the two methods and evaluate which gave the best results.

4.4.2 Technical specification of the Maximum Likelihood Method

The Maximum Likelihood Method requires the resolution of complex equations or the use of numerical optimisation schemes. The principle of this method is to choose values of the distribution parameters that maximise the chance of an observation falling within a small range around the data value. The following methodology is taken from Hosking (1985) based on Prescot and Walden (1983).

If a distribution $y = f(x, a, b, \ldots)$ with parameters $a, b, \ldots$ is considered, then $f(x, a, b, \ldots)\, dx$ is the chance of an observation falling in the range dx. In a sample of n data, the chance that there will be n_1 of them in the range dx_1, n_2 in the range dx_2, and so on is:

$$\frac{n!}{\prod_p (n_p!)} \prod_p \left(f(x_p, a, b, \ldots) dx_p \right)^{n_p} \tag{4.32}$$

Thus, to maximise this chance, the value of the parameters must be chosen in order to maximise:

$$\prod_p \left(f(x_p, a, b, \mathrm{K})dx_p \right)^{n_p} \tag{4.33}$$

or:

$$\sum_p n_p \ln(f(x_p, a, b, \mathrm{K})dx_p) \tag{4.34}$$

which is called the log likelihood function

Considering a sample data $x = (x_1 \mathrm{K}\ x_p)^t$ and a parameter vector $q = (a\ b\ \mathrm{K})^T$ the log likelihood function is then:

$$L(x,q) = \sum_p \ln(f(x_p, q)) \tag{4.35}$$

For a generalised extreme value distribution, with the data $x = (x_1 \mathrm{K}\ x_p)^t$ and the parameter vector $q = (x\ \ a\ \ k)^T$, the log-likelihood function would then be:

$$L(x,q) = -p\ln(a) - (1-k)\sum_{i=1}^{p} y_i - \sum_{i=1}^{p} e^{-y_i} \tag{4.36}$$

The maximum-likelihood parameter estimates will then be calculated via the likelihood equations that are defined by:

$$\frac{\partial L}{\partial q} = 0 \tag{4.37}$$

These likelihood equations have to be solved iteratively. The method used was the Newton-Raphson method, which solves the likelihood equations by the following iteration:

$$q_{j+1} = q_j + d\,q$$

$$d\,q = H^{-1}u \tag{4.38}$$

$$u = \left.\frac{\partial L}{\partial u}\right|_{q=q_j},\ H = \left.\frac{-\partial^2 L}{\partial q\ \partial q^T}\right|_{q=q_j}$$

The derivatives *u* and *H* (Hessian matrix) were calculated analytically using the formulae given by Prescott and Walden (1983). The fact that this is an iterative method implies that a good initial parameter estimates greatly increases the speed of the algorithm. The choice

$$
\begin{cases}
\mathrm{k} = 0 \\
a = \mathrm{s}\dfrac{\sqrt{6}}{p} \\
x = \bar{\mathrm{x}} - g\, a
\end{cases}
\tag{4.39}
$$

where $\bar{x}$ and s are the sample and standard deviation and $g = 0.57721566$ is Euler's constant, usually gives rapid convergence; these values are moment estimators of ? and a for the generalised extreme-value distribution with $k = 0$, which corresponds to a Gumbel distribution.

4.4.3 Technical specification for L-Moments

General case

The L-Moments method is a variant of the method of moments, which involves fitting a distribution so that the distribution mean and variance, and so on, match the sample mean and variance, and so on. The parameters are estimated by equating the L-Moments of the distribution to those of the sample. The following description is based on that of Hosking and Wallis (1997).

For a distribution that has a quantile function of x(u), the L-Moments, $?_1$, $?_2$, $?_3$ and $?_4$ are given by:

$$
\begin{cases}
l_1 = a_0 = b_0 \\
l_2 = a_0 - 2a_1 = 2b_1 - b_0 \\
l_3 = a_0 - 6a_1 + 6a_2 = 6b_2 - 6b_1 + b_0 \\
l_4 = a_0 - 12a_1 + 30a_2 - 20a_3 = 20b_3 - 30b_2 + 12b_1 - b_0
\end{cases}
\tag{4.40}
$$

where
$$
\begin{cases}
a_r = \int_0^1 x(u)(1-u)^r\, du \\
b_r = \int_0^1 x(u)\, u^r du
\end{cases}
\tag{4.41}
$$

and in general

$$l_{r+1} = (-1)^r \sum_{k=0}^{r} \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!} a_k = \sum_{k=0}^{r} \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!} b_k \qquad (4.42)$$

As it is more convenient to deal with dimensionless variables the L-Moments ratios have been used:

$$t_r = \frac{l_r}{l_2} \qquad \text{r} = 3,\ 4,\ \text{K} \qquad (4.43)$$

Expression for the Generalised Logistic Distribution

Once a distribution has been chosen it is then possible to express L-Moments and L-Moments ratios in terms of the distribution parameters. As an example the expression for a Generalised Logistic Distribution is presented in this section. The expressions, for the GEV, Generalised Pareto and Pearson Type-III distributions are given in Appendix 5.

The four L-Moments (?$_1$: ?$_2$, t_3, t_4) are given by

$$\begin{cases} l_1 = x + a\left(\dfrac{1}{k} - \dfrac{p}{\sin(kp)}\right) \\ l_2 = \dfrac{a\,kp}{\sin(kp)} \\ t_3 = -k \\ t_4 = \dfrac{1+5k^2}{6} \end{cases} \qquad (4.44)$$

Accordingly the distribution parameters, a, ? and *k* are given by

$$k = -t_3 \qquad (4.45)$$

$$a = \frac{l_2 \sin(k\,p)}{k\,p} \qquad (4.46)$$

$$x = l_1 - a\left(\frac{1}{k} - \frac{p}{\sin(kp)}\right) \qquad (4.47)$$

Once the L-Moments have been defined for each distribution, they may be calculated from the sample data and used to estimate parameter values. Consider a sample of size n arranged in ascending order $x_{1:n} \le x_{2:n} \le \ldots \le x_{n:n}$. An unbiased of the probability weighted moment b_r is then given by:

$$b_r = n^{-1}\binom{n-1}{r}^{-1}\sum_{j=r+1}^{n}\binom{j-1}{r}x_{j:n} \qquad (4.48)$$

which may also be written as:

$$b_r = n^{-1}\sum_{j=r+1}^{n}\frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)}x_{j:n} = n^{-1}\sum_{j=r+1}^{n}\frac{(j-1)!(n-r-1)!}{(j-r-1)!(n-1)!}x_{j:n} \qquad (4.49)$$

The sample L-moments l_r, unbiased estimator of l_r, are then equal to:

$$l_{r+1} = \sum_{k=0}^{r} p^*_{r,k} b_k \qquad (4.50)$$

where $\quad p^*_{r,k}\sum_{k=0}^{r}\frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$

Analogously, the sample L-moment ratios (which are the natural estimators of t_r) are given by:

$$t_r = l_r / l_2 \qquad (4.51)$$

4.4.4 Comparison of maximum likelihood and L-moments

The performance of the maximum likelihood and L-moments were compared using the annual minima time series derived for stations 9002 and 39016 for durations of 1day, 30 days and 180 days. Starting with a GEV parent family, parameters were estimated using both the maximum likelihood and L-moment methods, and subsequently used to derive Q-T relationships for each combination. The estimated parameters and associated estimates of goodness of fit are given in Table 4.1 (a more detailed discussion of goodness of fit measures is given later in Chapter 5).

Table 4.1: Comparison of parameter estimation methods

	Maximum Likelihood Test			L-Moments Test		
	D=1	D=30	D= 180	D=1	D=30	D= 180
9002						
a	4.634	6.486	14.880	4.794	6.642	16.640
?	19.430	23.260	43.270	19.530	23.380	44.350
k	-0.101	-0.156	-0.172	-0.063	-0.122	-0.036
$?^2$	1.089	4.225	8.326	1.688	6.795	10.270
RMSE	0.807	1.997	4.498	0.822	1.949	3.057
39016						
a	5.265	8.998	15.52	5.03	7.82	16.70
?	23.66	29.02	50.89	23.50	28.50	51.50
k	-0.08794	0.04015	-0.0142	-0.14	-0.07	0.06
$?^2$	1.926	5.535	3.049	1.74	3.84	3.13
RMSE	1.82	2.192	2.163	1.12	1.96	1.96

Table 4.1 shows that there are some differences between the two sets of estimates. Both methods produce acceptable fits to the data; the L-moments method seems to produce the best goodness of fit estimates for 39016, whilst the maximum likelihood estimates are 'better' in the case of 9002. One important result is that, for station 39016, the Maximum Likelihood parameter estimates suggest that the 30-day duration curve conforms to an EVII distribution (i.e. $k > 0$), whilst the parameters derived using the L-moment method suggest an EVIII distribution ($k < 0$). Similarly, using L-moments an EVII distribution is produced for the 180-day series, whereas an EVIII distribution is produced applying the Maximum Likelihood method. Figure 4.3 and 4.4 illustrate the Q_T-T relationships for 9002 and 39016 respectively.

The Q_T-T plots illustrate that the prescriptive performance of the methods is similar when short durations are used, but that the difference between the two becomes important when longer durations are used, in which case the L-moments seem to produce the best results.

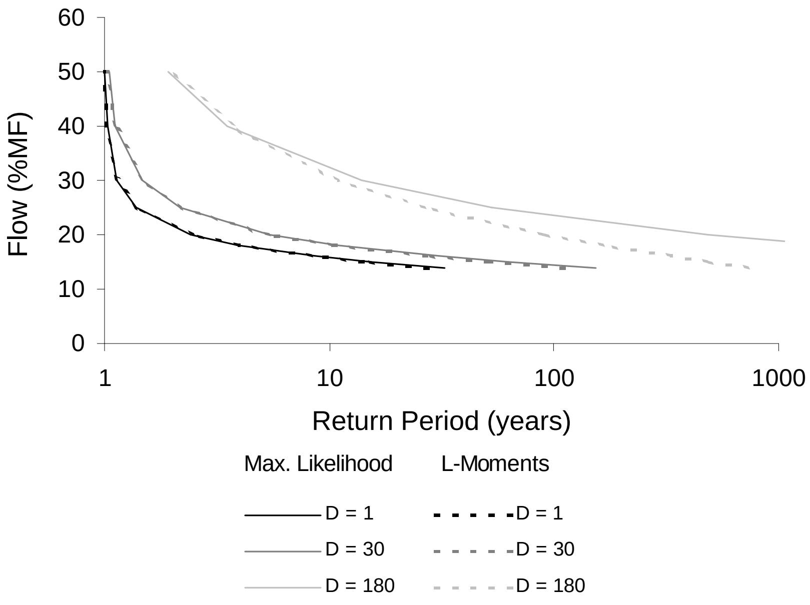


Figure 4.3: Flow–return period relationships for station 9002

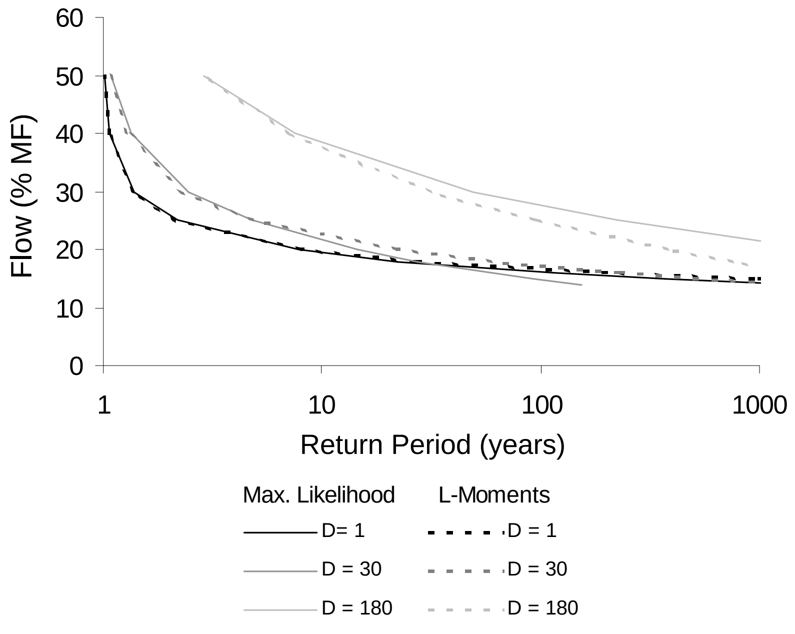


Figure 4.4: Flow-return period relationships for 39016

4.5 Assessment of the Effects of Hydrometric Error

4.5.1 Sources of hydrometric error

In frequency analysis it is important to bear in mind that the recorded flow values may be slightly different to the flows that actually occurred. Such differences are mainly due to measurement errors and may manifest in the flow record as a systematic bias (e.g. via a not-quite-correct rating curve), or as random errors. A second source of error arises due to the effect of external influences in the stream, such as weed growth around the gauging structure. Granularity within the annual minima series can be considered as a third cause of error. This arises when the annual minima fall into discrete ranges, resulting in the curve having a step-like shape, and is due to the measurement precision of the recording process (i.e. flows are rounded-up). The effect of granularity on the frequency curves is considered in section 4.5.3. Poor measurement at flows close to zero means that some flows are recorded as zero flows. A method for adjusting for zero flows is discussed in section 4.5.4. The step-like shape observed in many frequency curves may also be a result of the relatively small sample sizes involved, which, when the plotting position method is applied, result in similar flow values being assigned different probabilities. This is a separate issue to that of measurement precision and is thus discussed in section 4.6. Errors may also be introduced if the time series of interest is one that has been naturalised. However the application of LFFA to naturalised data is not described further in this report.

4.5.2 Random hydrometric errors

Representation of random errors in the data

As a full assessment of the effects of different error structures was beyond the scope of the project, the investigation was focussed on characterising the influence of random errors on the flow / probability relationship. Artificial random errors were introduced to the flow records of stations 39016 and 72004, being applied both to the annual minima flow series and the daily mean flow time series.

Three levels of error were added to the annual minima: errors up to ±5% of the annual minima, errors up to ±10% of the annual minima, and errors up to ±20%. However where errors were added directly to the flow record, new series of annual minima had to be generated. To apply errors of realistic size, the magnitude of the synthetic errors were constrained by the observed range of annual minima flows. As before, three different levels of error were considered. In the first case random errors of a maximum size of ±5% of the range of minima were applied, in the second case errors to within ±10% were applied, and in the third case a errors of up to ±20% were added to the data. To illustrate the method, the annual minimum series for a duration of D=1 at station

39016 is presented in Table 4.2, the corresponding values for errors of 5, 10 and 20% are 0.245 m^3s^{-1}, 0.485 m^3s^{-1} and 0.97 m^3s^{-1} respectively.

Table 4.2: Annual minima flow values (m^3s^{-1}) for d=1 at station 39016 for the original data set

Ranking position	Annual minimum flow(m^3/s)	Ranking position	Annual minimum flow(m^3/s)
1	5.78	20	3.89
2	5.64	21	3.81
3	5.49	22	3.8
4	5.24	23	3.74
5	5.1	24	3.71
6	4.93	25	3.7
7	4.9	26	3.68
8	4.8	27	3.57
9	4.79	28	3.53
10	4.55	29	3.49
11	4.43	30	3.43
12	4.38	31	3.25
13	4.37	32	3.17
14	4.22	33	3.1
15	4.2	34	3.08
16	4.11	35	3.07
17	4.09	36	2.92
18	4.08	37	2.34
19	4.03	38	0.93
Annual minimum flow range = 4.85 m^3s^{-1}			

Modification of the annual minima series

Figure 4.5 illustrates the adjusted probability plots when errors are introduced to the annual minima series for a duration of D=1, at station 39016.

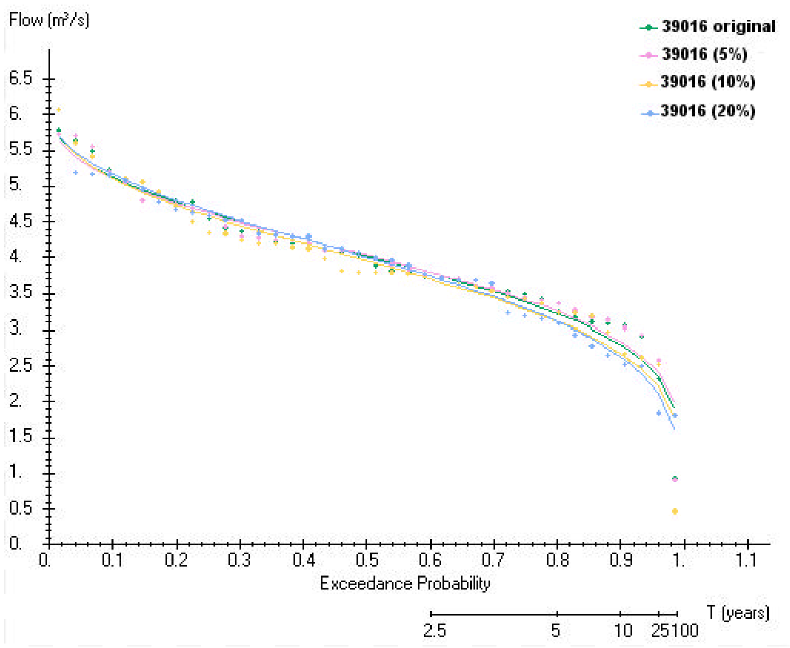


Figure 4.5: Probability plot showing the influence of hydrometric errors introduced within the 1-day annual minima series at Stations 39016, fitted with a GEV distribution

As shown in Figure 4.5 introducing errors to the annual minima series causes some changes in the flow-exceedance probability relationships with the largest changes occurring at the extremes of the curve (0.92 < p < 0.1). At the upper end of the probability scale a small change in flow can result in a large change in return period, and therefore the results illustrate that it is critical to have an understanding of hydrometric errors affecting recorded flows in this range.

Modification of the daily flow record

Figures 4.6, 4.7 and 4.8 illustrate adjusted probability plots for station 39016, for durations of D=1 and D=60 respectively. Here errors have been introduced directly in to the daily flow series from which the annual minima series is later derived. In the latter two cases the curves shown are flow-return period plots, using a linear scale to represent return period.

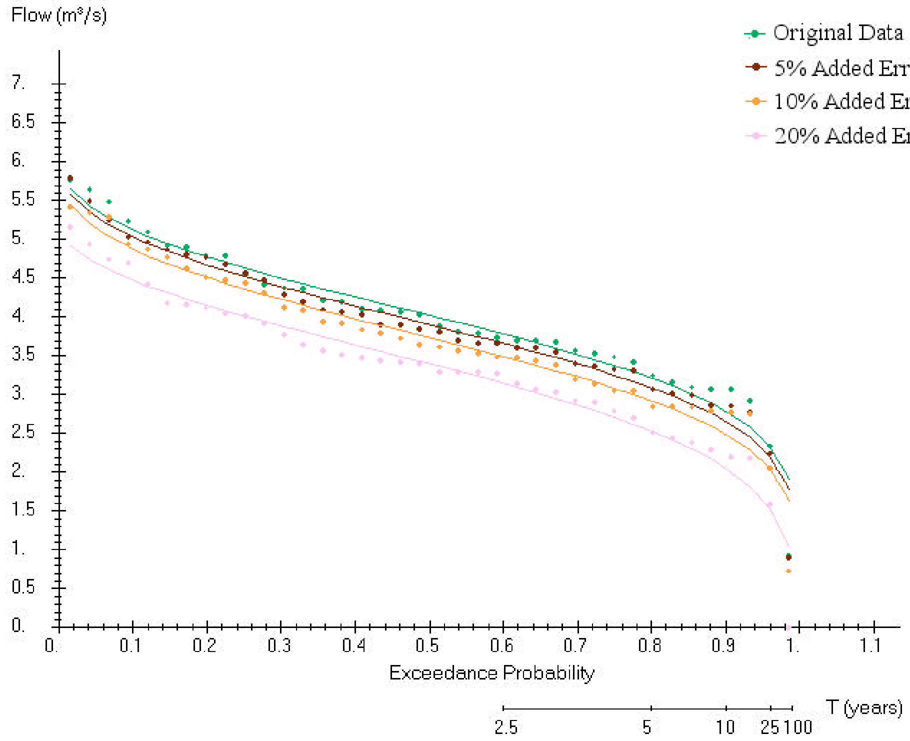


Figure 4.6: Probability plot showing the influence of hydrometric errors added to daily flows at Stations 39016, fitted with a GEV distribution, D= 1

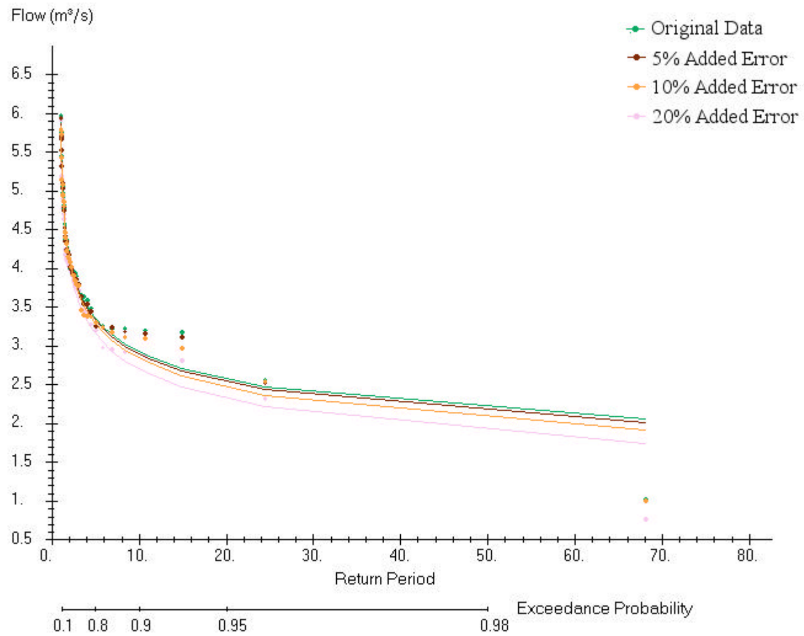


Figure 4.7: Flow- return period plot showing the influence of hydrometric errors added to daily flows at Stations 39016, D= 7

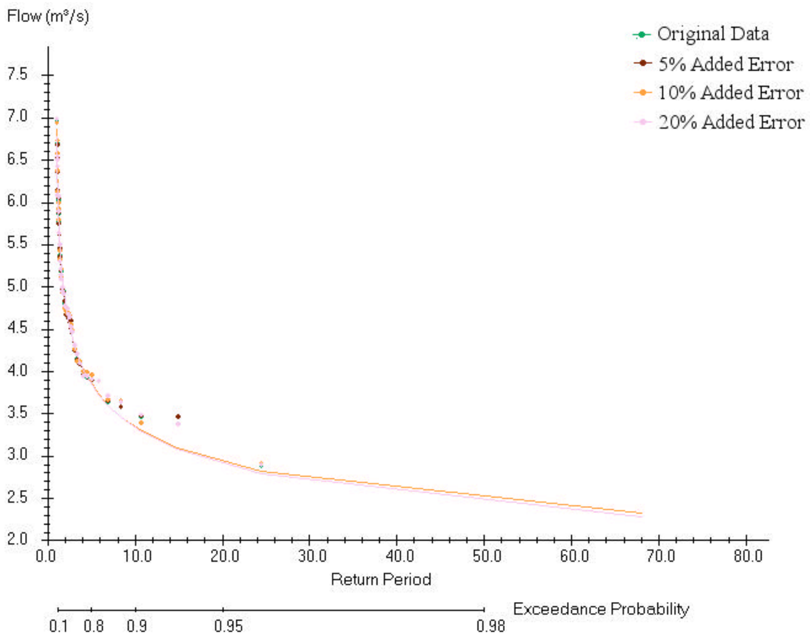


Figure 4.8: Flow- return period plot showing the influence of hydrometric errors added to daily flows at Stations 39016, considering duration D= 60

Modifying the daily flow values has an impact on both the position and shape of the curves. In the examples shown the curve is shifted further along the y-axis as the error added is increased. The impact of hydrometric errors is smallest where D=60, as the effect of each error is averaged out over 60 flow values.

Figures 4.5 to 4.8 indicate that, in general, hydrometric errors below ±10% have little influence on the shape of the flow-probability relationship. This is equally applicable whether the errors occur in the daily mean flow series or in the annual minima series. However where larger errors are considered (±20 %) there can be some considerable change in the flow-return period relationship (giving more than a 20-year uncertainty in return period), particularly where the duration considered is shorter than 30 days. At longer durations the effects of hydrometric errors in the original flow series are small.

4.5.3 Discretization effects due to granularity

Discretization of daily flow records may occur where the variation between flows on two consecutive days is lower than the measurement precision of the gauging instrumentation, causing a series of flows to be rounded up to the same value. Discretization is therefore more likely during low flow periods where there is little change in the flow from day to day. For example if the precision of the gauging process is 0.01, it is readily apparent that all the values recorded will be multiples of 0.01

cumecs, with the effect that the data-values to be modelled will have a discrete distribution. Discretization is discussed in more detail in Chapter 2.

In some cases, observations that are extremely low, but still above zero, are rounded to or reported as zero (the flow at which this occurs is usually a function of the measurement precision and is termed the ‘perception threshold’). Such data sets may be treated as censored samples. Zero flows may also, of course, appear naturally within the data set. However without circumstantial evidence, it is difficult to determine whether zeros are ‘real’ or not from the flow record alone (a more detailed discussion is given in Chapter 2). The next section investigates a method for adjusting the frequency curve to take into account zero flows. The impact on the fitted flow-frequency curves where records are subject to discretization is discussed in Section 4.6.

4.5.4 Treatment of zero flows as censored data

A specialised plotting position formula for censored data

Stedinger *et al.* (1992) proposed a specialised plotting position formula for use with censored data. Their argument for using a specialised formula supposed that among *n* flow records a perception threshold would be exceeded *r* times. The natural estimator of the exceedance probability, p_e, of this threshold is *r*/N. If the *r* values which exceeded the threshold were indexed by i=1… r , reasonable plotting positions approximating the exceedance probabilities within the interval (0, p_e) would then be:

$$p_i = p_e\left(\frac{i-a}{r+1-2a}\right) = \frac{r}{N}\left(\frac{i-a}{r+1-2a}\right) \tag{4.52}$$

where *a* is the plotting-position parameter.

Comparing results from censored data method and normal method

In order to assess the effects of applying the Stedinger Plotting Position Formula, the behaviour of two flow records known to contain zero flow values were investigated. For the first catchment, the Enrick at Mull of Tor (6008) annual minima values of zero occur for durations of 1 and 7 days. The annual minima series for the second catchment, the Enwenny at Keepers Lodge, contains zeros where D=1 (note that neither of these catchments are included in the list of 25 study catchments detailed in Chapter 3). In each case plotting positions were assigned using Stedinger’s formula, and parameter estimation carried out using L-Moments.

Figure 4.9 shows the resulting frequency curves in the case where D=1 for 6008. The curves derived using the Gringorten plotting position formula are also shown for comparison (zero flows are given the same rank (1)). Table 4.3 compares the parameter values for based on the censored and non-censored methods for durations of D=1 and D=7.

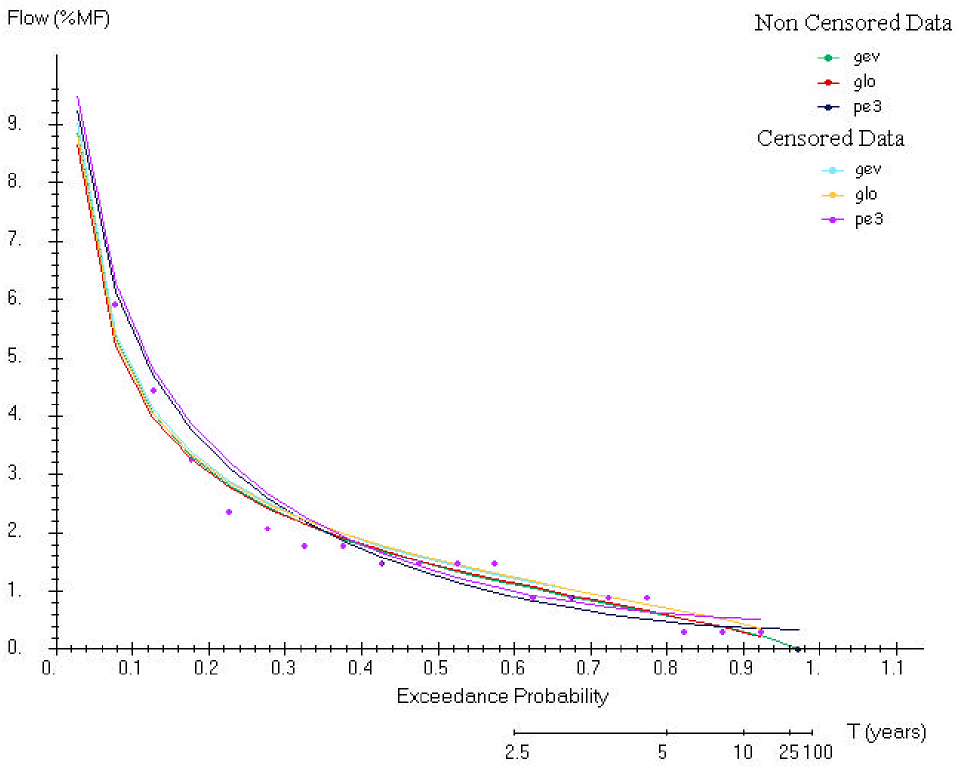


Figure 4.9: Frequency curve for 9008, where D=1

Table 4.3: Comparison of parameter values for 6008

Duration (days)	Parameter	GEV		GL		PE3	
		Non Cens.[1]	Censored	Non Cens.	Censored	Non Cens.	Censored
1	a	0.99	0.96	0.81	0.80	2.57	2.61
1	?	1.01	1.11	1.42	1.52	2.21	2.33
1	k	-0.40	-0.42	-0.45	-0.47	2.75	2.85
7	a	1.16	1.14	0.95	0.94	3.02	3.07
7	?	1.12	1.24	1.61	1.72	2.53	2.67
7	k	-0.40	-0.41	-0.45	-0.46	2.75	2.83

[1] Non Cens. Refers to non-censored

Similarly Figure 4.10 shows the resulting frequency curves in the case where D=1 for 58009, whilst Table 4.4 compare the parameter values for based on the censored and non-censored methods where D=1. Although the curve shape changes slightly due to the revised plotting positions for zero flows, this causes little change in the form of the fitted p.d.f.'s for the different distributions considered. The main influence seems to be on the the location parameter, ?.

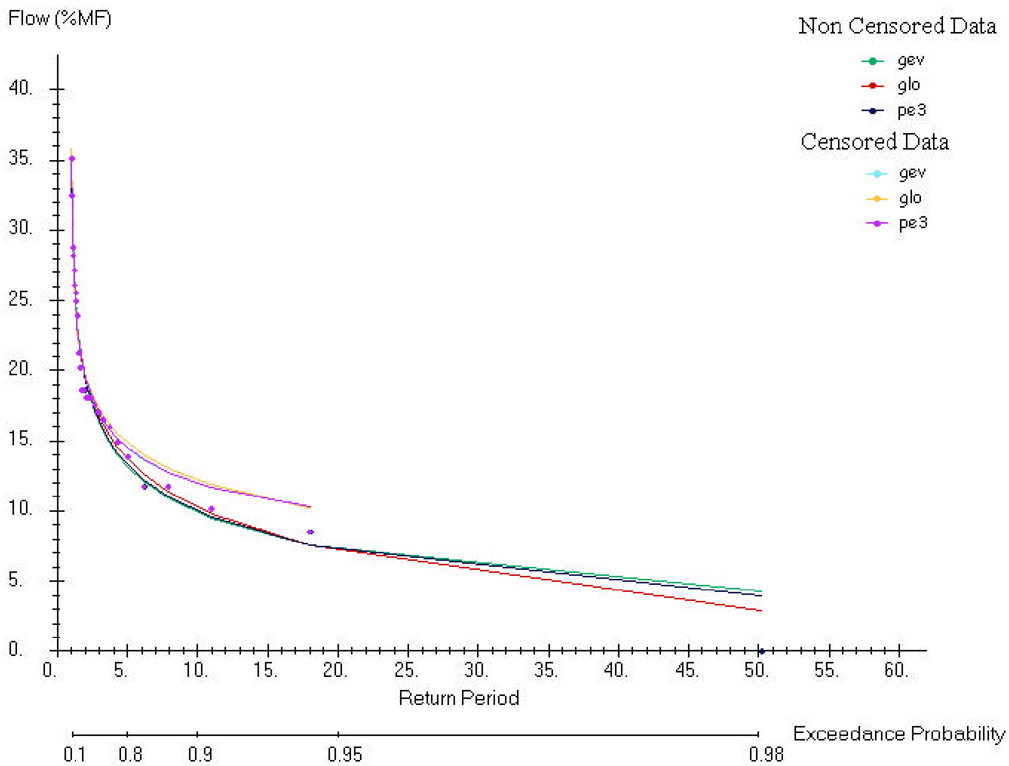


Figure 4.10: Frequency curve for 58009 where D=1

Table 4.4: Comparison of parameter values for 59008, D= 1

Parameter	GEV		GL		PE3	
	Non Cens. [1]	Censored	Non Cens.	Censored	Non Cens.	Censored
A	7.41	6.14	4.13	3.68	7.32	6.58
?	16.94	17.51	19.52	19.75	19.40	20.12
K	0.32	0.18	0.02	-0.06	-0.11	0.37

[1] refers to non-censored data

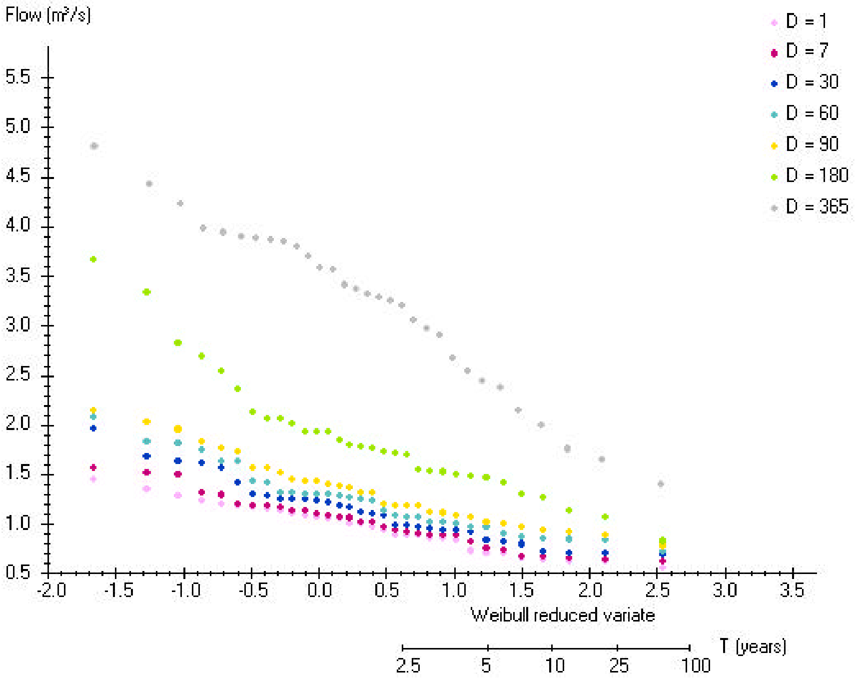
4.6 Discretization effects due to sampling error

4.6.1 Stepped probability plots

Discretization occurs when several of the annual minima have the same or very similar values resulting in the formation of steps within the frequency-curve. The probability plots for annual minima flows observed at stations 14001, and 72004, shown in Figure 4.11a and b, do not have a smooth curve. Rather, they have a stepped form.

Whilst the examples shown in Figures 4.11a and b are rather extreme cases, few of the flow records examined have completely smooth probability curves. Some other examples are presented within Section 5.2. The occurrence of steps may be explained in a number of ways. Granularity caused by measurement precision could result in discretization where short durations are considered (D=1 7, 30 days). However granularity is unlikely to have an effect for long durations, where each annual minima is an average value of many daily flows, and is unlikely to explain why annual minima frequency curves often show stepped features.

a)



b)

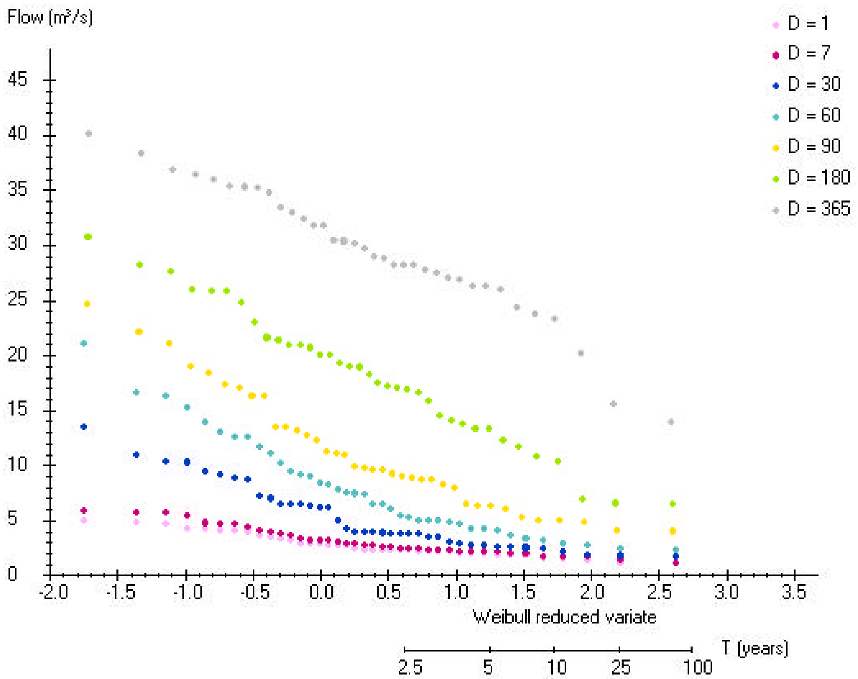


Figure 4.11b: Probability plots for annual minima flows observed at a) Station 14001 and b) Station 72004

Having excluded the possibility of discretization at near-zero flows, other ways to explain these steps must be sought. Sampling error where the record length is short may explain the occurrence of steps. For the station 14001, 31 years of data are available, and for station 72004, 38 years of data are available. Although we consider these as "long records", they reflect only a short part of the climate cycle and a limited range of flow extremes (i.e. years of similar drought behaviour are being sampled). This effect is exaggerated by the way exceedance probabilities are assigned according to rank order, so that the probability intervals are equally spaced whilst flow values are not. A longer record length will represent a bigger part of the climate cycle and a wider, more representative range of flow extremes and thus, the shape of the probability curve (for annual minima flows) will be smoothed. In light of these findings, further research was conducted, to determine whether the presence of stepped discrete data can influence the parameters and accuracy of fit of the distribution.

4.6.2 Influence of a step shape on the fitted distribution

In order to study the influence of a stepped curve on the fitted distribution (and hence the return period / flow relationship) two flow records were considered: those of stations 14001 and 72004. Two different types of distribution: a GEV distribution and a general logistic distribution are considered. Since the process of fitting a curve to the annual minima series is independent of the duration, only one duration, D=1, was considered.

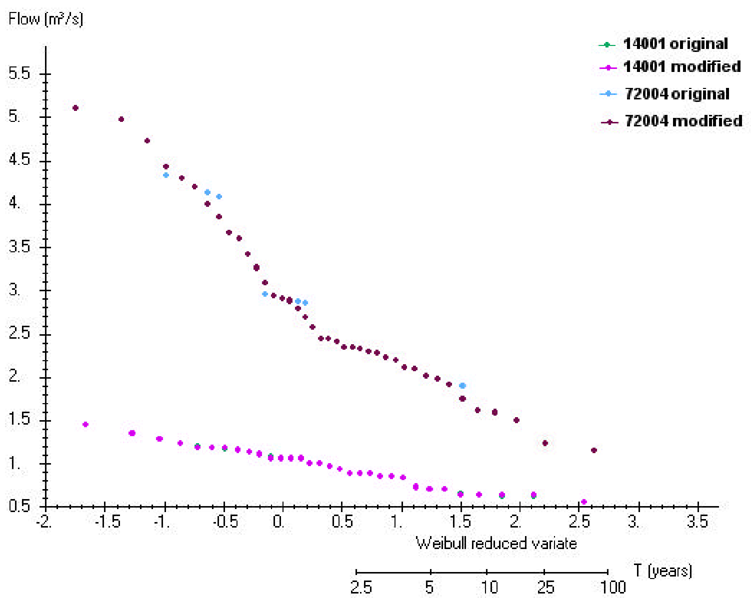
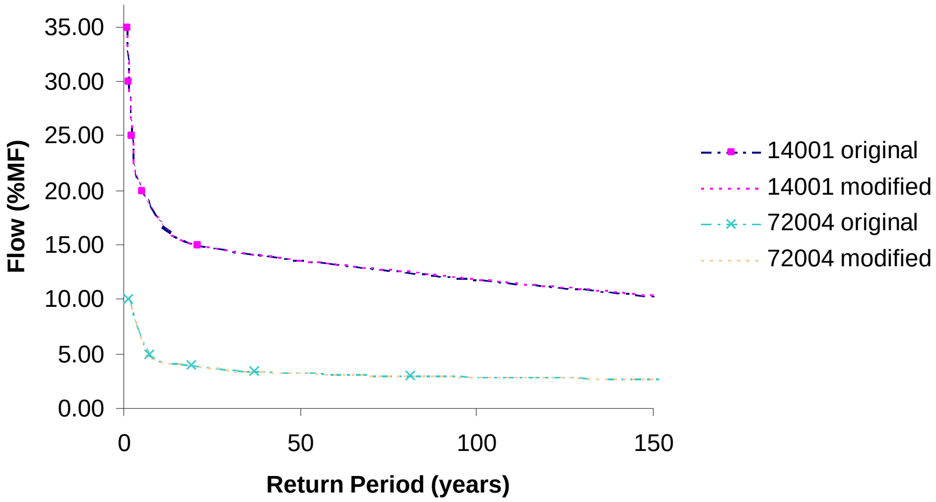


Figure 4.12: Probability plots for original and modified annual minima flows at Stations 14001 and 72004

Figure 4.12 shows the flow frequency curves (labelled “original”) for each station. That of station 14001 has a stepped form, whereas that for station 72004 is a good example of a smooth curve. For each station a modified daily data set was created: the probability plot for station 14001 was smoothed whilst that for station 72004 (Figure 4.12) was given some artificial steps. These modifications were carried out by adjusting the curves “by eye”. Each of these series was fitted using the generalised extreme value (GEV) and generalised logistic (GL) distributions: Figures 4.13a and b illustrates the resulting flow, and indicate that there is little difference in the goodness of fit for the original and artificial data.

a)



b)

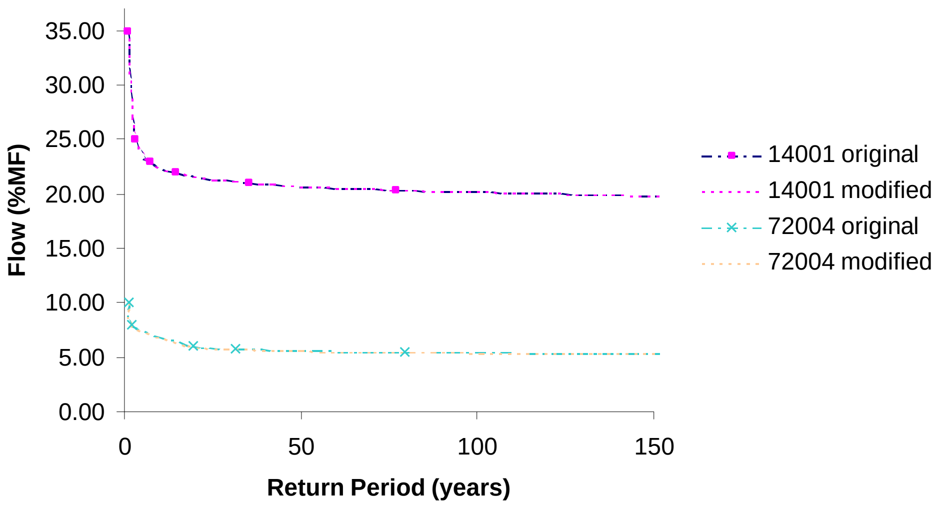


Figure 4.13: Effect of a step shape on the flow - return period relationship for stations 14001 and 72004 for a period of d=1, fitted with the a) GEV and b) GL distribution

Generalising the results presented within Figure 14.13, suggests that in the case of natural streams within the UK, a stepped probability curve does not result in a worse fit or poorer parameterisation.

5. EVALUATION

5.1 Introduction

This chapter presents the results of the modelling of the observed low flow frequency distribution, in which the parameters of three candidate distributions (Generalised Extreme Value, Generalised Logistic, and Pearson Type-III) and one ‘control’ distribution (Generalised Pareto) have been estimated using L-Moments. Parameters were derived for each of the catchments as detailed in Chapter 4.2. For illustrative purposes the results are illustrated using five flow records; the Deveron at Muiresk (9002), the Tweed at Boleside (21006), the Kennet at Theale (39016), the Wye at Ddol Farm (55026) and the Lune at Caton (72004), which represent a range of catchment types. The reader is referred to Chapter 3 for more details regarding these catchments.

Firstly, the observed probability distributions are presented in Section 5.2, in the form of frequency curves using the Gringorten plotting position method on a Weibull reduced variate plot. The results of fitting four different distributions to the curves are presented in Section 5.3. The suitability of each of the four distribution families for describing the annual minima frequency curves is then evaluated. The goodness of fit between the modelled and observed probability curves (i.e. descriptive ability of the distribution) is discussed in Section 5.4, whilst the robustness of the flow-return period relationship (i.e. prescriptive properties) is described in Section 5.5. The applicability of the distribution for different durations, and adequate representation of flows beyond the observed range (e.g. flows at or close to zero) were also considered in the assessment.
The relationship between catchment characteristics and low flow frequency behaviour for the 25 catchments is developed and discussed in Section 5.6. These results are closely related to the methods and findings of the Low Flow Studies Report (Institute of Hydrology, 1980), and are important with regard to choosing which distribution to use with data from different catchment types.

5.2 Observed Low Flow Distributions

5.2.1 Form of observed frequency curves

The following probability plots (Figures 5.1 to 5.5) show flow-probability relationships for annual minima series of the five main gauging stations. In each case probabilities of non-exceedance have been derived using the Gringorten plotting position formula (note that here the probability of exceedance is plotted (i.e. 1-*p*) so that the lowest flows appear at the upper end of the x-axis). To linearise each plot, probability is expressed using a reduced variate scale, in this case the Weibull reduced variate. In each case, the curves for a range of durations (1, 7, 30, 60, 90, 180 and 365 days) are shown. To facilitate comparison between catchments, whilst maintaining differentiation between different durations, the curves are standardised by mean flow.

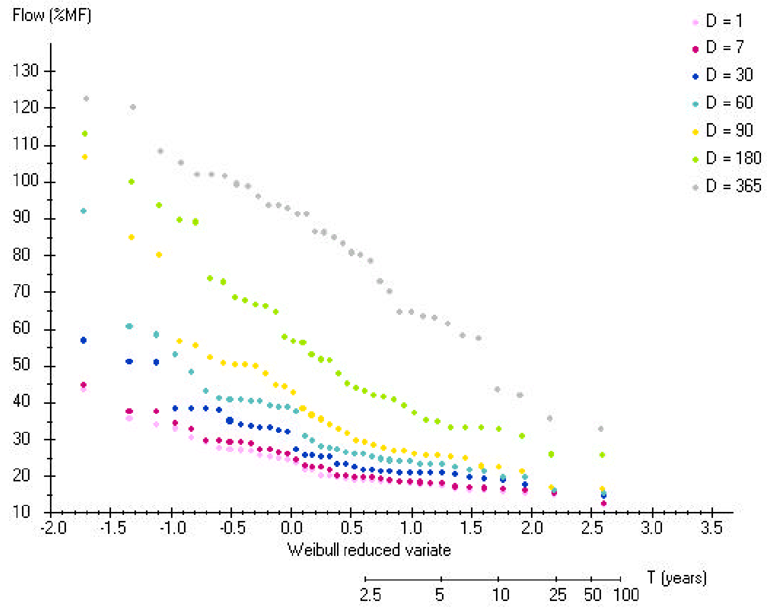


Figure 5.1: Probability plot for annual minima flows observed at Station 9002

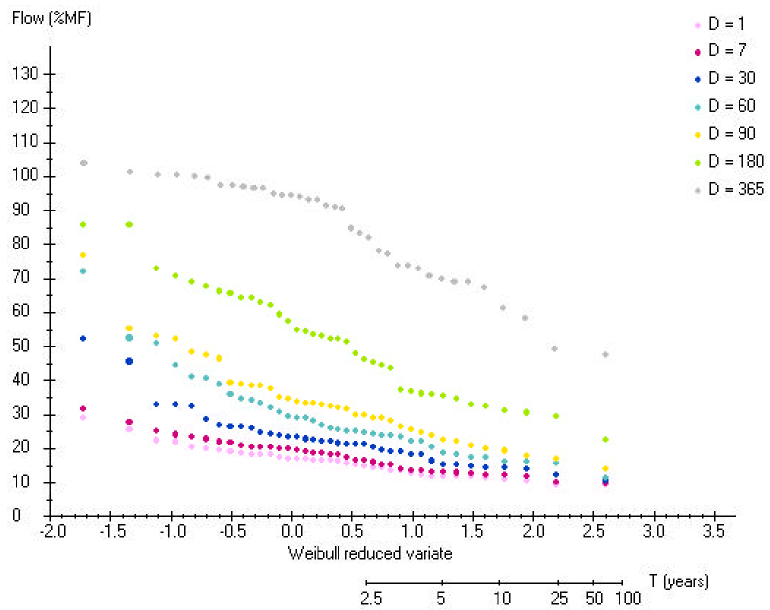


Figure 5.2: Probability plot for annual minima flows observed at Station 21006

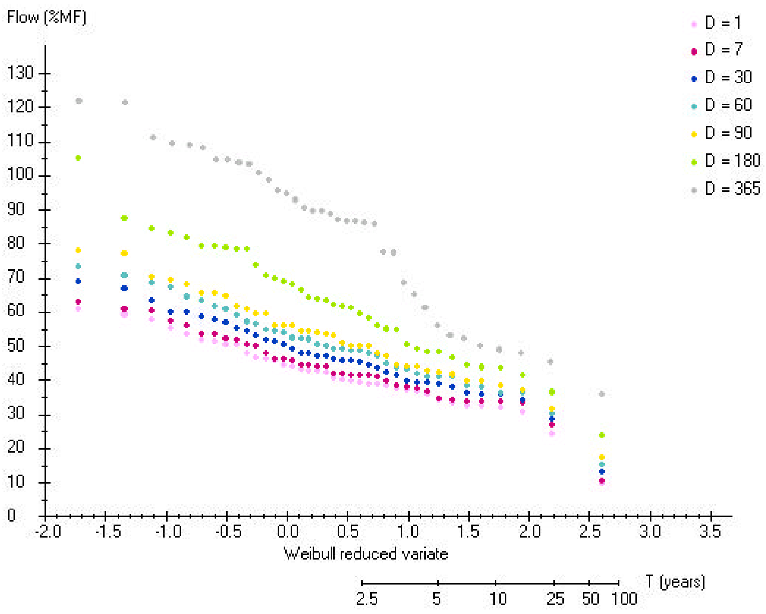


Figure 5.3: Probability plot for annual minima flows observed at Station 39016

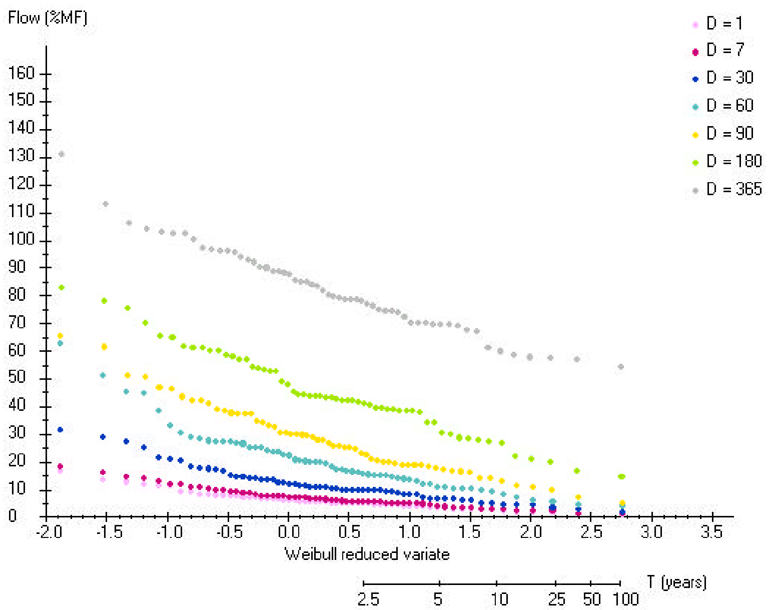


Figure 5.4: Probability plot for annual minima flows observed at Station 55026

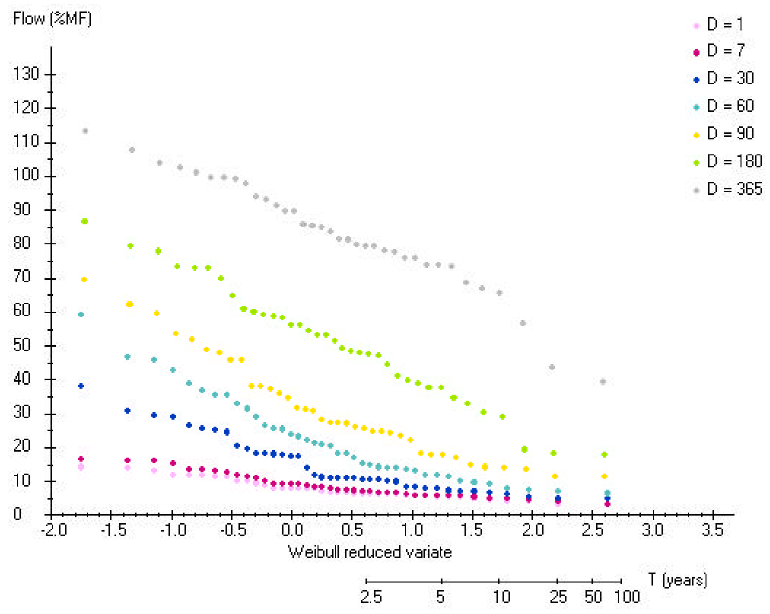


Figure 5.5: Probability plot for annual minima flows observed at Station 72004

Figures 5.1 to 5.5 illustrate that, for a given duration, the flow decreases as higher return periods are considered. Note that the appearance of a (semi)-linear decrease in flow results from the use of the reduced variate scale along the probability axis (x-axis). However, in some cases, there are changes in gradient along the curve, particularly within the central part of the curve (e.g. as shown in Figure 5.3).

In some cases distinct steps are seen in the frequency curve: a good example is that of station 19002 (Figure 5.6). These steps occur when several annual minima have the same value. In the case of 19002, the sampled annual minima conform to a discrete distribution, rather than a continuous one. Possible explanations for this include years of similar drought behaviour being sampled in the flow record (this is more likely where the record length is shorter) or, for longer durations, where the annual minima represents the average flow over a long period. It may also occur due to the effect of measurement precision at low flows (i.e. flows are rounded up), which is also more likely at short durations. An analysis and discussion of the effect of stepped curves on distribution fitting was given in Section 4.6. The presence of steps within the data was demonstrated to have little effect on the curve-fitting process.

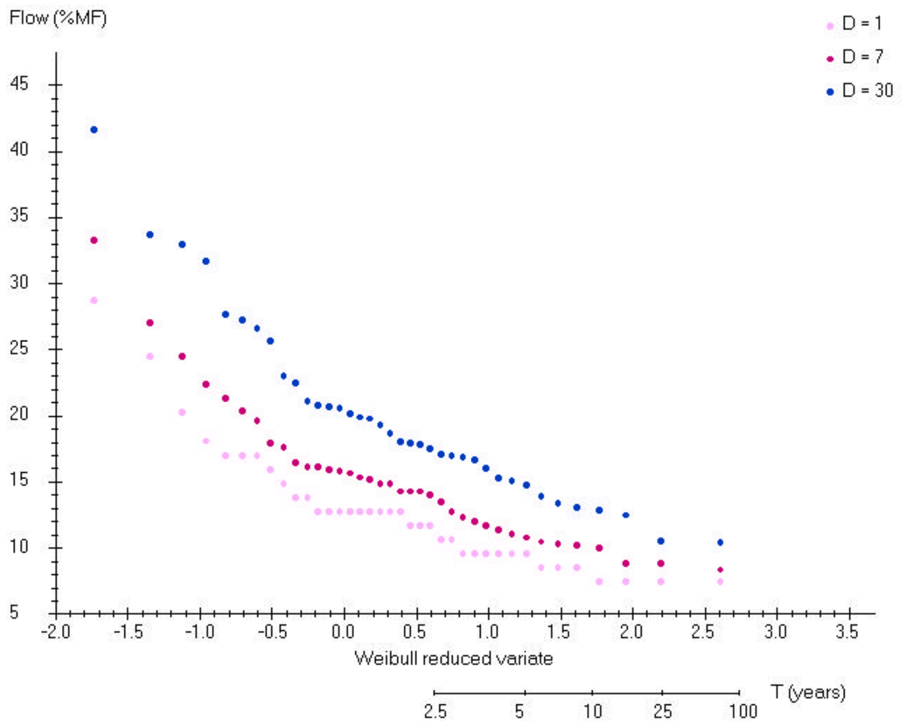
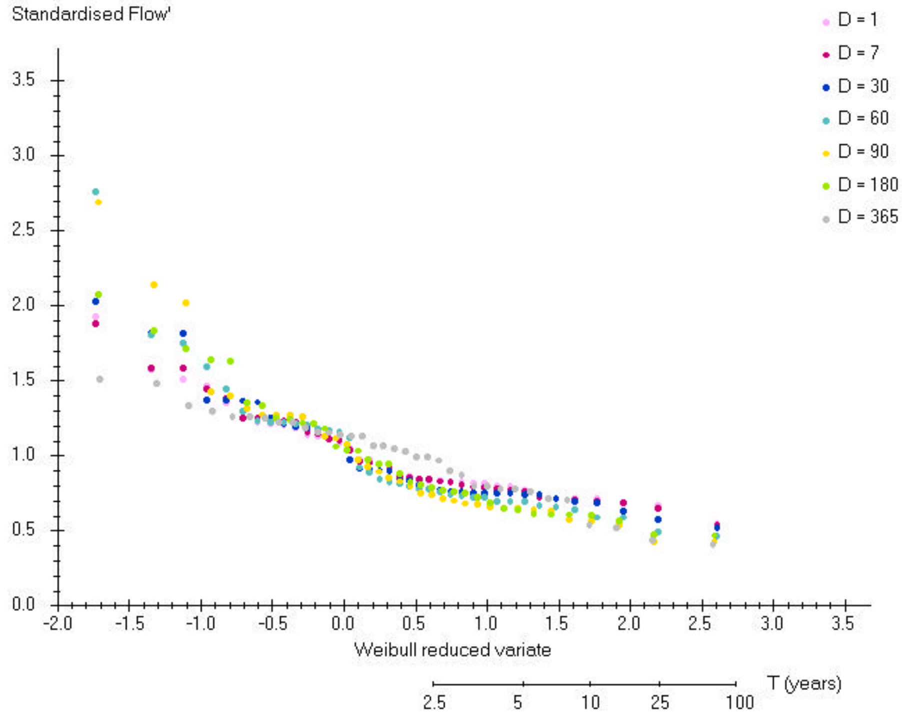


Figure 5.6: Probability curve for 19002, showing granular frequency curves

5.2.2 Variation of annual minima with duration

For a given station the gradient and form of the curves change as the duration, D, becomes longer. Where short durations are considered the annual minima are smaller, and occupy a much narrower range of values than where longer durations are considered. For example the curves for the case where D=1 have very shallow gradients, whilst the steepest curves are obtained where D=365. This was illustrated by the Institute of Hydrology (1980) and is a consequence of averaging within the longer durations (the effect of extreme low flows are averaged out causing higher values of the annual minima to become more frequent). The dominant influence is on location rather than shape of the curve (i.e. for longer durations the annual minima represent a larger proportion of the long-term mean flow, thus the curves begin at higher positions along the y-axis). This influence can be normalised by expressing the flow relative to the mean annual minima value for the particular duration, MAM(D), as shown in Figures 5.7a and 5.7b, and is discussed in more detail in Section 5.6. The curves for different durations are coincident along the central part of the curve (particularly at a return period of 2 years, which is where the MAM(D) value usually occurs), but deviate at the extremes. This is possibly a function of the increased uncertainty associated with the curve position at extreme flows resulting from extreme events being inadequately represented in the flow records.

a)



b)

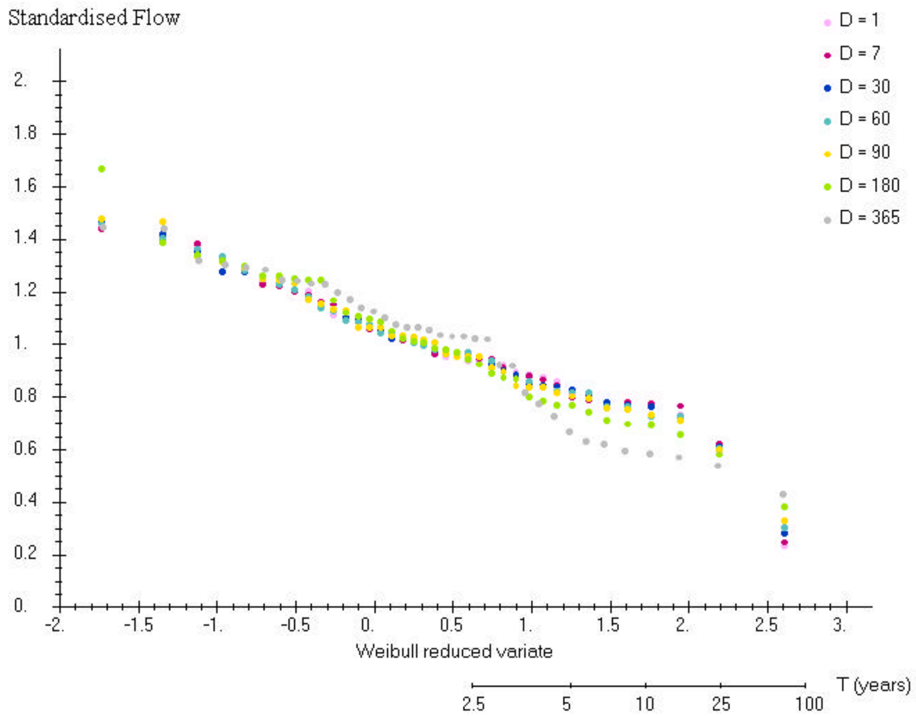


Figure 5.7: Probability plots for a) 9002 and b) 39016 where flow is expressed as a standardised value (i.e. relative to the mean annual minima value)

5.3 Modelling and Parameterisation of Frequency Distributions

5.3.1 The choice of distribution

Following the recommendations given in Chapter 4, the method of L-Moments was used to estimate the parameters of the different ‘model’ distributions. Four model distributions were used: the Generalised Extreme Value (GEV) distribution, the Generalised Logistic (GL) distribution, the Generalised Pareto Distribution (GP) and the Pearson Type-III (PE3) distribution. Parameters were derived for each of the 25 stations based on the annual minima series in units of cumecs. The results are illustrated using the five catchments, 9002, 21006, 39016, 55026, and 72004, however the full set of parameter values is given in Appendix 6.1.

5.3.2 Use of probability plots for distribution evaluation

Probability plots can be useful pre-fitting tests when deciding which of a number of alternative families of distributions is most appropriate. In this method probability plots are generated based on different reduced variate scales: the probability curves would be expected to produce straight lines if the distribution family on which the reduced variate was based is the ‘correct’ one.

Figures 5.1 to 5.7 used the Weibull reduced variate scale to express probability, that is the relationship between the probability p and Q_T is based on the Weibull distribution with arbitrary values for the parameters α, ξ and k. Similarly it is possible to use other distributions for the basis of this relationship. Figure 5.8 demonstrates probability curves for the five stations, for a duration (chosen arbitrarily) of 30 days. In Figure 5.8a a Weibull reduced variate scale is used, in Figure 5.8b a logistic reduced variate is used, in Figure 5.8c an exponential reduced variate scale in used whilst Figure 5.8d shows the use of a Gumbel reduced variate scale.

The Weibull reduced variate (EVIII distribution) and the logistic reduced variate (GL distribution) both produced relatively linear curves. The Gumbel reduced variate scale (representing the EVI distribution) is less successful (i.e. the resulting curve is not linear), whilst use of the exponential reduced variate (representing the GP distribution) produced very poor results. These results suggest that the EVIII and GL are better candidate families than the EVI or GP.

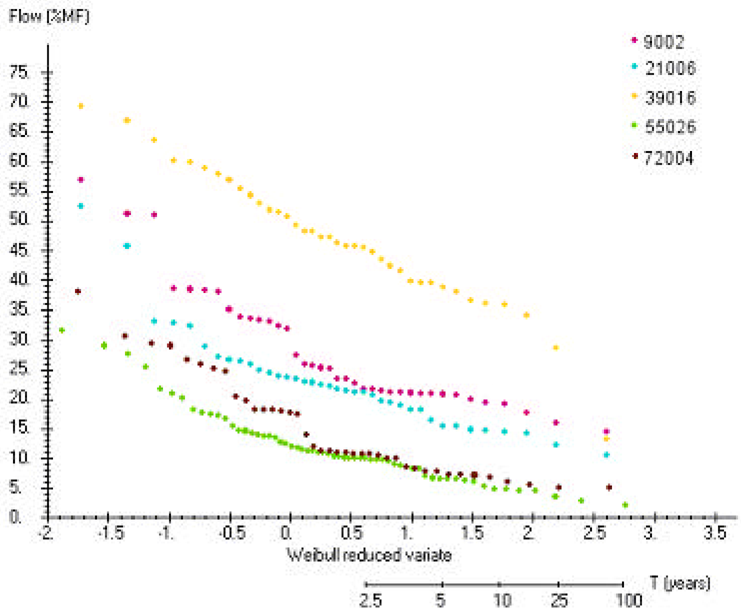


Figure 5.8a: Probability plots using the Weibull reduced variate for the five catchments for D=30

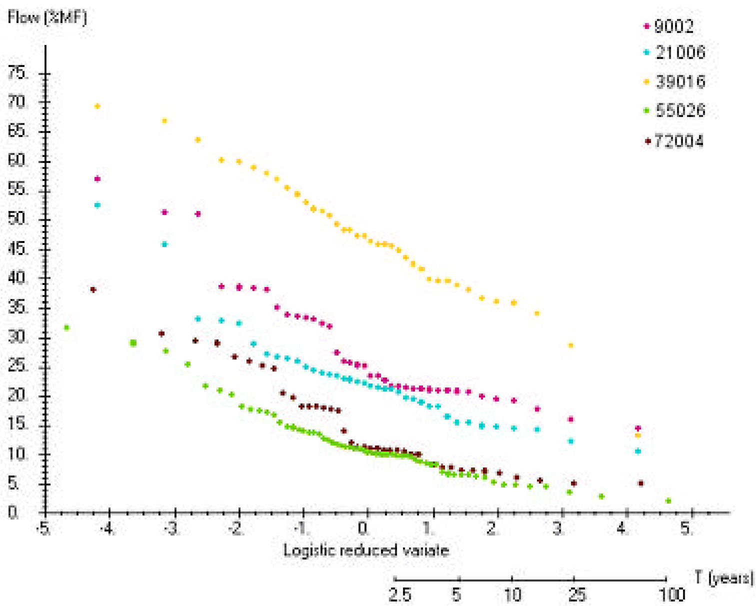


Figure 5.8b: Probability plots using the Logistic reduced variate for the five catchments where D=30

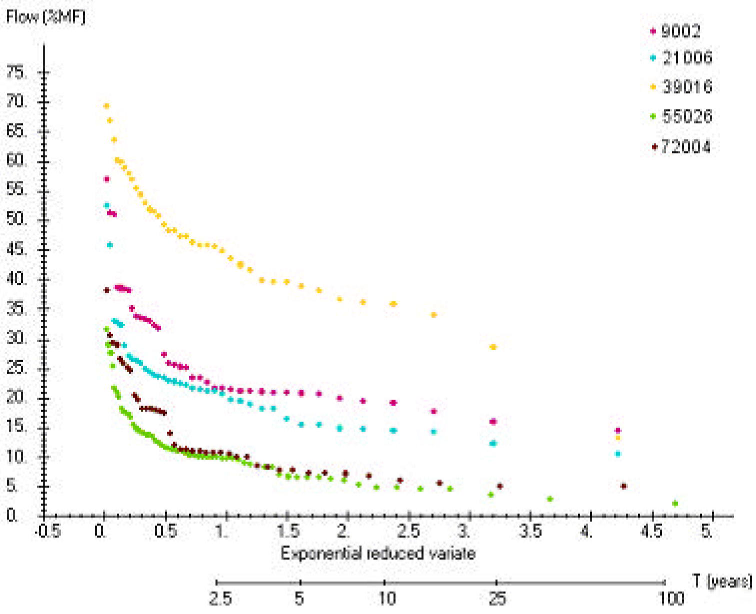


Figure 5.8c: Probability plots using the Exponential reduced variate for the five catchments for D=30

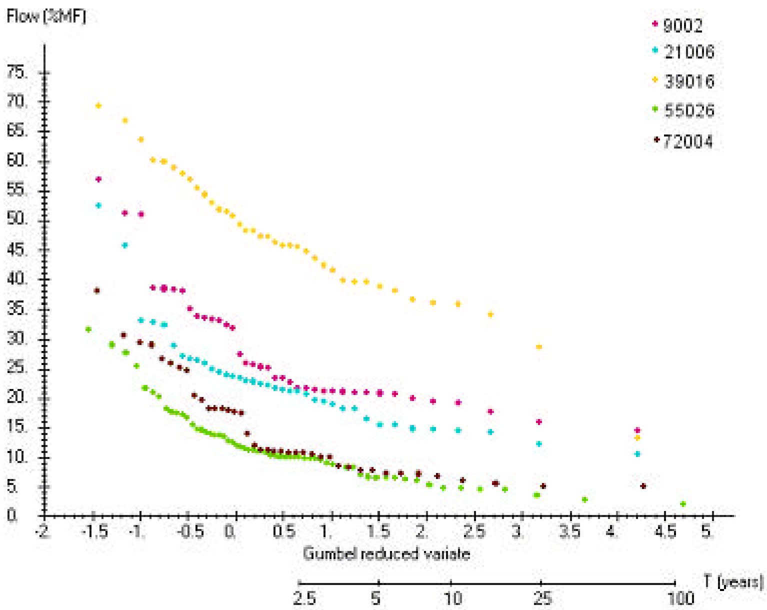


Figure 5.8d: Probability plots using the Gumbel reduced variate for the five catchments where D=30

5.3.3 Comparison of parameter values for different distributions

Figures 5.9 to 5.12 show the parameter values derived for each of the four distributions (the GEV, GL, GP and PE3, respectively. The computations were based on annual minima expressed in terms of the percentage mean flow of the flow record so that the curves for different durations plot as distinct curves. In each case the variation of the three parameters a (scale), ? (location) and *k* (shape) is shown.

The parameter magnitudes are similar for the GEV and GL families. In both cases α takes values of 20 or less, ξ takes values up to 80 and *k* varies between 0.8 and –0.4. However it is the correspondence in variation in parameter values with duration between the two sets, rather than the range, that is most striking. For both distributions the α parameter is very low where D=1, but increases as D increases, levelling off somewhat where D is above 200 days, whilst the ξ parameter increases steadily with duration. The behaviour of the *k* parameter is more complex: *k* becomes increasingly negative as D increases, until D=60, then rises steadily as D increases thereafter. In the case of the GEV distribution, this means that an EVII/III is generally favoured when the duration considered is between 7 and 90 days.

When the GP distribution is compared to the GEV and GL some differences in α and *k* parameters are apparent. The α parameter is generally much higher, and increases with increasing duration but tends not to level out at at high durations. The *k* parameter is always positive and, for the stations considered, takes values in the range 0 to 2.5. The PE3 distribution differs most in the behaviour of the *k* parameter. Although *k* takes both positive and negative values in the range, -1.5 to 2.0. in contrast to the other distributions the value of k increases with duration, reaching a point of inflection between the durations of D=60 and D=90, whereafter it decreases steadily with duration.

In each of the graphs the permeable catchment (39016) stands out from those of the other catchments which have moderate to low permeabilities. There is much less variation with duration, the curves being much flatter. In particular the value of k never drops below zero, i.e. an EVI/III distribution is always fitted. Potential contrasts between the behaviour of different catchment types are discussed further in Section 5.6 using MAM(D) to normalise for the effect of duration.

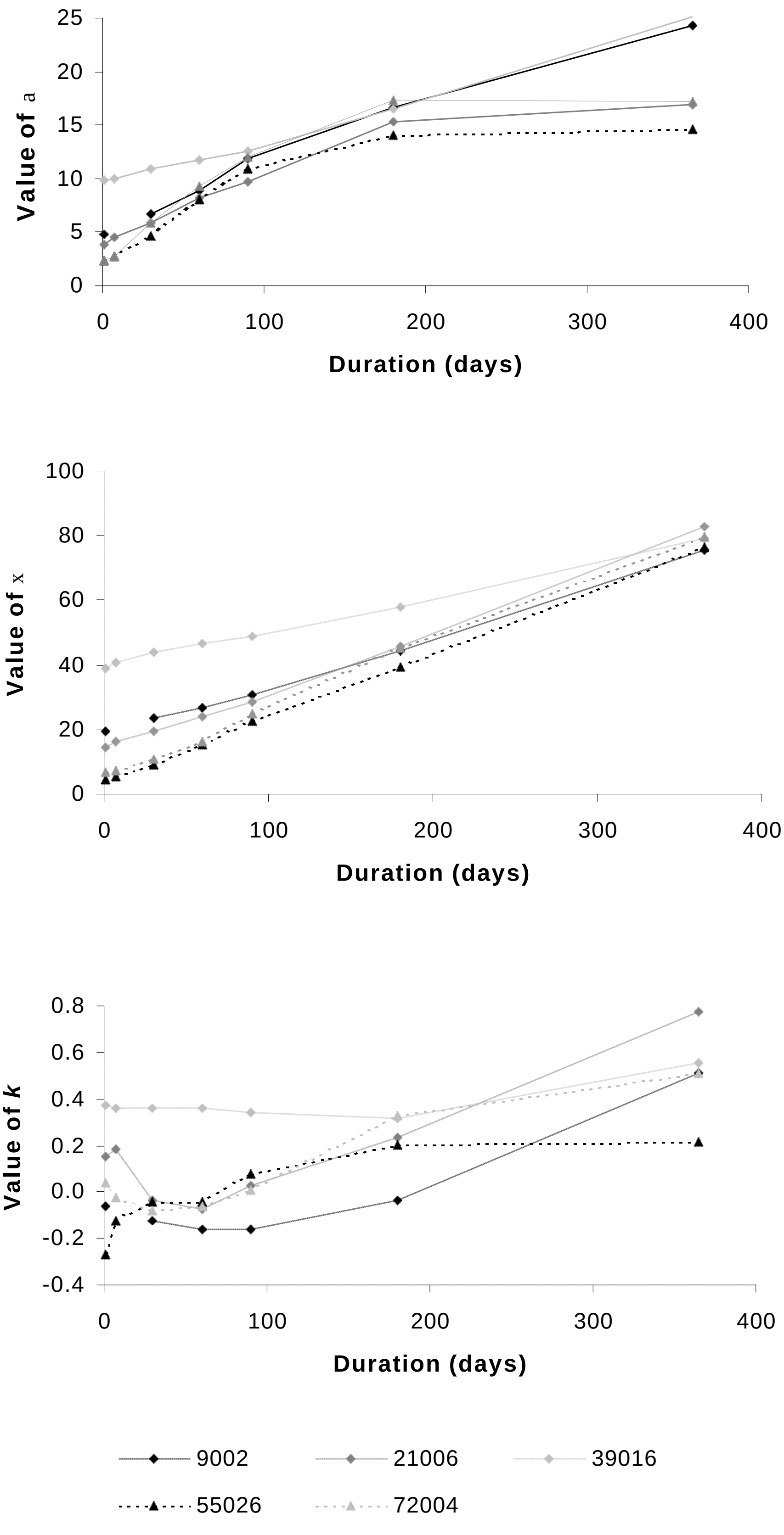


Figure 5.9: Variation in parameter values (GEV distribution)

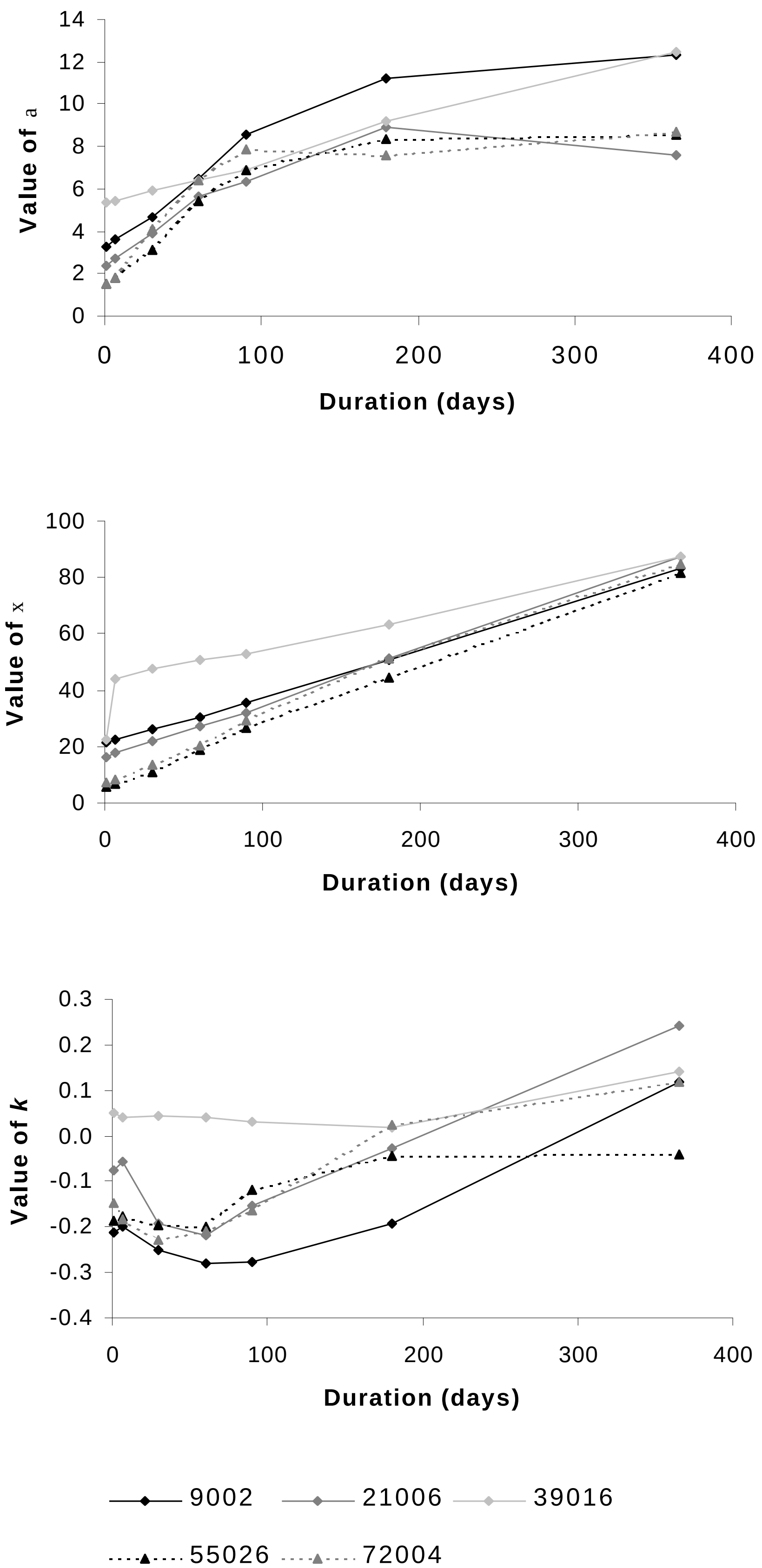


Figure 5.10: Variation in parameter values (GL distribution)

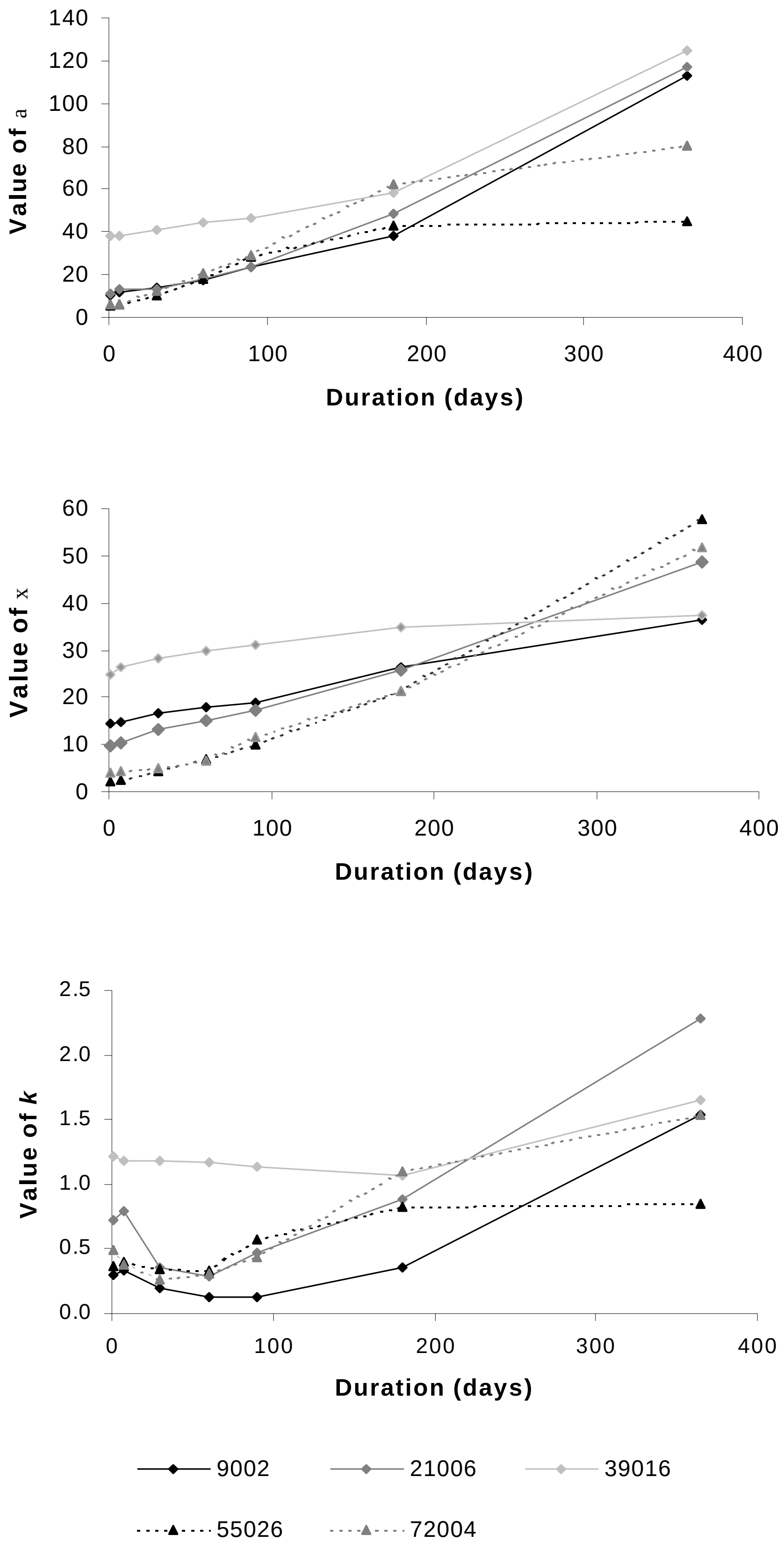


Figure 5.11: Variation in parameter values (GP distribution)

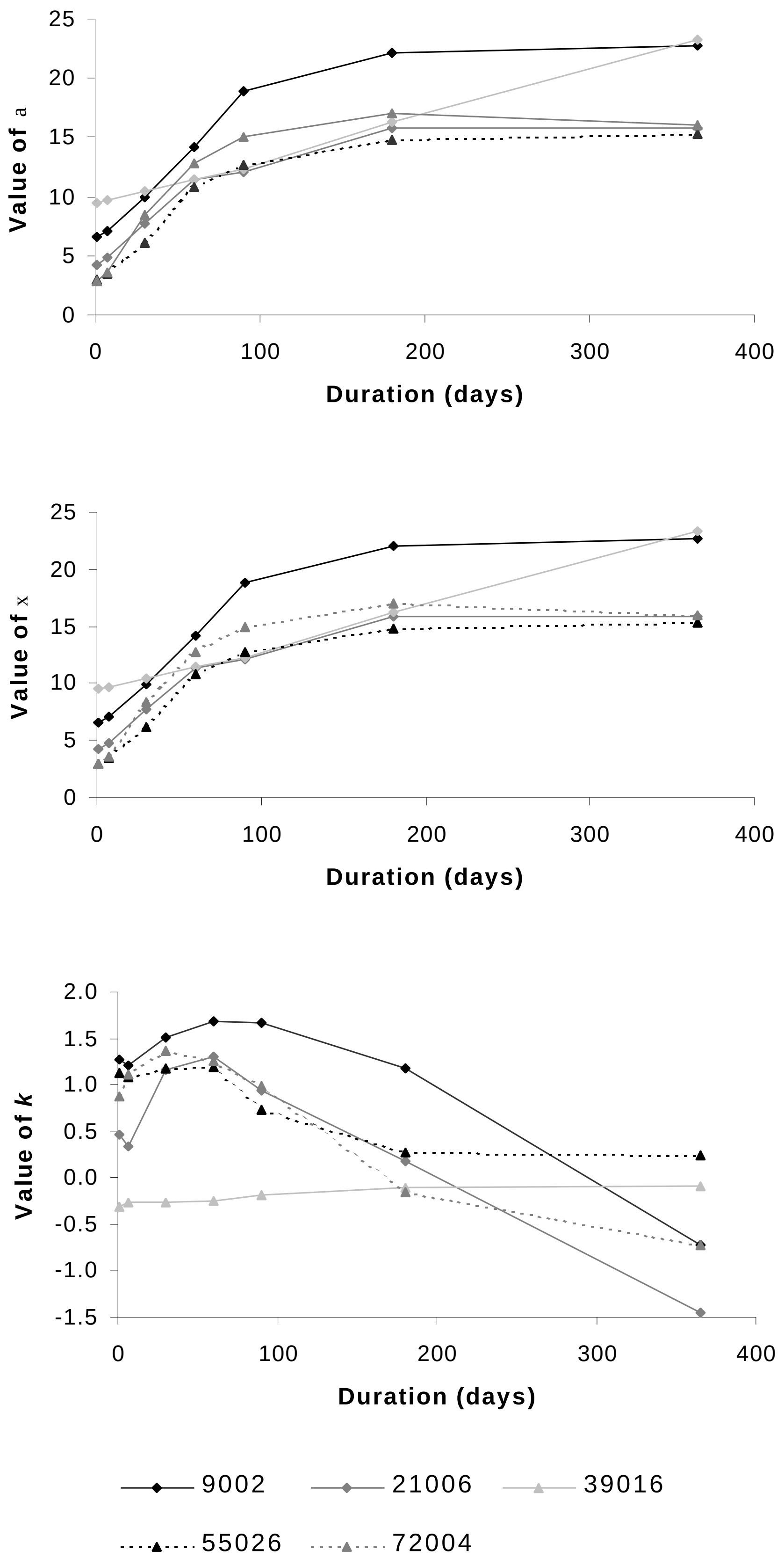


Figure 5.12: Variation in parameter values (PE3 distribution)

5.3.4 Visual comparison of observed and modelled frequency curves

Visual comparison is often the most straightforward way of determining whether the modelled curve is a good representation of the observed data! Figures 5.13 to 5.16 illustrate, as example, fitted and modelled curves for a variety of durations. The GEV distribution is shown by the green line, the GP model is described by the red line, the GL model is shown in black, whilst the blue line indicates the PE3 model.

The figures indicate that, whilst the GP distribution is generally not a good model to describe low flows, there is little to differential between the GEV, GL and PE3 distributions, which all describe the main body of the data very well.

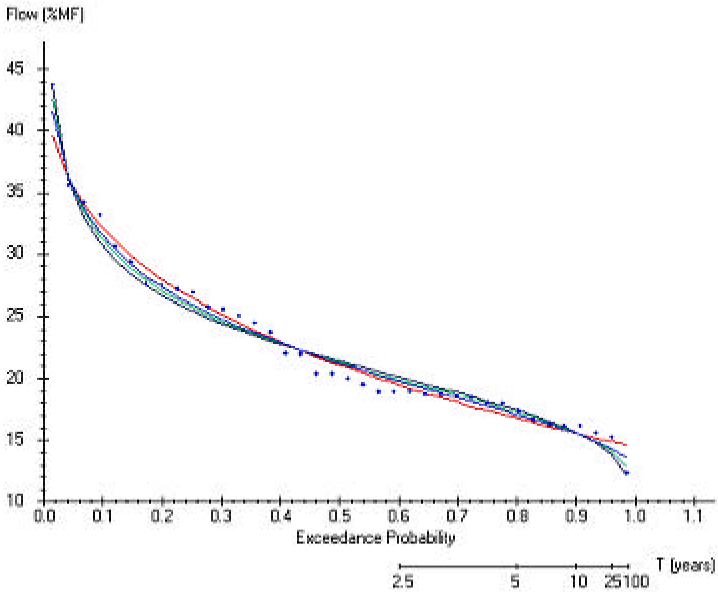


Figure 5.13: Observed and modelled curves for the Deveron (9002) in the case where D=1 *(GEV-green, GP-red, GL-black, PE3-blue)*

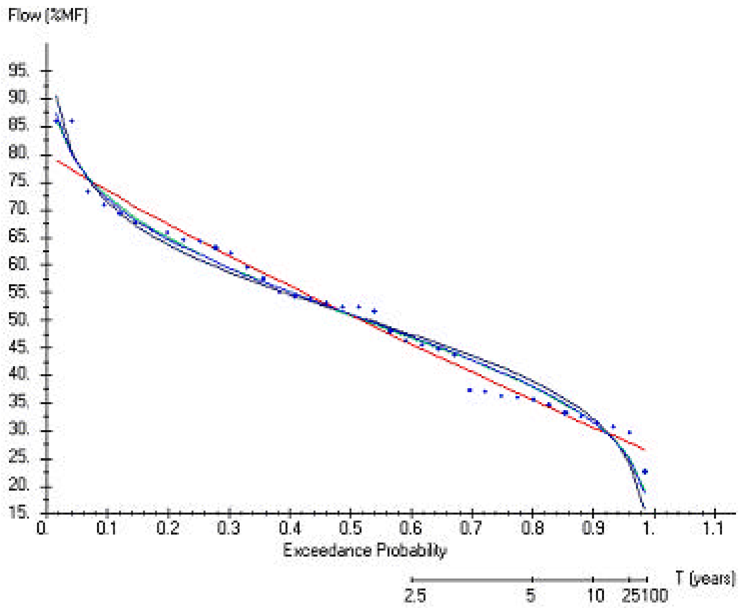


Figure 5.14: Observed and modelled curves for the the Tweed (21006) in the case where D=180 (GEV-green, GP-red, GL-black, PE3-blue)

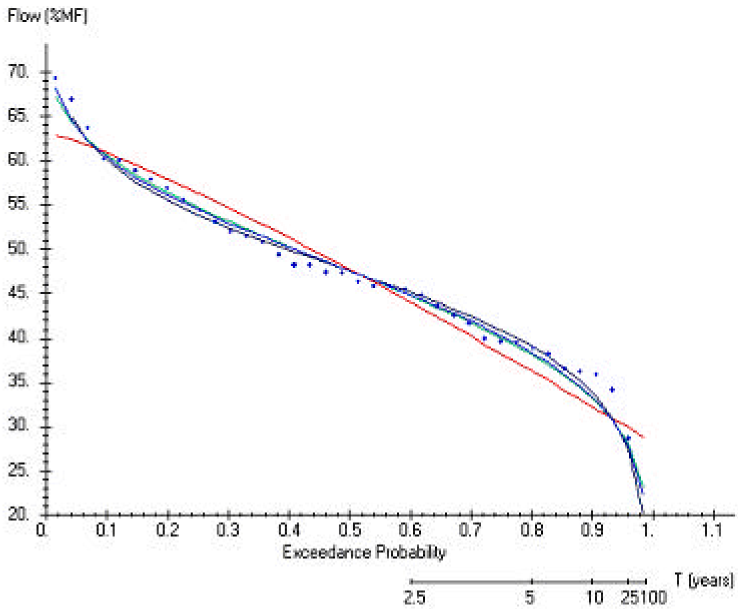


Figure 5.15: Observed and modelled curves for the Kennet (39016) in the case where D=30 (GEV-green, GP-red, GL-black, PE3-blue)

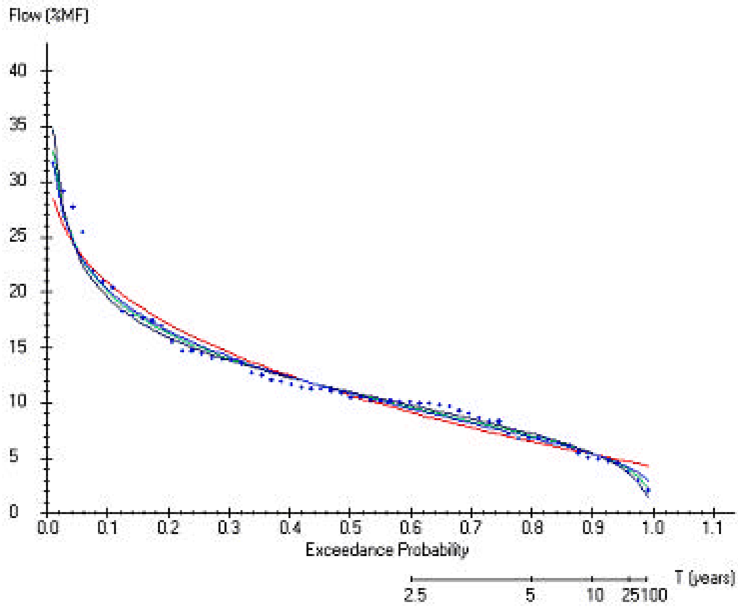


Figure 5.16: Observed and modelled curves for the Wye (55026) in the case where D=30 (GEV-green, GP-red, GL-black, PE3-blue)

5.4 Evaluation of the p.d.f. Models

5.4.1 Tests to assess the fit of the frequency curve

The parameterised distributions were evaluated in terms of how well they modelled the observed frequency curves using goodness-of-fit tests, and by examining the root-mean-square errors (RMSE) between the observed and fitted frequency curves.

Tests of goodness of fit tests can be used to verify whether the match between the modelled and observed frequency curves is satisfactory. These tests work by determining the extent by which the distribution function (p.d.f.) of the modelled curve represents the observations. Two goodness of fit tests were applied to the data: the Chi-square (c^2) test and the Kolmogorov Smirnov test. The c^2 goodness of fit index is defined as follows:

$$c^2 = \sum_{i=1}^{N} \frac{(o_i - e_i)^2}{e_i} \qquad (5.1)$$

where o_i is the observed frequency and e_i is the expected frequency.

The Kolmogorov-Smirnov goodness of fit index is given by

$$D_N = \max_i \left(\left| O_i - E_i \right| \right) \tag{5.2}$$

where O_i is the observed cumulative frequency and E_i is the expected cumulative frequency.

A disadvantage of goodness-of-fit tests is that they are dependent on the plotting positions used and, theoretically, the p.d.f. should be hypothetical rather than being estimated from the sample (Ashkar & Bobée, 1991). As a second verification of the match of the two curves, the root mean square errors between observed and modelled curves were also derived. The following formula was used:

$$RMSE = \sqrt{\frac{(x_p - x_o)^2}{N}} \tag{5.3}$$

where x_p is the predicted value, x_o is the observed value and N is the number of data in the examined sample.

The χ^2, Kolmogorov-Smirnov Values and the RMSE errors were derived for each model produced. Critical values of the test statistics were taken from Neave (1988) based on a confidence level of 95%.

5.4.2 Results of the goodness of fit tests

The results of the Kolmogorov-Smirnov Test were not useful for differentiating between distributions, as they indicated a satisfactory fit between all modelled and observed curves. Instead the size of the χ^2 value was used to rank the four distributions in order of goodness of fit. Again few distributions could be rejected outright based on the χ^2 value. This is due to the sample sizes (which are relatively small) – at small sample sizes these kinds of test are usually not powerful enough to discriminate between distributions. The results of the Chi-Square Test applied at durations of D=1, 7, 30, 60, 90, 180 and 365 for all 25 stations are also given in Appendix 6.2. Table 5.1 summarises the results of the Chi-Square test, indicating the distribution with the lowest χ^2 values for durations of D=7 and D=90, the distribution that most frequently has the lowest χ^2 when all durations are considered and the distribution that most frequency has the lowest χ^2 when all durations lower than D=90 are considered. The results shown for each catchment in Table 5.1 are translated to their geographic locations in Figures 5.17a to d.

Table 5.1: Summary of Chi-Square Test

Station	Distribution with lowest ?^{2} value			
	D=7	**D=90**	**Most frequent (Overall)**	**Most frequent (D=90)**
9001	PE3	GEV	PE3	PE3
9002	GP	GEV	GEV/GP	GEV/PE3/GP
14001	GEV	PE3	PE3	PE3
19002	GEV	PE3	GEV	GEV
19004	GEV	PE3	GL/GEV/GP	GEV/GL
20001	PE3	GP	PE3	PE3
20003	GEV	GP	GEV	GEV
20005	GP	PE3	GEV	GEV/PE3
21006	GL	GL	GL	GL
21012	PE3	GPA	GP	GP/PE3
21013	GEV	PE3	GEV	GEV
21015	PE3	PE3	GEV/PE3/GP	GEV/PE3
21017	PE3	PE3	GPA	PE3
28031	PE3	GP	GEV/PE3	PE3
34003	GEV	GEV	GEV	GEV
39016	GL	GL	GL	GL
39028	GEV	GEV	GEV	GEV
43005	GL	GL	GL	GL
43006	GL	GL	GL	GL
48010	GEV	GL	GL	GL
51001	GEV	GL	GEV	GEV/GL
55016	PE3	GP	PE3	PE3
55026	GL	GEV	GEV	GEV
60002	PE3	GEV	GEV	GEV/PE3
72004	GP	GEV	GP	GP

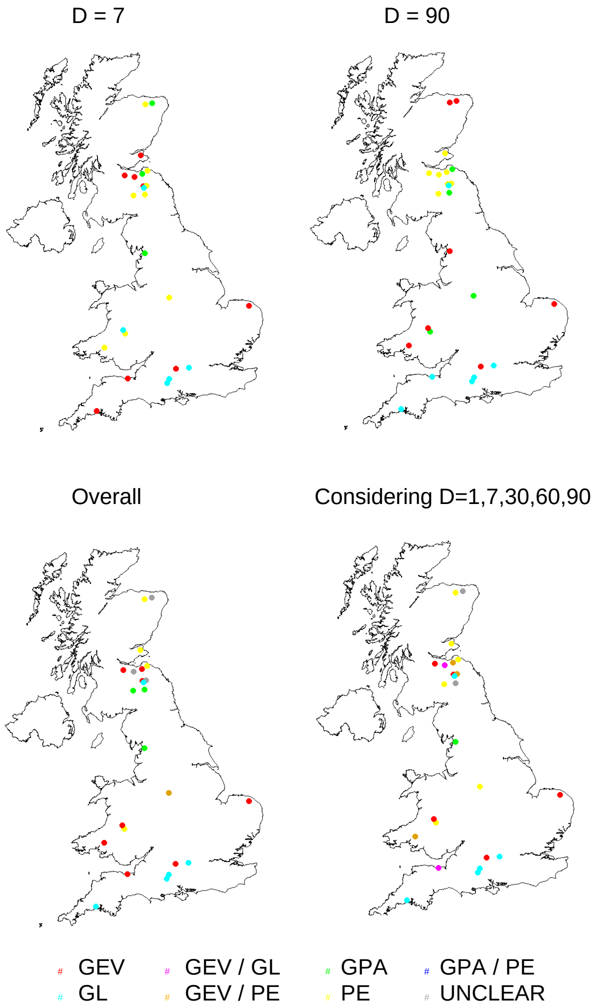


Figure 5.17: Distributions with lowest c^{2} values under different criteria

The results of the Chi-Square Test are quite complex. Based on the critical ?^{2} for a sample of given size, all four parameterised distributions nearly always model the observed data adequately giving an acceptable goodness of fit. Using the size of ?^{2} value to discriminate between distributions (i.e. that with the lowest value is considered as the one having the 'best fit') produces different results when different durations are considered. However several patterns are evident. Firstly, there seems to be a distinct difference in the 'best distribution' for durations of 90 days or less and those of 180/365 days. For example, for short durations the PE3 or GEV distributions tend to be favoured, whilst the GP distribution is often better when longer durations are considered. Secondly for permeable catchments (e.g. 39016, 43005 etc) the GL distribution is applicable across a wide range of durations probably as a consequence of the much lower variation in minima between different durations.

5.4.3 Results of the RMSE analysis

Root mean square errors between the modelled and the observed data points were calculated for each of the 25 stations for durations of D=1, 7, 30, 60, 90, 180 and 365 days. The full results are given in Appendix 6.2. However Table 5.2 summarises the RMSE errors for the 25 catchments, indicating the distribution with the lowest RMSE values for durations of D=7 and D=90, the distribution that most frequently has the lowest RMSE (when all durations are considered) and the distribution that most frequently has the lowest RMSE when only the results for durations lower than D=90 are considered. The results shown for each catchment in Table 5.2 are translated to their geographic locations in Figure 5.18.

Table 5.2: Summary of RMSE results

Station	Distribution with lowest RMSE value			
	D=7	D=90	Most frequent (Overall)	Most frequent (D=90)
9001	PE3	GEV	GEV/PE3	GEV/PE3
9002	PE3	GEV	GEV/PE3	PE3
14001	GEV	GEV/PE3	GEV	GEV
19002	PE3	GEV	GEV/PE	GEV/PE3
19004	PE3	GEV	GL	GL
20001	GEV/PE3	GP	PE3	PE3
20003	GEV/PE3	GP/PE3	PE3	PE3
20005	GP	PE3	GPA	PE3
21006	GEV/PE3	GL	GEV/GL	GL
21012	PE33	GL	GP	PE3
21013	GEV/PE	GP	GP	GEV
21015	GP/GL	PE3	GP	PE3
21017	GEV/PE3	GEV	PE3/GEV	PE3/GEV
28031	GEV/PE3	GP	GEV	GEV/PE3
34003	GEV	GEV	GEV	GEV
39016	GL	GL	GL	GL
39028	GEV/PE3	GEV/PE3	GEV/PE3	PE3
43005	GL	GL	GL	GL
43006	GL	GL	GL	GL
48010	GEV/GL	GL	GL	GL:
51001	GEV	GL	GEV	GEV
55016	PE3	GP	PE3	PE3
55026	GEV	GEV/PE3	GEV/PE	GEV/PE3
60002	GEV/PE3	GEV	GEV	PE3
72004	GP	GP	GP/PE3	GP

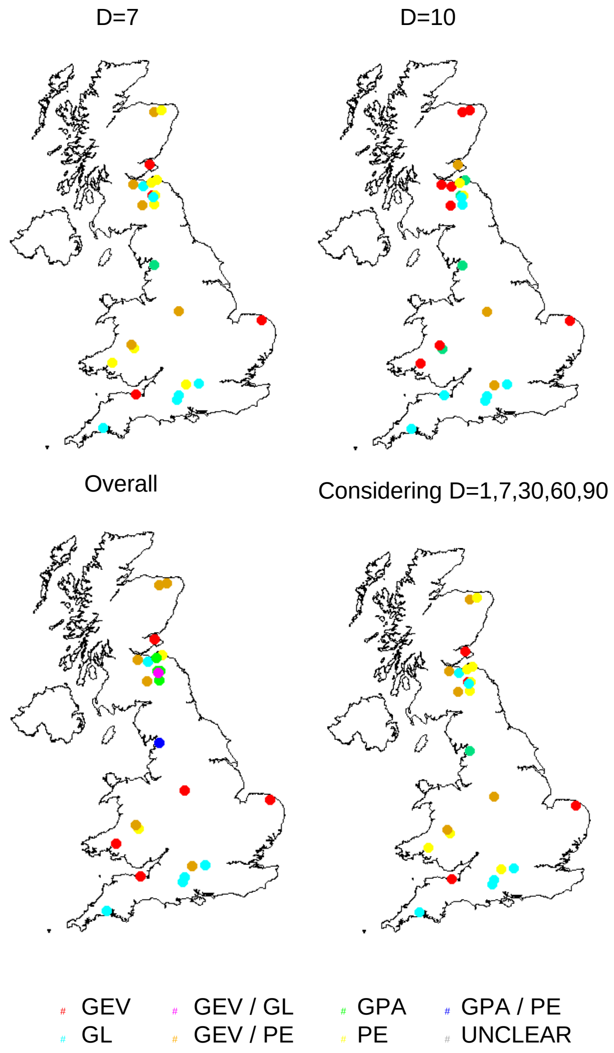


Figure 5.18: 'Best' distributions indicated by RMSE values under different criteria

As observed with the results of the ?[2] test, for a given catchment, there is little to differentiate between the performance of the four distributions. All give similar values for RMSE, with the exception of the GP models, which tends to produce larger errors. Using the size of the RMSE to discriminate between distributions (i.e. that with the lowest value is considered as the one having the 'best fit') as shown in Table 5.2 indicates that different distributions are favoured for different catchments and for different durations. The PE3 performs better at short durations, but tends to be replaced by the GEV as longer durations are considered. Permeable catchments, such as 39016 and 43005 located in southern England, which have a damped flow signature for both short and long durations tend to be best represented by the GL distribution, whilst, with a couple of exceptions, the GP distribution gives the biggest errors, particularly at short durations. These trends are discussed further in Section 5.6.

As well as calculating a single RMSE value for the whole sample, it is possible to determine the RMSE for observations occurring in different portions of the curve. RMSE values were also therefore determined for different portions of the curve. Figure 5.19 shows how the RMSE error varies for different probability intervals for stations 9002, 21006 and 39016, with the model fitting based on the GEV distribution. Figure 5.19 illustrates that in general, observations in the probability range $0.3 \le P \le 0.7$ will be fitted better than those outside this range, i.e. the modelled curves do not represent the extremes of the observed frequency distribution very well. This is a common problem in modelling (i.e. identifying a model that can explain the fill variance of the observed data set is difficult).

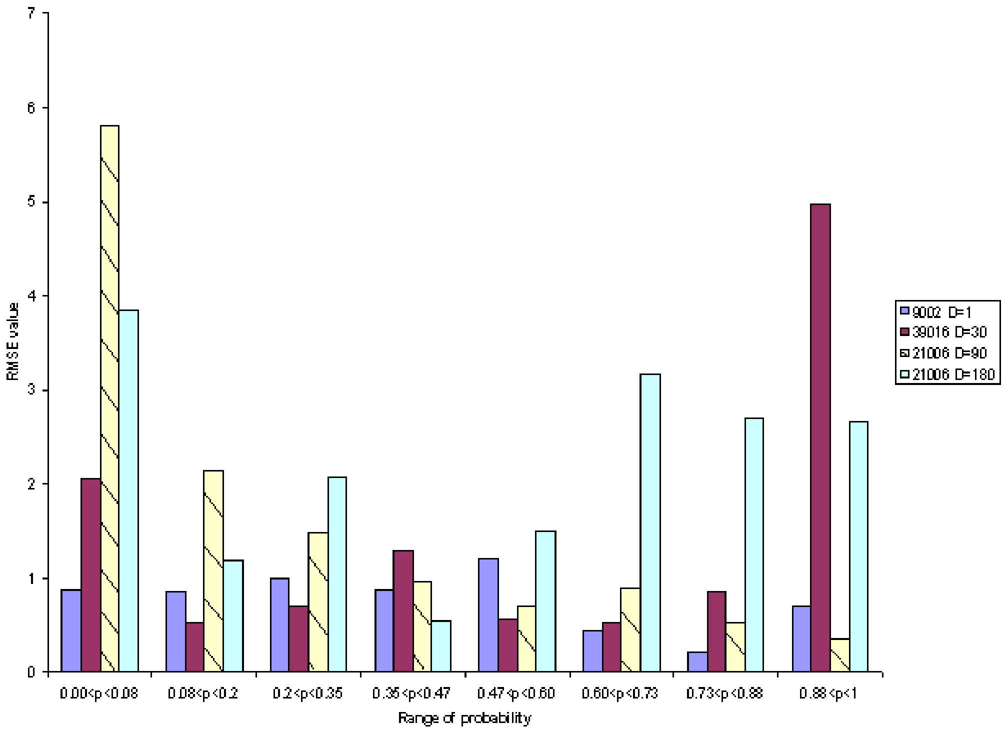


Figure 5.19: Variation in RMSE for different probability intervals

5.4.4 Selection of the 'best' distribution

The previous two subsections described how the size of the χ^2 value and the RMSE were used to discriminate between distributions, in each case the distribution with the smallest χ^2 or RMSE value being chosen. Unfortunately the 'best distribution' identified by the χ^2 and RMSE tests are not always identical. For instance Table 5.3 shows the favoured distribution according to the two tests for five flow records. In each case the assessment is based on average χ^2 and RMSE values (see Appendix 5.2) for durations of 1, 7, 30, 60, 90,180 and 365 days.

Table 5.3: 'Best' distributions as suggested by c^2 and RMSE tests

Station	**?** ^{2}	RMSE
9002	GEV	PE3
55026	GEV	GEV
72004	PE3	PE3
39016	GL	PE/GL
21006	GEV	GEV

Table 5.3 shows that, for 9002, the GEV is the best distribution according to the χ^2 values, whilst the RMSE test indicates that the PE3 is the better. A similar situation occurs for 39016. Figures 5.20 and 5.21 illustrate how the χ^2 and RMSE errors vary according to duration, for 9002 and 39016 respectively.

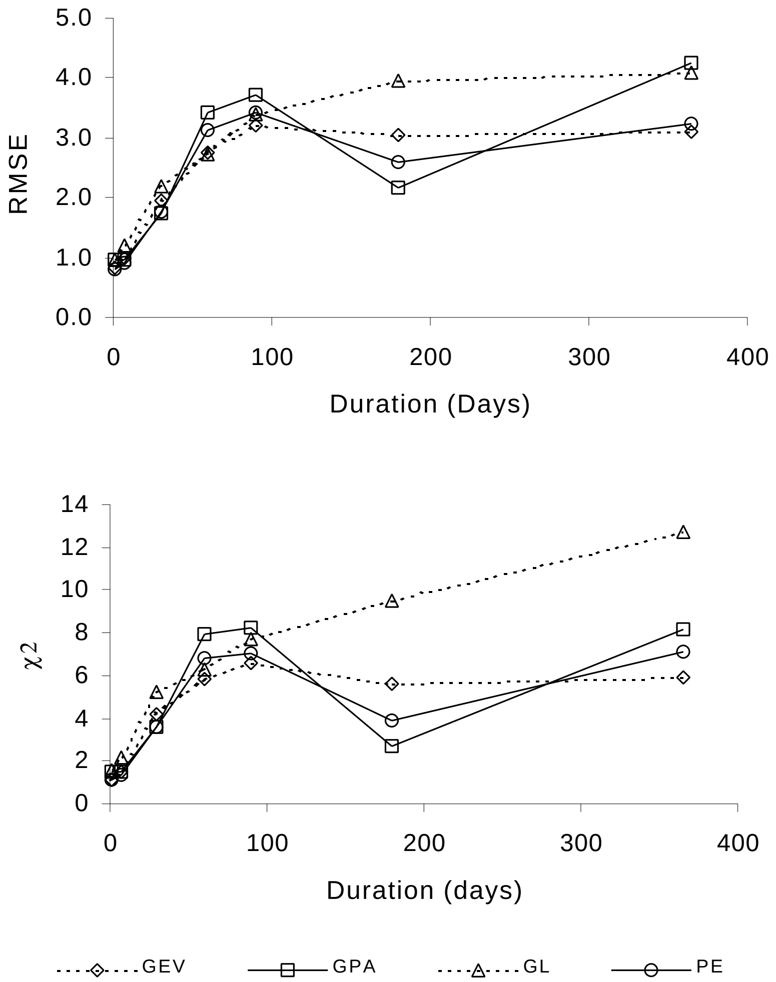


Figure 5.20: Variation in a) RMSE and b) c^2 values with duration for 9002

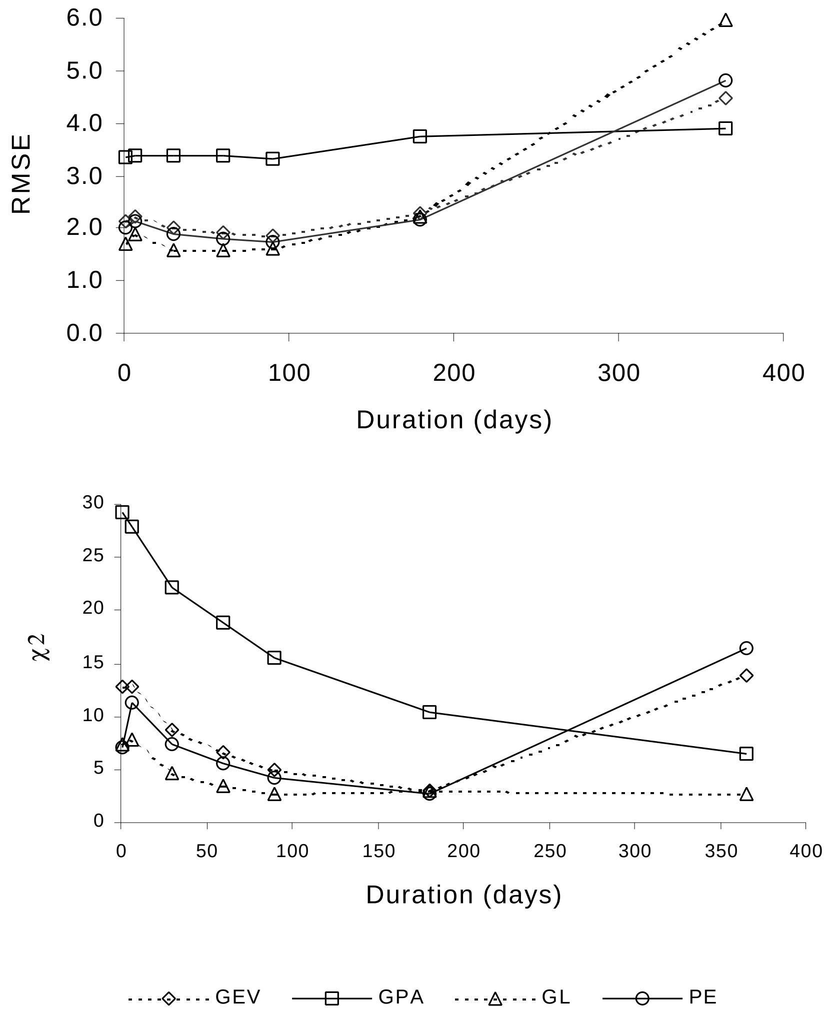


Figure 5.21: Variation in a) RMSE and b) c^2 values for gauging station 39016

The best distribution was based on assigning a rank to each of the distributions based on its c^2 value (i.e 1 for the lowest c^2 value and 4 for the highest c^2 value). The ranking exercise was repeated based on the RMSE data. In each case the distribution with the lowest rank sum was considered the 'best'. Table 5.4 summarises the 'best' distributions for the 25 stations, when both values are taken into account for durations of 7, 90 and 365 days.

Table 5.4: Summary of distribution types resulting in CDF models with lowest errors for AMS based on durations of 7, 90 and 365 days, considering both RMSE and c^2 test results

Station	D=7	D=90	D=365
9001	PE3	GEV	GEV
9002	GP & PE3	GEV	GEV
14001	GEV	PE3	GEV
19002	GEV & PE3	GEV & PE3	GEV
19004	GEV & PE3	GEV & PE3	GP
20001	PE3	GP	GP & PE3
20003	GEV	GP	GEV
20005	GP & PE3	PE3	GEV & GP
21006	All Equal	GL	GP
21012	PE3	GP & GL	
21013	GEV	GA & PE3	GEV
21015	All equal	PE3	GP
21017	PE3	GEV & PE3	GP
28031	PE3	GP	GEV
34003	GEV	GEV	GP
39016	GL	GL	GP
39028	GEV	GEV	GP & PE3
43005	GL	GL	GEV
43006	GL	GL	GEV & PE3
48010	GEV	GEV	GPA
51001	GEV	GL	PE3
55016	PE3	GP	PE3
55026	GEV	GEV	PE3
60002	PE3	GEV	GP
72004	GP	GEV & GP	GL

Table 5.4 indicates that, for a given station, different distributions produced the best results when different durations were considered. For many catchments the PE3 or GEV were favoured for short durations, whilst the GP was often the better model where longer durations, such as 365 days, were considered. These results agree with the *a-priori* expectation that the GP would not perform well (at short durations the GP distribution only produced the 'best' results for the Lune at Caton (72004)). The GP performs better when longer durations are considered because those series are less likely to have extreme value-like properties. This does not necessarily suggest that the GP distribution should be applied preferentially when minima of interest are of longer durations, rather, it can be said that the GEV, GL and PE3 are less suitable for describing long duration minima than they are for short records.

As four parameterised distributions nearly always model the observed data adequately, there is no clear indication that a single distribution forms the best model of low flow behaviour. However Figures 5.17 to 5.20, together with Tables 5.1 to 5.4, reveal some broad trends related to catchment properties (see Table 3.1 or Appendix 2 for details of catchment characteristics). For the subset of permeable catchments included in the study, the GL distribution gave the 'best' models across all durations (note however that it is important to recognise that only five permeable lowland catchments were included in the study). For the remaining low storage catchments, the PE3 or GEV distributions gave the best results particularly at short durations. This relationship between 'best distribution' and catchment characteristics is explored further in section 5.6. The Institute of Hydrology (1980) investigated relationships between low flow frequency curve form and catchment characteristics. This subsequently formed the basis of the regionalisation method published by Institute of Hydrology, (1980) and Gustard *et al*. (1992).

5.5 The Flow-Return Period Relationship

5.5.1 Prescriptive ability of the modelled curves

In addition to the goodness of the curve fit, the behaviour of the flow-return period relationship was also considered when evaluating the performance of each distribution family. A modelled curve may fit the observed data well, but it may be unsatisfactory in terms of the uncertainty of the flow estimates it prescribes for high return periods i.e. it may not be robust. This section examines the errors associated with quantile estimates at high return periods and errors in return period when extreme low flow quantiles are considered. To do this the estimated distribution parameters were used to calculate F(q) and q(F) and hence develop flow-return period curves for each series of annual minima.

Figure 5.22 illustrates the flow-return period relationships for 9002, for the duration of D=1, derived using fitted parameters for the GEV, GP and PE3 distributions (that for the GL is not shown but closely follows that of the GEV). Similarly Figure 5.23 illustrates the flow return period relationships for 39016, for the duration of D=30, derived using fitted parameters for the GEV, GP and PE 3 distributions. Figures 5.22 and 5.23 represent return period on a linear scale as the behaviour at high return periods (> 30 years) is of interest.

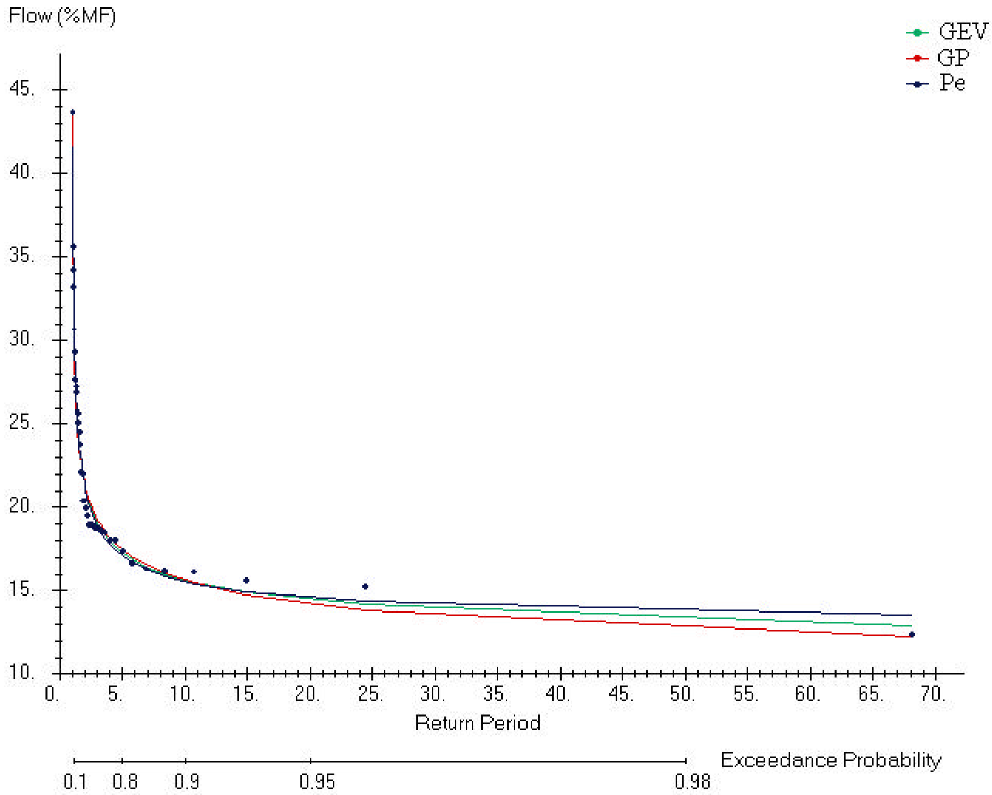


Figure 5.22: Flow – return period relationship for Station 9002 where D=1

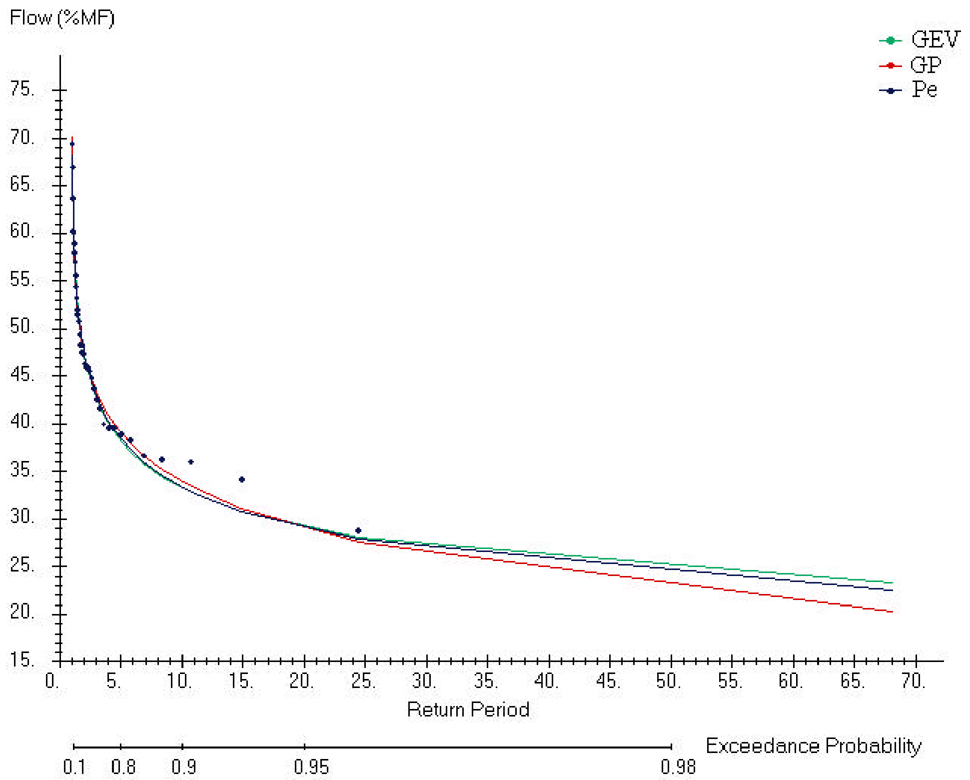


Figure 5.23: Flow – return period relationship for 39016 where D=30

There are two main reasons why estimates for high return periods may be poor. Firstly there is greater uncertainty at the extremes of the frequency curve. The level of uncertainty depends upon the number and the values of the sample minima for extreme years. Figures 5.22 and 5.23 are particularly good illustrations of this point: as both have the same number of data points (38), the lowest ranking flow in each case is assigned the same plotting position and will have the same return period (in this case 68 years). The size of this single flow value strongly influences the shape and gradient of the curve at high return periods (25 to 70 years range). Note that this effect should be minimised if unbiased plotting positions have been used (i.e. the resulting quartile estimates will have smaller error) – for further explanation the reader should refer back to Chapter 2). Secondly the differing characteristics of the different distribution families means that although the forms of the corresponding Q_T -T relationships are very similar around the median they differ at the tails of the distribution (e.g. at high return periods). Again this is illustrated well in Figures 5.22 and 5.23. These differences may be magnified when the relationship is extrapolated beyond the observed range.

5.5.2 Uncertainty in quantile estimates

As the 25 flow records considered in the study have a range of different lengths and the observed range of return period is different for each. Table 5.5 gives details of the range of flow (Q_T), probability (p) and return period (T) values for each station.

Table 5.5: Range of return periods values observed for each station

Station	Maximum Observed Values			Minimum Observed Values		
	T	*p*	Q_T	T	*p*	Q_T
9001	71.64	0.986	14.95	1.01	0.014	51.33
9002	68.07	0.985	12.37	1.01	0.015	43.71
14001	57.36	0.983	14.73	1.02	0.017	37.72
19002	68.07	0.985	7.46	1.01	0.015	28.76
19004	71.64	0.986	6.52	1.01	0.014	39.74
20001	68.07	0.985	11.98	1.01	0.015	42.85
20003	62.71	0.984	10.15	1.02	0.016	39.89
20005	62.71	0.984	7.41	1.02	0.016	30.68
21006	68.07	0.985	9.35	1.01	0.015	29.03
21012	64.50	0.984	5.00	1.02	0.016	18.97
21013	62.71	0.984	6.59	1.02	0.016	30.74
21015	59.14	0.983	7.25	1.02	0.017	30.76
21017	60.93	0.984	3.70	1.02	0.016	16.92
28031	55.57	0.982	8.97	1.02	0.018	36.44
34003	71.64	0.986	34.91	1.01	0.014	77.17
39016	68.07	0.985	9.82	1.01	0.015	61.02
39028	55.57	0.982	26.72	1.02	0.018	59.06
43005	60.93	0.984	5.00	1.02	0.016	49.66
43006	60.93	0.984	17.28	1.02	0.016	50.44
48010	53.79	0.981	11.88	1.02	0.019	39.61
51001	55.57	0.982	7.51	1.02	0.018	33.89

Station	Maximum Observed Values			Minimum Observed Values		
	T	*p*	Q_T	T	*p*	Q_T
55016	53.79	0.981	0.49	1.02	0.019	13.24
55026	109.14	0.991	1.19	1.01	0.009	16.84
60002	68.07	0.985	1.90	1.01	0.015	18.44
72004	71.64	0.986	3.27	1.01	0.014	14.39

A shown by Table 5.5 the maximum observed return periods range from T=53 to T=109. Similarly the flow values (as %MF) of the observed values of lowest probability of non-exceedance range from as little as 13% to 77%. These differences illustrate that some models will be more prescriptive in the extrapolated upper tail of the probability distribution, whilst others will perform better in the extrapolated lower tail of the distribution.

The flow–return period relationships in the range 1=T=200 were derived for each of the 25 catchments. Those where D= 1 are illustrated in Figure 5.24.

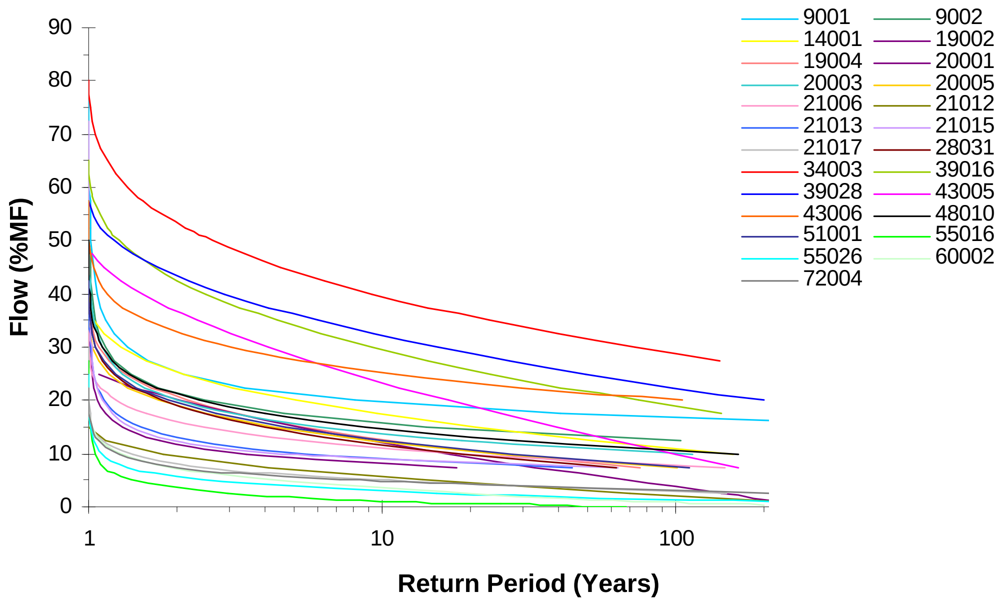


Figure 5.24: Flow-retun period relationships for D=1.

Table 5.6: Flows (expressed as %MF) corresponding to a 50-year event

Station	D=1	D=7	D=30	D=60	D=90	D=180	D=365
			Generalised Extreme Value				
9002	13.26	13.43	15.03	15.73	16.08	22.20	27.59
21006	8.71	9.18	11.73	13.22	14.76	21.05	41.71
39016	21.54	23.07	24.57	26.05	26.95	29.60	28.15
55026	2.15	2.12	3.08	4.85	6.99	17.22	53.49
72004	3.30	3.65	3.36	4.24	8.31	15.64	45.58
			Generalised Logistic				
9002	12.71	12.79	14.35	14.92	14.98	20.15	22.48
21006	8.03	8.35	11.03	12.25	13.38	18.38	38.59
39016	N/A	21.00	22.45	23.70	24.46	26.50	22.53
55026	1.41	1.52	2.53	3.84	5.34	14.77	50.86
72004	2.99	3.29	2.80	3.18	6.69	20.37	41.94
			Pearson – Type III				
9002	13.766	13.979	16.077	17.51	18.325	23.748	26.135
21006	8.664	9.102	12.323	14.167	15.228	20.578	41.292
39016	20.951	22.415	23.877	25.282	26.317	28.848	26.553
55026	1.89	2.073	3.477	5.664	7.28	16.855	53.047
72004	3.394	3.86	4.184	5.27	9.128	4.365	44.576

Table 5.7: Flows (expressed as %MF) corresponding to a 100-year event

Station	D=1	D=7	D=30	D=60	D=90	D=180	D=365
			Generalised Extreme Value				
9002	12.55	12.62	14.12	14.59	14.55	19.63	19.30
21006	7.92	8.22	10.83	12.01	13.11	17.56	33.24
39016	18.77	20.33	21.57	22.82	23.59	25.37	19.02
55026	1.89	1.75	2.37	3.62	4.99	14.14	50.25
72004	2.89	3.21	2.50	2.84	6.32	11.10	39.70
			Generalised Logistic				
9002	11.77	11.71	13.21	13.53	13.12	16.67	8.16
21006	6.82	6.85	9.80	10.67	11.05	12.85	23.69
39016	N/A	16.41	17.46	18.33	18.91	19.58	6.59
55026	0.92	0.92	1.58	2.19	2.41	9.93	45.77
72004	2.42	2.69	1.70	1.32	3.92	14.44	31.78
			Pearson Type III				
9002	13.358	13.48	15.7	17.149	17.832	22.117	16.566
21006	7.879	8.106	11.753	13.512	13.954	16.823	32.151
39016	17.771	19.256	20.428	21.571	22.498	24.075	16.071
55026	1.66	1.773	3.032	4.919	5.523	13.596	49.617
72004	3.062	3.57	3.754	4.463	7.657	9.588	37.808

Tables 5.6 and 5.7 give the flows corresponding to the 50-year and 100-year low flow events for the five catchments, over a range of annual minima durations. The values are derived from the p.d.f. for each of the modelled curves, applying Eqns 2.3 and 2.4 to

determine the Q_T-F(q) and hence the Q_T-T relationships (equally, they could have been read off the appropriate flow-return period plot, such as that shown in Figure 5.24 for D=1). Note that for stations 9002, 21006, 39016 and 72004, which all have 38 years of data, the maxmium return period for observed annual minima is about 68 to 70 years. For these stations the 50-year event should be predicted fairly accurately as it is within the observed range. In contrast the 100-year event is beyond the observed range for these records, but can be calculated using the modelled p.d.f.s. The accuracy of the predicted quantile for the 100-year event depends on the degree of robustness of the modelled p.d.f.. For station 55026, the lowest rank value occurs at a return period of about 105 years, and therefore both the 50 and the 100-year events are within the observed range. The p.d.f.s should be more constrained to the observed data at this range, and the quantile estimates for the four distributions should be comparable.

For the 50-year event, the quantile estimates vary by less that 1% MF at D=1 to about 4%MF at D=365. This variability is probably acceptable to the end-user. For the 100-year event the quantiles predicted using the different p.d.f. forms for 55026 vary from about 1%MF at D=1 to about 4% MF at D=365. In constrast the quantile estimates for the remaining four catchments vary by 2%MF at D=1 to about 10% of the MF at D=365. However the size of the error does vary depending on the catchment. This suggests that particular catchment – p.d.f. combinations are more likely to produce more robust quantile estimates at higher return periods than others. However, in the absence of a ‘bench mark’ observed flow with which to compare these quantile estimates, it is difficult to discriminate between the performance of different distributions in this respect.

5.5.3 Uncertainty in return period estimates

In the same way as it is important to obtain robust quantile estimates for a particular return period, it is also important to assess the error asssociated with return periods for prescribed flows. Table 5.8a details return periods for a prescribed flow level of 15%MF, estimated by fitting a GEV distribution. Table 5.8b gives the absolute error between return period estimates derived from fitting GEV and GL distributions respectively, whilst Table 5.8c compares the estimates for the GEV and Pearson Type-III distributions. The values given are the difference, in years, between the return period estimates for the stated distributions, given a flow equal to 15% of the mean flow.

Table 5.8 indicates that in terms of quantile estimates, and for short durations, the three distributions give similar results. Differences between distributions are largest when longer durations are considered. Clearly, in addition to catchment type, the duration considered has some control over the return periods and their corresponding flows.

Table 5.8: Comparison of return periods.

Station	D=1	D=7	D=30	D=60	D=90	D=180
a) Estimates based on fitting a GEV distribution						
9002	13.64	15.93	34.99	48.96	48.68	144.27
21006	2.46	3.64	8.84	21.55	45.53	176.69
39016	98.85	124.06	139.92	151.90	162.63	157.81
55026	1.02	1.03	1.32	3.04	7.80	48.78
72004	1.03	1.06	1.68	3.42	10.01	39.02
b) Difference between GEV and GL distributions						
9002	-0.42	-1.83	-14.34	-28.25	21.65	-350.07
21006	-0.10	-0.32	-0.25	2.84	12.89	100.94
39016	-196.63	-330.97	-472.31	-573.15	-685.12	-622.37
55026	0.00	-0.02	-0.02	0.14	0.38	-34.17
72004	1.03	1.06	1.68	3.42	10.01	39.02
c) Difference between GEV and PE distributions						
9002	1.18	N/A	-77.21	-27.03	-494.34	-37.35
21006	-0.27	-0.85	8.04	23.51	69.17	233.99
39016	-180.06	-257.50	-340.70	-393.84	-337.98	343.90
55026	-0.03	-0.31	-0.10	-0.23	-5.76	N/A
72004	0.00	0.01	-0.14	-0.23	-5.42	-2100.78

In the example given in Table 5.8 an annual minima equivalent to 15%MF represents a more extreme event for catchment 39016 than for the others, and has a very long return period. In fact a 15%MF event is not within the observed range of annual minima values. This has resulted in the difference between the estimates for different distributions being much higher: running into 100's of years, even where D=1. In contrast, the estimates for 55026 at D=1 differ by less than a year. The difference between estimates is also dependent on the duration considered. The increase in the difference between estimates with duration occurs because a 15%MF event becomes progressively more extreme, and at some point falls beyond the observed range.

The duration at which a 15%MF event falls beyond the observed range depends on the catchment in question. It is unlikely that a good estimate of the return period for an annual minima equal to 15% of the mean flow could ever be obtained for permeable catchments, where daily flows hardly ever drop below 20% of the mean flow. In highly permeable catchments, with extremely flashy regimes, daily flows may drop to less than

5% of the mean flow. However for durations of 30 days or more, where the annual minima represents an averaged value such flows are unlikely to occur except during exceptionally prolonged droughts.

Table 5.9 indicates whether sensible estimates of return period can be obtained for annual minima flows equivalent to 10% (representing a fairly low annual minima) and 40% (representing quite a large annual minima) of the long-term mean flow.

Table 5.9: Sensible return period estimates, for flow level at 10%MF

	Flow = 10%							Flow = 40%						
	Duration, D (days)							Duration, D (days)						
	1	7	30	60	90	180	365	1	7	30	60	90	180	365
9001	X	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
9002	X	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
14001	**Y**	**Y**	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
19002	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**
19004	**Y**	**Y**	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
20001	**Y**	**Y**	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
20003	**Y**	**Y**	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
20005	**Y**	**Y**	X	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**
21006	**Y**	**Y**	**Y**	X	X	X	X	X	X	**Y**	**Y**	X	X	X
21012	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**
21013	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**
21015	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
21017	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**
28031	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**
34003	X	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	X	X
39016	X	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
39028	X	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
43005	**Y**	**Y**	**Y**	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
43006	X	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**
48010	**Y**	X	X	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	X
51001	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	X	**Y**	**Y**	**Y**	**Y**	**Y**
55016	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	X	**Y**	**Y**	**Y**	**Y**
55026	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	X	X	**Y**	**Y**	**Y**	X
60002	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	X	**Y**	**Y**	**Y**	X
72004	**Y**	**Y**	**Y**	**Y**	**Y**	**Y**	X	X	X	X	**Y**	**Y**	**Y**	X

Y indicates a sensible estimate is obtained

X indicates that a sensible estimate cannot be obtained

In general, sensible estimates are obtained for D=1 and D=7 where the prescribed flow is 10%MF, and for D=30 or more where the prescribed flow is 40% MF. The pattern is, of course, more complex than this as there is a strong influence with the catchment type. For example return period estimates for chalk catchments such as 39016 and 39028 are not robust at 10%MF, whilst for flashy catchments, such as 60002, return period estimates at 40% MF are not robust for short durations because all of the observed minima are much lower than 40% MF.

5.6 Relationship Between Probability Distribution & Catchment Caracteristics

5.6.1 Introduction

The results presented in Sections 5.4 and 5.5 indicated a geographical control on the observed form of the frequency curves and on the distribution family found to best match this curve. For example the results of the goodness-of-fit and error tests (Figures 5.17/5.18) indicated a geographical control on the form of the curve best describing the observed probability distribution. The two most likely geographical controls are the amount of rainfall, and the storage behaviour of the catchment, which usually depends on the catchment geology. Here the 1961-1990 standardised annual average rainfall (SAAR) is used to characterise the catchment rainfall, whilst catchment storage behaviour is represented numerically, by the baseflow index (BFI).

As mentioned previously, the Institute of Hydrology (1980) and Gustard *et al*. (1992) developed relationships between frequency behaviour and catchment characteristics for British catchments with the aim of providing a mechanism for predicting the frequency behaviour of rivers that are ungauged or that have short records unsuitable for frequency analysis. These studies used the MAM10 and MAM7 values respectively (under the assumption that a good estimate of MAM7 or MAM10 can be determined from most hydrometric records) as an indicator of frequency behaviour, given known catchment characteristics. The methodology of Gustard *et al*. (1992), based on flows given in units of %MF, applied the following:

1. A regression relationship between MAM7 and catchment characteristics
2. A regression-derived relationship between MAM7, rainfall (SAAR) and the rate of change of MAM(d)/MAM7 with D (the latter usually termed ‘GRADMAM’). This relationship, referred to as the ‘External Relationship’, is used to predict GRADMAM for any catchment provided MAM7 and SAAR values are known.
3. The theoretical linear relationship between MAM(D), MAM7 and GRADMAM, known as the ‘Internal Duration Relationship, which can be used to predict MAM(D) provided GRADMAM and MAM7 are known.
4. Flow Frequency Type Curves, based on categorising the frequency curves of British Rivers Given an estimate of MAM(D) these curves allow the Annual minima at a particular return period for duration D, AMP(D), to be estimated. This relationship is referred to as the ‘Internal Frequency Relationship’.

Here a slightly different approach is taken, in which catchment characteristics are linked directly to the p.d.f. form which best represents the observed data.

5.6.2 Observed curves

Figures 5.25 to 5.29 illustrate the frequency curves for the catchments 9002, 21006, 39016, 55026 and 72004 respectively. In each case a range of durations (D= 1, 7, 30 and 90) are shown, and the flows are standardised by the relevant MAM(D) value.

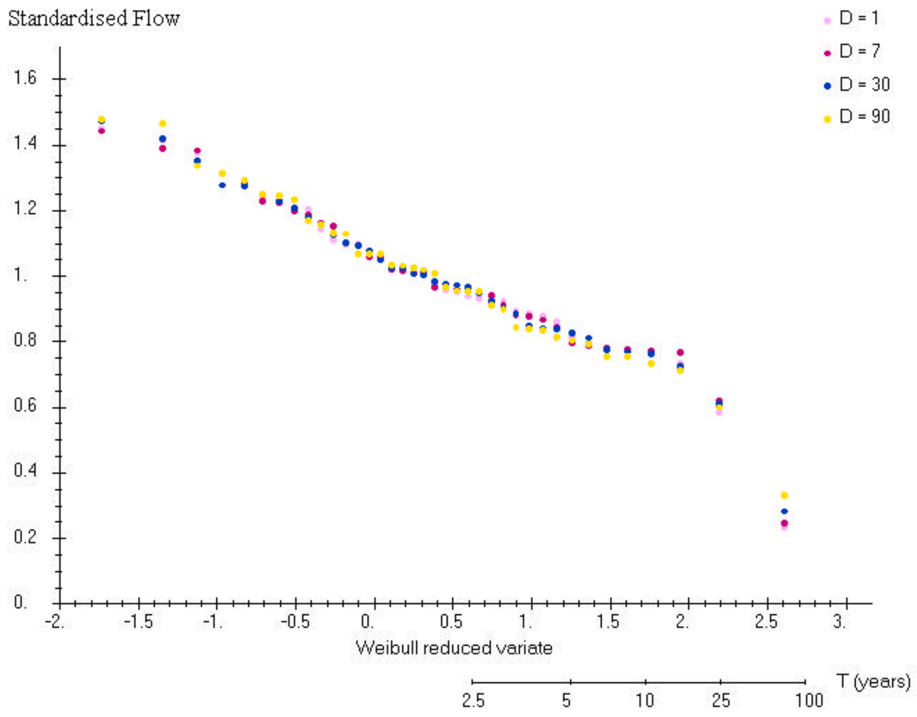


Figure 5.25: Frequency curve for 9002

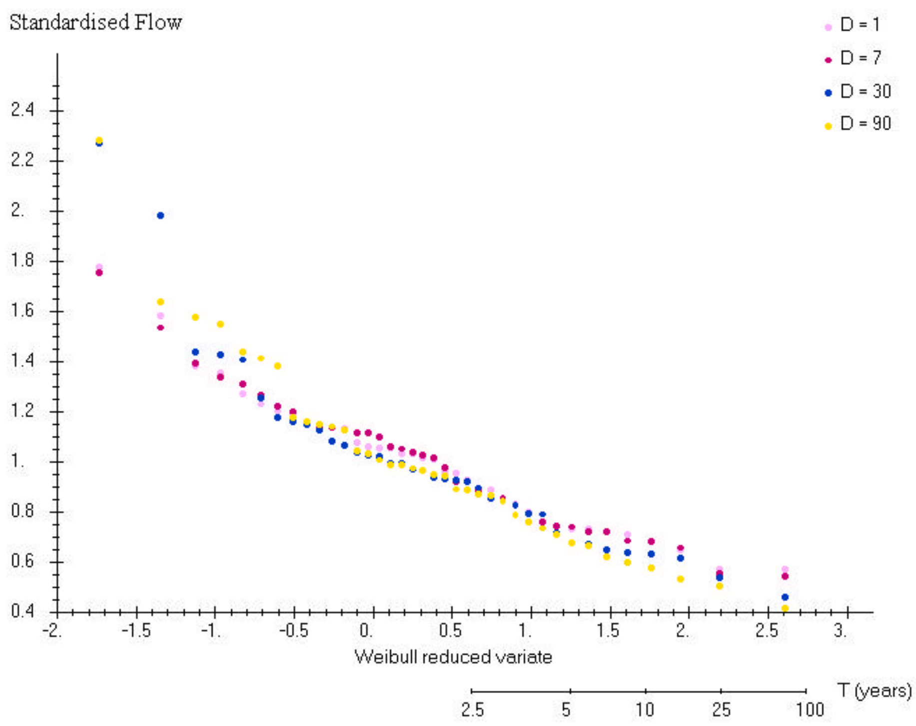


Figure 5.26: Frequency curve for 21006

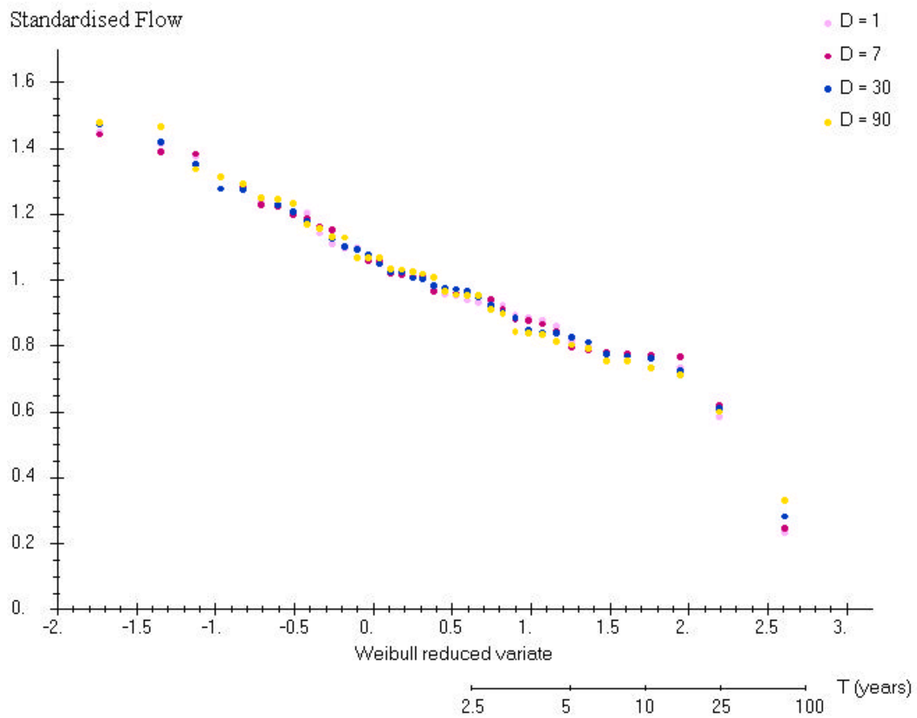


Figure 5.27: Frequency curve for 39016

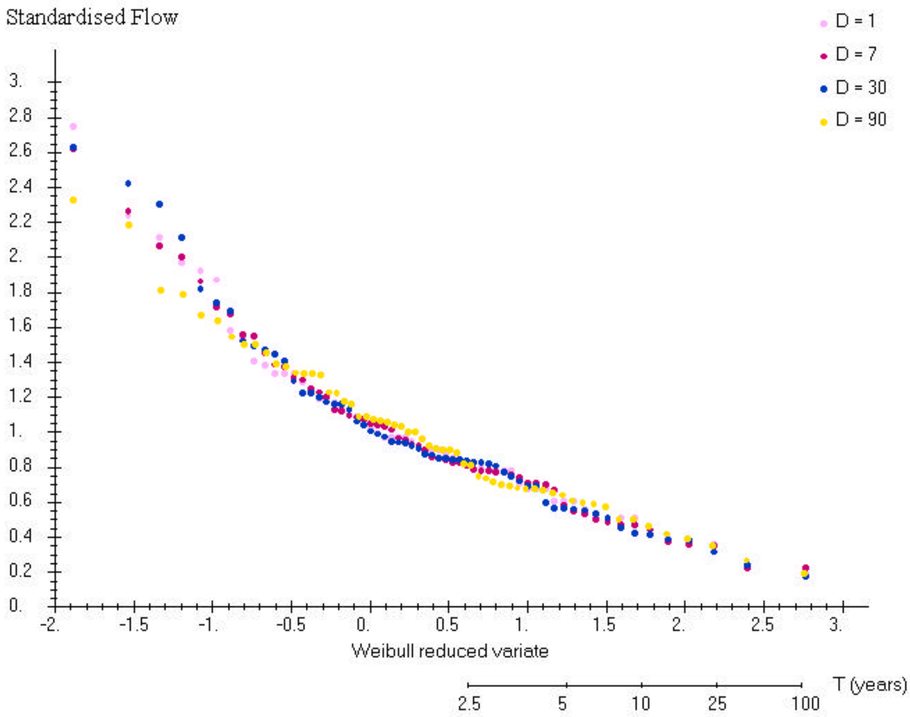


Figure 5.28: Frequency curve for 55026

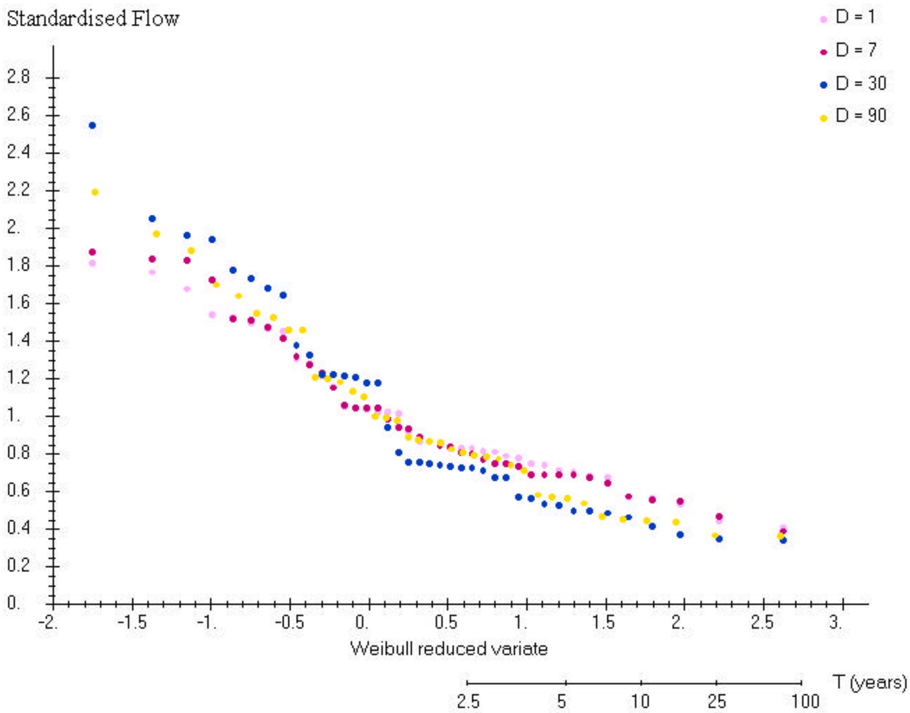


Figure 5.29: Frequency curve for 72004

Figures 5.25 to 5.29 illustrate a very strong relationship between the form of the probability curve and catchment type. The Kennet (39016), which has a BFI of 0.87 is very distinct in form, having a fairly steep gradient, and dropping off very suddenly at the end of the curve. There is good agreement between the normalised forms for the different durations shown, whilst values for the flow-MAM(D) ratio vary from 0.2 to 1.5 (i.e. the MAM(D) is quite big relative to the observed flows). The Deveron (9002), which is moderately permeable (BFI=0.59) has similar behaviour. Conversely the curves for the impermeable catchments (the Lune having a BFI of 0.32, the Wye having a BFI of 0.37, and the Tweed having a BFI of 0.52) are represented by curves that have fairly steep gradients, with the data points representing a different wider range of annual minima-MAM(D) ratios (for example the annual minima-MAM(D) ratio for the Lune at Caton varies between 0.4 and 2.4). This is probably a result of the flashy behaviour of these catchments (i.e. more extreme events are observed) which causes the distribution of annual minima to be skewed. This causes the MAM(D) to be small compared to range of the observed flows giving larger AM(D)/MAM(D) ratios, particularly where D is short. This also explains where there is also poorer agreement between the normalised curves for different durations than observed in the permeable catchments: the flashy nature of low BFI stations is averaged out more at longer durations so that extreme events have less influence on the MAM(D).

5.6.3 Relationship between 'best' distribution & catchment characteristics

The results of the goodness of fit and errors test (Figures 5.18/19) indicated a geographical or catchment characteristics control on the form of the curve best describing the observed probability distribution. The histograms shown in Figure 5.30 illustrate the variation in 'best' distribution with catchment storage. The best distribution is determined taking into account both RMSE and the Chi-Square value of each p.d.f. model. The storage is represented by the BFI, with permeable catchments having high BFI values. Figure 5.30 shows that when the 180 and 365 day curves are ignored the PE3 distribution is strongly favoured at low to mid BFI values, whereas there are fewer stations that can be described by the GP or GEV distributions. Figure 5.30 also shows that the GL distribution seems to perform most well for permeable catchments, whereas more responsive catchments seem to be best described using the PE3 or GEV distributions. Note that a bias is introduced because few highly permeable catchments were included in the study.

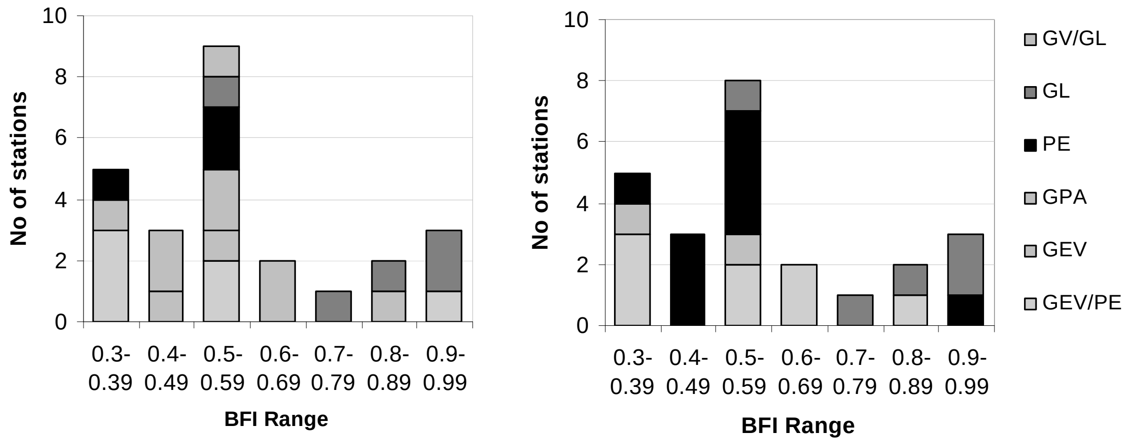


Figure 5.30: Distributions obtained when considering durations of a) 1, 7, 30, 60, 90, 180 and 365 days, b) 1, 7, 30, 60, 90 days only

The histograms shown in Figure 5.31 illustrate how the best distribution varies with catchment rainfall, represented by the 1961-90 SAAR . Figure 5.31 suggests that PE3 and GL distributions are favoured by catchments with low SAAR, whilst there is a trend towards the GEV distribution for catchments with high rainfall. When long durations are considered the GP distribution becomes more common at low SAAR.

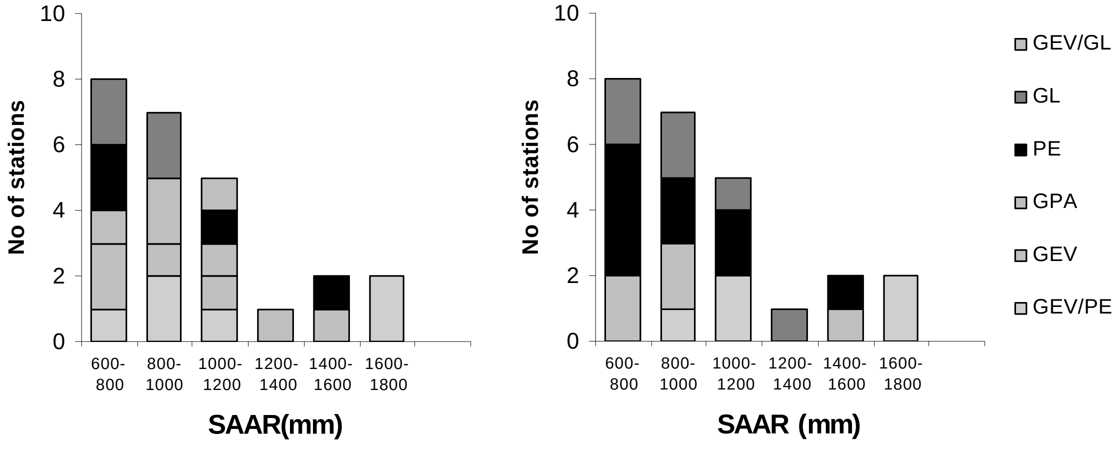


Figure 5.31: Distributions obtained when considering durations of a) 1, 7, 30, 60, 90, 180 and 365 days, b) 1, 7, 30, 60, 90 days only

5.6.4 Parameter values

If similar catchment types have similar low flow frequency responses they should also have similar parameter values (provided these are determined using flow data normalised by the MAM(D)). Thus it may be possible to determine type curves for parameter values, so that once a particular distribution family has been selected, suitable parameter values might be determined without having to use a formal method of parameter estimation such as L-Moments or Maximum Likelihood. Unfortunately such an investigation is beyond the scope of this study, but could be a subject for further work. However the next section 5.6.5 investigates how the form of a fitted p.d.f. varies according to the catchment properties.

5.6.5 The flow-return period relationship

In this section the behaviour of the flow-return period relationship for different catchments is investigated. The results are based on the p.d.f. derived by fitting a Generalised Extreme Value distribution to the observed data. The estimated distribution parameters were used to calculate F(Q) and Q(F) and hence develop flow-return period curves for each series of annual minima. In order to compare the curves for different regimes, the annual minima are standardised by the corresponding MAM(D) values. Figure 5.32 illustrates the flow return period relationships for all 25 catchments where D=30 (as normalising by MAM(D) removes the effect of duration the relationships for the other durations will be very similar to this are therefore not shown). The catchments are coloured according to their Base Flow Index value. Similarly Figure 5.33 illustrates the same flow-return period relationships with the catchments classified according to the average annual rainfall, SAAR.

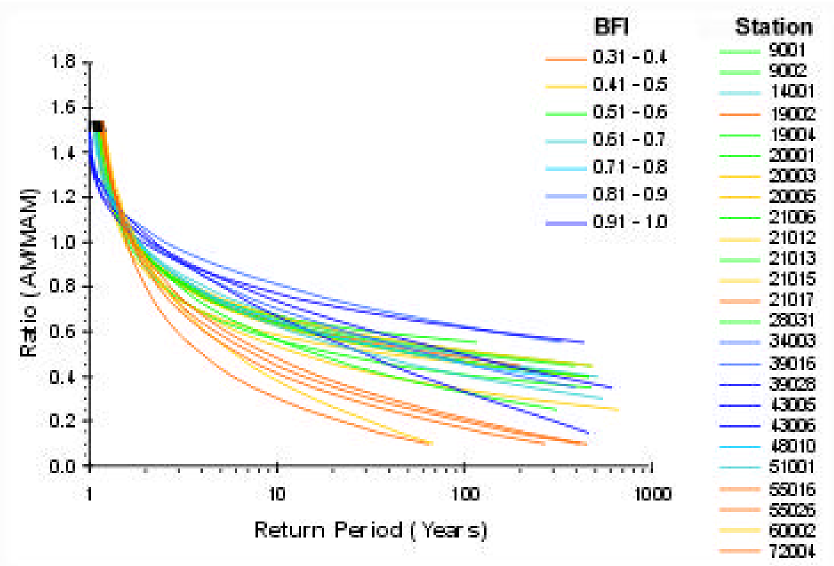


Figure 5.32: Flow-return period relationships classified by BFI

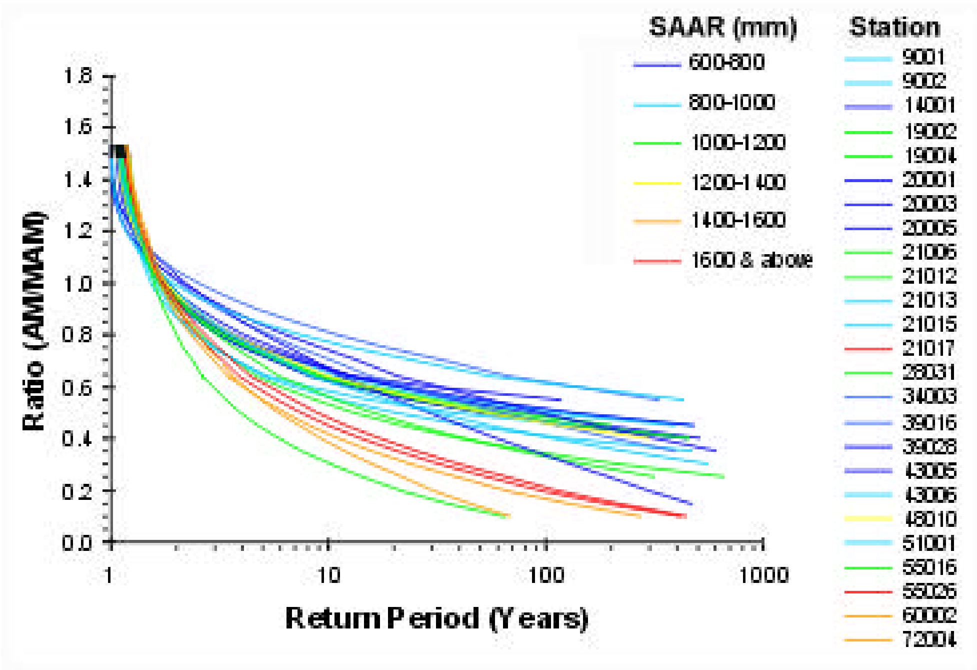


Figure 5.33: Flow – return period relationship classified by SAAR

Figures 5.32 and 5.33 illustrate that there are strong trends with in the form of the quantile-return period relationships. The forms may be classified into two groups: those for high rainfall, low storage catchments and those for high storage, low rainfall catchments. Without further work it is difficult to differentiate whether storage or rainfall is the dominant catchment characteristic due to the strong negative correlation between baseflow and SAAR for British catchments).

5.6.6 Summary

These initial investigations have shown some trends in the ‘best distribution’ with both BFI and SAAR (i.e. the PE3 and GEV are more suited to upland catchments, with the GL preforming better for lowland catchments). In the UK there is a strong negative correlation between BFI and SAAR as the highly permeable catchments are found mainly in the southeast where rainfall levels are comparitively low, whilst wet regions such as the north west are main dominated by impermeable geologies. As a result, without, further work, it is difficult to make any firm conclusions regarding which of these catchment characteristics has the dominant effect on distribution shape.

6. PROTOTYPE METHODOLOGY FOR SHORT RECORDS

6.1 Introduction

6.1.1 Potential methodologies for frequency analysis of Short Records

In an at-site analysis of low flow frequencies it is unrealistic to expect to accurately predict the probability distribution and flow – return period relationship when the number of data observation is restrained, i.e. the flow record is short. A series of annual minima derived from a short flow record has a number of features which increases the inherent uncertainty in any probability distribution derived from it including:

- The range of flow values obtained may be unrepresentative of the long-term low flow behaviour of the site. The range of observed flow values will be highly dependent on the period during which the observations were made (e.g. a 15 year record from 1970-1984) may contain more extreme annual minima than a record running from 1978-1991 etc.). This means that the distribution of annual minima is more likely to be biased or skewed as the full range of conditions characteristic of the site will not be sampled.
- The plotting position formulae, which are used to estimate the probability of each point in the annual minima series, produce highly biased results for short periods. This occurs because the plotting position formulae are not robust for small sample sizes (Cunnane, 1978), as they are based on a rank-order system. In other words, plotting positions are assigned regardless of the range or skew of the sampled flows resulting in similar low flow events values being assigned very different exceedance probabilities.
- Where flows of longer duration are considered, even fewer annual minima values will be generated due to the imposition of missing data criteria. This is especially problematic for discontinuous records with several short gaps.

Frequency estimation techniques have to be adapted to deal with such problems at sites where the flow record is short. Three main approaches may be considered; (i) generalised estimation where frequency relationships are estimated from catchment descriptors, (ii) estimation of flow frequency based on rainfall frequency, or (iii) the uncertainty of the derived frequency relationships can be lessened by incorporating low flow data from analogue sites.

i) Generalised estimation

In this approach key variables describing frequency behaviour at the subject site are estimated from catchment descriptors. This was the basis of the method presented by the Institute of Hydrology (1980) and later developed by Gustard *et al*. (1992), in which relationships between key variables MAM7 and GRADMAM, and catchment descriptor SAAR were established using regional regression techniques with data from 865 gauged sites in Britain. Other potential catchment descriptors include soil drainage type, storage attenuation, and baseflow index. Such methods have the advantage that once catchment descriptors have been established the flow behaviour of a site can be estimated quickly. Generalised estimation methods are most commonly used for flow estimation at the ungauged site, but are also applicable where the flow record at the subject site is extremely short or unreliable.

ii) Estimation based on rainfall frequency

In this approach the flow frequency behaviour at the subject site is estimated from rainfall frequency. This is a complex method requiring the use of rainfall-runoff model to describe the temporal characteristics of the runoff response to rainfall. This model must take into account streamflow recession and catchment storage as well as any spatial variation in runoff-generating mechanisms within the catchment, including base flow. Good rainfall records are also required and some consideration must be given to how rainfall is spatially distributed across a catchment. This approach is most useful where the catchment has unusual features (these can be considered explicitly in the model) and in small, impermeable catchments where runoff is primarily generated by rainfall.

iii) Data transfer methods

Data transfer is the traditional method of improving frequency analyses for short records. Transferring data involves taking into account flow data from analogue catchments (analogue catchments have similar hydrological behaviour) during statistical frequency analysis of the flow record at the subject site. This could involve the flow record being extrapolated in some way or assumptions about characteristics of the frequency curve being made. Data transfer allows more information to be incorporated into the analysis, which not only reduces sampling uncertainty, but also generally results in a reduction of model uncertainty by facilitating the choice of distribution.

In general there are two approaches to data transfer. In the first, the regionalisation approach, data from a group of analogue catchments is pooled with that from the subject site, whilst the second method involves the transposition of data from a single long-record analogue site. These methods are detailed below.

a) Regionalisation approach

The regionalisation approach works on the basis that a ‘region’ or group of catchments can all have similar distributions, or particular aspects of distributions, and that estimation of these aspects jointly across sites leads to improved estimation at individual sites. Regions are defined in terms of data space (i.e. regions consist of catchments with similar hydrological behaviour regardless of their spatial distribution). Such regional pooling methods are common practice for flood frequency analysis in the UK (e.g. Institute of Hydrology, 1999), where data from several sites (roughly between 5 and 50 sites) are used to estimate event frequencies at the short record site.

b) Geographically-restrained regionalisation approach

As droughts tend to behave in a spatially coherent manner, a geographical restraint may be placed on catchments forming a pooling group. This means the region consists only of nearby analogue catchments, and thus the number of catchments included in the group will probably be smaller. Four possible ways of using nearby analogue sites to give an indication as to ‘how bad’ the low flow or drought condition is in the vicinity of the station of interest were considered:

i) Annual minimum analysis may be applied to the sum of flows across a group of nearby analogue sites (i.e. the ‘regional’ total flow). The resulting estimated quantiles would be meaningful with respect to the site in question. Problems might include identifying enough representative analogue catchments of suitable length.

ii) Scale the flows at each analogue site by some appropriate quantity, (such as mean annual flow, mean annual minimum flow etc.) and form an average of these scaled values. Finally calculate the annual minima of these values. This gives an index of the severity of the drought or low flow situation in the vicinity of the station of interest.

iii) As in (ii), but use regional minima of scaled values (i.e. for a given time step base the group minima on that at the ‘worse’ analogue site). Unfortunately if a particular site experiences zero flows, this is likely to dominate the group.

iv) Scale the flows at each analogue site, but do this in a probabilistic sense, where each flow-value is replaced by the probability of point on some appropriate distribution. This leaves the overall distribution of flows or the distribution of annual minimum flows (for each site separately). At each time point the minimum probability across the group of nearby analogues would be found and finally the annual minima of these would be calculated. In this case the effect of sites at which positive values occur has less impact, since these zero-flows would be assigned some positive probability, and a lower probability might be assigned to the value at another site.

c) Transfer of data from a donor (analogue) site.

A donor site is an analogue site that is sufficiently close to the subject site to make its gauged flow data during the low flow period of special relevance. An upstream or downstream site on the same river would usually be an ideal donor catchment, although a catchment on an adjacent tributary might also be a good donor catchment provided that the catchment characteristics, such as the storage behaviour, are broadly similar. Again this relies on the fact that droughts tend to be spatially coherent entities, and that two nearby sites are likely to experience the same severity of drought at the same time (and is in fact a special case of b) where region size is restricted to a single site). A number of ways in which data transfer may take place were considered:

i) The daily flow record at the donor site may be used to extend the shorter daily flow record using regression or modelling techniques. This would require a sufficient period of overlap between the two sets of records.

ii) The flows at the donor site are scaled by some appropriate quantity, (such as mean annual flow or mean annual minimum flow). The probability distribution at the donor site provides, when appropriately re-scaled, an estimate of the probability distribution at the short-record site. This is the most commonly used method of data transfer.

iii) Re-scale the flows at each site, but do this in a probabilistic sense, where each flow-value is replaced by the probability of point on some appropriate distribution. In other words, for each year within the short record, the probability of the corresponding annual minima at the donor site is used to estimate the probability of the corresponding annual minima at the short record. This method relies on the assumption that the selected distribution is applicable to both records.

iv) Re-scale the flows at each site, but do this based on the plotting positions. In other words, for each year within the short record, the plotting position of the corresponding annual minima at the donor site is used to estimate the plotting position of the corresponding annual minima at the short record. The resulting probability plot would then be used as the basis for distribution fitting in the usual manner. The number of plotting positions 'borrowed' would be dependent on length of short record, which may influence the error associated with distribution fitting.

6.1.2 Selected methods for prototyping

Comparison of all the methods outlined in Section 6.1.1 is beyond the scope of this study. Therefore only the two approaches based on data transfer from a donor catchment (as detailed in points c-ii and c-iv) were implemented. These were thought to be the most relevant methods, given the length of the short records of interest (between 10 and 20 years) and the remit of the study, since they are theoretically robust and relatively straightforward to implement.

The two methods, transposing the p.d.f. from the donor catchment to the subject site and transposing exceedance probabilities of flows in individual years from the donor to the subject site, were tested using flow event durations of up to 90 days. In each case the approach adopted was to:

a) select suitable pairs of analogue catchments, one of which is used as the donor site (usually that with the longer record), and the other as the subject site.
b) resample the flow record of the subject site to make 'short records' of 10 and 15 years in length.
c) determine annual minima series from the flow records at the donor site and subject site where applicable.
d) apply the data transfer technique.
e) assess the performance of the data transfer technique (the frequency curve based on the complete flow record at the sample site was used as a reference curve, by which the performance of the two data transfer methods was assessed).

The study did not specifically consider records less than 10 years in length. Although the results may be applicable to short records of, say, five years in length, a more robust investigation is required to provide definitive answers where record lengths are so short. Although beyond the scope of this prototyping study, such investigation could form the basis of further work.

Section 6.2 demonstrates how pairs of analogue catchments were identified based on a comparison of hydrological and geological characteristics. The analogue pairs were selected from the good quality long flow records identified in Section 3.1. The application of the two methods (transposing the distribution function from the donor catchment to the subject site and transposing exceedance probabilities of flows in individual years from the donor to the subject site) to each pair of analogues (points b to e above), are described in Sections 6.3 and 6.4 respectively. According to the findings in Chapter 4, in application of the data transfer methods and assessment of their results the Gringorten Plotting Position formula, the L-Moment method and the GEV distribution are used throughout.

6.2 Selection of Donor/Short Record Catchment Pairs

6.2.1 Definition of analogue catchments

Catchments are said to be analogous if they have similar flow regimes. Catchment characteristics are given in Table A2.2. In this study particular importance is placed on the similarity in hydrological behaviour during the low-flow period. Hydrological regimes may be compared in a number of ways, including looking at flow frequency curves and flow duration curves. However this may become time-consuming if a large number of catchments are to be compared. The relation between flow regime and catchment characteristics is well established in hydrology and it is thus fairly reasonable to assume that catchments having similar characteristics might be good analogue. Here pairs of analogous catchments were identified based on similarity of their hydrogeological properties.

6.2.2 Selection criteria

A set of quantitative criteria were applied to the set of 37 long flow records identified as suitable for use in the study (see Table 3.1) in order to identify pairs of analogue catchments. The 37 catchments all received AA grading (i.e. good hydrometric quality and low artificial influence on both the Q95 and mean flow values (details are given in Appendix 2). The following criteria were applied:

- Analogues should receive similar (± 10%) rainfall volumes.

  Streamflow drought occurrence is strongly linked to rainfall occurrence. The standard average annual rainfall (SAAR) from 1941 to 1970 was used to describe catchment rainfall.

- Catchments should have similar storage properties.

  As baseflow often sustains flows in high-storage catchments during the low flow period, catchments should have similar storage properties The Baseflow Index (BFI) values derived both from streamflow using the IH Method (Gustard *et al*., 1992) and from the representation of different HOST (Hydrology of Soil Types Method) classes were used as quantitative assessments of catchment storage. In each case a difference of less than 10% was required for a pair of catchments to qualify as analogues.

- Catchments should be located within the same hydrometric area.

  As there is strong spatial coherence during droughts, it is also important to restrain the geographical location of analogue catchments. A further condition was therefore imposed: analogue pairs have to be located in the same hydrometric area.

6.2.3 Suitable analogue pairs

Table 6.1 details the station pairs that met the applied criteria. From these three pairs of analogue stations (9001 / 9002; 19004 / 21013; 39019 / 39028) were selected for investigation in more detail, and are used to illustrate the general findings of the study. The three sets were selected to represent three different scenarios: the 9001/9002 pair represents an upstream – downstream pair, the 39019/39028 pair are located on adjacent tributaries within permeable systems, whilst the 19004 / 21013 pair are actually located on different rivers within the same hydrometric region. In each case one of the pair was used as the subject site, whilst the other was used as the donor catchment.

Table 6.1: Suitable pairs of analogue catchments (differences in values between the pair are shown)

Station		**Record Length**		**% BFI**	**% HOSTBFI**	**% SAAR**
Subject	Donor	Subject	Donor			
9001	9002	38	38	1.7	-1.2	6.1
19004	21013	38	35	5.6	4.8	2.1
20001	20003	38	34	5.7	-6.0	-1.5
20001	20005	38	34	7.5	-8.6	-6.7
20003	20005	34	34	2.0	-2.5	-5.1
21013	21015	35	33	2.0	-6.0	8.2
39028	39019	38	31	6.2	8.3	-6.8

6.2.4 Comparison of the flow-return period relationship for analogue catchments

Analogue pairs were further assessed by comparing their frequency curves. For each station the distribution parameters were estimated based on standardised flow data and the flow-return period curves for different series of annual minima were then developed. This standardisation was used to remove the various complications associated with minima derived for different durations.

Figure 6.1 illustrates the flow return period relationships at the analogue stations 9001 and 9002 for durations D=7 and D=180 days. Similarly, Figures 6.2 and 6.3 illustrate the flow-return period relationships at the analogue stations 19004 and 21013, and 39019 and 39028.

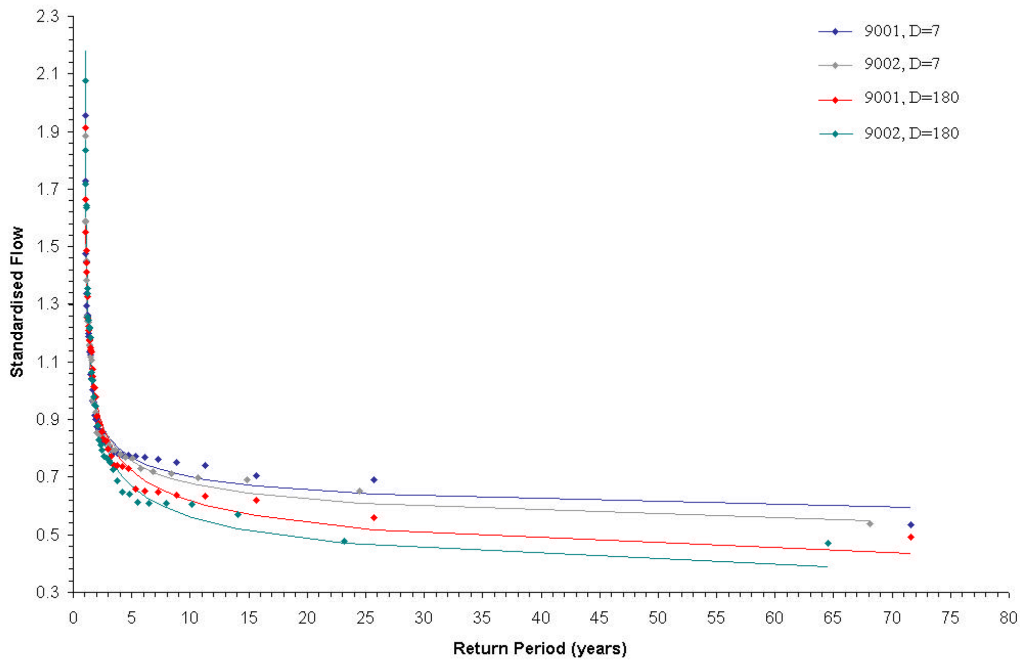


Figure 6.1:Flow- return period plots for Stations 9001 and 9002

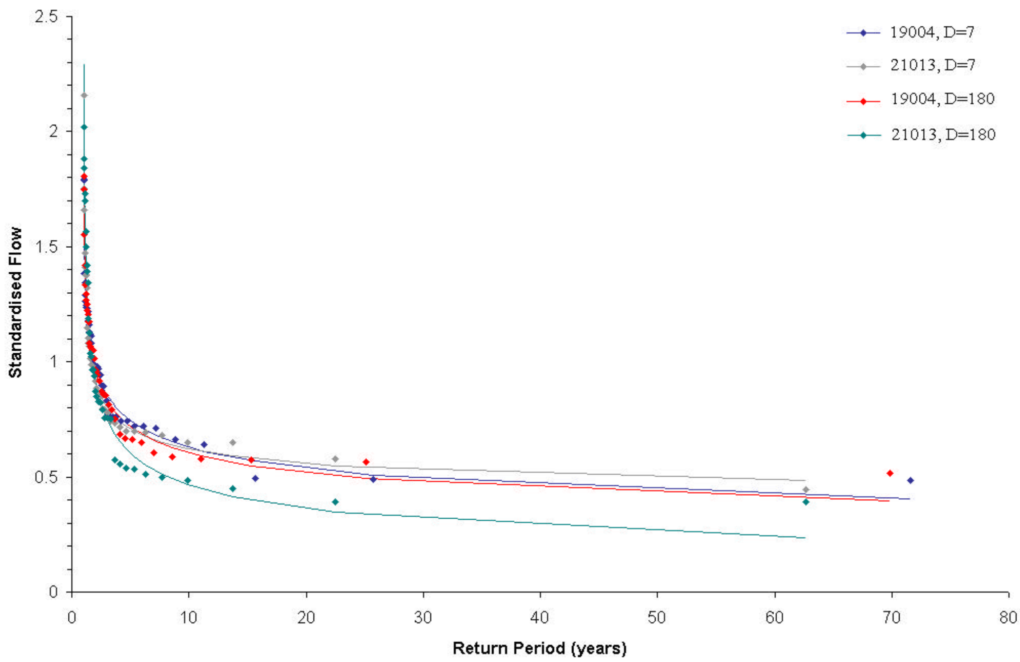


Figure 6.2:Flow- return period plots for Stations 19004 and 21013

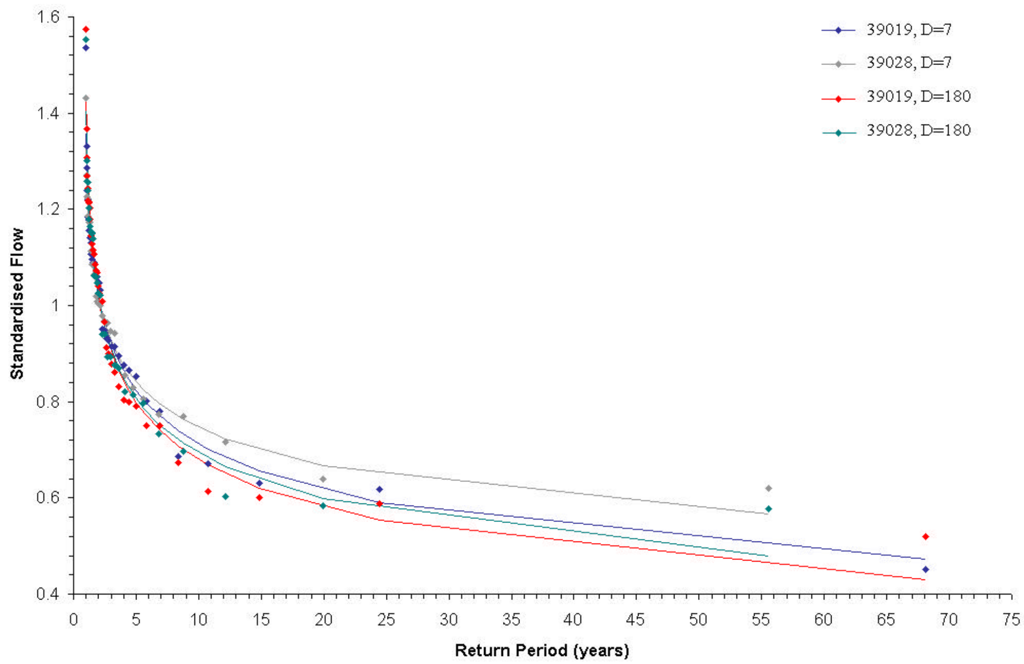


Figure 6.3: Flow- return period plots for Stations 39016 and 39028.

Figures 6.1 to 6.3 illustrate that, even when flows are expressed as standardised values, there are still some differences in the flow – return period relationships for the selected analogue pairs, but that these differences are generally smaller in the case where D=7 than where D=180 days.

As the aim is to use one catchment as a donor to the other, it is important to quantify the difference in flow estimates derived from analogue curves and to assess how this error varies according to the duration and the return period range considered and vice versa. The error, with respect to flow E_Q, is given by the following equation

$$E_Q(T) = \frac{|Q_1(T) - Q_2(T)|}{Q_1(T)} \qquad (6.1)$$

where $Q_1(T)$ is the annual minimum flow at the subject site for a return period of T years and $Q_2(T)$ the flow at the donor station for the same return period T.

Similarly the error with respect to return period, E_T, is defined by:

$$E_T(Q) = \frac{|T_1(Q) - T_2(Q)|}{T_1(Q)} \qquad (6.2)$$

where $T_1(Q)$ is the flow at the subject site for a return period of T years and $T_2(Q)$ the flow at the donor station for the same annual minima flow Q.

Figure 6.4 shows how the errors between flow estimates for the three pairs of analogue sites vary with return period, in each case durations of 7 and 180 days are considered. Small errors are observed for the 9001/9002 pair, whilst errors are highest for the 19004/21013 (e.g. the error is above 10% for return periods of 3 years and more, for a duration of 180 days). This reflects the fact that 19004 and 21013 are located on different river systems, whilst 9001 and 9002 are nested catchments.

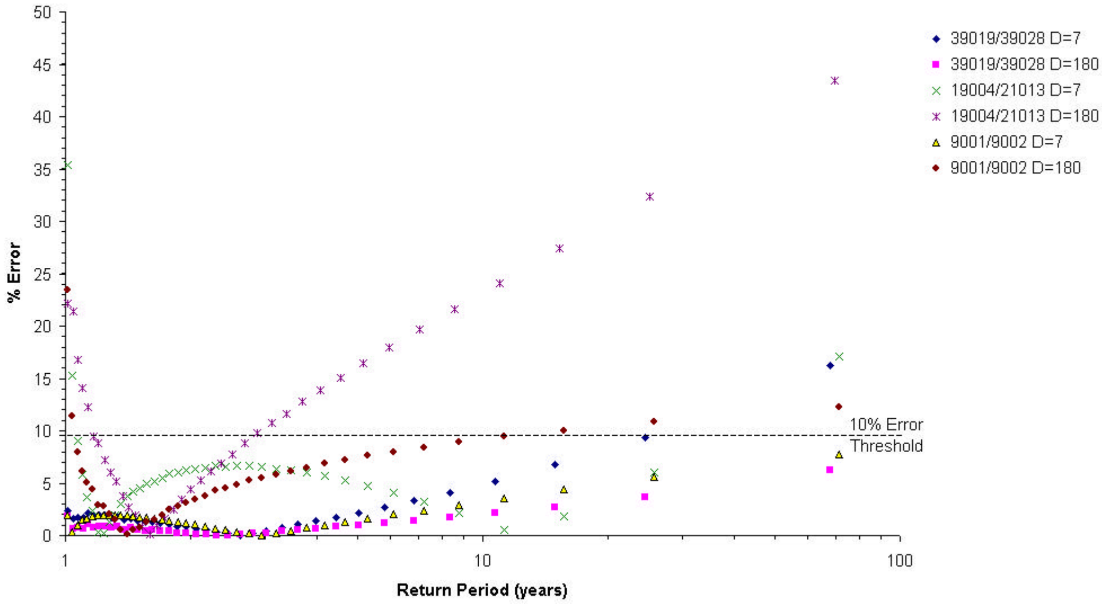


Figure 6.4: Error for the standardised flow-return period relationships for the analogue pairs

Assuming an acceptable range of error of between 0 and 10%, the range in which suitable estimates of return period can be made is 2 years to 25 years. If only short duration events are considered (e.g. D = 7) the acceptable range is wider, between 2 years and 40 years. The varying range of validity probably occurs because the annual minima series for longer durations tend to reflect the intrinsic geological characteristics of the catchments whereas the minima for shorter durations tend to reflect variations in rainfall.

6.3 Transposition of Flow-Frequency Characteristics

6.3.1 Rationale

In the first of the methods considered, the probability density function of the donor catchment is transposed to the subject site. This method assumes that, for a given duration (where D=90), the flow-return period relationships of the pair should be similar (within a 10% tolerance – see 6.2) provided that the scale of the flows are taken into consideration. Here an appropriate scaling is to standardise by the mean annual minima flow or MAM(D). Then the flow – return period relationship for the long record (based on standardised data) can be used to estimate that at the subject site, by re-scaling it by the short record MAM(D).

This method is dependent upon the sampling error for MAM(D) being insensitive to record length i.e. the estimate of MAM(D) for the short record should be representative of the long-term MAM(D) value for the site. The Institute of Hydrology (1980) suggested, as the mean value of the annual minima usually occurs at a return period in the range of 1.9 – 2.5 years, it is possible to estimate mean annual minima robustly for relatively short records. To confirm this observation the stability of MAM(D) with record length was investigated, as presented in 6.3.2.

6.3.2 Stability of the mean annual minima flow

As the basis of this method is the standardisation of the flow by the mean annual minima, it is vital to have a very good estimate of the MAM(D) for the subject site. It is therefore crucial to determine whether good estimates of MAM(D) can, in fact, be estimated at sites where the record length is reduced. In the interests of time, the stability of the mean annual minima was examined only for two stations, 9001 (38 years in length) and 55026 (61 years in length). The longer record (55026) was included to examine whether the level of error is higher where the initial MAM(D) is more certain (i.e. derived from more years). A more comprehensive study would also consider a larger number of stations.

Firstly the ‘true’ MAM(D) value for the subject site was determined. Then, the flow record the subject site was artificially shortened by removing data from the time-series. A series of artificial short records were created for each station. These were generated by removing a number of years, x, from each record, where x was such that the total record length varied from 8 to 20 years. For each value of x considered, ten different ‘short’ records were generated (with data being removed randomly in each case), their MAM(D) values were compared to the ‘true’ MAM(D) and the average difference across the ten was calculated. The results for D=7 and D= 180 are shown in Table 6.2.

Table 6.2: Errors associated to the estimation of the mean annual minima

	D=7		D=180	
Station	Length of record (years)	Average % Error	Length of record (years)	Average % Error
9001	20	1.53		
9001	15	3.11	15	4.08
9001	10	5.82		
9001	8	8.14		
55026	15	3.80	15	7.86
55026	10	10.64		

Table 6.2 indicates that the percentage of error of the MAM(D) estimate increases as duration increases and as the record length decreases. Note that as the record length decreases the error becomes more highly dependent on the distribution of flows used to derive the MAM(D). However generally observed errors were less than 10%, and the MAM(D) appears to be relatively stable for records of length 10-20 years. Therefore it is applicable to derive the flow-return period relationship of a short record station from the relationship obtained at a donor catchment possessing a relevant record length (at least 30 years worth data)

6.3.3 Methodology

Calculation of the mean annual minima

Having ascertained that good estimates of MAM(D) can be derived for short records, the validity of transferring the standardised frequency curve from donor site to subject site was investigated. This methodology is illustrated with the analogue pair 9001 / 9002. In this case the 'short record' for 9001 consists only of flow data between 01/01/76 and 31/12/90. The method was applied using four different durations, D = 1, 7, 30 and 60 days, using the following methodology.

Firstly the annual minima flow series were derived for the artificial short record 9001 and the long record 9002, from which MAM(D) values were derived. These data are shown in Tables 6.3 and 6.4 respectively.

Table 6.3: Annual minima flow series at station 9001 (short record))

Flow (% MF)			
D = 1	**D = 7**	**D = 30**	**D = 60**
49.54	54.54	63.58	101.0
39.36	41.22	46.77	61.83
37.58	40.44	46.09	50.77
34.44	37.29	43.82	47.08
34.22	35.13	40.60	46.98
30.19	31.65	39.03	44.66
29.64	31.41	38.55	43.46
28.41	29.51	32.64	38.96
22.48	23.05	26.97	31.86
21.25	21.70	26.29	29.93
20.69	21.55	25.10	29.86
20.35	21.47	25.01	28.60
20.02	20.64	24.13	27.34
19.01	19.67	22.10	22.60
14.43	14.94	17.22	18.43
Mean Annual Minima Values (%MF)			
28.11	29.61	34.53	41.56

Table 6.4: Annual minima flow series at station 9002(long record)

Flow (% MF)			
D = 1	**D = 7**	**D = 30**	**D = 60**
43.71	44.82	57.02	92.32
35.67	37.81	51.23	60.64
34.23	37.72	51.09	58.61
33.2	34.45	38.62	53.29
30.68	32.91	38.58	48.36
29.36	29.82	38.5	43.44
27.68	29.76	38.24	41.36
27.56	29.54	35.19	41.09
27.26	29.44	34.07	41
26.96	29.12	33.64	40.74
25.82	27.54	33.51	40.56
25.64	27.45	33.32	39.38
25.1	26.55	32.52	39.06
24.5	26.27	32.08	38.99
23.78	24.72	27.43	37.7
22.1	23.03	25.94	30.94
22.04	22.76	25.77	29.84
20.42	22.61	25.42	28.28

Flow (% MF)			
D = 1	**D = 7**	**D = 30**	**D = 60**
20.42	22.03	25.36	27.72
19.99	20.3	23.64	27.35
19.51	20.25	23.47	26.69
18.97	20.09	22.74	26.19
18.97	20.08	21.86	26.18
18.97	19.8	21.79	25.41
18.79	19.63	21.56	24.87
18.79	19.27	21.42	24.48
18.61	18.91	21.28	24.35
18.49	18.85	21.2	24.26
18.07	18.52	21.19	23.44
18.01	18.31	21.11	23.42
17.41	18.17	20.96	23.41
16.69	17.34	20.85	22.52
16.33	17.12	20.1	22.06
16.21	16.93	19.61	21.62
16.15	16.58	19.32	19.81
15.61	16.44	17.76	19.8
15.25	15.52	16.2	16.45
12.37	12.8	14.67	15.55
Mean Annual Minima Values (%MF)			
22.61	23.77	28.11	33.45

Transposition of the p.d.f.

The flow – return period relationship for the donor catchment 9002, expressed in terms of standardised flow, was calculated. The methodology for deriving this relationship remains the same whatever fitting distribution is considered, in this case all samples were fitted with a Generalised Extreme Value distribution (more details are given in Chapter 4.2) The quantile function for the GEV distribution is defined by the equation:

$$x(F) = \begin{cases} x + a\,\dfrac{1-(-\log F)^k}{k} & k \neq 0 \\ x - a\,\log(-\log F) & k = 0 \end{cases} \qquad (6.3)$$

where ξ is the location parameter, α is the scale parameter and *k* the shape parameter. Table 6.5 details the parameter values for the GEV distribution at (9002) based on flow values in cumecs.

Table 6.5: Parameter values at Station 9002 (long record) for the flow – return period relationship in cumecs

	D=1	D=7	D=30	D=60
α	4.79	5.31	6.64	8.90
ξ	19.5	20.051	23.4	26.6
k	-0.0626	-0.0446	-0.012	-0.164

By substituting these parameters, Eqn. 6.3 becomes:

$$x(F) = \begin{cases} 19.5 + 4.79\dfrac{1-(-\log F)^{-0.06}}{-0.06} & D = 1 \\ 20.5 + 5.31\dfrac{1-(-\log F)^{-0.04}}{-0.04} & D = 7 \\ 23.4 + 6.64\dfrac{1-(-\log F)^{-0.12}}{-0.12} & D = 30 \\ 26.6 + 8.90\dfrac{1-(-\log F)^{-0.16}}{-0.16} & D = 60 \end{cases} \qquad (6.4)$$

By re-scaling using the MAM(D) values for the subject site(Tables 6.3 / 6.4), Eqn. 6.4 then becomes

$$x(F) = \begin{cases} \dfrac{22.61}{28.11}\left(19.5 + 4.79\dfrac{1-(-\log F)^{-0.06}}{-0.06}\right) & D = 1 \\ \dfrac{23.77}{29.61}\left(20.5 + 5.31\dfrac{1-(-\log F)^{-0.04}}{-0.04}\right) & D = 7 \\ \dfrac{28.11}{34.53}\left(23.4 + 6.64\dfrac{1-(-\log F)^{-0.12}}{-0.12}\right) & D = 30 \\ \dfrac{33.45}{41.56}\left(26.6 + 8.90\dfrac{1-(-\log F)^{-0.16}}{-0.16}\right) & D = 60 \end{cases} \qquad (6.5)$$

6.3.4 Results

The resulting re-scaled flow – return period relationships for station 9001 at durations of 1, 7, 30 and 60 days are shown in Figure 6.5, whilst Figure 6.6 illustrates the differences between these curves and those generated using the full record from station 9001.

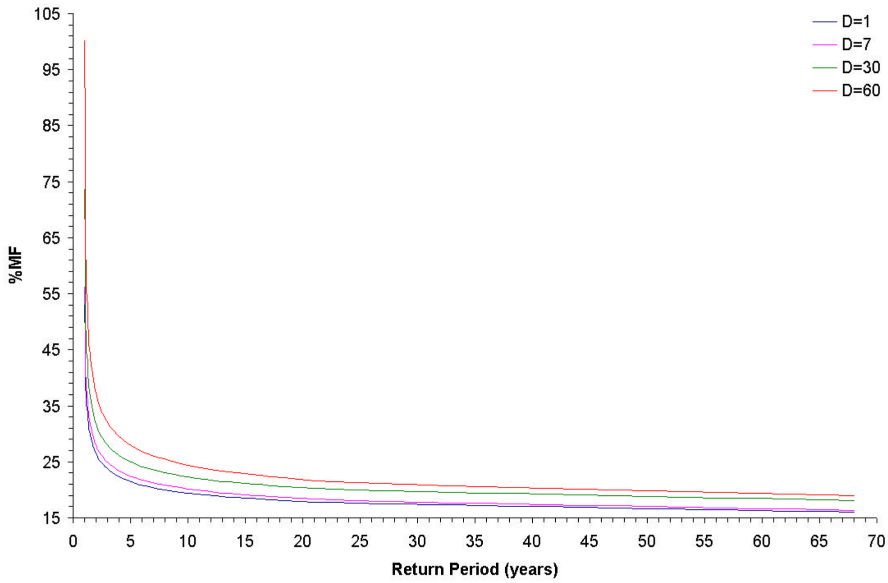


Figure 6.5: Estimated flow–return period relationship for Station 9001

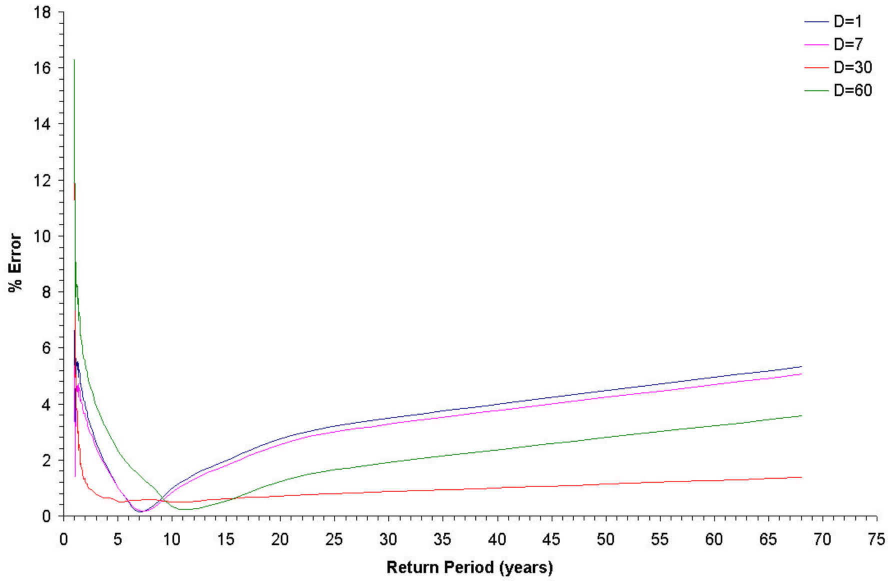


Figure 6.6: Error in estimation of the flow–return period relationship for 9001

Figure 6.6 shows that, in this case, the smallest errors are obtained where D=60. The level of error is higher where shorter durations are considered (between 15% and 25%). In this case the worst errors are obtained for a duration of 30 days. In general higher errors are observed at short return periods (i.e. at the least extreme flows). This suggests that while extreme events occur at similar frequencies of occurrence at the two catchments, the annual minima in 'wet' years may be quite different. This may reflect the fact that in wet years the spatial distribution of precipitation will influence the spatial distribution of standardised minima. In dry years, when rainfall is very low, this influence will be minimal; hence there will be a greater degree of homogeneity driven by release of water from storage.

6.4 Transfer of Plotting Positions

6.4.1 Rationale

In a short record, as the range of data available is reduced, the plotting positions derived from the ranks of the observed data points do not return accurate probabilities due to sampling error. The method developed within this sub-chapter consists of "borrowing" the plotting position for a specific year, regardless of the magnitude of the annual minima for that year, from the annual minima flow series of the donor catchment. The borrowed plotting positions are used when constructing the probability plot for the subject site, to which the distribution of interest can then be fitted in the usual way. This method assumes that for a pair of analogue catchments the chronological order of severity of droughts will be the same. For example, if the worst drought for the long-record donor occurred in 1976 it is reasonable to expect that 1976 was also the worst drought for the subject site. Hence, the major constraint of this method is that the flow record of the donor catchment must include the period of record of the short record. Furthermore, as more hydrological information is retained from the subject site there is thus an immediate benefit of this method over the p.d.f. method described in Section 6.3.

The assumption of a strong correlation between the annual minima series of the donor and subject site was tested prior to application of the method as described in the next section. The methodology was then tested using a series of artificial short records derived by sampling the flow record for the subject sites.

6.4.2 Methodology

Generation of artificial short records

As when investigating the stability of the MAM(D) value, the flow record of the subject site was artificially shortened by removing a number of years, x, from the record. For each analogue pair included in the study three 'short records' were created: two records with 15 years and one record with 10 years. For the first of these, Short Record 1 (SR1), years were removed randomly giving rise to a discontinuous record. In the remaining two cases, years were removed randomly in such a ways as to leave a continuous block of data 15 and 10 years in length respectively. Table 6.6 shows the selected pairs and identifies the long and the short record detailing the years contained within each artificial record created.

Table 6.6: Characteristics of the short records artificially generated

Subject Site	Donor Station	Donor	Years Included		
			Short Record 1 (15-years)	Short Record 2 (15-years)	Short Record 3 (10-years)
9001	9002	1961 - 1999	1964; 1971, 1972, 1973, 1974; 1975; 1978, 1979; 1980; 1986; 1988; 1991; 1995; 1998; 1999	1976 -1990	1974 - 1983
39028	39019	1960 - 1999	1970; 1971; 1972; 1975; 1977; 1980; 1983; 1986; 1987; 1989; 1991; 1994; 1995; 1996; 1997	1970 - 1984	1987 - 1996
21013	19004	1963 - 2000	1968; 1969l 1970; 1971; 1973; 1976; 1977; 1979; 1981; 1982; 1986; 1989; 1990; 1994; 1995	1971 - 1985	1977 - 1986

Generation of annual minima series

The study considered all the short records generated artificially, as detailed in Table 6.6. However to illustrate the methodology only the results for 39028 (subject site and 39019 (donor site) are presented. The annual minima flows were calculated according to the general rules applied (detailed in Chapter 3). Table 6.7 shows the annual minima flow series for Short Record 3 for four different durations: D=1, 7, 30, 60 and 90 days.

Table 6.7: Annual minima flows in percentage mean flow - Short Record 3

	Annual Minima				
Year	**D=1**	**D=7**	**D=30**	**D=60**	**D=90**
1987	47.78	47.99	52.21	55.27	59.17
1988	53.75	55.88	57.23	58.45	62.04
1989	35.83	36.26	37.33	38.2	39.57
1990	32.85	33.7	34.74	35.73	36.11
1991	37.33	39.03	40.01	41.13	41.66
1992	43.3	45.43	47.38	48.87	49.24
1993	49.27	51.19	53.35	55.59	59.82
1994	46.28	47.35	49.62	50.56	51.79
1995	41.8	44.58	45.89	49.2	50.86
1996	37.33	37.97	39.27	39.54	40.15

Table 6.8 shows the annual minima series and corresponding plotting positions for the donor site 39019. Note that only the years contained within the short record are relevant for the purpose of this method (as shown).

Table 6.8: Annual minima flows (%MF) at Station 39019[1]

Year	D=1		**D=7**		**D=30**		**D=60**		**D=90**	
	PP	**Flow**	**PP**	**Flow**	**PP**	**Flow**	**PP**	**Flow**	**PP**	**Flow**
1987	0.49	53.9	0.36	57.5	0.28	61.0	0.28	63.5	0.33	64.7
1988	0.46	53.9	0.49	55.2	0.54	56.0	0.59	56.4	0.59	56.5
1989	0.88	34.9	0.85	41.2	0.83	44.2	0.83	45.2	0.83	49.2
1990	0.96	29.8	0.91	35.4	0.88	39.1	0.88	44.1	0.85	47.1
1991	0.77	41.2	0.83	42.3	0.85	44.1	0.85	44.5	0.88	45.3
1992	0.91	32.1	0.96	32.6	0.93	36.0	0.93	37.1	0.93	37.8
1993	0.70	46.4	0.67	48.3	0.59	52.9	0.51	60.6	0.38	63.1
1994	0.28	57.3	0.22	60.9	0.20	63.9	0.17	67.7	0.15	70.5
1995	0.43	53.9	0.43	56.6	0.49	58.8	0.38	61.3	0.36	63.5
1996	0.83	36.7	0.78	45.7	0.75	48.0	0.75	49.5	0.75	51.1

[1]NB plotting positions are based upon the long record for 39019

Correlation tests

The plotting position method relies on “borrowing” probabilities from one set of data to another, and therefore the two sets of data need to be fairly well correlated to ensure relevant results. This section examines the level of correlation between the annual minima series derived from donor and short records. The process of transferring plotting position probabilities from donor catchment to short record is explained within section 6.4.2.4, and illustrated with the 39019 / 39028 analogue pair.

The level of correlation between the annual minima series of the short and long records was assessed by two types of correlation test: Spearman's Rank Correlation Test and the Pearson Correlation test (Tables 6.9 and 6.10).

Table 6.9: Spearman correlation coefficients

Station	Record	Years	D=1	D=7	D=30	D=60	D=90
	SR 1	15	0.85	0.91	0.93	0.99	1.00
9001	SR 2	15	0.79	0.92	0.95	0.96	0.99
	SR 3	10	0.90	0.88	0.92	0.90	0.94
	SR 1	15	0.61	0.74	0.79	0.83	1.00
21013	SR 2	15	0.69	0.83	0.84	0.94	0.91
	SR 3	10	0.33	0.55	0.72	0.94	0.83
	SR 1	15	0.87	0.86	0.85	0.72	1.00
39028	SR 2	15	0.80	0.77	0.77	0.81	0.80
	SR 3	10	0.69	0.59	0.56	0.63	0.60

Table 6.10: Pearson correlation coefficients

Station	Record	Years	D=1	D=7	D=30	D=60	D=90
	SR 1	15	0.93	0.95	0.98	0.99	0.99
9001	SR 2	15	0.92	0.97	0.99	0.99	0.99
	SR 3	10	0.94	0.95	0.97	0.98	0.98
	SR 1	15	0.58	0.74	0.79	0.97	1.00
21013	SR 2	15	0.69	0.77	0.85	0.96	0.93
	SR 3	10	0.41	0.56	0.78	0.93	0.92
	SR 1	15	0.86	0.89	0.88	0.85	0.98
39028	SR 2	15	0.87	0.88	0.88	0.90	0.90
	SR 3	10	0.74	0.61	0.63	0.63	0.60

In the Spearman Test the rank order of the values in the two series are compared (further details can be found in Section 5.5), whereas the Pearson test compares the magnitude of corresponding values (i.e. assesses whether flows for year *Y* are similar). For the two sets to be considered correlated the correlation coefficients must exceed a critical value (governed by the sample size). Thereafter a larger coefficient indicates a better correlation. Critical values for sample sizes of 15 and 10 years are 0.4464 and 0.5636 respectively for a 10% significance level (Neave, 1981).

The correlation tests showed that the more years contained within the short record the better the correlation. Furthermore the correlation is increased if the years selected within the short record form a continuous period. The degree of correlation also depends on the duration considered, as in general, the higher the duration is the better the correlation. The results of the correlation test show that, in general, there is satisfactory correlation (most of the results are above 0.8). The 21013/21015 pair has the worst correlation being below the critical values for Short Record 3.

Transposing plotting positions

Having ensured a sufficient degree of correlation between the annual minima flow series of donor and subject catchments, it is then possible to "borrow" the plotting position from the donor. For each year considered at the subject site, the plotting position associated with that year is taken from the corresponding year in the donor. Table 6.11 illustrates the implementation of this method, using 39028 SR3 as the subject site (see Table 6.7) and 39019 as the donor site (see Table 6.8).

Table 6.11: Annual minima flow series at Station 39028 Short Record 3

D=1		D=7		D=30		D=60		D=90	
Plotting Position	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**
0.46	53.75	0.49	55.88	0.54	57.23	0.59	58.45	0.59	62.04
0.70	49.27	0.67	51.19	0.59	53.35	0.51	55.59	0.38	59.82
0.49	47.78	0.36	47.99	0.28	52.21	0.28	55.27	0.33	59.17
0.28	46.28	0.22	47.35	0.20	49.62	0.17	50.56	0.15	51.79
0.91	43.30	0.96	45.43	0.93	47.38	0.38	49.2	0.36	50.86
0.43	41.80	0.43	44.58	0.49	45.89	0.93	48.87	0.93	49.24
0.83	37.33	0.83	39.03	0.85	40.01	0.85	41.13	0.88	41.66
0.77	37.33	0.78	37.97	0.75	39.27	0.75	39.54	0.75	40.15
0.88	35.83	0.85	36.26	0.83	37.33	0.83	38.2	0.83	39.57
0.96	32.85	0.91	33.7	0.88	34.74	0.88	35.73	0.85	36.11

The major issue arising from "borrowing" plotting positions is that the borrowed plotting positions do not necessarily decrease in the same order than the flow. This issue is inherent to this method, and has little influence on the fitting procedure provided the number of outlying points is small (Figure 6.7a). The problem is to quantify whether this phenomenon has an impact on the fitting of the annual minima curve where the number of outlying points is relatively large (Figure 6.7b).

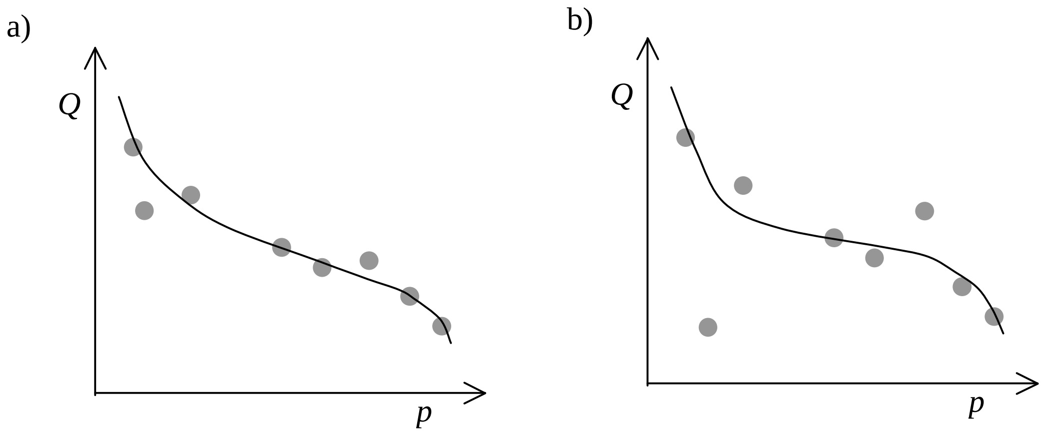


Figure 6.7: Outlying plotting positions with a) little effect on the curve shape and b) a large effect on the curve shape.

Two methods by which the same decreasing order for both variables could be re-established were considered. The first option involves erasing outlying data points from the series. However, removing outlying data points is not acceptable, as this would leave too few points on which to model the probability behaviour and thus increasing the uncertainty. The second option, which requires the modification of borrowed plotting positions to guarantee the same decreasing order for the plotting positions and the flow, is preferable. The advantage of this procedure is to keep the same decreasing order between flows and plotting positions without discarding much of the available data.

Modifying the annual minima flow series

The problem of wrongly ordered plotting positions may occur because the donor catchment behaves differently to the catchment of interest and experiences extreme annual minima in different years. It may also occur because when plotting positions are assigned according to the rank of the annual minima flow, i.e. the contrast in annual minima between two adjacent ranks is not considered. This means that the flows corresponding to consecutive ranks/plotting position could actually be very similar. For example, the annual minima for a particular year, say 1981, may have a rank of 8 and a value of 25%MF, whilst that of 1994 may have a rank of 9, and a value of 25.2%. In a donor catchment 1981 may have a flow of 28%MF, and have a rank of 9, whilst 1994 may have a rank of 8 and a value of 27.4 %MF. At both stations the flows for 1994 and 1981 are very close, yet for each, the two years are assigned different plotting positions.

Here the idea is to swap the “borrowed” plotting position for that of a different year, provided that the annual minimum flows recorded on the new and original vary by less than 10%. To illustrate this method Table 6.12 presents the rearranged annual minima flow series for 39028 SR 3.

Table 6.12: Modified plotting positions - Station 39028 SR3

D=1		D=7		D=30		D=60		D=90	
Plotting Position	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**	**Plotting Position**	**Flow (%MF)**
0.28	53.75	0.22	55.88	0.28	57.23	0.17	55.27	0.15	59.82
0.46	47.78	0.36	47.99	0.49	53.35	0.28	50.56	0.33	59.17
0.49	46.28	0.49	47.35	0.54	52.21	0.59	49.2	0.36	51.79
0.78	37.33	0.78	39.03	0.59	45.89	0.75	41.13	0.38	50.86
0.83	37.33	0.83	37.97	0.75	39.27	0.83	39.54	0.75	41.66
0.88	35.83	0.85	36.26	0.83	37.33	0.85	38.2	0.83	40.15
0.96	32.85	0.91	33.7	0.88	34.74	0.88	35.73	0.85	39.57
								0.88	36.11

6.4.3 Results

Figure 6.8 compares the estimated flow-return period plot for SR3 at D=90 days based on the borrowing of plotting positions with the reference curve derived from the annual minimum series based on the full record for 39028. The curve derived by modifying the probabilities using the method outlined in 6.4.2.4 is also shown.

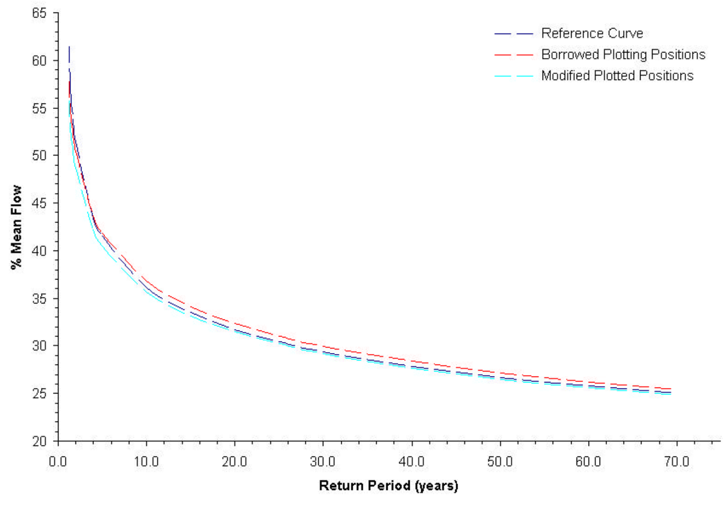


Figure 6.8: Flow – return period relationships at Station 39028 for D=90.

Figure 6.8 shows that there is a very good agreement between the reference curve and that derived by borrowing plotting positions. In this example modifying the plotting positions has resulted in a curve with even better fit, but this was the exception to the rule. For the other analogue pairs the modified curves did not provide a better estimation of the flow –

return period relationship. Therefore as rearranging the flows does not always improve the accuracy of the predictions and is an arbitrary process that is time consuming, it is better to stick to the original values even if the flows and plotting positions do not increase in the same order. Although not shown, equally good agreement between the short record and reference curve were also observed when the method was applied to the other analogue pairs, and seems to be valid for all durations.

6.5 Comparison of the Data Transfer Methods

6.5.1 General observations

This section examines the relative ability of the two data transfer methods in predicting the flow-return period relationship at the subject site. A selection of results from the three subject sites considered as part of the study are shown in Figures 6.9 through 6.12. In each case three different flow-return period curves are shown:

- That based on the original full record (reference curve).
- That based on borrowing of the plotting positions from the donor site.
- That based on transfer of the standardised probability function from the donor site.

The root mean square error (RMSE) between the predicted curves and the reference curve is indicated in each case.

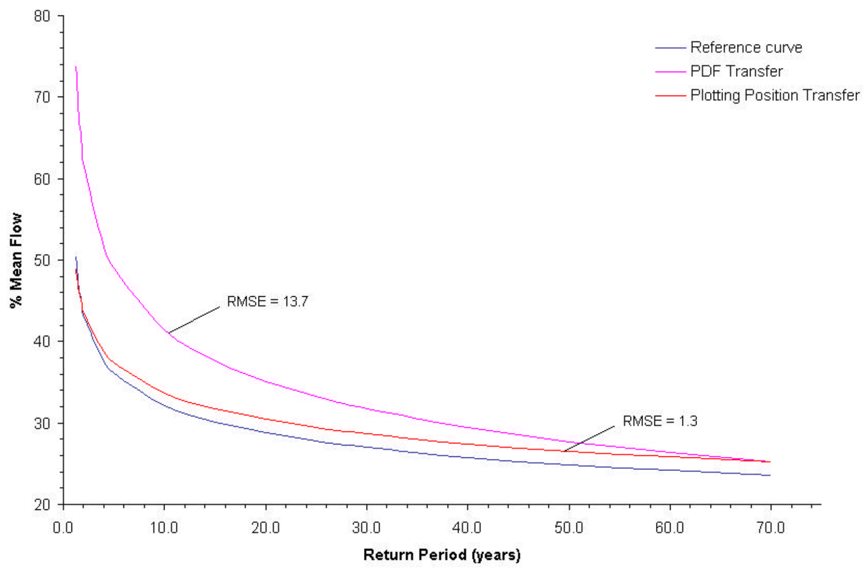


Figure 6.9: Flow – return period relationship at Station 39028, SR3, for D=1.

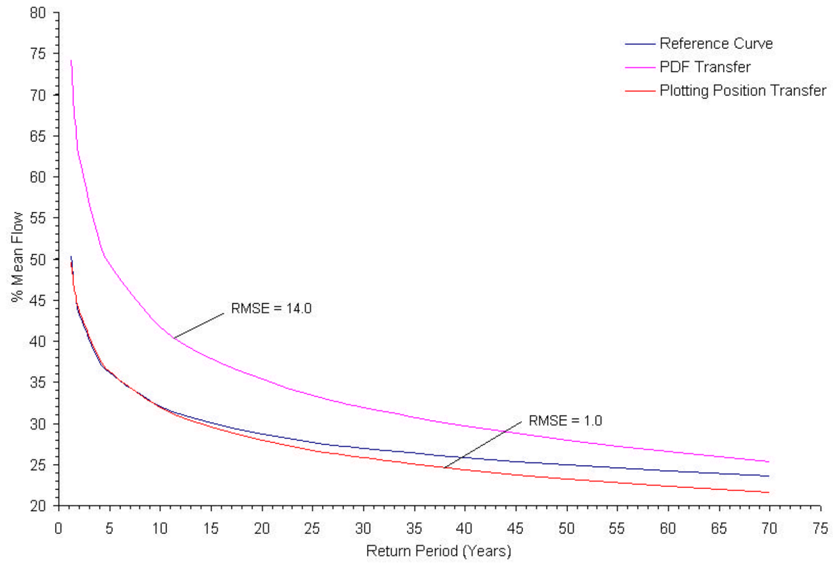


Figure 6.10: Flow – return period relationship at Station 39028, SR2, for D=1.

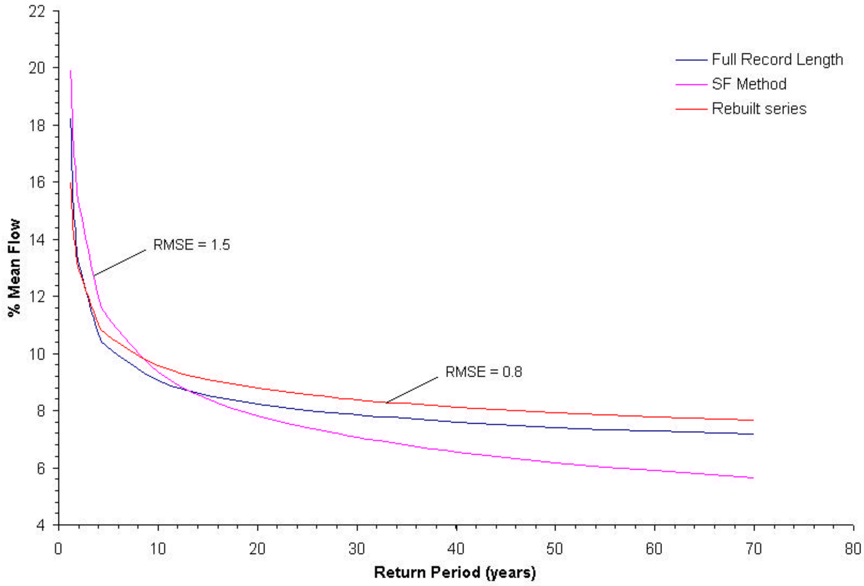


Figure 6.11: Flow – return period relationship at Station 21013, SR2, for D=1

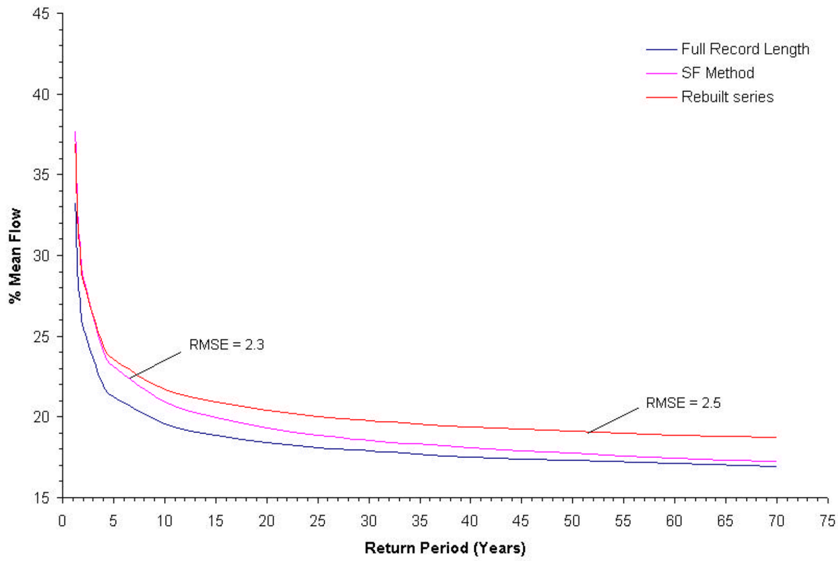


Figure 6.12: Flow – return period relationship at Station 9001, SR1, for D=1.

These figures show that, although the p.d.f. method gave good results in many cases, in general, the transfer of plotting positions provides a better estimation of the flow – return period relationship than the transposition of the standardised p.d.f. from donor to subject site. This reflects the fact that the plotting position method utilises more information from the short record, whereas the standardised flow method uses only the

MAM(D) value of the short record, and thus is highly dependent on the choice of donor station.

Having established that the transfer of plotting positions provides the better estimation, the next logical step is to assess whether there are further controls on the estimation procedure, such as the influence of record length or catchment type. It is particularly important to assess whether a higher degree of correlation between the annual minima series of the subject site and its chosen donor results in improved estimation of the probability distribution.

6.5.2 Influence of record length and catchment type

Figure 6.13 shows the flow – return period plots for the three 'short records' generated from the flow record for 39028. As shown, each short record represents a different set of 'observed' annual minima values; the range of flow values in the observed set may influence the curve-fitting process. Here SR3 (10 years) contains few values below 34% of the mean flow, whilst SR2 contains more extreme values. As a result of these biases the curve derived for SR3 overestimates the flow for a given return period, whilst that for SR2 produces estimates of flow that are too low. The greatest level of accuracy occurs in the range 0 – 15 years; this range contains the MAM(D) and most of the observed flow values.

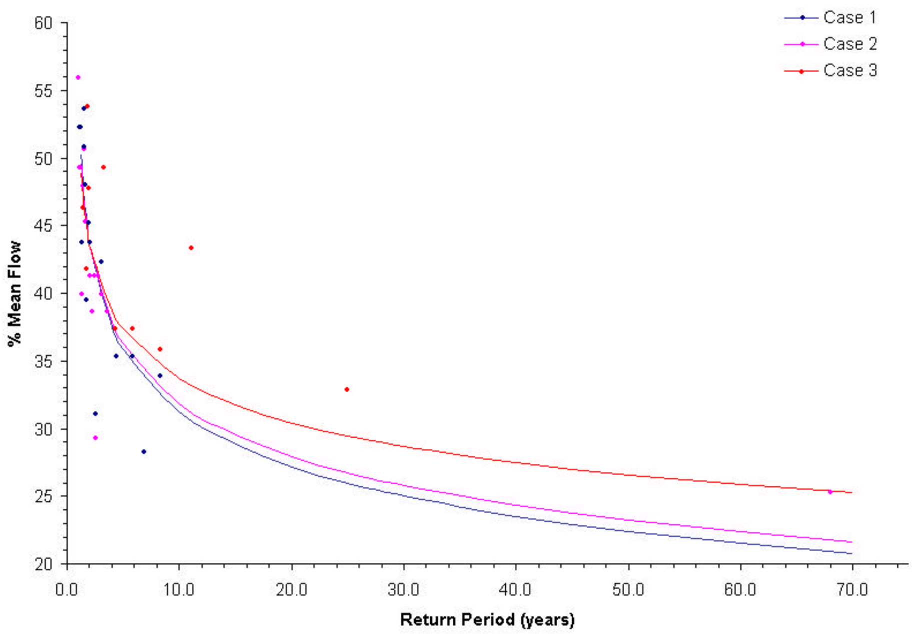


Figure 6.13: Flow – return period relationship at Station 39028 for D=1, using the short records SR1 (15 years), SR2 (15 years) and SR3 (10 years)

Figure 6.14 examines the level of error for each of the short records artificially generated for all catchments. The percentage error in the predicted flow values at different return periods are shown, in the case where D=1.

Figure 6.14 indicates that the lowest errors are observed in the range 2-10 years, i.e. good estimates can be obtained for short return periods. Thereafter the error increases as the return period of interest increases. Note that this is still an improvement on the results obtained by re-scaling/ standardisation method.

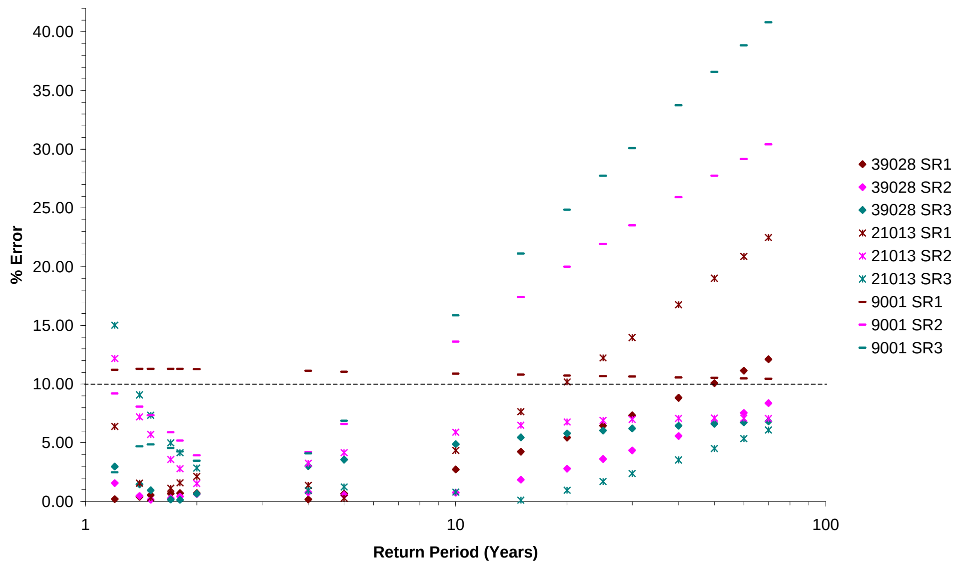


Figure 6.14: Error for the flow – return period relationship for all artificial short record generated for D=1

As only three analogue pairs have been considered it is impossible to make any statements regarding the influence of catchment type on the estimation procedure. However in Figure 6.14 the errors for the analogue pair 9001/9002 are the highest, despite being located on the same river, and showing the highest levels of correlation between the AMS of the donor catchment and the short records. This may reflect the flashy nature of the flow regimes in these catchments.

6.5.3: Influence of correlation between analogues

The correlation between the annual minima series of the analogue pairs was discussed. The aim of this section is to establish whether there is a relation between the correlation and the error associated with the estimated probability distribution. Table 6.13 compared the Spearman Rank correlation coefficients with the RMSE error for each of the short-record – donor site combinations.

Table 6.13: Comparison of correlation and RMSE

Subject Site	Short Record	Donor Site	Spearman Coefficient	RMSE
39028	SR2	39019	0.8	1.0
39028	SR3	39019	0.69	1.3
9001	SR1	9002	0.85	2.5
21013	SR2	21015	0.69	0.8

The results shown in Table 6.13 do not indicate a clear relationship between correlation coefficient and root mean square error. For instance if the plotting position derived Q_T-T curves for 39028 SR2 and 9001 SR1 are compared (Figures 6.10 and 6.12) it is clear that greater accuracy is achieved by 39028. However of the two, 9001 shows the greatest level of correlation with its donor catchment. These inconsistencies indicate that the correlation value cannot be used to predict the level of accuracy that might be obtained in the predicted flow return period relationship.

6.6 Concluding Remarks

Of the two methods considered the transfer of plotting positions proved to be the most accurate at predicting the low flow frequency behaviour of the ‘short record’.

The transposition of the standardised probability distribution from donor to subject site performed well where the donor record was that of an upstream (or downstream) gauging station. The error associated with the predicted curves varied according to the duration considered. For short durations (e.g. D=1, 7, and 30 days) the method was fairly accurate (no greater than ±10%) for return periods between 2 and 40 years. For higher durations of D=60 and 90 days the method was fairly accurate for a return period range between 2 and 25 years.
The method based on transfer of plotting positions generally provided better predictions of the flow frequency curves for the ‘short record’. The main constraint of the method is

that the flow record of the donor catchment selected must overlap the data provided within the short record. However the accuracy of the method is highly dependent on the range of flows observed in the short record, and the period over which the short record was collected. This may result in the plotting positions for some annual minima being over or underestimated, which may influence the ability to accurately reproduce the flow-return period relationship at longer return periods.

7. SUMMARY AND CONCLUSIONS

7.1 Methodology for Long Records

7.1.1 Overview

Models of the probability distribution functions of annual minima flows have been determined for 25 British Rivers. Based on theoretical considerations and previous work these models have been based on parameterising four candidate distributions: the Generalised Extreme Value distribution (GEV), the Generalised Logistic distribution (GL), the Generalised Pareto distribution (GP) and the Pearson Type III distribution (PE3).

In each case the following methodology has been applied:

i) Derive annual minima series for each flow record, based on D-day running averages with the duration, D, taking values of 1, 7,30,60,90,180 and 365 days.
ii) Determine the degree of independence and the level of stationarity in the annual minima series.
iii) Determine the probability of non-exceedance of each annual minimum using a Plotting Position formula
iv) Determine the parameters of the four candidate distributions based on the flows and probabilities of the observed data
v) Assess the descriptive ability of models by assessing the goodness of fit method between model and observed data
vi) Assess the predictive ability of the models by deriving the corresponding flow-return period relationship for each.

Each element of the methodology was scrutinised as reported in Chapters 3 to 5.

7.1.2 Time series analysis

The results of the time series tests showed that the flow records can be assumed to be stationary, as no strong trends were observed in the data sets. The level of dependency in the data was also small, although plots of the partial auto-correlation function showed that annual minima values were correlated for a lag time of 2 years. The level of correlation increased as annual minima for periods of increased duration were considered.

7.1.3 Evaluation of plotting position formulae

An evaluation of the different plotting position formulae that may be used to estimate the probability of non-exceedance of each observed flow was reported in Section 4.3. The Gringorten Plotting Position Formula and Hazen Formulae produced similar results whilst the Weibull plotting formula was found to be less robust. Accordingly, the Gringorten plotting position formula was chosen as the default for the analysis.

A special plotting position for data series containing zero values (that advocated by Stedinger *et al.*,1993) was investigated. However, applying this formula for annual minima series containing observations of zero did little to improve the curve fitting procedure.

7.1.4 Evaluation of fitting technique

Two curve fitting techniques, L-Moments and Maximum Likelihood Estimation, were compared. For this test each was fitted to observed data using the GEV distribution. The descriptive and prescriptive abilities of each method were considered in this comparison. Both methods modelled the observed data to within acceptable goodness of fit parameters. However, although the prescriptive performance of the two methods was similar when short durations are considered, they became quite different when longer durations are considered. Given that the Maximum Likelihood method does not provide 'better parameters', whilst L-Moments codes are readily available for a number of distribution families, the L-Moment technique was chosen as the default for the study.

7.1.5 Assessment of hydrometric errors

Hydrometric errors

The influence of hydrometric errors on annual minima series was studied by adding random errors to both the annual minima flow series and the original gauged daily flow records. Errors added to the annual minimum series have greatest influence on the shape of the tails of the probability distribution, where $0.92< p < 0.1$. Errors in the gauged daily flow record have negligible effect on the flow – return period relationships, except where the duration considered is 1 or 7 days. Given that in the context of low flows, typical hydrometric errors are within a range of ±10%, it can be assumed that the influence of true hydrometric errors on low flow frequency estimation is generally negligible.

Discretization

The effect of discretization of values within the annual minima series, which manifests in the form of steps in the probability plots and may result from poor measurement precision at low flows, was examined. Although annual minima frequency curves often show stepped features, the study has demonstrated that these are unlikely to be caused by discretization, and have little impact on the curve-fitting procedure.

7.1.6 Evaluation of distribution families

General approach

Two approaches were used to evaluate which of the four distribution families is most suitable for representing the low flow probability behaviour of British rivers. In order to apply these methods, the parameters of each of the candidate distributions were estimated using the L-Moments method of parameter estimation, having used the Gringorten Plotting Position formula to estimate probabilities of the observed annual minima flow series.

Firstly, goodness-of-fit tests were used to determine the level of agreement between the modelled and observed probability plots. The root mean square error was also considered. Secondly, the robustness of the flow-return period relationship derived from the modelled probability distribution function was also considered (i.e. whether sensible and accurate estimates of return period can be determined for flows levels in the extrapolated tail of the distribution, and vice versa). Unfortunately, as there are few benchmarks against which to assess the predictive ability of the flow-return period relationships, this latter method provides little opportunity to discriminate between the different distributions.

Goodness of fit results

The results of the Chi-square tests indicated that there was a satisfactory fit between all modelled and observed curves. In order to differentiate between distributions the size of the Chi-Square Statistic and RMSE were considered – the distribution giving the lowest Chi-Square and RMSE values was assumed to be the most appropriate. This showed that, for a given flow record, the most applicable distribution was dependent on the period of duration over which the annual minima were calculated. In general, the GEV and PE3 distributions performed better when short durations were considered, whilst the GP distribution performed better at longer durations (of 180 days or more). The catchment type also influenced which distribution was most applicable. High permeability catchments, such as Chalk catchments, were best represented by the GL distribution. Moderately permeable catchments (i.e. BFI between 0.4 and 0.7) were best represented by the PE3 very short durations, by the GEV distribution at durations of 30 and 60 days, and by the GP distribution at long durations. These differences are caused by averaging out of extremes when longer durations are considered. In impermeable catchments, many of the extreme events are lost when the duration considered is greater than 7 days. For such catchments the annual minima for a period of duration of 365 days would reflect the range of variation of mean flow at the site rather than low flow behaviour. Thus whereas the GEV is better for short durations, at long durations a Pareto-like distribution is slightly more appropriate. In permeable catchments, extreme behaviour is always damped by the input of base flow into the river system.

Flow – return period relationships

Here the goal was to evaluate the level of the uncertainty of the flow estimates prescribed by the modelled curve for high return periods i.e. the robustness of the model. The errors associated with quantile estimates at high return periods and errors in return period atextreme low flow quantiles were considered. The ability to make sensible predictions regarding quantiles at high return periods was found to depend on the number of observations – for those series with more data points the observed range of return periods is wider, and therefore there is less uncertainty when extrapolating beyond the observed range. However, the differing characteristics of the different distribution families means that, although the forms of the corresponding flow – return period relationships are very similar around the median, they differ at the tails of the distribution (i.e. at high return periods). The size of the observed flows may also influence the shape of the curves. Models for some catchments (usually the low permeability catchments) were more prescriptive in the extrapolated upper tail of the probability distribution, whilst others (most permeable catchments) performed better in the extrapolated lower tail of the distribution.

7.1.7 Low flow frequency and catchment characteristics

The two factors most likely to influence the shape of the probability distribution are rainfall and catchment storage. The relationships between standardised annual average rainfall (SAAR) and Baseflow Index (BFI) and distribution form were examined. However SAAR and BFI are correlated for British catchments. This means that whilst it is possible to identify some trends, it is difficult to make any firm conclusions regarding them without conducting more detailed research.

The form of the probability curve was found to depend on catchment type. The results of the goodness of fit and errors test showed that for low to mid BFI values, the PE and GEV distributions were favoured, whereas for catchments with high BFI rating the GL was favoured. PE3 and GL distributions are favoured by catchments with low SAAR, whilst there is a trend towards the GEV distribution for catchments with high rainfall. When long durations are considered the GP distribution becomes more common at low SAAR.

There are also strong trends within the form of the quantile-return period relationships. When expressed in terms of the AM/MAM(d) ratio the forms may be classified into two groups: those for high rainfall, low storage catchments (having steep gradients) and those for high storage, low rainfall catchments (having shallow gradients).

7.2 Prototyping a Methodology for Short Records

7.2.1 Methodology

Possible methods for short records

In an at-site analysis of low flow frequencies, it is unrealistic to expect to accurately predict the flow – return period relationship when the number of data observation is restrained, i.e. the flow record is short. Although a number of different regionalisation techniques were considered, given the project budget, a simple approach was adopted in which the probability plot at the subject site was derived from that at a donor station. Two variations of this method were considered: transposition of the standardised distribution function from the donor to the subject site, and transfer of plotting positions from donor to subject site. Both were applied to three analogue pairs.

Transposition of standardised probability distribution

This method was based on assuming that analogue pairs having very similar hydrological behaviour should have similar probability distributions, provided the flow values are standardised. The annual minima for the donor site were standardised by expressing them as a ratio of the MAM(D) value at the donor site. The flow - return period relationship for the subject site was derived by re-scaling that derived for the donor catchment using the MAM(D) value from the short record at the subject site.

Transfer of plotting positions

Assuming a similar hydrological behaviour, it is possible to calculate the probability of occurrence a flow minima occurring in any particular year by matching it with the plotting position of the corresponding year in the longer series derived for the donor catchment. Hence, the major constraint of this method is that the flow record of the selected donor catchment must overlap the data provided within the short record.

7.2.2 Evaluation of applied methodologies

Of the two methods investigated the transfer of plotting positions proved to be the most accurate at predicting the low flow frequency behaviour of the 'short record'.

The standardised p.d.f. method performed well where the donor record was that of an upstream (or downstream) gauging station. The error associated with the predicted curves varied according to the duration considered. For short durations (e.g. D=1, 7, and 30 days) the method was fairly accurate (no greater than ±10%) for return periods between 2 and 40 years. For longer durations, of between 60 and 90 days, the method was fairly accurate for a return period range between 2 and 25 years.

The method based on transfer of plotting positions generally provided better predictions of the flow frequency curves for the 'short record'. The main constraint of the method is that the flow record of the donor catchment selected must overlap with the short record. However, the accuracy of the method is highly dependent on the range of flows observed in the short record, and the period over which the short record was collected. This may result in the plotting positions for some annual minima being over or underestimated, which may influence the ability to accurately reproduce the flow-return period relationship at longer return periods. Fixing the plotting positions to deal with outliers did not improve the prediction of the frequency for the short record.

7.3 Guidance Document

The recommendations for the analysis of long records are summarised in the accompanying guidance document (Zaidman *et al*., 2002). In order to encourage a greater level of consistency in probability estimates for low flow events in the UK, a single parametric approach for low flow frequency analysis of annual minimum flows is advocated, based on the use of the Pearson Type III distribution. Emphasis is also placed on the uncertainties associated with frequency analysis, both in general, and in respect to the methodology outlined here.

8. REFERENCES

8.1 References (Main Review)

Acreman, M.C. & Sinclair, C.D. 1986. Classification of drainage basins according to their physical characteristics; an application for flood frequency analysis in Scotland. Journal of Hydrology, 84, 365-380.

Arnold, S.F. 1993. Gibbs Sampling. In: C.R. Rao, C.R. (Ed.), Computational Statistics (Handbook of Statistics, Vol. 9*)*, North-Holland, London, 599-625.

Arora, K & Singh, V.P. 1987. An evaluation of seven methods for estimating parameters of EV1 distribution. *In*: V.P. Singh (ed.), Hydrologic Frequency Modelling, pp 383-394, Reidel Publishing, Dordrecht.

Bobée, B. and Ashkar, F. 1991. The Gamma Family and Derived Distributions Applied in Hydrology. Water Resources Publications, Colorado, USA. 205pp.

Boorman, D.B., Hollis, J.M. and Lilly, A. 1995. Hydrology of soil types: a hydrologically-based classification of the soils of the United Kingdom. IH Report No. 126, Institute of Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK.

Box, G.E.P and Tiao, G.C. 1973. Bayesian Inference in Statistical Analysis. Addison-Wesley, London. (Reprinted in 1992 by John Wiley & Sons, New York.).

Bulu, A. & Onoz, B. 1997. Frequency analysis of low flows by the PPCC test in Turkey. *In*: FRIEND: Regional Hydrology, Concepts and Models for Sustainable Water Resource Management (Proceeding of a conference at Postojna, Slovenia, September-October 1997). IAHS Publication no 246. 133-140.

Bulu, A. 1997. Statistical Analysis of low flows with zero discharges. FRIEND: Flow Regimes from Artificial and Network Data, Third Report, Cemagref, 167-170.

Burn, D.H. & Boorman, D.B. 1993. Estimation of hydrological parameters at ungauged catchments. Journal of Hydrology, 143, 429-454.

Chung, C. and Salas, J.D. 2000. Drought Occurrence Probabilities and Risks of Dependent Hydrologic Processes. Journal of Hydrologic Engineering, 5(3), 259-268.

Cohen, J.P. 1982. Convergence rates for the ultimate and penultimate approximations in extreme-value theory. Advances in Applied Probability, 14, 833-854.

Cox, D.R. and Hinkley, D.V. 1974. Theoretical Statistics. Chapman and Hall, London.

Cunnane, C. 1978. Unbiased plotting positions – a review. Journal of Hydrology, 37, 205-222.

Cunnane, C. 1989. Statistical distributions for flood frequency analysis. Operational Hydrology Report No. 33, World Meteorological Organisation, Geneva.

DeGroot, M.H. 1970. Optimal Statistical Decisions. McGraw-Hill, New York.

Dracup, J.A., Lee, K.S. and Paulson, E.G. Jr. 1980. On the definition of droughts. Water Resources Research, 16 (2), 297-302.

Durrans, S.R., Quard, TB, Rasmussen, P.F. and Bobee, B. 1999. Treatment of zeroes in tail modeling of low flows. Journal of Hydrological Engineering, 4(1), 19-27.

Evans, M., Hastings, N.A.J. and Peacock, J.B. 1993. Statistical Distributions (2nd Edition). John Wiley, Chichester.

Galambos, J. 1978. The Asymptotic Theory of Extreme Order Statistics. Wiley, New York.

Godambe, V.P. 1960. An optimum property of regular maximum likelihood estimation. Ann. Math. Stat., 31, 1208-1211.

Godambe, V.P. 1976. Conditional likelihood and unconditional optimum estimating equations. Biometrika, 63 (2), 277-284.

Gottschalk, L., Tallaksen, L.M. & Perzyna, G. 1997. Derivation of low flow distribution functions using recession curves. Journal of Hydrology, 194, 239-262.

Gustard, A., Bullock, A. and Dixon, J.M. 1992. Low Flow Estimation in the United Kingdom. IH Report No. 108, Institute of Hydrology, Wallingford.

Hinkley, D.V. and Reid, N. 1991. Statistical Theory. *In*: "Statistical Theory and Modelling", Ed. D.V. Hinkley, N. Reid and E.J. Snell, pp. 1-29. Chapman and Hall, London.

Hisdal, H. & Tallaksen, L.M. (eds.). 2000. Drought Event Definition. ARIDE Technical Report No 6, University of Oslo, Norway.

Holmes, M.G.R., Young A.R. and Sekulin, A. 2000. Low Flow Studies 2000 – Technical Report 13: Estimation of monthly mean flow statistics in the UK. Centre for ecology and hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK (*in press*).

Hosking, J.R.M. 1985. Maximum-Likelihood-Estimation of the Parameters of the Generalized Extreme-Value Distribution. Applied Statistics-Journal Of The Royal Statistical Society- Series C., 34(3), 301-310.

Hosking, J.R.M. 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. R. Statistic. Soc., B, 52(1), 105-124.

Hosking, J.R.M. and Wallis, J.R. 1997. Regional frequency analysis: an approach based on L-moments. Cambridge University Press, Cambridge, UK.

Institute of Hydrology. 1980. Low Flow Studies. Institute of Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK.

Institute of Hydrology. 1999. Flood Estimation Handbook. Institute of Hydrology, Crowmarsh Gifford, Wallingford, Oxfordshire, OX10 8BB, UK.

Jakob, D, Reed, D.W. and Robson, A. J. 1999. Selecting a pooling group (Chapter 16). *In*: Institute of Hydrology (1999). Flood Estimation Handbook, 153-180.

Leadbetter, M.R., Lindgren, G. and Rootzen, H. 1983. Extremes and Related Properties of Random Sequences and Series. Springer Verlag, New York.

Lindley, D.V. 1980. Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2, Inference. Cambridge University Press, Cambridge.

N.R.A. 1995. The drought of 1995. A Report to the Secretary of State for the Environment. Nation Rivers Authority, Head Office, August 1995.

Neave, H.R. 1981. Elementary Statistics Tables. Unwin Hyman, Ltd. London.46pp.

NERC. 1975. Flood Studies Report. Natural Environmental Research Council, London.

O'Hagan, A. 1994. Bayesian Inference (Vol. 2B of Kendall's Advanced Theory of Statistics). Edward Arnold, London.

Owen, A.B. 1995. Nonparametric Likelihood Confidence Bands for a Distribution Function. Journal of the American Statistical Association, 90, 516-521.

Prescott, P. and Walden, A.T. 1983. Maximum-likelihood estimation of the parameters of the three parameter generalised extreme value distribution from censored samples. Journal of Statistics Computer Simulation.

Raiffa, H. and Schlaifer, R. 1961. Applied Statistical Decision Theory. MIT Press, Cambridge, USA.

Robson, A. J. 1999. Distributions for flood frequency analysis (Chapter 15). *In*: Institute of Hydrology, 1999. Flood Estimation Handbook, 139-152.

Round, C. and Young, A.R. 2001. Low Flows 2000 – Technical Report No 3. Centre for Ecology and Hydrology, Wallingford, OX10 8BB, UK.

Sen, Z. 1980. Regional drought and flood frequency analysis. Theoretical Consideration. Journal of Hydrology, 46, 265-279.

Shahin, M., van Oorschot, H.J.L., de Lange, S.J. 1993. Statistical Analysis in Water Resources Engineering. Applied Hydrology Monographs, 1. Balkema, Rotterdam, Netherlands.

Silverman, B.W. 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Singh, V.P. and Guo, H. 1995. Parameter estimation for 3-parameter generalised pareto distribution by the principle of maximum entropy (POME). Hydrological Sciences Journal, 40 (2), 165-181.

Smakhtin, V.U. 2001. Low flow hydrology: a review. Journal of Hydrology, 240, 147-186.

Smith, J.Q. 1988. Decision Analysis: A Bayesian Approach. Chapman and Hall, London.

Stedinger, J.R., Vogel, R.M. and Foufoula-Georgiou, E. 1992. Frequency analysis of extreme events. In: Handbook of Hydrology, Maidment, D.R. (ed.), chapter 18. McGraw Hill, New York.

Takara, K.T. and Stedinger, J.R. 1994. Recent Japanese contributions to frequency analysis and quantile lower bound estimators. *In*: K.W. Hippel (ed.). Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Vol 1, 217-234. Kluwer Academic Publishers, Netherlands.

Tallaksen, L.M. & Hisdal, H. 1997. Regional analysis of extreme streamflow drought duration and deficit volume. *In*: FRIEND'97-Regional Hydrology: Concepts and Models for Sustainable Water Resource Management. Proceedings of the Third International Conference on FRIEND, held at Postojna, Slovenia, 30 September - 4 October 1997. IAHS Publication no 246. 141-150.

Tallaksen, L.M. 2000. Streamflow drought frequency analysis. *In*: Drought and Drought Mitigation in Europe (ed. By J.V. Vogt & Somma, F), pp103-117. Kluwer Academic Publishers, Dordrecht, Netherlands.

Tanner, M.A. 1993. Tools for Statistical Inference. Springer-Verlag , New York.

Vogel, R.M. and Kroll, C.N. 1990. Low-flow frequency analysis using probability-plot correlation coefficients. J. Water Resources Planning and Management, 115 (3), 338-357.

Zaidman, M.D., Keller, V. & Young, A.R. 2002. Low flow frequency analysis : Recommended Practice. Environment Agency, Bristol, UK. 27pp.

8.2 Citations from Table A1

Caffey, J.E., Orsborn, J.F., Schaake, J.C., Singh, K.P, Wallace, J.R. and Riggs, H.C. 1980. Characteristics of Low Flows. J. Hydraulics Division (Proc ASCE), 106, HY5, 717-731.

Deininger, R.A. and Westfield, J.D. 1969. Estimation of the parameters of Gumbel's third asymptotic distribution by different methods. Water Resources Research, 5 (6), 1238-1243.

Durrans, S.R. 1996. Low-flow analysis with a conditional Weibull tail model. Water Resources Research, 32 (6), 1749-1760.

Gumbel, E.J. 1963. Statistical forecast of droughts. Bull. Int. Association for Scientific Hydrology, 8 (1), 5-23.

Guo, S.L., Kachroo, R.K. and Mngodo, R.J. 1996. Nonparametric kernel estimatation of low flow quantiles. Journal of Hydrology, 185, 335-348.

Gustard, A., Bullock, A. and Dixon, J.M. 1992. Low Flow Estimation in the United Kingdom. IH Report No. 108, Institute of Hydrology, Wallingford, UK

Gustard, A. and Gross, R. 1989. Low Flow Regimes of Northern and Western Europe. In: FRIENDS in Hydrology (Ed. Roald, L., Norsdeth, K. and Hassel, K.A.), IAHS Publication No. 187, 205-212.

Joseph, E.S. 1970. Probability Distribution of Annual Droughts. J. Irrigation and Drainage Division (Proc. ASCE), 96, IR4, 461-474.

Kobold, M. and Brilly, M. 1993. Low flow discharge analysis in Slovenia. *In*: "FRIEND: Flow Regimes from International Experimental and Network Data", IAHS Publication No.221, 119-131.

Krokli, B. 1989. Low flow analysis. In: FRIENDS in Hydrology (Ed. Roald, L., Norsdeth, K. and Hassel, K.A.), IAHS Publication No. 187, 443-451.

Kumar, A. and Devi, R. 1982. Statistical determination of design low flows - a comment. Journal of Hydrology, 58, 175-177.

Lawal, S.A. and Watt, W.E. 1996. Non-zero lower-limit in low flow frequency analysis? Water Resources Bulletin, 32 (6) 1159-1166.

Leppäjärvi, R. 1989. Frequency analysis of flood and low flow. In: FRIENDS in Hydrology (Ed. Roald, L., Norsdeth, K. and Hassel, K.A.), IAHS Publication No. 187, 435-442.

Loganathan, G.V., Kuo, C.Y. and McCormick, T.C. 1985. Frequency analysis of low flows. Nordic Hydrology, 16, 105-128.

Loganathan, G.V., Mattejat, P., Kuo, C.Y. and Diskin, M.H. 1986. Frequency analysis of low flows: hypothetical distribution methods and a physically based approach. Nordic Hydrology, 17, 129-150.

Matalas, N.C. 1963. Probability distribution of low flows. Geological Survey Professional Paper 434-A, US Government Printing Office, Washington, DC.

Nathan, R.J. and McMahon, T.A. 1990. Practical aspects of low-flow frequency analysis. Water Resources Research, 26 (9), 2135-2141.

Pearson, C.P. 1995. Regional frequency analysis of low flows in New Zealand rivers. Journal of Hydrology (NZ), 33 (2), 94-122.

Perzyna, J.G. and Gottschalk, L. 1995. Derived frequency distribution for low flows. *In*: "Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Volume 1 (Extreme Values: Floods and Droughts)", K.W.Hipel (ed.), Kluwer, Dordrecht, 139-149.

Pilon, P.J. 1990. The Weibull distribution applied to regional low flow frequency analysis. *In*: "Regionalization in Hydrology", IAHS Publication No. 191, 227-237.

Polarski, M. 1989. Fitting distributions to annual minimum flows of different durations. *In*: FRIENDS in Hydro logy (Ed. Roald, L., Norsdeth, K. and Hassel, K.A.), IAHS Publication No. 187, 97-104.

Prakash, A. 1981. Statistical determination of design low flows. Journal of Hydrology, 51, 109-118.

Rao, D.V. 1980. Log Pearson Type 3 Distribution: a generalised evaluation. J. Hydraulics Division, Proc ASCE, 106, HY5, 853-872.

Riggs, H.C. 1965. Estimating probability distributions of drought flows. Water & Sewage Works, 112 (5), 153-157.

Tasker, G.D. 1987. A comparison of methods for estimating low flow characteristics of streams. Water Resources Bulletin, 23 (6), 1077-1083.

Vogel, R.M. and Kroll, C.N. 1990. Low-flow frequency analysis using probability-plot correlation coefficients. J. Water Resources Planning and Management, 115 (3), 338-357.

Vogel, R.M. and Kroll, C.N. 1990. Generalized low-flow frequency relationships for ungauged sites in Massachusetts. Water Resources Bulletin, 26 (2), 241-253.

Wang, S.X. and Singh, V.P. 1995. Frequency estimation for hydrological samples with zero values. J. Water Resources Planning and Management, 121 (1), 98-108.

APPENDIX 1: THEORETICAL ASPECTS

1.1 Extreme Value Theory

The subset of statistical theory, known as "extreme value theory relates to the statistical behaviour of maxima and minima of large collections of (not necessarily independent) random variables. This theory strongly suggests that the GEV (Generalised Extreme Value) family of distributions will be appropriate for this type of data. The following discussion, given in terms of maxima (since this is the case usually treated in the theory), outlines the main points of extreme value theory.

Suppose Z_M is defined for a sample size M, by

$$Z_M = \max(X_1, \ldots, X_M), \qquad \text{(A1.1)}$$

where $\{X_1, \ldots, X_M\}$ is a collection of random variables such that the notion of an increasing sample size is appropriately defined. The distribution function $F_M(z)$ of Z_M is then such that there are sequences of constants a_M and b_M , for which the re-scaled random variable Y_M has a distribution function that is arbitrarily close to one of the standardised GEV distributions. The random variable Y_M is defined by:

$$Y_M = (Z_M - a_M)/b_M,$$

where

$$\begin{aligned} \Pr\{ Y_M \le y \} &= \Pr\{ Z_M \le b_M y + a_M \} = F_M(b_M y + a_M), \\ &\to \exp\{ - (1- ky)_+^{1/k} \} \qquad \text{as } M \to \infty. \qquad \text{(A1.2)} \end{aligned}$$

Here k is the shape parameter of the GEV distribution and the case k = 0 is interpreted via the limit as $k \to 0$:

$$\begin{aligned} (1- ky)^{1/k} &= \exp\{ (1/k) \log(1 - ky) \}, \\ &= \exp[-(1/k) \{ ky + (ky)^2/2 + \ldots \}], \\ &= \exp[-y - ½ky^2 - \ldots], \\ &\to \exp(-y) \ \text{ as } k \to 0. \end{aligned}$$

Note also that the subscript "+" in expression (A1.2) has the following interpretation:

$$\begin{aligned} x_+ &= x, \text{ if } x \ge 0, \\ &= 0, \text{ if } x < 0. \end{aligned}$$

The result (A1.2) holds given certain conditions on the basic random variables $\{X_1, \ldots ,X_M\}$. The random variable need not be independent and identically distributed. However

they must satisfy conditions which ensure that the variables approach being independent as the distance between them in the sequence grows and which ensures that the maximum value is not dominated by one random variable (or a finite subset of random variables) in the sequence. In addition, conditions are required on the distribution functions. These condition are, firstly, to ensure that there is no discrete component at the upper bound of the random variables which would lead to the limiting distribution being concentrated at that point and, secondly, to ensure that the distribution function is not 'log-tailed'.
Thus equation (A1.3) is acceptable while (A1.4) is not.

$$F_X(x) \sim 1 - x^{-p} \qquad \text{as } x \to \infty \ (p>0)$$

(A1.4)

$$F_X(x) \sim 1 - (\log x)^{-p} \qquad \text{as } x \to \infty \ (p>0) \qquad \text{(A1.5)}$$

One aspect of the statistical theory mentioned above is that if the properties of the basic random variables $\{X_1, \ldots ,X_M\}$ are known, then appropriate values for the sequences of constants a_M and b_M can be found (these are not uniquely determined), and the value of the shape parameter k can be fixed. This aspect of the theory is of little use since the required properties are not known. However, the result implies that a single random variable Y, which is constructed as the maximum of an effectively large number of underlying data-values, should be such that there will be constants a and b for which (Z-a)/b is close to being a standard GEV distribution. This means that that Z can be modelled as having a GEV distribution with three unknown parameters a,b and k.

A second set of theoretical results, derived by Cohen (1982), indicates that, even if the shape parameter of the asymptotic distribution can be determined theoretically, it may still be best to treat it as being unknown when actually fitting distributions in applications. These theoretical results imply that a better sequence of approximations to the distribution function can be obtained by allowing the shape parameter k in expression (A1.2) to vary with M, as well as the scaling parameters a and b: thus, when fitting a distribution, k should not be set to its asymptotic value.

For random variables related to minima, the above discussion can be applied to minus one times the required random variables. This leads to the conclusion that if Z is constructed as the minimum of an effectively large number of underlying data-values, then there should be values a and b such that

$$F_Z(a+bz) \approx 1 - \exp\{ - (1+ kz)_+^{1/k} \} \qquad \text{(A1.6)}$$

where the right hand side of this expression is the distribution function of the standardised GEV distribution for minima.

1.2 Graphical Estimation with 2 or more Parameters.

Graphical techniques can also be applied for more general types of distribution beyond those in the location-scale class. For example, suppose that the idea of the standard distribution F_S, above, is extended to allow it to have a "shape" parameter γ. Then the distribution and quantile functions are $F_S(x;\ \gamma)$ and $x_S(p;\ \gamma)$, respectively. The idea is to form an initial estimate of the extra parameter γ and then to proceed as with the one-parameter case to estimate the parameters α and β, substituting the estimated value of γ where required. One approach is to create a statistic whose distribution does not depend upon α and β, but only upon γ and to use this to specify the estimate for γ. In particular, let i,j and k be any three integers satisfying $1 \le i<j<k \le N$: then, with the location-scale assumption:

$$\{x_{(k)}-x_{(j)}\}/\{x_{(j)}-x_{(i)}\} \approx \{x_S(p_k;\ \gamma)-x_S(p_j;\ \gamma)\}/\{x_S(p_j;\ \gamma)-x_S(p_i;\ \gamma)\}. \qquad (A1.7)$$

Then the estimate of γ is specified to be the value at which exact equality occurs in equation (A1.7): this value may be found either graphically, by plotting the right hand side of (A1.7) as a function of γ, or by a simple root-finding search procedure. The values chosen for i,j and k should be such that $x_{(j)}$ is close to the middle of the ranked values and $x_{(i)}$ and $x_{(k)}$ are close to, but not right at the lower and upper ends of the range: a possible choice would be to take them to be near the 20% and 80% points of the ordered sample.

1.3 The Method of Maximum Likelihood

The first step in defining the maximum likelihood estimator for a problem is to define the **likelihood function** for the recorded observations. This arises directly from the statistical model being fitted, and can be determined for independent or dependent data. If the observations are of essentially continuous random variables the likelihood function is given by the probability density function evaluated at the observations: thus

$$l(\theta) = f\,(x_1, x_2, \ldots , x_M). \qquad (A1.8)$$

Where the model has been modified to take into account discretization effects or to represent the presence of zero flows, the likelihood function adopted should also relate to the modified model. For example, for a model with parameters denoted by θ applied to observations of random variables $(X_1,\ X_2,\ \ldots,\ X_M)$, which happen to have taken the essentially discrete values $(x_1, x_2, \ldots , x_M)$ the likelihood function $l(\theta)$ is given by

$$l(\theta) = \Pr(\ X_1 = x_1, X_2 = x_2, \ldots , X_M=x_M). \qquad (A1.9)$$

While it may seem that the definitions in Equations (A1.8) and (A1.9) are on incommensurate scales, since the first involves a probability density and the second an ordinary probability, they are similar since in both cases they represent the derivative of the probability measure with respect to its supporting measure. Similarly if, for a particular estimation problem, there are grounds for restricting the parameters to obey certain constraints, then such constraints should be used to restrict the range of parameters over which the likelihood function is considered.

It is usual to work with the logarithm of the likelihood function, the **log-likelihood function** $L(\theta)$, given by

$$L(\theta) = \log\{ l(\theta) \}. \qquad \text{(A1.10)}$$

There are in fact arbitrary factors involved in the definition in Equation (A1.10), related to the possibility of changing the units of measurement in which each observation is recorded: these factors appear as additive constants in Equation (A1.10) and are unimportant. In fact, it is possible to drop any collection of terms that do not depend upon the unknown parameters from an expression for the log-likelihood. In writing down an expression for the log-likelihood function of a model, it is important to keep track of the range of validity of the expressions. For example, if the model is such that the observations $\{x_i\}$ are bounded below by a parameter, b, then the likelihood function will be zero for values of b greater than the smallest observation. In this case, the interesting range for the likelihood or log-likelihood functions would be $b < \min \{x_i\}$. If, for a particular estimation problem, there are grounds for restricting the parameters to obey certain constraints, then such constraints should be used to restrict the range of parameters over which the likelihood function is considered.

The **maximum likelihood estimate** for a particular problem is defined to be the value that maximises the likelihood function, or equivalently, the log-likelihood function. Here the maximisation of the log-likelihood function is restricted to the region of the set of parameters defined by any constraints on the parameter values derived from either modelling or more general considerations. The most directly interpretable justification for maximum likelihood estimation as a general estimation method can be framed as follows. Varying the parameters of the model can be thought of as shifting and changing the horizontal scale of the probability density function that is derived from the model. A "good" model should be one which has a high density function in regions of the observation space where there many observed data points and a low density where there are few data points. The objective function used to define maximum likelihood estimation, in a sense, measures the merit of a given set of model parameters as a match to the observed data. This merit value can be thought of as having two effects: to reward models which have a high density function at most of the observed data points and to penalise models which have an extremely low, or zero, density function at any of the observed data points. In contrast to other objective-function based methods of estimation, the objective function to be optimised for maximum-likelihood estimation is not based on constructing a distance between a set of model-related quantities and a "target" set derived from the observed data, for which a zero-distance would represent a "perfect fit".

In some circumstances, it is convenient to treat maximum likelihood estimation as being equivalent to solution of the **likelihood equations**. For this approach to be applicable, it is necessary that the likelihood function should be differentiable with respect to the parameters. If this applies, then the maximum likelihood estimate is potentially defined as the solution to the likelihood equations:

$$L'() = 0, \qquad \text{(A1.11)}$$

where $L'(\theta)$ denotes the (vector of) derivatives of the log-likelihood function. While this looks relatively simple, in order to use the likelihood equations to define the maximum likelihood estimate, it is strictly necessary to ensure:

(i) that a given solution of Equation (A1.11) corresponds to a maximum and not to a minimum or other turning point of the function;

(ii) that the given solution lies inside the allowable range of parameters;

(iii) that there are no points on the boundary of the range of parameters with a larger value of the likelihood function.

However the likelihood equations are useful because, in certain cases, they can be manipulated so as to provide a simple, explicit expression for the maximum likelihood estimate.

Various problems can arise when attempting to apply maximum likelihood estimation in practice. These can be categorised as follows:

Non-existence of maximum.

Here the allowable region of the parameter space is not closed and, for any parameter vector within the set, it is always possible to find another vector (still in the set) for which the likelihood function is higher. In cases where a numerical search procedure was being applied to maximise the likelihood function, this problem might lead to trial values of the parameters drifting off to infinity in one or more dimensions. There can also be situations in which the likelihood function continues to grow as the parameter vector approaches a finite boundary point of the allowable parameter region. In some instances the problem can be overcome by effectively forming the allowable parameter region into a closed set: this involves extending the formulation of the model so that points on the boundary of the parameter region can be represented within the model, possibly as special forms of the model. There are many well-known cases in which a particular distributional form is a special limiting case of a more general distribution that has an apparently rather different structure: for example, the exponential distribution (which has a finite non-zero density at zero) occurs as the limiting form of two classes of gamma distributions - those with zero density at the origin and those with infinite density at the origin.

Several local maxima.

Here, cases in which the objective function takes identical values at different local maxima are excluded, being dealt with in the next paragraph. Excluding such cases means that the maximum likelihood estimate is well defined (as the global maximum), so that from a theoretical stand point no problem arises. There can however be practical problems relating to finding the maximum likelihood estimate. If a numerical search procedure is used, a strategy designed to overcome this possible problem may need to be devised whereas, if an algebraic derivation of the maximum likelihood estimate is attempted, this would lead to several candidate solutions for which further tests would need to be made.

Several equal-valued maxima.

The problem of having several local maxima at which the likelihood function takes the same value is commonly associated with the problem of parameter identifiability and this can usually be avoided by placing suitable restrictions on the parameters. For example, suppose that the model is that observations are independent and identically distributed from a two component mixture model, based on a known standardised density function g, and with an unknown mixing proportion p. Then the modelled density for an individual observation would be

$$f(x) = p\, b^{-1}\, g\{ (x-a)b^{-1}\} + (1-p)\, d^{-1}\, g\{ (x-c)d^{-1}\}. \qquad \text{(A1.12)}$$

It is clear that, for any parameter vector $\theta_0 = (p_0, a_0, b_0, c_0, d_0)^T$, the model with parameter vector $\theta_1 = (p_1, a_1, b_1, c_1, d_1)^T$ will have the same likelihood value as that for θ_0 if

$$p_1 = (1-p_0)\ ;\ a_1=c_0\ ;\ b_1=d_0\ ;\ c_1=a_0\ ;\ d_1=b_0.$$

This problem can be overcome by imposing an appropriate constraint on the parameters, for example, $a \le c$. Note that this constraint does not entirely overcome all problems in this example, since the density is identical for all values of p if $a = c$ and $b=d$.

Non-distinct maxima.

A problem related to that in the last paragraph can arise if the likelihood function does not have distinct maxima: there may be a set, with dimension larger than zero, of points which share the same likelihood value and for which no point can be found having a larger likelihood. Such a problem is again related to parameter identifiability. It can arise if there are insufficient data to estimate the full set of model parameters. Typically, one requires there to be at least one data point for each separate model parameter, and there is an implication that the problem will be overcome once enough data are accumulated. However, as in the example at the end of the previous paragraph, there can be cases in which a particular structure of model is not identifiable and no amount of extra data will make it so. Thus, in the case of the mixture model in which the sub-populations are identical, Equation (A3.5) with a=c and b=d, the parameter p is not identifiable: the likelihood function would be constant for all values of p.

Methods for finding maximum likelihood estimates can be categorised as follows:

(a) *Graphical or tabulation methods*.

If the dimension of the parameter space to be explored is only one or two dimensional, it can be convenient to effect the optimisation by constructing a plot of the function, either as a line-plot or as a contour plot, and then judging visually the location of the optimum. Underlying this approach is the evaluation of the likelihood function on a regular grid effectively covering a reasonable part of the parameter space, so that the possibility of there being multiple optima, or other problems, is being considered at the same time. While such "non-technical" approaches are not often adopted (more fully automatic methods are preferred), plots of the likelihood function can be extremely useful since they implicitly contain information about the uncertainty associated with the estimates of the parameters. In particular, a good confidence region for the parameters can be constructed as the region of parameters for which the values of the log-likelihood function $L(\theta)$, satisfy the following closeness bound with respect to the value at maximum likelihood estimate:

$$L(\theta) > L() - ½\,\chi^2(p;d), \qquad \text{(A1.13)}$$

where $\chi^2(p;d)$ denotes the quantile function at probability p of the chi-squared distribution with d degrees of freedom and where d is the dimension of the parameter space being optimised (i.e. d = 1 or 2 for line-plots or contour plots respectively). A 95% confidence region is therefore given by $L(\theta) > L() - 1.92$, or by $L(\theta) > L() - 3.00$, for the 1- or 2-dimensional cases, respectively. If a line-plot or contour-plot of the objective function has already been prepared, it is clearly easy to read-off from this the interval covered by the confidence region.

(b) *Numerical procedures for optimisation*.

Various widely available subroutine-libraries contain procedures for searching for the optimum of a objective function. The more sophisticated of these can deal with the imposition of bounds on the parameter values as part of the optimisation problem. The requirements of these procedures for problem-specific programming can range from supplying just the function-value at any given set a parameters, to supplying the function-value together with first and second derivatives. Certain routines may automatically provide for a search for a global optimum. Where a procedure implements a search procedure based on a limited local search for the optimum, it would be best to adopt a scheme in which the local search is started at a number of different locations, unless experience with similar problems has shown this to be unnecessary.

(c) *Numerical procedures for root-finding*.

As for numerical optimisation, standard subroutine-libraries often contain procedures for root-finding: these could be applied to the likelihood equations as a means of finding the maximum-likelihood estimates. However, as a general strategy, this course is to be avoided since, in order to check whether a root corresponds to a local maximum, or to check which of a number of potential solutions gives the largest likelihood, values of the log-likelihood function need to be evaluated. One of the apparent benefits of the root-finding approach is that, after a certain amount of algebraic manipulation, the likelihood equations can often be put into a form considerably simpler than the log-likelihood function, and thus less programming effort may be required. This benefit is eliminated if the log-likelihood must be evaluated for checking purposes.

(d) *Algebraic solutions.*

For various standard distributing-fitting problems it is possible to manipulate the likelihood equations in order to derive a set of explicit expressions for the maximum likelihood estimates. While this is, of course, extremely convenient, such explicit solutions are typically not available for less standard problems. However, it would be sensible, when faced with a new problem, to attempt to find such a solution since, even if unsuccessful overall, the work involved can contribute to reducing the computational complexity of the work required in applying other methods (for example, by reducing the dimension of the parameter space to be searched by numerical optimisation): see the next paragraph.

(e) *Mixed approaches*.

As noted above, it is common to manipulate the likelihood equations in order to arrive at a partial solution to the overall problem. Thus various steps involving combination and cross-substitution between the individual likelihood equations are attempted, with the intention of reducing the number of "free" parameters. Then, either a numerical optimisation or a numerical root-finding procedure is applied. The steps of combination and cross-substitution between the equations can be rather uncontrolled, and it may be better practice to adopt the following more formal procedure. It is convenient to introduce the idea of a **profile likelihood** here. Suppose that the overall parameter vector, θ, can be divided into two parts, α and β. Suppose that the likelihood function is such that, taking α as having a fixed value, it is easy to find the value of β which maximises the likelihood function: this would yield the function (α), representing the maximum likelihood estimate of β for the given value of α. If the overall log-likelihood function is denoted by $L(\alpha, \beta)$, the profile log-likelihood function for α, $L_P(\alpha)$, is given by

$$L_P(\alpha) = L(\alpha,(\alpha)). \qquad \text{(A1.14)}$$

Thus the idea is to substitute, into the log-likelihood function, the expression for β which would maximise the log-likelihood for a fixed value of α, and then to maximise the result as a function of α. The advantage of the approach via profile likelihoods is that these functions can be treated in much the same way as the ordinary likelihood function. In particular, the approach to obtaining confidence regions outlined under (a) above can be applied to the profile likelihood: in this case it would provide a confidence region for α. In this case, when Equation (A1.14) is applied, the value for the degrees of freedom, d, is the dimension of the reduced parameter vector α. If the profile likelihood approach can be applied to reduce the number of free parameters to only one or two, then plotting the profile likelihood as a line-plot or contour-plot can provide a convenient way of checking from possible multiple local optima, or others of the problems described above.

As mentioned at the beginning of this sub-section, results are available from the statistical theory of maximum likelihood estimation which deal with the properties of these estimates. These results show that, provided that certain conditions hold and provided that the sample size is large enough, and assuming also that the model being fitted actually does hold, the maximum likelihood estimates will be close to the notional true values of the parameters, θ_0. Furthermore, the errors in these estimates, $-\theta_0$, can be treated as if this vector were (multivariate) Normally distributed with zero mean and covariance matrix equal to the inverse of the observed precision matrix j(), where

$$j(\theta) = - L''(\theta) \qquad \text{(A1.15)}$$

and where $L''(\theta)$ denotes the matrix of second derivatives of the log-likelihood function and where, specifically, $j(\theta)$ is evaluated at the maximum likelihood estimate . This result can be used to create confidence regions for the parameter vector and for subsets of them, and these can be extended to deal with non-linear functions of the parameters. Although such regions are undoubtedly convenient, better statistical performance can be obtained by making use of the result that

$$\{ L() - L(\theta_0) \} \sim \tfrac{1}{2} \chi^2(d), \qquad \text{(A1.16)}$$

where "~" means "is distributed as" and where $\chi^2(d)$ denotes a random variable from the chi-squared distribution with d degrees of freedom. This result is the basis of the confidence region suggested in Equation (A3.6) and it can be applied also to the profile log-likelihood function to derive confidence regions for subsets of the parameter vector, as noted following Equation (A1.14). The distributional results quoted here are at best approximate, where the approximation error will be reduced as the sample size becomes larger. As noted, various conditions are needed to justify these results, among which are the requirement that the number of parameters in the model should not increase as the sample size increases and that the true parameter value, θ_0, should not lie on the boundary of the parameter space being searched. While other conditions are also required for a full theoretically-based justification, it is usually assumed that the results can be applied provided that the maximum-likelihood estimates can be found without problems of the kind outlined above related to solving the optimisation problem numerically.

1.4 Approaches for Estimating Equations

Estimating equations are used in the range of techniques that can be broadly classed as 'exact –matching ' techniques. Suppose S denotes a vector of sample-derived estimates of statistical properties of the underlying population, and θ is the vector of parameters indexing the family of distributions. If $s(\theta)$ denotes the vector-function describing how the values of these properties for the family of distributions vary with θ, the estimated parameter values are defined to be equal to the solution to the set of estimating equations:

$$S = s(\theta). \qquad \text{(A1.17)}$$

Here, it is understood that the solution sought is one that lies within an allowable set of parameter values. There are two main ways of arriving at a set of estimating equations appropriate to a particular problem. Either a set of sample statistics in the vector S, would be used to deduce a corresponding vector of "population-values", $s(\theta)$, or given the vector $s(\theta)$, a reasonable set of sample statistics, S, would be sought, which in some sense measure the same sort of properties as measured by $s(\theta)$.

In the first type of approach, starting with S, the function $s(\theta)$ can be given a formal specification in the following ways. One can consider the sample statistic in the vector S as a function of the sample size, N say, and of the vector of observed data $X=(X_1, X_2, \ldots, X_N)^T$. In this sense, S is a function of the sample size N and the data-values X: $S = S_N(X)$. One can consider evaluating the vector of statistics S for a sample of data $Z(\theta)$ which arises from a given member, indexed by θ, of the family of distribution being fitted: the value of S in this case would be $S = S_N\{Z(\theta)\}$. Here, N is again a sample size, but this need not be the same as that for the original sample. In one variant of the approach, it is convenient to consider larger and larger samples sizes in order to define the "population value" of the sample statistics for the selected member of the family of distributions. Thus,

$$s(\theta) = \lim_{N \to \infty} S_N\{Z(\theta)\}, \qquad \text{(A1.18)}$$

where the limit is assumed to exist in an appropriate probabilistic sense. In some situations it may be necessary to introduce a scaling factor, as a function of the sample size N, in order to ensure that the limit in Equation (A1.18) has a non-trivial value. A second variant of the approach defines $s(\theta)$ by replacing the definition in Equation (A1.18) with one which considers only a fixed sample size and which uses a measure of central location to define the "population value": for example,

$$s(\theta) = s_N(\theta) = E\,[\,S_N\{Z(\theta)\}\,]. \qquad \text{(A1.19)}$$

Although alternative measures of location (such as the median) might be used, these are considerably more difficult to work with than the expectation, and hence are not much used in the present context. As made explicit in Equation (A1.19), this approach leads to the right hand side of Equation (A4.1) being formally a function of the sample size. Both of these variants rely on the argument that S is "a good estimate" of $s(\theta)$ for their intuitive justification. Hence, if one believes that the "unbiased" derivation via Equation (A1.19) means that S is a better estimate of $s_N(\theta)$, from that equation, than it is of $s(\theta)$ from Equation (A1.17), one might hope that the parameter estimates obtained from the "unbiased" estimating equations would be correspondingly better. However, this is not necessarily true, and the parameter estimates would, in general, not themselves be unbiased. Where several variants of similar sets of estimating equations are available, these would usually need to be compared on the basis of simulation experiments.

The second type of approach, starting with $s(\theta)$, can arise in two ways. The first of these occurs as a follow-up to the first type of approach, in which one has an initial S and $s(\theta)$ and seeks an improved version of S, S^* say, with "better" properties as an estimate of $s(\theta)$. Such improved versions of S can sometimes be constructed in straightforward ways. For example, one might set $S^* = kS$ and choose the factor k according to some requirement: that S^* should be unbiased, or that S^* should have minimum mean square error. Alternatively, if S is unbiased, one might construct S^* as a U-statistic derived from S (Cox & Hinkley, 1974). This leaves the case where one starts with $s(\theta)$ and requires to form an estimate S for this vector. If this were always easy to achieve, one would just start with $s(\theta)=\theta$. Instead, the vector $s(\theta)$ is often specified as a functional of the candidate distribution function:

$$s(\theta) \quad = s[\,F(\,.\,;\theta)\,]. \qquad \text{(A1.20)}$$

For example, vector $s(\theta)$ might contain as individual elements, measure of location, spread and shape (e.g. mean, variance and skewness). Then one possible sample statistic can be derived by applying this functional to the sample distribution function:

$$S_N \quad = s[\,F_N(\,.\,)\,]. \qquad \text{(A1.21)}$$

Where the functional can be expressed in the form of an expectation of some function of the data-values, this can lead to a direct sample-average expression for the corresponding sample statistic. Once an initial estimate is available, it might be possible to improve it in the ways indicated at the beginning of this paragraph. However, it should be remembered that, as indicated elsewhere in this section, "optimal" properties of the sample estimates S are not necessarily carried over into "optimal" properties for the parameter estimates, nor into "optimal" properties for other quantities being estimated.

For certain purposes it can be helpful to slightly extend the notion of estimating equations, as described above, to encompass **estimating functions**. In the notation used above, the estimating equations can be written in the form

$$S(X) = s(\theta), \qquad \text{(A1.22)}$$

which makes explicit the data-dependence of the left-hand-side and the parameter-dependence of the right-hand-side. The slightly extended version of this is to consider a vector-valued estimating function $g(x,\theta)$ and to define the parameter estimates as being the solution of the equations

$$g(X, \theta) = 0. \qquad \text{(A1.23)}$$

Equations of this form can be called estimating equations without too much confusion.

It is clear that a necessary condition for the estimating equation approach to provide a reasonable estimation procedure is that the number of different statistical properties used in the procedure, which in turn specifies both the dimension of S and the number of separate equations in an element-by-element expansion of Equation (A1.20), must be the same as the number of parameters being estimated. Of course, this condition is not enough to ensure that the set of equations does have a solution or that, if it does, the solution is unique. Similar conclusions apply to estimating functions.

To a considerable extent the choice of which statistical properties are used to construct the estimating equations is rather arbitrary. The choice for S or $s(\theta)$ represents a set of statistics which will be "reproduced" by the estimation technique: for example, samples generated from the fitted family of distributions would have properties exactly matching those of the observed data sample, in either a large-sample sense (if the approach via Equation (A1.18) is used), or in an expectation sense (if the approach via Equation (A1.19)) is used). Hence it can be argued that the statistics used should be "important" ones to be reproduced in some practical sense, reflecting what the fitted distribution will be used for. However, in practice, this type of argument does not tend to be helpful in choosing statistics for use in the estimating equations.

APPENDIX 2:
SUMMARY OF GAUGING STATION AND CATCHMENT CHARACTERISTICS

Table A2.1: Suitability of 'AA' graded stations

Station	River	Gauge Name	Measuring Authority	Suitability	Measuring Authority's Comment (if received)
6008	Enrick	Mill of Tore	SEPA-North		
7001	Findhorn	Shenachie	SEPA-North		
7002	Findhorn	Forres	SEPA-North		
8006	Spey	Boat o' Brig	SEPA-North	Suitable	Artificial influence by major upper catchment transfer estimated at 15 m^3/s over the year (Q95 18.75).
8009	Dulnain	Balnaan Bridge	SEPA-North		
9001	Deveron	Avochie	SEPA-North	Suitable	Influence 0% of MF and 2% of Q95
9002	Deveron	Muiresk	SEPA-North	Suitable	Influence 0% of MF and 1% of Q95
9003	Isla	Grange	SEPA-North	Suitable	Artificial influence (discharge from wastewater plant) would be 1% of MF and 3% of Q95.
11002	Don	Haughton	SEPA-North		
11003	Don	Bridge of Alford	SEPA-North		
12001	Dee	Woodend	SEPA-North		
12003	Dee	Polhollick	SEPA-North		
12005	Muick	Invermuick	SEPA-North		
12006	Gairn	Invergairn	SEPA-North	Suitable	Artificial Influence by small upstream abstraction (15l/s) would be o% MF and 2% Q95
14001	Eden	Kemback	SEPA-North	Suitable	CEH assessment is reasonable, support grade A assessment
17005	Avon	Polmonthill	SEPA-East	Suitable	Support grade A classification, influence 12% MF, 14% Q95
18001	Allan Water	Kinbuck	SEPA-East	Suitable	Support grade A classification
18005	Allan Water	Bridge of Allan	SEPA-East	Suitable	Support grade A classification
18008	Leny	Anie	SEPA-East	Suitable	Support grade A classification
19002	Almond	Almond Weir	SEPA-East	Suitable	Support grade A classification
19004	North Esk	Dalmore Weir	SEPA-East	Suitable	Support grade A classification
19011	North Esk	Dalkeith Palace	SEPA-East	Unsuitable	Actually grade B, due to control being a natural rock, gravel bar which is insensitive at low flows
20001	Tyne	East Linton	SEPA-East	Suitable	Support grade A classification
20003	Tyne	Spilmersford	SEPA-East	Suitable	Support grade A classification
20005	Birns Water	Saltoun Hall	SEPA-East	Suitable	Support grade A classification
20007	Gifford Water	Lennoxlove	SEPA-East	Suitable	Support grade A classification
21006	Tweed	Boleside	SEPA-East	Suitable	Suitable for use
21012	Teviot	Hawick	SEPA-East	Suitable	Suitable for use
21013	Gala Water	Galashiels	SEPA-East	Suitable	Suitable for use
21015	Leader Water	Earlston	SEPA-East	Suitable	Suitable for use
21017	Ettrick Water	Brockhoperig	SEPA-East	Suitable	Suitable for use
21018	Lyne Water	Lyne Station	SEPA-East	Unsuitable	Unsuitable, compensation discharges used to augment low flows
21022	Whiteadder Water	Hutton Castle	SEPA-East	Unsuitable	Unsuitable, compensation discharges used to augment low flows
21023	Leet Water	Coldstream	SEPA-East	Unsuitable	Algal growth occurs in periods of low flow, resulting in marked diurnal variation in river levels.
21024	Jed Water	Jedburgh	SEPA-East	Suitable	Suitable for use

Station	River	Gauge Name	Measuring Authority	Suitability	Measuring Authority's Comment (if received)
21027	Blackadder Water	Mouth Bridge	SEPA-East	Suitable	Suitable for use
22001	Coquet	Morwick	EA –North East		
22009	Coquet	Rothbury	EA –North East		
23004	South Tyne	Haydon Bridge	EA –North East		
23006	South Tyne	Featherstone	EA –North East		
23008	Rede	Rede Bridge	EA –North East		
23011	Kielder Burn	Kielder	EA –North East		
24004	Bedburn Beck	Bedburn	EA –North East		
25006	Greta	Rutherford Bridge	EA –North East		
27034	Ure	Kilgram Bridge	EA –North East		
27041	Derwent	Buttercrambe	EA –North East		
27049	Rye	Ness	EA –North East		
27050	Esk	Sleights	EA –North East		
27054	Hodge Beck	Cherry Farm	EA –North East		
27055	Rye	Broadway Foot	EA –North East		
27057	Seven	Normansby	EA –North East		
28008	Dove	Rocester Weir	EA -Midlands	Suitable	Suitable for use
28031	Manifold	Ilam	EA -Midlands	Suitable	Suitable for use
33014	Lark	Temple	Anglian		
33019	Thet	Melford Bridge	EA- Anglian		
34002	Tas	Shotesham	EA- Anglian		
34003	Bure	Ingworth	EA- Anglian		
34006	Waveney	Needham Mill	EA- Anglian		
34007	Dove	Oakley Park	EA- Anglian		
36002	Glem	Glemsford	EA- Anglian		
36005	Brett	Hadleigh	EA- Anglian		
36007	Belchamp Brook	Bardfield Bridge	EA- Anglian		
38022	Pymmes Brook	Silver St, Edmonton	EA -Thames	Unsuitable	Not suitable, < 80% urban, baseflow rising steadily due to artificial influences
39016	Kennet	Theale	EA -Thames	Suitable	Very consistent low flow behaviour
39028	Dun	Hungerford	EA -Thames	Suitable	Consistent low flow behaviour
39054	Mole	Gatwick Airport	EA -Thames	Unsuitable	Not suitable, current rating is underestimating flows and a side weir was built in 1984.
42010	Itchen	Highbridge + Allbrook	EA -Southern	Unsuitable	Drowns out in summer due to weed growth
43008	Wylye	South Newton	EA – South West	Unsuitable	Significantly impact by abstraction, impact > 10%.
47008	Thrushel	Tinhay	EA – South West		
47009	Tiddy	Tideford	EA – South West		
48010	Seaton	Trebrownbridge	EA – South West		
51001	Doniford Stream	Swill Bridge	EA – South West		
52003	Halse Water	Bishops Hull	EA – South West	Unsuitable	Wide flumes so probably insensitive at low flows
52004	Isle	Ashford Mill	EA – South West	Unsuitable	Weed growth resulting in uncertainty about the calculated flows
52010	Brue	Lovington	EA – South West	Unsuitable	Weed growth resulting in uncertainty about the calculated flows

Station	River	Gauge Name	Measuring Authority	Suitability	Measuring Authority's Comment (if received)
53005	Midford Brook	Midford	EA – South West	Unsuitable	Weed growth resulting in uncertainty about the calculated flows
53006	Frome (Bristol)	Frenchay	EA – South West	Unsuitable	Possible weed growth
53013	Marden	Stanley	EA – South West	Unsuitable	Possible weed growth
53026	Frome (Bristol)	Frampton Cotterell	EA – South West	Unsuitable	Wide flume for size of low flow so probably insensitive at low flows
54025	Dulas	Rhos-Y-Pentref	EA -Midlands	Unsuitable	Unsuitable – gravel buildup!
54034	Dowles Brook	Dowles	EA -Midlands	Suitable	Suitable for use
55013	Arrow	Titley Mill	EA - Wales	Unsuitable	Unsuitable, large PWS abstraction with high daily rate which can take up to 52 days per year. Large influence at medium-low flows.
55014	Lugg	Byton	EA – Wales	Suitable	Influence about 16% if licensed rates used, but suitable for use as actual abstractions are much less than licensed
55016	Ithon	Disserth	EA – Wales	Suitable	Natural upland catchment – minimal artificial influence
55026	Wye	Ddol Farm	EA - Wales	Suitable	Natural mountain upland catchment – minimal artificial influence
55028	Frome	Bishops Frome	EA - Wales	Unsuitable	Unsuitable, impact between 3.5% (current annual rates) and 15% maximum licensed daily rates.
55029	Monnow	Grosmont	EA – Wales	Unsuitable	Unsuitable, large PWS boreholes and SI in lowland part.
56013	Yscir	Pontaryscir	EA – Wales	Suitable	Natural catchment – minimal artificial influence
56015	Olway Brook	Olway Inn	EA – Wales	Unsuitable	Too much artificial influence at low flows.
58005	Ogmore	Brynmenyn	EA – Wales	Unsuitable	Trend due to cessation of mining? Silt build up at low flows
58006	Mellte	Pontneddfechan	EA - Wales	Unsuitable	Ystradfellte Reservoir intercepts about 20% of the flow, which have significant impact on flow regimes during autumn refill.
58009	Ewenny	Keepers Lodge	EA – Wales	Suitable	Minimal influence, but changes at high flows due to urbanisation, road projects etc.
60002	Cothi	Felin Mynachdy	EA – Wales	Suitable	Support CEH classification
60006	Gwili	Glangwili	EA – Wales	Suitable	Support CEH comments
66011	Conwy	Cwm Llanerch	EA - Wales	Unsuitable	Unsuitable, not considered a good station at low flows.
68005	Weaver	Audlem	EA – NW	Unsuitable	Actually quality Grade B, due to weed growth in summer
71011	Ribble	Arnford	EA – NW	Suitable	Suitable for use
72004	Lune	Caton	EA – NW	Suitable	Suitable for use
72005	Lune	Killington New Bridge	EA – NW	Suitable	Suitable for use
73008	Bela	Beetham	EA – NW	Unsuitable	Heavy weed growth in summer, otherwise OK
76014	Eden	Kirkby Stephen	EA – NW	Unsuitable	Deterioration between 1979-1989, OK thence onward.
77002	Esk	Canonbie	SEPA-West		
77003	Liddel Water	Rowanburnfoot	SEPA-West		
78003	Annan	Brydekirk	SEPA-West		
78004	Kinnel Water	Redhall	SEPA-West		
78005	Kinnel Water	Bridgemuir	SEPA-West		
79003	Nith	Hall Bridge	SEPA-West		
79004	Scar Water	Capenoch	SEPA-West		
79005	Cluden Water	Fiddlers Ford	SEPA-West		
79006	Nith	Drumlanrig	SEPA-West		
80001	Urr	Dalbeattie	SEPA-West		
83004	Lugar	Langholm	SEPA-West		
83005	Irvine	Shewalton	SEPA-West		
83010	Irvine	Newmilns	SEPA-West		
84003	Clyde	Hazelbank	SEPA-West		

Station	River	Gauge Name	Measuring Authority	Suitability	Measuring Authority's Comment (if received)
84004	Clyde	Sills	SEPA-West		
84005	Clyde	Blairston	SEPA-West		
84012	White Cart	Hawkhead	SEPA-West		
84014	Avon Water	Fairholm	SEPA-West		
84015	Kelvin	Dryfield	SEPA-West		
84018	Clyde	Tulliford Mill	SEPA-West		
84020	Glazert Water	Milton of Campsie	SEPA-West		
85004	Luss Water	Luss	SEPA-West		

Table A2.2: Characteristics of the 25 selected catchments

Station	River	Site	Area km^2	N years	Start Date	End Date	Period of Record Mean Flow m^3/s	SAAR (61-90) mm	BFI
9001	Deveron	Avochie	441.6	38	01/10/59	31/12/99	8.56 – 8.63	988	0.59
9002	Deveron	Muiresk	954.9	38	01/10/60	31/12/99	16.3 – 16.65	928	0.58
14001	Eden	Kemback	307.4	31	01/10/67	31/12/99	3.86	799	0.62
19002	Almond	Almond Weir	43.8	37	01/01/62	31/12/99	0.94	1017	0.34
19004	North Esk	Dalmore Weir	81.6	38	01/01/60	31/12/99	1.53	951	0.54
20001	Tyne	East Linton	307	38	01/01/61	31/12/99	2.75	713	0.52
20003	Tyne	Splimersford	161	34	01/01/65	31/12/99	1.37	725	0.49
20005	Birns Water	Saltoun Hall	93	34	01/01/65	31/12/99	0.94	759	0.49
21006	Tweed	Boleside	1500	38	01/10/61	31/12/99	36.5 – 36.99	1166	0.51
21012	Teviot	Hawick	323	36	01/10/63	31/12/99	8.69 – 8.8	1151	0.44
21013	Gala Water	Galashiels	207	35	01/10/64	31/12/99	3.61 – 3.64	930	0.52
21015	Leader Water	Earlston	239	33	01/10/66	31/12/99	3.37 – 3.44	853	0.49
21017	Ettrick Water	Brockhoperig	37.5	34	01/10/65	31/12/99	1.88	1733	0.34
28031	Manifold	Iiam	148.5	31	01/05/68	31/12/99	3.48	1096	0.54
34003 *	Bure	Ingworth	164.7	39	01/01/59	01/01/00	1.09	669	0.83
39016	Kennet	Theale	1033	38	01/10/61	31/12/99	9.53 – 9.47	759	0.87
39028	Dun	Hungerford	101.3	31	01/04/68	31/12/99	0.71	786	0.95
43005 *	Upper Avon	Amesbury	323.7	33	01/01/65	31/12/99	3.41	745	0.91
43006 *	Nadder	Witton	220.6	33	01/01/66	31/12/99	2.84	875	0.82
48010 *	Seaton	Trebrown Bridge	39.1	31	01/01/57	31/12/99	1.02	1328	0.73
51001 *	Doniford Strm	Swill Bridge	75.8	32	01/01/67	31/12/00	1.06	908	0.64
55016	Ithon	Disserth	358	31	01/10/68	01/03/00	8.05 – 8.24	1066	0.38
55026	Wye	Ddol Farm	174	61	01/10/37	01/03/00	6.65 – 6.71	1637	0.36
60002	Cothi	Felin Mynachdy	297.8	38	01/10/61	01/05/00	11.3 – 11.55	1551	0.43
72004	Lune	Caton	983	38	01/01/59	31/05/01	35.51	1523	0.32

APPENDIX 3:
ANNUAL MINIMA SERIES

Table A3.1: Annual Minima Series - Station 9001

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1960	24.68	27.84	34.28	38.49	41.83	73.65	
1961	21.9	23.99	28.47	31.84	38.56	48.02	77.91
1962	34.07	37.48	42.65	50.02	61.59	84.96	108.2
1963	28.39	34.36	42.66	50.64	65.28	93.31	97.47
1964	23.87	24.58	28.49	30.24	34.09	58.86	70.23
1965	27.46	34.2	41.48	46.4	53.21	75.35	82.69
1966	35.8	36.53	45.22	65.19	74.37	86.78	112.1
1967	25.26	27.81	29.84	35.66	37.54	60.94	91.27
1968	29.66	30.19	35.19	41.41	52.45	69.08	103.6
1969	25.84	26	28.62	29.99	31.32	43.91	96.95
1970	33.49	34.68	41.27	48.34	57.39	89.52	93.75
1971	21.44	22.94	24.75	25.5	26.77	37.21	61.16
1972	21.44	21.68	22.29	22.64	23.35	33.61	40.58
1973	22.25	22.63	23.88	26.57	30.4	39.14	42.42
1974	23.64	24.83	25.74	29.21	32.09	39.73	78.72
1975	24.68	25.33	31.07	34.6	39.16	46.47	62.15
1976	14.95	15.48	17.84	19.1	21.06	38.1	58.29
1977	31.28	32.79	42.06	46.27	49.1	60.69	105.4
1978	40.79	41.9	48.46	52.61	58.96	70.84	103.7
1979	35.69	38.63	47.76	64.06	68.08	79.74	107.9
1980	38.93	42.71	45.41	48.68	60.12	68.23	108.5
1981	20.74	22.25	26.01	29.64	38.35	44.56	100.1
1982	21.09	21.39	27.94	30.93	35.04	51.54	94.04
1983	23.29	23.89	25.92	31.01	41.12	44.59	86.22
1984	21.44	22.33	27.24	28.33	30.52	49.63	91.57
1985	51.33	56.51	65.87	104.7	111.7	115.1	122.9
1986	29.43	30.57	33.82	40.37	42.44	53.16	78.14
1987	35.46	36.4	39.94	48.78	58.9	72.84	88.44
1988	30.71	32.54	40.43	45.03	48.18	68.65	74.63
1989	19.7	20.38	22.9	23.41	24.72	29.55	44.43
1990	22.02	22.48	25	33.01	42.82	44.33	46.37
1991	21.9	23.59	24.85	26.37	31.25	50.18	64.38
1992	20.97	22.05	24.78	28.99	33.88	53.41	62.53
1993	25.84	29.05	38.42	41.06	43.9	54.89	76.95
1994	19	19.95	20.86	22.96	26.27	38.92	67.34
1995	21.44	22.43	24.02	30.33	50.02	70.53	75.93
1996	22.36	22.74	24.48	29.47	29.61	38.25	62.04
1997	24.8	24.98	28.5	30.28	35.28	64.57	76.35
1998	47.27	50.01	63.49	75.32	91.17	100.1	96.94
1999	25.14	26.43	30.72	32.54	41.49	63.18	90.53

Table A3.2: Annual Minima Series - Station 9002

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1961	20.42	22.61	25.36	27.35	31.82	42.09	
1962	27.68	29.44	32.52	39.38	48.11	72.91	102.2
1963	24.5	27.54	34.07	41.36	52.28	89.7	99.02
1964	18.01	18.17	21.2	23.41	26.22	51.86	70.49
1965	25.64	27.45	33.51	41.09	50.34	67.94	84.85
1966	35.67	37.72	51.23	58.61	80.2	93.76	120.3
1967	20.42	22.03	25.42	30.94	33.03	56.72	91.5
1968	25.82	26.55	32.08	38.99	50.03	66.41	105.2
1969	19.99	20.09	22.74	24.48	26.02	41.75	96.04
1970	25.1	26.27	33.32	40.56	50.69	89.17	93.7
1971	18.79	19.27	20.85	21.62	22.89	33.09	57.57
1972	15.25	15.52	16.2	16.45	17.03	26.07	33.02
1973	18.49	20.25	21.86	24.87	26.99	33.27	35.66
1974	22.1	22.76	23.64	26.69	28.52	35	78.68
1975	18.97	19.8	25.77	28.28	35.61	43.35	63.35
1976	12.37	12.8	14.67	15.55	16.92	31.15	58.31
1977	29.36	29.82	38.5	43.44	44.42	53.42	101.6
1978	33.2	34.45	38.62	48.36	55.6	66.65	93.89
1979	30.68	32.91	38.58	53.29	56.74	73.87	102
1980	26.96	29.54	33.64	37.7	50.4	58.06	108.3
1981	18.07	18.52	20.1	22.52	27.17	35.39	91.46
1982	17.41	18.31	23.47	26.19	29.83	44.27	86.39
1983	16.21	16.93	19.32	24.26	34.08	39.54	80.84
1984	16.33	17.34	21.42	23.44	25.18	41.05	86.48
1985	43.71	44.82	57.02	92.32	106.9	113.2	122.7
1986	22.04	23.03	25.94	40.74			
1988	27.56	29.12	35.19	39.06			
1989	15.61	16.58	19.61	19.8	21.5	25.72	42.23
1990	18.79	18.85	21.56	27.72	36.8	37.51	43.62
1991	19.51	20.3	20.96	22.06	25.58	45.33	63.49
1992	18.97	20.08	21.79	24.35	27.89	48.02	61.61
1993	27.26	29.76	38.24	41	45.06	51.62	80.34
1994	16.69	17.12	17.76	19.81	22.72	33.35	64.64
1995	18.97	19.63	21.19	26.18	42.77	64.64	73.2
1996	18.61	18.91	21.11	25.41	25.75	33.32	64.95
1997	16.15	16.44	21.28	23.42	29.51	68.56	83.49
1998	34.23	37.81	51.09	60.64	85.16	100.1	99.49
1999	23.78	24.72	27.43	29.84	38.59	56.53	92.77

Table A3.3: Annual Minima Series - Station 14001

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1968	27.38	28.45	32.53	34.32	40.61	65.87	
1969	30.23	30.59	33.76	34.28	36	50.25	76.82
1970	27.38	27.75	32.44	33.74	36.43	52.2	75.2

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1971	26.09	27.64	30.33	33.02	40.52	44.09	65.77
1972	21.7	23.14	24.6	25.14	26.35	32.84	42.79
1973	16.53	16.72	18.56	19.05	20.38	21.67	36.6
1974	16.79	17.68	20.83	22.09	23.96	28.06	51.89
1975	19.12	19.74	21.61	22.93	25.3	34.02	45.58
1976	17.31	17.46	18.98	21.79	24.5	38.45	55.46
1977	24.28	25.39	29.05	32.32	34.23	50.09	109.4
1978	29.45	30.96	32.76	36.7	39.23	44.38	102
1979	25.32	26.42	28.92	37.13	35.56	46.38	101.1
1980	35.13	39.08	42.37	47.48	47.44	55.08	95.65
1981	26.09	26.68	28.27	29.67	30.91	46.66	86.01
1982	22.22	23.36	24.76	26.62	28.99	39.62	92.44
1983	28.93	29.38	30.95	33.33	37.4	48.04	98.29
1984	22.48	23.25	24.47	26.04	28.01	38.27	99.82
1985	31.26	34.21	40.78	45.45	52.64	94.74	114.8
1986	27.9	28.86	31.97	33.87	37.11	49.99	100.2
1987	32.03	33.62	41.87	42.34	44.79	61.28	100.6
1988	33.58	40.56	43.51	46.95	50.72	73.02	92.86
1989	14.73	16.28	18.21	22.19	23.11	29.43	61.8
1990	18.34	19.38	22.05	25.4	28.94	40.45	79.14
1991	23.25	24.32	25.27	28.04	30.98	39.8	63.34
1992	18.34	21.52	24.01	27.74	31.25	53.59	69.36
1993	30.74	31.48	36.96	42.21	45.87	86.45	124.3
1994	22.99	23.62	25.64	28.15	28.44	36.73	85.04
1995	16.53	17.23	18.59	23.69	30.7	45.04	87.2
1996	23.25	24.21	25.76	26.64	26.78	38.93	83.04
1997	30.48	30.74	33.41	34	34.47	45.73	84.37
1998	37.72	39.45	50.78	54.07	55.8	69.83	103.1
1999	28.42	29.67	31.68	32.38	37.76	53.49	88.33

Table A3.4: Annual Minima Series - Station 19002

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1962	28.76	33.33	41.65	42.9	44.42		
1963	17.05	17.96	26.63	38.48	60.29	74.6	98.86
1964	12.78	14.91	17.54	23.51	34.73	55.83	73.4
1965	20.24	24.5	33.02	48.97	51.88	69.65	104
1966	15.98	22.37	27.66	61.82	66.5	88.36	106.3
1967	10.65	14.31	20.84	25.67	43.29	52.24	99.31
1968	10.65	15.68	23.01	27.61	30.56	70.28	84.3
1969	11.72	12.02	20.17	31.59	32.21	53.66	69.56
1970	12.78	14.31	17.12	20.92	25.96	46.25	83.66
1971	13.85	14.91	19.81	24.52	27.44	49.75	71.11
1972	12.78	16.13	16.73	17.97	18.18	32.95	53.53
1973	12.78	15.22	17.86	19.57	22.86	29.75	52.87
1974	12.78	15.37	19.92	22.42	22.64	32	64.47
1975	8.523	10.04	14.74	19.87	21.98	30.47	61.55
1976	12.78	14.31	16.05	19.12	21.44	36.84	72.5

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1977	17.05	20.39	25.67	26.81	34.65	58.13	89.92
1978	18.11	21.31	27.27	33.58	34.78	55.09	94.16
1979	12.78	15.98	18.71	22.64	31.65	40.33	90.22
1980	7.457	8.827	10.4	13.12	13.94	26.75	78.15
1981	7.457	11.41	15.13	20.37	21.84	38.01	93.15
1982	11.72	12.78	17.97	26.72	29.12	40.51	92.73
1983	9.588	11.11	12.85	15.09	18.43	43.36	88.33
1984	9.588	10.2	16.9	18.34	20.58	25.83	79.55
1985	12.78	17.65	22.55	29.05	29.49	62.82	88.17
1986	14.91	16.44	19.32	40.02	57.82	69.21	110.8
1987	12.78	16.13	21.09	32.63	33.78	57.3	103.9
1988	13.85	15.83	20.67	24.96	33.45	61.5	85.7
1989	11.72	13.54	18.04	21.45	21.97	32	69.86
1990	9.588	14	20.56	21.95	29.82	32.49	84.12
1991	9.588	12.33	13.39	18.07	31.52	38.62	86.03
1992	7.457	10.35	13.92	14.75	19.74	63.33	90.44
1993	7.457	8.37	12.5	19.3	20.62	39.67	78.05
1994	8.523	10.81	15.31	19.9	21.55	27.82	106.7
1995	9.588	10.5	13.1	16.07	16.32	25.12	59.44
1996	8.523	8.827	10.55	13.17	13.99	25.72	50.13
1997	9.588	11.72	16.97	25.14	31.01	42.78	72.16
1998	17.05	19.63	33.7	44.09	57.01	75.91	87.72
1999	24.5	27.09	31.75	38	49.23	53.24	90.76

Table A3.5: Annual Minima Series - Station 19004

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1960	20.85	20.85	24.15	27.7	30.71		
1961	25.41	26.9	29.19	33.59	35.88	49.59	84.16
1962	25.41	29.88	31.77	35.68	41.59	73.43	91.12
1963	29.97	32.02	42.67	50.44	71.33	88.04	99.64
1964	27.36	28.85	32.49	36.95	42.1	57.55	68.21
1965	19.55	22.71	50.71	56.28	65.11	80.51	89.88
1966	25.41	28.2	33.03	38.85	45.22	71.96	94.91
1967	25.41	29.32	34.77	40.72	48.7	69.36	102.2
1968	35.83	41.51	50.77	66.99	66.56	99.19	104.5
1969	20.2	25.04	31.03	32.39	33.92	48.82	75.62
1970	20.2	22.52	26.34	30.19	41.89	60.6	78.02
1971	20.85	25.78	31.03	34.62	37.39	52.15	75.35
1972	14.98	15.36	17.42	19.2	20.7	29.25	45.07
1973	13.68	14.89	16.83	19	22.97	32.59	44.82
1974	15.64	16.75	24.02	25.9	26.57	33.43	57.25
1975	9.121	11.35	15.46	20.97	21.99	37.66	58.22
1976	15.64	16.66	19.13	20.81	23.19	37.83	61.39
1977	22.8	23.36	27.08	30.27	36.88	70.96	98.15
1978	22.8	25.97	27.78	36.14	41.54	59.52	93.25
1979	20.85	21.87	24.67	27.94	28.82	45	97.8
1980	23.45	28.67	32.99	35.96	36.25	54.25	98.37

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1981	25.41	27.27	29.06	32.83	35.52	48.55	95.9
1982	19.55	22.9	25.93	35.16	40.56	46.21	95.53
1983	14.33	17.22	20.09	22.88	27.9	53.66	95.01
1984	9.121	11.26	15.66	17.6	19.88	32.08	74.45
1985	24.76	26.15	37.79	49.76	50.88	75.75	83.43
1986	18.89	22.99	28.71	47.67	56.63	66.58	102.1
1987	24.11	31.09	41.22	54.5	58.98	66.71	96.36
1988	14.98	17.22	22.98	32.25	48.31	68.43	88.34
1989	16.94	17.87	19.89	23.01	25.83	32.75	66.72
1990	18.24	19.27	27.28	35.12	48.1	54.84	98.61
1991	14.98	16.47	21.52	25.5	30.24	38.99	85.11
1992	6.515	11.45	18.85	23.03	28.51	60.1	92.13
1993	26.71	28.57	31.86	38.82	44.23	80.55	103.6
1994	17.59	20.76	23.63	24.65	27	36.79	100.5
1995	16.94	17.68	20.52	22.93	24.92	42.74	70.4
1996	16.94	17.59	19.81	22.24	23.87	34.44	62.68
1997	22.15	22.8	31.58	34.55	35.06	61.51	98.31
1998	39.74	41.42	71.97	83.44	93.67	102.6	104.9
1999	26.06	28.57	34.1	37.64	48.05	59.74	104.6

Table A3.6: Annual Minima Series - Station 20001

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1961	20.68	22.13	28.41	30.75	31.69		
1962	19.59	20.42	23.01	26.65	29.22	43.73	81.3
1963	27.21	30.95	41	50.43	66.63	97.05	110.1
1964	17.42	19.59	22.81	25.32	24.96	37.94	65.13
1965	25.4	28.67	34.36	39.34	46.64	79.78	69.19
1966	25.4	28.25	32.85	38.89	43.55	75.91	120.3
1967	21.41	25.04	27.66	33	35.39	65.42	93.59
1968	33.38	35.4	41.64	48.09	73.35	106	101.2
1969	11.97	15.03	18.66	21.67	22.88	35.05	77.77
1970	23.95	24.73	27.55	44.63	57.37	68.55	79.29
1971	16.33	17.57	18.58	20.99	22.78	31.97	70.78
1972	14.88	15.71	16.33	16.87	17.54	20.01	23.87
1973	12.34	15.55	16.63	18.54	19.77	21.19	23.67
1974	12.34	16.48	20.84	22.27	25.78	31	47.42
1975	19.23	20.48	22.52	24.41	25.65	47.92	57.03
1976	12.7	14.1	16.1	16.97	19	30.96	43.01
1977	22.5	24.42	27.2	30.73	32.27	57.36	94.76
1978	23.95	28.51	31.56	35.45	38.13	39.8	91.24
1979	19.59	22.45	23.78	26.8	29.52	46.49	106.1
1980	25.76	27.27	33.13	39.6	39.91	46.82	101
1981	16.69	18.19	19.72	22.64	25.4	36.25	87.05
1982	14.51	18.14	23.08	24.85	27.07	33.98	83.28
1983	19.96	23.27	28.87	34.56	39.98	48.21	109.6
1984	14.51	16.12	17.61	19	20.35	33.34	98.79

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1985	28.67	30.48	34.25	39	45.39	99.46	102.2
1986	27.21	28.72	30.77	38.77	46.34	52.75	100.5
1987	33.38	37.01	45.39	55.86	61.43	78.52	102
1988	21.05	23.17	27.44	34.7	41.31	47.29	54.7
1989	19.23	19.49	19.86	21	21.23	23.37	36.89
1990	19.23	20.53	22.23	24.58	30.05	31.73	49.63
1991	14.15	16.23	17.88	20.62	22.17	30.1	58.96
1992	21.05	21.67	23.66	25.45	26.59	39.79	76.39
1993	27.21	27.58	30.38	32.67	36.42	82.63	97.12
1994	20.32	21.41	22.78	23.5	24.59	30.46	66.58
1995	13.43	14.1	15.09	17.95	22.36	29.12	61.49
1996	16.33	17.83	19.43	20.91	21.22	27.49	61.16
1997	27.21	28.56	31.21	32.37	34.89	72.27	96.18
1998	42.82	44.89	55.75	63.23	70.17	86.85	103
1999	25.4	27.01	30.54	31.95	36.17	44.37	82.83

Table A3.7: Annual Minima Series - Station 20003

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1965	36.26	39.78	43.85	47.89	52.03		
1966	31.18	33.05	37.61	43.8	49.54	78.05	117.8
1967	21.76	25.69	29.25	32.73	36.96	64.54	93.9
1968	32.63	36.36	40.39	43.51	69.81	104.5	99.74
1969	16.68	20.31	22.46	26.17	26.05	40.43	67.3
1970	16.68	18.54	22.12	24.72	27.41	38.19	66.67
1971	18.13	19.79	22.96	24.38	25.3	33.54	71.44
1972	17.41	17.41	18.32	19.4	20.26	22.44	25.69
1973	13.78	16.27	17.16	17.3	18.29	20.01	24.72
1974	14.5	15.23	16.97	19.13	19.9	25.89	48.67
1975	13.05	13.05	14.67	17.11	18.7	39.91	56.72
1976	10.15	10.15	13.2	14.79	16.43	31.02	46.7
1977	23.21	24.24	26.13	29.58	32.61	54.58	85.68
1978	23.21	24.14	27.63	31.74	34.64	36.44	80.68
1979	23.21	24.24	24.73	28.09	29.06	45.02	93.63
1980	20.31	26.94	32.92	39.56	37.93	42.68	93.85
1981	18.13	19.17	20.5	22.98	24.08	35.58	85.22
1982	20.31	20.93	23.21	24.49	28.58	33	88.94
1983	24.66	26.73	28.4	31.23	40.86	49.58	113
1984	15.23	16.58	18.71	20.87	22	33.25	95.67
1985	23.21	24.14	29.66	36.09	43.17	92.29	97.68
1986	26.11	27.77	32.78	34.44	43.3	76.67	107.6
1987	31.91	33.05	42.3	53.62	64.35	75.96	108.4
1988	27.56	30.56	34.47	42.34	48.7	53.28	61.58
1989	17.41	18.75	20.48	21.68	21.76	22.72	41.51
1990	14.5	15.54	16.8	18.54	26.88	29.88	60.15
1991	20.31	20.82	21.59	23.99	25.48	33.01	64.99
1992	18.13	19.89	21.49	23.87	27.36	46.1	79.23

1993	19.58	20.72	23.57	27.76	33.92	76.36	97.96
1994	18.86	20.1	21.37	22.64	23.49	29.38	74.88
1995	14.5	15.64	16.29	18.71	22.13	28.31	61.71
1996	14.5	15.23	16.22	17.66	19.11	25.24	60.03
1997	23.21	23.72	25.48	29	31.79	78.52	102.1
1998	39.89	42.37	54.34	66.6	76.76	88.41	107.3
1999	21.76	23	25.24	27.33	31.98	44.13	78.58

Table A3.8: Annual Minima Series - Station 20005

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1965	28.57	29.93	34.28	39.69	45.74		
1966	26.45	32.19	36.33	45.27	50.67	82.15	124
1967	15.87	21.01	26.06	27.44	29.34	60.72	81.77
1968	19.04	19.95	23.7	27.72	57.52	84.07	88.65
1969	7.406	14.06	17.39	21.6	22.78	34.06	69.94
1970	14.81	17.08	20.07	24.19	27.17	41.19	69.93
1971	14.81	15.57	21.16	24.56	27.18	34.84	77.71
1972	14.81	15.11	16.08	17.03	17.43	21.08	28.3
1973	12.7	13.15	14.07	14.69	17.23	20.06	28.2
1974	14.81	15.27	17.1	18.69	19.13	26.25	53.62
1975	13.75	14.36	17.1	20.47	21.15	41.63	58.69
1976	9.522	12.39	15.73	16.45	16.89	33.49	48.02
1977	17.99	20.4	22.08	28.5	32.56	57.11	95.17
1978	21.16	25.39	27.61	31.78	34.5	36.4	89.32
1979	22.22	23.28	24.12	28.83	29.74	45.01	98.13
1980	14.81	17.84	26.38	36.08	35.41	45.38	96.21
1981	17.99	18.89	20.31	22.91	24.96	35.54	88.26
1982	19.04	20.4	23.35	24.55	29.73	33.08	90.38
1983	23.28	23.28	26.06	29.82	39.29	52.01	106
1984	16.93	17.23	18.73	20.49	21.54	30.89	92.04
1985	24.33	25.85	31.49	36.89	43.17	90.67	98.47
1986	24.33	25.09	28.6	39.85	44.01	56.8	95.18
1987	30.68	32.8	44.75	59.92	65.55	75.6	96.49
1988	27.51	29.02	30.65	35.57	45.25	53.88	69.83
1989	17.99	17.99	18.48	20.01	19.97	21.46	42.72
1990	16.93	17.38	18.52	20.47	28.08	28.78	61.67
1991	20.1	20.25	21.16	23.01	24.33	29.59	64.1
1992	21.16	22.37	23.81	24.88	27.83	49.79	78.95
1993	22.22	23.43	24.62	27.84	32.75	76.66	95.64
1994	16.93	17.99	19.04	20.19	21.33	29.16	72.75
1995	12.7	13.6	14.57	17.18	19.97	27.07	64.38
1996	13.75	14.96	16.19	17.44	19.76	25.21	62.86
1997	23.28	24.94	28.43	30.24	33.56	79.46	100.7
1998	29.62	30.83	45.46	64.24	73.97	84.93	105.4
1999	21.16	21.77	25.36	28.78	33.43	46.72	74.55

Table A3.9: Annual Minima Series - Station 21006

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1962	13.63	16.12	23.61	25.46	30.11	57.63	
1963	18.68	24.28	27.26	32.14	52.35	86.08	99.63
1964	14.54	15.62	19.48	24.25	33.43	59.55	70.11
1965	22.57	25.3	45.79	51.35	53.3	66.43	91.44
1966	16.54	18.39	26.04	44.61	47.71	67.8	100.7
1967	17.36	20.44	26.85	40.85	46.64	73.26	97.75
1968	15.63	17.72	25.07	33.62	34.93	63.16	84.92
1969	12.22	13.46	22.39	25.55	29.41	45.53	69.12
1970	11.62	13.12	18.31	24.96	32.93	52.46	73.97
1971	15.73	16.75	21.72	29.82	33.47	46.43	73.16
1972	9.354	9.894	10.72	11.71	14.11	32.62	49.5
1973	13.84	14.27	16.56	18.66	23.97	29.62	47.57
1974	13.08	13.8	19.17	20.52	21.01	31.48	67.67
1975	12.46	13.48	15.68	22.38	22.91	34.78	58.32
1976	9.354	10.12	12.51	16.35	19.57	44.79	61.35
1977	11.17	11.94	14.63	17.44	28.55	54.44	94.76
1978	15.19	16.68	21.42	24.18	26.71	43.78	82.21
1979	19.49	20.68	23	34.22	39.17	52.49	94.84
1980	17.22	18.63	19.73	26.21	29.29	51.67	93.45
1981	14.63	15.51	18.39	31.13	34.06	55.09	97.18
1982	16.92	19.19	22.95	29.27	32.58	37.41	96.84
1983	16.87	19.9	21.3	22.25	24.89	47.99	83.31
1984	12.03	13.07	15.01	16.02	17.05	22.66	69.32
1985	20.82	21.75	28.97	34.88	38.53	64.31	78.07
1986	18.6	20.24	24.63	41.17	48.63	71.01	96.62
1987	25.84	27.83	33.2	52.65	55.37	64.53	94.33
1988	18.49	20.67	23.75	29.43	38.85	65.89	90.65
1989	11.87	13.82	15.62	18.99	22.53	33.25	77.22
1990	20.11	23	26.55	28.39	35.28	37.16	99.99
1991	16.63	18.81	20.73	25.83	30.1	36.01	93.24
1992	18.9	20.23	21.54	23.47	31.88	69.33	101.3
1993	22.14	23.79	32.52	35.95	38	62.23	100.7
1994	17.27	19.08	22.5	24.58	25.75	36.33	104.1
1995	10.65	12.39	14.27	17.55	20.34	30.74	74.04
1996	12.03	12.45	14.78	16.51	18.02	35.61	70.81
1997	19.65	20.88	33	39.09	39.69	53.87	90.99
1998	29.03	31.76	52.5	72.26	77.06	85.98	97.48
1999	17.65	22.14	23.95	26.98	32.16	53.2	94.66

Table A3.10: Annual Minima Series - Station 21012

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1964	14.77	16.12	22.9	25.13	38.02	56.58	
1965	18.41	20.86	36.76	42.29	53.27	63.36	91.05
1966	14.54	15.53	21.09	39.74	47.87	69.17	100.5

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1967	16.7	20.26	27.12	53.66	67.32	89.23	99.39
1968	9.772	10.52	16.83	22.81	28.8	51.49	78.46
1969	8.976	9.09	13.77	18.49	28.03	47.36	68.89
1970	7.954	8.976	12.8	18.71	24.77	40.06	71.51
1971	8.635	8.96	15.89	18.11	19.04	34.58	67.91
1972	6.817	7.256	8.48	9.444	10.38	34.24	52.88
1973	9.544	10.1	11.96	14.5	19.83	25.72	44.49
1974	7.385	7.889	11.45	12.35	13.87	27.8	67.65
1975	8.408	8.879	11.73	19.01	19.47	33.88	56.4
1976	6.817	7.174	8.753	10.19	12.45	32.64	52.98
1977	8.749	9.268	11.17	14.26	26.95	52.06	82.97
1978	5.795	6.217	9.79	11.44	14.13	29.75	75.39
1979	10.79	13.26	14.74	27.41	43.85	54.7	95.02
1980	12.84	13.18	14.65	21.63	36.92	51.23	94.81
1981	10.57	11.1	13.52	26.07	30.66	57.44	98.64
1982	12.84	13.81	17.9	26.64	30.48	34.89	95.75
1983	6.022	6.785	8.075	11.2	15.12	46.32	81.12
1984	6.363	6.947	7.848	8.569	9.051	14.51	70.16
1985	13.07	14.66	24.56	38.41	38.47	72.79	86.22
1986	12.16	13.94	17.18	39.95	47.2	71.95	102.9
1987	14.77	16.09	21.19	41.03	46.31	59.19	94.11
1988	12.5	13.36	17.29	26.6	41.02	73.61	94.6
1989	4.999	5.259	6.761	9.139	11.32	30.39	81.8
1990	10.45	11.26	13.38	14.83	21.78	25.14	95.76
1991	6.817	8.148	9.196	13.2	17.18	25.7	88.27
1992	9.431	9.999	12.31	14.94	23.25	64.11	97.05
1993	8.635	9.22	15.23	18.57	22.12	42	93.98
1994	9.203	10.91	17.27	18.54	18.44	29.72	102
1995	4.999	5.405	6.2	8.508	10.81	18.88	70.64
1996	7.272	7.467	8.847	10.96	13.21	35.6	71.27
1997	12.27	14.79	17.69	26.63	33.01	47.93	86.49
1998	18.97	21.85	38.07	67.36	73.43	79.23	93.08
1999	8.976	10.16	12.07	14.46	21.33	42.87	89.28

Table A3.11: Annual Minima Series - Station 21013

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1965	24.42	25.44	34.06	42.47	46.73	82.67	
1966	20.58	21.6	28.25	44.6	45.76	81.31	115.5
1967	15.64	22.54	26.92	30.15	36.13	71.68	91.42
1968	18.66	21.17	26.67	56.72	63.05	96.58	99.87
1969	14	14.7	18.92	20.93	22.73	36.26	72.66
1970	14	14.74	18.67	23.29	30.21	44.87	72.53
1971	16.19	16.54	19.55	26.05	31.64	39.55	73.39
1972	11.25	11.25	12.04	12.41	13.12	18.81	36.58
1973	10.15	10.74	12.16	13.69	17.4	23.88	36.49
1974	11.25	11.96	15.25	19.18	18.46	26.52	63.25
1975	9.331	9.958	12.33	16.85	17.89	37.77	53.91

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1976	8.507	8.86	10.46	11.99	15.98	39.31	56.15
1977	10.43	10.98	12.95	17.53	21.15	53.85	94.86
1978	14.54	16.39	22.48	26.39	28.84	40.61	92.6
1979	13.72	15.13	16.24	21.5	20.55	36	94.44
1980	14	16.94	20.22	29.83	33.03	46.26	95.21
1981	12.35	13.13	14.37	21.44	25.59	41.71	94.86
1982	14.27	15.56	19.35	24.84	31.96	36.08	90.27
1983	11.8	12.19	13.46	15.61	23.41	48.86	94.33
1984	9.605	10.43	11.25	12.51	13.43	21.53	81.56
1985	19.76	21.09	28.84	34.22	38.93	88.05	93.53
1986	19.48	20.27	24.99	44.37	55.23	67.79	104
1987	24.7	26.85	33.41	57.87	62.54	66.62	104
1988	16.47	17.6	20.85	28.12	42.69	64.24	79.54
1989	9.605	10.66	11.44	13.79	14.65	18.82	55.71
1990	12.62	13.05	15.14	17.4	24.15	27.48	75.14
1991	10.98	11.41	13.14	17.97	21.37	25.42	77.13
1992	10.7	12.15	13.25	14.55	19.6	49.63	86.46
1993	13.17	14.03	19.58	23.27	28.03	74.97	103.5
1994	9.605	10.7	12.04	14.12	16.29	25.69	102.8
1995	6.586	6.821	8.004	10.13	13.91	24.46	69.89
1996	9.879	9.958	11.14	12.94	13.07	23.27	69
1997	12.9	13.6	16.42	17.17	18.68	56.84	89.62
1998	30.74	33.05	57.64	68.79	73.09	89.96	101.1
1999	14.27	15.05	17.12	20.23	27.29	46.18	80.87

Table A3.12: Annual Minima Series - Station 21015

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1967	17.12	17.45	22.98	28.61	30.73	74.36	
1968	15.96	16.96	21.54	32.13	48.01	80.62	102.2
1969	11.9	12.15	14.01	16.12	16.71	28.34	83.02
1970	12.19	12.98	16.38	17.65	22.78	34.66	82.69
1971	10.74	11.4	15.85	17.85	21.22	35.61	75.08
1972	10.45	10.7	11.29	11.6	12.84	17.26	28.56
1973	9.867	9.991	10.63	11.8	14.68	19.1	28.47
1974	9.867	10.24	12.15	15.68	15.95	25.79	59.96
1975	9.576	9.991	12.59	14.1	15.01	35.55	50.05
1976	7.835	8.167	9.538	11.34	13.29	36.4	50.48
1977	12.19	12.81	14.72	18.27	19.72	41.91	95.78
1978	15.38	16.87	23.39	30.01	33.46	40.39	107.7
1979	14.22	14.63	17.24	25.07	26.96	47.56	104.4
1980	18.28	21.6	24.1	33.52	34.09	38.49	101.3
1981	11.32	11.57	12.93	16.77	20.18	37.49	89.78
1982	12.48	13.93	15.47	17.79	23.03	26.69	83.75
1983	13.93	15.84	18.53	21.98	29.09	43.15	97.46
1984	9.286	10.12	11.71	13.91	14.74	21.76	88.96
1985	26.7	28.44	34.97	43.28	48.4	98.75	101.2
1986	18.57	19.94	25.22	38.1	55.24	64.39	104.1

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1987	28.73	30.72	39	58.62	61.69	72.49	100.4
1988	17.12	17.95	20.96	29.96	42.11	62.38	70.91
1989	11.03	11.98	13.07	14.77	14.88	16.79	42.95
1990	11.03	11.03	12.18	13.99	20.99	23.86	58.31
1991	9.867	9.908	11	13.43	15.21	19.14	58.92
1992	12.19	13.22	14.56	15.17	17.93	42.17	74.26
1993	12.48	12.81	14.95	17.58	22.15	63.91	92.75
1994	10.16	10.9	12.04	13	14.29	20.43	74.88
1995	7.255	7.379	8.367	10.22	12.8	18.14	56.53
1996	11.03	11.03	11.86	13.39	13.53	21.25	69.09
1997	13.64	14.01	15.29	16.37	17.6	51.72	85.21
1998	30.76	32.21	51.06	65.07	77.5	91.32	96.9
1999	14.8	15.75	19.22	24.67	28.99	44.4	78.42

Table A3.13: Annual Minima Series - Station 21017

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1966	9.518	10.35	16.41	36.3	53.98	67.57	
1967	13.75	15.94	22.37	54.45	67.66	81.96	94.72
1968	6.345	6.949	11.39	19.57	19.21	45.93	74.49
1969	7.931	8.687	13.5	18.31	25.39	42.72	63.41
1970	7.403	8.309	14.89	22.83	36.94	54.88	73.92
1971	10.05	11.48	16.51	32.23	36.21	51.32	74.31
1972	6.345	6.345	7.72	8.636	15.52	45.39	59.57
1973	11.1	11.48	16.85	20.45	28.76	34.54	52.58
1974	7.931	9.895	13.32	15.22	21.53	38.68	71.12
1975	6.874	8.082	13.13	18.69	24.59	31.79	62.08
1976	4.759	4.985	6.363	11.56	17.33	47.66	67.4
1977	4.759	5.514	8.513	10.8	17.11	58.01	88.63
1978	5.816	6.043	9.606	10.13	14.28	42.06	81.43
1979	10.58	12.01	16.14	43.9	48.82	73.36	97.32
1980	7.403	7.856	9.535	17.8	28.17	58.11	99.36
1981	8.989	10.27	15.65	27.06	28.96	60.55	100
1982	12.16	13.45	21.06	32.62	32.76	47.72	104.5
1983	4.759	5.439	6.856	9.791	16.79	51.21	78.86
1984	3.701	3.777	5.376	5.949	6.95	17.64	64.05
1985	13.75	16.39	25.24	39.3	40.34	61.86	78.47
1986	10.05	11.78	13.24	30.1	56.25	79.23	101
1987	10.58	12.31	22.16	43.87	50.08	65.19	86.55
1988	8.46	9.14	12.07	20.09	34.58	72.83	98.7
1989	5.288	5.363	8.372	11.77	14.57	45.72	88.67
1990	11.1	12.09	20.16	33.89	38.72	47.98	96.82
1991	7.403	8.007	10.82	22.97	29.16	40.16	93.75
1992	6.345	6.496	7.879	10.59	27.7	75.09	101.4
1993	8.989	9.971	25.5	41.29	40.86	50.23	93.74
1994	8.46	9.82	12.43	24.33	27.21	44.91	100.3
1995	4.23	4.532	6.61	11.57	13.77	25.67	77.37
1996	5.816	6.194	7.579	8.592	12.77	45.14	79.05

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1997	13.22	15.49	25.8	30.23	38.99	56.73	95.47
1998	16.92	19.19	31.92	70.54	78.58	92.32	102.6
1999	8.989	11.26	17.33	28.25	32.95	58.26	94.05

Table A3.14 Annual Minima Series - Station 28031

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1969	17.93	18.92	21.81	22.92	24.26	45.82	84.04
1970	19.38	20.29	23.91	25.39	30.86	49.37	88.14
1971	22.27	22.56	25.91	37.02	41.18	60.54	81.54
1972	20.82	22.27	25.49	29.51	36.45	60.84	78.74
1973	22.56	23.43	26.83	34.02	44.99	53.99	77.59
1974	17.93	18.72	21.1	22.2	26.07	44.05	84.93
1975	13.3	13.39	13.93	14.81	16.85	21.22	54.84
1976	8.965	9.461	10.46	11.38	12.62	28.07	54.69
1977	14.17	15.16	17.73	18.08	19.88	35.63	93.35
1978	21.69	23.34	29.09	40.68	48.04	53.37	92.43
1979	21.69	22.52	24.94	29	35.5	59.14	103.1
1980	23.14	24.66	25.8	31.6	34.98	57.32	103.5
1981	21.69	22.02	27.34	40.05	40.52	80.39	113.2
1982	23.14	24.75	27.28	29.91	40.24	50.83	100
1983	18.8	19.62	21.23	25.17	31.46	51.06	97.52
1984	13.59	13.8	15.95	17.84	18.98	25.12	66.82
1985	32.97	34.7	43.12	64.5	65.78	79.56	86.01
1986	21.98	22.76	24.16	29	36.51	54.28	106.9
1987	36.44	38.84	51.27	67.67	65.37	82.47	113.5
1988	27.76	29.29	41.33	46.47	56.36	76.09	94.4
1989	11.28	11.82	12.92	13.97	15.31	27.26	72.91
1990	12.44	12.85	15.01	16.08	21.51	27.42	73.34
1991	12.15	12.39	13.2	13.95	15.57	21.7	65.64
1992	17.35	17.68	21.13	24.81	30.95	49.8	75.06
1993	23.14	24.09	29.27	40.79	54.44	63.98	81.96
1994	16.48	16.81	19.06	22.62	24.62	48.37	117.9
1995	11.86	12.19	13.24	14.69	16.47	20.52	48.96
1996	10.12	10.37	11.83	13.81	14.39	20.99	48.18
1997	20.24	20.82	25.27	32.72	33.97	48.56	72.19
1998	28.05	30.08	37.49	42.51	50.74	66.86	100.7
1999	17.06	18.43	24.34	27.34	33.02	56.01	88.16

Table A3.15 Annual Minima Series - Station 34003

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1960	41.34	43.18	49.09	52.18	54.85	65.61	82.92

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1961	58.8	61.55	67.37	72.52	73.43	78.36	104.4
1962	56.96	59.32	61.06	67.17	71.25	77.63	90.41
1963	49.61	56.96	63.58	76.15	77.36	82.9	90.64
1964	39.51	45.02	48.14	49.06	51	60.02	74.56
1965	53.29	56.7	62.32	63.88	68.31	77.02	74.54
1966	56.96	58.14	61.19	65.57	68.13	72.3	94.6
1967	48.69	51.19	54.27	54.43	55.84	67.76	81.22
1968	49.61	51.32	55.71	60.99	64.43	72.18	81.35
1969	71.66	74.02	77.33	79.48	84.99	93.78	121.2
1970	61.55	62.47	66.94	68.08	68.68	74.05	100.2
1971	54.2	57.09	61.03	67.59	71.41	75.4	99.14
1972	58.8	60.37	61.71	64.39	64.57	70.31	75.37
1973	45.94	47.38	49.06	53.96	61.6	65.48	73.26
1974	41.34	43.18	47.22	49.18	50.5	53.84	70.55
1975	50.53	52.24	54.73	58.14	60.95	68.93	73.01
1976	34.91	36.88	39.99	43.46	44.55	53.21	70.31
1977	56.04	58.27	61.71	66.27	67.58	71.19	93.46
1978	57.88	59.72	65.08	65.09	67.32	74.28	93.89
1979	54.2	55.39	58.34	61.36	61.55	68.88	95.57
1980	59.72	61.16	65.05	69.44	72.52	74.19	96.06
1981	60.64	62.6	64.98	71.83	79.2	85.88	91.65
1982	53.29	54.99	57.97	63.53	68.5	74.03	89.38
1983	61.55	64.05	67.28	71.48	73.9	84.58	104.6
1984	66.15	67.85	74.51	80.79	82.06	92.67	106.7
1985	68.9	71.66	79.16	85	88.75	91.19	107.8
1986	62.47	66.02	69.33	72.21	75.03	84.84	107.1
1987	72.58	73.37	82.53	87.69	90.57	104.3	109.5
1988	77.17	78.88	82.32	86.67	90.69	95.04	93.72
1989	47.77	49.87	59.5	60.19	61.82	65.1	79.9
1990	39.51	40.16	43.21	47.82	50.39	55.52	69.61
1991	37.67	38.98	40.79	43.26	45.76	52.86	65.28
1992	36.75	38.19	45.94	49.08	49.98	53.89	66.9
1993	41.34	43.18	48.51	53.76	53.51	62.15	79.19
1994	58.8	60.5	64.22	71.15	75.45	95.61	118.6
1995	56.04	57.09	59.47	67.63	68.67	74.24	76.44
1996	36.75	38.72	41.86	43.99	45.6	53.06	70.41
1997	44.1	44.62	47.59	48.37	51.04	60.56	73.37
1998	48.69	50.14	55.31	59.4	64.14	75.9	94.73
1999	56.96	59.06	68.35	81.9	83.57	95.96	100.3

Table A3.16: Annual Minima Series - Station 39016

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1962	38.85	39.87	43.66	47.4	48.12	56.27	
1963	55.32	57.49	60.08	63.66	65.08	79.14	88.94
1964	39.17	41.22	42.55	43.3	44.6	50.63	56.31
1965	30.83	34.1	38.3	41.29	41.92	44.06	52.48
1966	61.02	63.11	69.41	73.57	77.41	105.4	111.3
1967	51.73	53.58	55.58	59.53	61.84	81.98	103.9

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1968	57.96	60.85	66.97	68.79	70.59	83.41	103.8
1969	39.48	41.9	47.4	52.23	54.52	63.78	89.85
1970	43.39	44.23	45.54	48.15	50.38	59.7	86.05
1971	59.54	60.52	63.77	67.33	69.46	79.56	109.7
1972	44.55	46.18	48.29	49.26	51.02	64.59	77.59
1973	37.69	38.46	41.72	43.91	44.15	49.53	65.4
1974	41.07	42.2	45.93	49.38	54.26	66.47	89.94
1975	40.22	42.24	44.86	49.03	50.32	48.52	45.45
1976	9.818	10.84	13.35	15.41	17.43	24.07	36.31
1977	53.84	56.13	58.93	60.92	65.95	73.81	108.3
1978	36.84	38.59	39.65	41.12	42.44	58.45	95.93
1979	39.06	41.73	46.34	48.88	50.48	64.1	94.96
1980	40.12	41.56	45.89	50.33	53.3	62.01	86.26
1981	52.04	53.7	60.24	70.8	78.1	87.66	98.93
1982	43.07	44.22	48.3	50.75	54.65	69.96	109.1
1983	42.54	44.65	47.47	52.66	53.85	62.32	93.08
1984	37.27	37.94	39.66	42.22	44.36	55.24	87.39
1985	50.57	51.95	54.44	57.4	59.67	70.86	101
1986	48.03	50.79	57	64.71	68.22	79.68	104.8
1987	46.77	50.43	53.19	54.71	59.82	78.59	105.1
1988	46.13	46.63	49.45	52.63	56.36	61.39	77.66
1989	32.51	34.02	36.67	38.46	39.96	44.86	68.93
1990	32.73	33.53	34.21	36.61	37.6	41.57	61.47
1991	34.31	34.94	36.29	38.52	39.8	44	49.23
1992	32.41	33.83	39.99	41.47	43.08	48.66	48.14
1993	50.67	52.53	58	62.01	65.73	84.71	121.6
1994	46.24	48.29	51.52	55.08	56.38	68.54	122
1995	36.21	36.86	38.99	45.07	47.4	54.9	86.92
1996	33.46	34.48	36.01	36.61	38.7	47.03	53.31
1997	24.7	27.15	28.78	30.45	31.7	36.76	50.27
1998	44.34	46.36	51.96	54.25	56.38	78.75	90.66
1999	43.18	44.58	50.78	56.77	61.08	69.25	86.93

Table A3.17: Annual Minima Series - Station 39028

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1969	45	48.21	49.08	52.05	54.58	63.81	93.23
1970	30.94	34.35	35.2	36.8	38.75	52.41	85.44
1971	52.03	54.24	58.08	59.98	61.86	72.41	97.93
1972	43.59	44.8	45.94	48.42	50.09	63.09	77.62
1973	40.78	42.99	43.73	44.72	45.97	49.5	65.41
1974	43.59	47.21	49.13	51.19	55.16	71.06	96.08
1975	42.19	48.21	51.19	53.72	53.19	52.77	52.39
1976	26.72	27.52	28.17	28.52	29.7	34.73	42.94
1977	52.03	54.44	57.47	60.66	64.59	75.72	111
1978	43.59	44.4	46.83	48.3	50.23	61.47	103.1
1979	50.63	52.03	53.44	55.99	58.41	74.6	108
1980	47.81	49.42	52.69	55.78	59.38	70.15	95.84
1981	59.06	63.48	67.17	73.97	80.16	93.61	108.5
1982	42.19	43.39	48.09	52.29	55.28	68.61	114.4
1983	53.44	56.45	58.41	59.88	60.45	69.46	103.2
1984	40.78	41.79	45	47.06	47.97	56.66	96.46
1985	50.63	52.03	53.48	56.95	59.55	69.4	103.3
1986	50.63	52.03	58.27	61.76	62.78	78.37	107.5
1987	45	45.2	49.17	52.05	55.73	69.3	96.24
1988	50.63	52.63	53.91	55.05	58.44	63.94	81.2
1989	33.75	34.15	35.16	35.98	37.27	41.95	70.17
1990	30.94	31.74	32.72	33.66	34.02	36.37	56.98
1991	35.16	36.76	37.69	38.74	39.23	44.19	50.28
1992	40.78	42.79	44.63	46.03	46.38	49	49.96
1993	46.41	48.21	50.25	52.36	56.34	75.85	112.8
1994	43.59	44.6	46.73	47.63	48.78	61.81	119.1
1995	39.38	41.99	43.22	46.34	47.91	53.83	85.29
1996	35.16	35.76	36.98	37.24	37.81	47.99	57.22
1997	28.13	28.33	29.3	29.77	30.14	35.11	51.78
1998	37.97	37.97	39.84	40.9	42.56	56.63	74.45
1999	37.97	38.57	42.09	44.58	46.14	53.82	75.18

Table A3.18: Annual Minima Series - Station 43005

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1966	43.79	44.25	49.93	54.48	56.94	82.82	115.4
1967	48.19	49.58	51.87	54.47	57	74.15	101.9
1968	49.66	52.31	56.81	60.43	65.06	68.83	99.95
1969	32.62	33.46	35.28	37.85	39.52	47.31	81.66
1970	31.15	32.03	33.64	36.95	38.05	47.25	80.03
1971	39.67	40.64	43.38	45.22	46.98	55.26	93.22
1972	31.74	32.91	34.9	36.23	37.89	50.9	72
1973	29.09	29.68	31.66	32.83	32.96	37.38	58.38

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1974	33.79	35.47	39.15	41.45	45.73	63.26	85.1
1975	32.91	33.79	35.9	37.88	37.97	36.92	35.56
1976	4.996	5.374	9.864	12	13.53	19.23	28.11
1977	49.37	51.43	54.44	63.18	64.01	72.67	118.3
1978	32.03	34	35.16	36.86	38.48	52.46	95.72
1979	42.02	43.11	45.57	46.59	47.49	61.33	104.5
1980	42.9	44.04	45.77	49.17	52.6	61.24	91.36
1981	42.32	43.03	46.26	52.03	56.68	69.3	95.7
1982	35.56	36.44	38.53	41.92	46.03	59.88	111.2
1983	40.55	40.85	42.98	46.43	47.33	55.5	92.03
1984	29.39	30.1	31.57	34.25	35.43	46.34	86.15
1985	41.73	42.02	45.83	47.55	48.6	57.21	99.32
1986	40.85	41.35	43.15	46.8	49.43	65.34	101.2
1987	39.08	40.18	42	43.2	48.2	62.77	99.83
1988	39.97	40.89	42.24	44.28	46.36	56.42	78.16
1989	24.1	25.19	25.67	26.95	28.58	34.62	68.51
1990	24.68	24.89	25.65	27.41	28.37	32.33	61.41
1991	32.62	32.7	35.35	37.04	39.45	45.35	52.08
1992	29.39	29.93	34.06	37.27	39.48	48.81	49.67
1993	40.55	40.93	44.87	47.6	52.26	75.46	117.5
1994	40.55	42.78	43.98	46.79	47.24	64.28	126
1995	28.51	29.6	30.76	35.85	37.12	45.39	89.35
1996	26.15	27.67	28.69	29.82	31.77	39.34	48.36
1997	20.57	20.82	23.44	25.12	25.95	30.57	45.06
1998	37.91	40.22	43.71	46.09	48.43	71.21	87.56
1999	40.26	41.81	46.64	50.98	54.32	64.26	82.52

Table A3.19: Annual Minima Series - Station 43006

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1966	37.74	39.25	41.49	45.66	50.61		
1967	42.32	48.92	50.16	53.79	58.63	76.63	94.14
1968	42.68	45.65	50.01	52.74	60.89	68.12	94.19
1969	29.63	31.39	33.64	37.1	38.15	46.35	85.96
1970	29.63	30.48	32.17	36.32	37.95	46.03	88.14
1971	31.74	34.11	36.13	37.36	39.57	46.36	80.51
1972	28.92	30.84	33.28	34.1	35.64	45.04	69.92
1973	26.1	26.55	28.77	29.03	29.63	33.61	53.47
1974	29.98	31.94	33.56	36.07	40.58	56.21	80.74
1975	31.74	32.8	35.42	38.07	39.69	40.77	40.93
1976	17.28	18.54	20.66	21.88	23.13	28.79	35.75
1977	50.44	51.64	56.57	64.67	64.77	74.96	127
1978	35.27	36.63	37.29	39.48	41.04	52.83	105.3
1979	42.68	43.48	45.93	48.39	50.91	64.58	112.1
1980	37.03	38.9	40.27	43.56	44.8	57.3	92.43
1981	41.27	41.57	43.01	48.9	52.48	61.42	96.29
1982	32.1	34.26	38.5	43.25	49.63	60.65	106.1
1983	35.62	39.3	40.32	43.35	43.68	52.34	89.61

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1984	29.63	30.84	31.78	33.3	34.42	43.5	85.42
1985	35.97	37.79	40.61	41.87	43.96	50.43	90.96
1986	31.39	32.1	33.56	42.42	44.12	59.56	97.49
1987	28.92	32.95	35.32	35.6	39.01	53.76	87.29
1988	33.15	34.01	36.6	38.9	42.09	51.81	67.41
1989	26.8	27.06	29.84	31.82	33.82	37.42	71.08
1990	24.34	25.29	27.57	28.13	29.46	32.75	62.94
1991	25.75	26.91	29.79	31.95	35.32	41.55	48.98
1992	27.16	27.86	30.53	37.47	38.64	48.03	48.72
1993	29.63	34.16	38.13	41.68	44.32	63.72	102.4
1994	38.09	38.44	40.94	44.26	46.8	63.39	126.8
1995	35.27	36.08	39.42	42.85	42.66	49.56	92.45
1996	32.8	34.46	35.25	36.14	39.13	47.12	69.72
1997	28.57	29.27	30.81	33.76	38.08	42.96	69.13
1998	34.56	35.72	40.09	40.92	43.24	61.05	99.52
1999	37.74	40.21	44.24	48.58	53.39	69.54	109.8

Table A3.20 Annual Minima Series - Station 48010

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1970	26.74	27.58	28.68	30.86	31.86	35.19	
1971	19.8	19.8	20.89	21.64	22.52	27.76	58.27
1972	22.77	22.77	25.94	27.99	30.5	44.26	75.35
1973	19.8	21.22	24.39	27.45	28.41	35.69	68.25
1974	20.79	21.64	27.46	31.62	35.41	58.9	104
1975	20.79	21.08	25.48	25.91	28.45	35.75	60.98
1976	12.87	13.16	14.72	17.26	20.17	33.23	64.48
1977	21.78	21.78	24.16	24.8	25.83	35.38	95.42
1978	14.85	15.7	15.91	16.69	18.01	24.37	84.75
1979	23.77	26.88	28.75	32.61	33.62	45.46	88.71
1980	25.75	27.02	29.64	32.03	33.45	48.09	98.54
1981	20.79	21.78	23.77	29.19	37.83	70.08	114.3
1982	20.79	21.5	25.22	28.06	32.31	42.33	111.3
1983	18.81	19.24	21.26	24.51	26.72	37.8	82.7
1984	12.87	13.72	14.92	16.32	16.86	23.57	81.05
1985	24.76	28.01	29.28	34.92	39.14	54.59	89.5
1986	39.61	41.73	51.26	67.27	70.84	80.75	95.87
1987	20.79	21.5	23.93	24.57	28.1	44.87	90.34
1988	28.72	30.27	32.78	38.37	40.75	57.44	80.69
1989	13.86	14.85	15.94	17.36	18.01	26.19	69.66
1990	15.84	16.41	17.56	19.95	21.61	25.81	72.19
1991	23.77	25.04	28.88	33.93	37.65	46.86	64.25
1992	18.81	20.09	25.32	27.53	31.16	52.35	59.9
1993	32.68	35.22	42.55	50.65	61.94	75.06	106.1
1994	20.79	24.33	26.14	26.8	30.74	47.08	118.1
1995	11.88	13.3	14.33	16.14	17.25	23.57	75.84
1996	17.82	18.11	20.79	24.41	25.37	39.5	74.61
1997	23.77	23.91	27.4	31.7	33.38	40.08	74.93

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1998	37.63	40.17	46.08	51.57	60.3	76.14	112.9
1999	22.77	24.05	27.2	28.86	32.49	50.72	87.07

Table A3.21: Annual Minima Series - Station 51001

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1967	22.53	24.01	25.48	26.62	27.82		
1968	27.23	28.84	34.08	36.9	40.7	65.49	86.05
1969	20.66	21.33	23.1	26.59	31.28	41.23	101.2
1970	15.96	17.17	18.5	21	26.27	40.48	95.09
1971	22.53	23.34	24.19	26.38	28.09	39.13	81.98
1972	20.66	21.33	22.75	23.43	24.04	40.71	60.73
1973	15.96	17.3	19.87	21.31	23.06	25.61	42.14
1974	17.84	18.64	21.06	22.58	24.59	37.05	81.6
1975	13.14	13.55	15.52	16.06	18.32	20.77	29.63
1976	7.511	8.048	8.638	9.295	10.71	20.47	28.09
1977	18.78	19.99	23.16	24.19	25.45	32.76	91.33
1978	14.08	14.89	15.65	17.03	17.97	24.35	79.6
1979	24.41	24.81	25.79	27.28	28.28	38.76	90.97
1980	19.72	20.39	21.66	23.52	26.03	39.18	77.98
1981	15.96	16.1	18.75	22.5	27.04	45.27	83.12
1982	15.02	15.83	17.71	21.24	23.31	30.17	96.22
1985	19.25	20.2	21.82	23.73	24.75	28.56	67.45
1986	21.78	22.87	25.93	29.61	29.52	41.04	78.42
1987	15.12	16.14	18.31	18.77	21.1	38.39	75.38
1988	25.07	27.86	30.5	34.94	36.93	41.24	56.86
1989	13.8	14.15	15.52	16.47	17.25	27.33	53.97
1990	12.02	12.41	13.45	13.97	14.49	18.8	53.64
1991	15.3	15.95	16.78	19.37	23.64	38.17	55.04
1992	12.68	13.69	15.37	16.56	17.39	23.47	48.14
1993	23.66	25.23	29.36	34.13	38.01	54.99	85.1
1994	26.2	30.85	32.69	35.51	37.65	55.12	127.9
1995	18.68	20.37	21.82	25.22	29.73	39.74	99.65
1996	23.47	24.33	25.8	27.18	29.31	43.99	69.36
1997	33.89	35.76	40.46	45.98	52.11	77.51	73.85
1998	29.58	30.47	35.85	38.02	40.77	64.47	118.9
1999	30.33	31.65	38.03	41.59			

Table A3.22: Annual Minima Series - Station 55016

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1969	5.581	6.292	10.1	14.47	21.83	43.98	
1970	2.548	2.669	5.221	7.424	13.02	33.84	
1971	4.004	4.229	5.367	9.871	12.04	17.97	
1972	5.217	5.581	11.31	12.91	16.53	44.18	
1973	5.096	6.015	9.549	21.04	34.88	17.97	
1974	4.853	6.639	9.068	10.67	12.23	25.81	
1975	3.276	3.467	4.586	5.363	6.444	15.06	
1976	0.485	0.589	1.116	1.771	2.826	10.41	
1977	2.669	2.877	4.174	7.07	8.427	25.67	
1978	3.883	4.021	5.533	6.516	7.586	10.56	
1979	2.912	3.675	4.845	8.835	8.311	26.37	
1980	5.46	6.188	7.664	23.29	24.96	52.41	
1981	2.063	2.236	3.143	4.73	8.623	48.83	
1982	2.184	2.617	4.688	6.795	16.73	17.63	
1983	1.82	2.323	2.989	5.013	7.784	33.06	
1984	1.092	1.317	2.022	2.688	3.269	10.87	
1985	8.979	11.06	18.63	40.77	41.31	56.68	
1986	2.427	2.756	4.518	5.992	12.37	33.04	
1987	3.761	4.229	5.865	9.288	14.35	25.59	
1988	9.949	11.75	22.12	39.86	40.13	63.97	
1989	0.607	0.953	1.703	2.006	3.083	7.287	
1990	1.092	1.404	2.495	2.983	6.278	11.06	
1991	3.883	4.299	5.86	11.59	13.61	26.52	
1992	7.765	8.285	17.03	19.92	30.54	54.61	
1993	8.372	9.551	17.01	25.08	36.92	53.03	
1994	1.335	2.479	4.93	5.571	8.306	29.99	
1996	2.111	2.467	4.08	5.067	5.715	23.3	
1997	4.186	4.892	8.189	16.21	18.01	39.24	
1998	13.24	14.43	21.03	30.04	39.3	49.97	
1999	1.844	2.363	5.166	9.833	10.18		

Table A3.23: Annual Minima Series - Station 55026

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1938	4.768	5.79	10.19	20.36	30.3	44.6	
1939	3.725	4.172	10.3	14.99	27.2	44.31	85.78
1940	4.768	5.407	10.08	15.9	26.08	34.51	75.28
1941	5.514	6.152	9.741	15.47	28.26	38.87	70.16
1942	5.514	5.514	12.11	27.59	33.2	42.54	74.75
1943	9.388	12.28	18.33	25.52	37.66	43.8	78.35
1944	8.196	8.6	11.93	23.73	25.42	27.14	69.47
1945	5.961	6.663	11.27	23.66	30.82	39.77	67.38
1946	7.004	8.089	13.99	15.1	19.06	42.24	74.11
1947	5.066	5.066	6.408	11.22	11.86	22.29	70.19
1948	12.07	14.31	29.15	45.31	50.56	53.95	78.78
1949	3.427	3.832	6.82	16.69	18.49	27.93	70.22

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1950	4.768	5.024	8.335	13.41	22.89	52.94	97.06
1951	5.066	8.025	14.79	15.6	19.89	34.57	85.12
1952	8.196	8.792	14.75	28.71	34.56	40.5	67.57
1953	8.494	11.15	15.56	22.83	30.17	57.19	72.61
1954	6.706	7.515	10.13	25.54	42.51	75.78	83.91
1955	3.129	3.598	6.12	8.563	16.96	38.84	57.95
1956	8.196	9.388	12.56	15.58	18.13	39.28	57.47
1957	5.961	6.919	10.97	19.45	18.87	65.43	90.23
1958	11.77	11.98	17.72	33.19	43.65	54.29	82.09
1959	3.129	3.385	3.81	4.975	14.21	27.74	58.62
1960	3.725	6.046	13.97	17.44	20.35	58.27	96.5
1961	2.384	2.682	6.8	10.7	19.08	44.08	79.01
1962	6.11	6.62	9.895	12.13			
1964	9.686	11.07	20.99	27.28	37.43	43.71	54.64
1965	13.71	16.18	21.9	38.46	39.23	53.34	85.22
1966	16.84	18.75	31.65	45.15	47.16	60.35	96.24
1967	7.6	10.39	16.96	27.77	46.34	61.34	95.62
1968	7.302	9.303	14.47	16.17	42.43	58.84	80.48
1969	5.812	7.472	11.42	20.52	32.72	49.43	76.46
1970	5.514	5.939	9.03	17.39	25.7	45.52	93.96
1971	4.917	5.514	9.279	20.89	25.48	29.94	69.68
1972	5.215	5.599	7.242	10.9	16.18	41.07	79.98
1973	5.066	5.663	9.969	27.54	41.06	60.41	77.25
1974	4.172	6.088	11.13	14.02	19.58	42.23	100.4
1975	3.129	3.363	4.987	10.15	13.16	28.83	61.4
1976	1.49	1.597	2.126	4.366	5.417	16.98	57.23
1977	5.961	7.77	11.35	19.98	23.18	48.15	89.11
1978	7.898	9.835	13.63	29.33	37.85	39.64	87.86
1979	6.259	7.259	10.5	19.88	37.66	62.11	102.8
1980	4.172	4.811	6.676	10.62	29.88	43.38	104.4
1981	4.768	5.918	11.73	16.67	16.65	43.58	103
1982	5.961	6.45	8.37	10.34	21.24	30.91	90.32
1983	1.192	1.618	2.931	6.288	17.23	61.6	97.28
1984	2.235	2.555	4.664	5.682	7.583	14.65	74.79
1985	12.96	14.79	27.72	51.23	61.66	83.26	88.94
1986	4.321	5.088	6.775	22.47	29.25	38.97	79.47
1987	2.533	3.172	10.54	24.27	28.24	38.28	88.58
1988	4.172	5.322	17.45	30.66	30.87	56.9	83.62
1989	3.129	3.491	5.066	9.537	10.02	20.33	69.84
1990	5.365	5.577	8.678	18.75	25.04	28.65	93.14
1991	6.706	7.408	12.79	25.57	29.44	42.05	102.5
1992	5.663	6.812	14.12	20.11	34.7	70.48	106.6
1993	11.47	13.33	20.42	26.79	51.1	65.3	92.35
1994	7.153	7.834	9.954	13.57	20.82	38.74	131.1
1995	3.725	3.96	4.6	7.334	11.12	20.97	60.06
1996	7.302	7.706	10.27	14.05	19.25	42.63	69.9
1997	8.643	9.92	17.97	26.73	38.86	53.1	79.09
1998	6.408	8.92	25.44	62.91	65.75	78.39	113.5
1999	2.235	2.533	5.504	12.31	14.19		

Table A3.24: Annual Minima Series - Station 60002

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1962	5.627	7.111	11.32	13.3	31.05	51.11	
1963	13.68	15.87	21.03	30.07	43.85	54.94	64.18
1964	10.3	11.96	24.12	27.15	27.91	40.04	52.8
1965	11.6	16.03	22.82	51.36	57.24	62.35	88.47
1966	13.59	15.16	28.18	37.69	53.94	74.2	101.3
1967	7.619	8.967	14.62	25.03	49.48	65.75	103.4
1968	10.13	11.45	18.94	21.11	45.11	66.03	89.63
1969	8.138	9.882	15.64	18.67	22.76	36.56	73.47
1970	7.186	8.002	10.94	12.22	19.49	43.32	85.71
1971	7.705	8.286	12.55	22.7	34.45	37.3	72
1972	5.541	5.912	7.14	10.81	20.12	55.2	72.35
1973	5.627	6.134	9.044	13.48	25.88	40.31	57.2
1974	9.263	12	16.72	22.01	21.85	42.71	92.42
1975	4.502	4.823	7.067	9.373	10.99	24.5	54.18
1976	2.164	2.263	2.978	4.013	4.986	14.94	51.8
1977	3.896	4.527	6.597	8.214	11.75	34.8	92.93
1978	8.398	9.931	11.6	16.5	23.96	29.63	83.08
1979	4.588	5.392	8.328	15.27	39.76	58.13	91.68
1980	7.013	7.816	9.01	12.79	19.67	42.16	88.97
1981	3.203	3.648	5.867	8.307	12.7	41.97	103.2
1982	6.839	7.779	10.3	14.56	17.08	23.36	94.95
1983	3.376	3.451	4.955	8.375	16.04	52.99	70.91
1984	1.905	2.288	3.345	4.056	4.432	8.388	65.06
1985	18.44	20.36	29.53	36.49	42.21	75.93	100.8
1986	7.272	7.408	10.99	26.28	36.19	59.59	90.84
1987	8.831	10.17	14.85	21.54	32.57	39.52	94.27
1988	10.04	10.14	18.44	28.45	33.29	61.6	79.19
1989	3.636	3.871	4.435	5.867	6.514	13.17	64.03
1990	7.013	7.619	10.22	11.19	18.41	18.71	70.42
1991	10.22	11.11	14.78	26.57	35.39	42.07	73.19
1992	9.004	9.845	13.38	17.16	26.26	63.92	71.81
1993	13.07	15.41	26.13	32.08	33.75	57.63	84.92
1994	9.696	11.22	20.19	22.86	25.95	36.24	102.6
1995	2.944	3.216	3.806	6.989	8.409	13.62	60.88
1996	5.974	6.468	8.536	13.78	19.66	39	69.18
1997	8.311	8.583	12.23	23.79	31.38	38.48	66.52
1998	8.484	9.115	16.37	42.95	58.87	65.41	109.1
1999	4.242	4.724	7.613	16.67	18.74	47.32	

Table A3.25: Annual Minima Series - Station 72004

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	D=7	D=30	D=60	D=90	D=180	D=365
1959	5.914	6.184	7.32	7.901	14.1		
1960	7.294	7.596	19.92	23.43	38.17	59.13	93.2
1961	14.39	16.8	24.73	35.61	59.65	78.03	113.4

Year	Annual Minimum D-Day Duration Flow (as % MF)						
	D=1	**D=7**	**D=30**	**D=60**	**D=90**	**D=180**	**D=365**
1962	12.14	16.41	29.5	35.68	46.19	70.01	79.52
1963	8.083	9.507	10.91	14.4	48.28	73.1	81.55
1964	9.21	11.07	29.14	47.06	46.13	54.44	67.14
1965	8.308	9.395	18.28	39.25	49.08	56.4	89.86
1966	13.32	15.49	30.83	59.41	69.51	79.54	99.75
1967	14.03	16.45	38.24	46.16	62.31	86.81	99.55
1968	6.59	6.94	10.12	18.41	36.02	60.22	79.56
1969	8.224	8.839	18.35	21.16	24.56	39.01	56.88
1970	5.351	6.035	8.039	14.97	27.41	59.02	79.79
1971	5.689	6.212	12.12	14.26	18.15	44.8	68.75
1972	5.605	6.212	8.446	9.353	18.48	60.97	75.97
1973	6.816	8.461	11.37	18.5	26.19	41.21	65.97
1974	5.407	5.782	6.99	8.326	11.68	37.89	76.29
1975	4.563	5.166	7.531	10.41	22.53	37.85	74.04
1976	3.52	4.208	5.242	9.712			
1979	10.34	11.84	17.75	26.68			
1980	6.281	6.743	7.479	11.69	25.7	53.5	108.1
1981	9.66	10.34	25.24	33.28	31.02	56.5	104
1982	6.9	8.007	11.02	21.43	27.71	39.89	94.22
1983	6.196	6.566	10.73	13.43	17.97	47.09	85.26
1984	3.267	3.508	5.15	6.663	14.4	18.26	73.41
1985	11.52	12.71	25.99	31.61	37.46	58.48	85.71
1986	8.111	8.397	10.91	25.58	31.39	48.28	83.82
1987	10.17	11.41	18.35	36.81	51.97	73.07	101.3
1988	6.9	7.516	11.11	17.12	25.06	64.79	98.01
1989	4.253	4.993	6.251	12.02	13.93	19.56	81.47
1990	6.618	7.17	10.17	14.22	27.3	29.19	77.68
1991	6.506	7.258	8.627	15.36	28.23	33.12	85.99
1992	6.449	6.731	7.975	12.17	17.08	51.59	91.49
1993	11.86	13.22	20.69	29.04	38.11	47.91	78.44
1994	8.139	9.399	14.18	25.97	24.92	48.52	102.6
1995	4.506	4.945	5.64	7.264	11.51	18.66	39.37
1996	5.971	6.204	11.39	20.79	23.49	30.54	43.87
1997	6.618	7.817	11.23	13.84	14.94	34.88	74.13
1998	12.22	13.56	26.74	43.09	53.84	73.4	89.84
1999	11.66	13.62	18.19	22.25	31.72	53.44	99.24
2000	8.365	9.358	17.7	23.97	34.98	49.38	

APPENDIX 4.1:
ANNUAL MINIMA SERIES - STATISTICS

Table A4.1.1: Characteristics of Annual Minima, where D = 1

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
9001	40	14.95	51.33	24.74	27.24	7.80	0.29	1.33	1.71
9002	38	12.37	43.71	20.20	22.61	6.69	0.29	1.15	1.37
14001	32	14.73	37.72	25.71	25.19	6.01	0.24	0.04	-0.79
19002	38	7.457	28.76	12.78	12.78	4.62	0.36	1.58	3.29
19004	40	6.52	39.74	20.53	20.73	6.64	0.32	0.44	1.13
20001	39	11.97	42.82	20.32	21.24	6.71	0.32	0.94	1.45
20003	35	10.15	39.89	20.31	21.20	6.83	0.32	0.98	0.74
20005	35	7.41	30.68	17.99	19.10	5.61	0.29	0.23	-0.36
21006	38	9.35	29.03	16.59	16.31	4.37	0.27	0.72	0.82
21012	36	5.00	18.97	9.32	10.20	3.63	0.36	0.74	-0.01
21013	35	6.59	30.74	13.17	14.18	5.15	0.36	1.40	2.22
21015	33	7.26	30.76	12.19	13.88	5.56	0.40	1.80	3.13
21017	34	3.70	16.92	8.20	8.52	3.11	0.36	0.67	0.22
28031	31	8.97	36.44	19.38	19.37	6.44	0.33	0.64	0.59
34003	40	34.91	77.17	54.20	53.22	10.57	0.20	0.14	-0.47
39016	38	9.82	61.02	41.81	42.04	10.00	0.24	-0.59	1.80
39028	31	26.72	59.06	43.59	42.60	7.89	0.19	-0.16	-0.40
43005	34	4.99	49.66	36.73	35.25	9.06	0.26	-1.05	2.40
43006	34	17.28	50.44	31.92	32.99	6.45	0.19	0.35	0.92
48010	30	11.88	39.61	20.79	21.88	6.56	0.3	0.99	1.43
51001	31	7.51	33.89	19.25	19.77	6.03	0.30	0.36	-0.14
55016	30	0.49	13.24	3.52	4.09	3.01	0.74	1.35	1.80
55026	61	1.19	16.84	5.51	6.13	3.07	0.50	1.20	1.92
60002	38	1.91	18.44	7.45	7.61	3.57	0.47	0.74	0.97
72004	40	3.27	14.39	6.90	7.92	2.91	0.37	0.64	-0.41

Table A4.1.2: Characteristics of Annual Minima, where D = 7

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
9001	40	15.48	56.51	25.67	28.91	8.63	0.30	1.30	1.79
9002	38	12.8	44.82	21.16	23.77	7.19	0.30	1.01	0.72
14001	32	16.28	40.56	26.55	26.53	6.63	0.25	0.35	-0.36
19002	38	8.37	33.33	14.61	15.28	5.27	0.35	1.44	2.81
19004	40	11.26	41.51	22.85	23.18	7.10	0.31	0.54	0.53
20001	39	14.10	44.89	22.13	23.26	6.85	0.29	0.99	1.29
20003	35	10.15	42.37	20.82	22.85	7.42	0.32	0.91	0.64
20005	35	12.39	32.80	20.25	20.72	5.66	0.27	0.57	-0.50
21006	38	9.89	31.76	18.51	18.08	4.93	0.27	0.54	0.28
21012	36	5.26	21.85	10.13	11.24	4.26	0.38	0.88	0.28
21013	35	6.82	33.05	14.03	15.33	5.69	0.37	1.24	1.60
21015	33	7.38	32.21	12.81	14.69	6.04	0.41	1.68	2.58
21017	34	3.78	19.19	9.48	9.56	3.72	0.39	0.64	0.05
28031	31	9.46	38.84	20.29	20.26	6.92	0.34	0.69	0.64
34003	40	36.88	78.88	57.03	55.29	10.60	0.19	0.08	-0.50
39016	38	10.84	63.11	43.23	43.73	10.21	0.23	-0.56	1.73
39028	31	27.52	63.48	44.60	44.38	8.49	0.19	-0.10	-0.22
43005	34	5.37	52.31	38.31	36.28	9.33	0.26	-1.00	2.38

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
43006	34	18.54	51.64	34.13	34.69	6.84	0.19	0.34	0.68
48010	30	13.16	41.73	21.71	23.06	7.06	0.31	1.03	1.27
51001	31	8.05	35.76	20.37	20.89	6.48	0.31	0.38	-0.28
55016	30	0.59	14.43	3.85	4.72	3.38	0.72	1.33	1.41
55026	61	1.60	18.75	6.45	7.15	3.56	0.50	1.05	1.29
60002	38	2.26	20.36	8.14	8.63	4.20	0.49	0.70	0.41
72004	40	3.51	16.8	7.912	8.95	3.55	0.39	0.80	-0.23

Table A4.1.3: Characteristics of Annual Minima, where D = 30

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
9001	40	17.84	65.87	29.23	33.57	10.99	0.33	1.17	1.32
9002	38	14.67	57.02	24.5	28.11	10.05	0.36	1.23	1.16
14001	32	18.21	50.78	28.99	29.55	8.13	0.27	0.69	0.15
19002	38	10.4	41.65	18.38	20.03	6.85	0.34	1.23	1.67
19004	40	15.46	71.97	27.53	29.14	10.99	0.38	1.81	5.00
20001	39	15.09	55.75	23.78	26.68	8.82	0.33	1.22	1.96
20003	35	13.20	54.34	23.21	25.81	9.30	0.36	1.21	1.42
20005	35	14.07	45.46	23.35	23.97	7.67	0.32	1.24	1.59
21006	38	10.72	52.50	22.06	23.06	8.33	0.36	1.66	4.10
21012	36	6.20	38.07	13.65	15.40	7.39	0.48	1.55	2.78
21013	35	8.00	57.64	16.42	19.10	9.47	0.50	2.22	7.06
21015	33	8.37	51.06	14.95	17.84	9.07	0.51	2.15	5.29
21017	34	5.38	31.92	13.28	14.48	6.65	0.46	0.75	-0.01
28031	31	10.46	51.27	24.16	23.92	9.51	0.40	1.04	1.36
34003	40	39.99	82.53	61.05	59.59	11.12	0.19	0.15	-0.42
39016	38	13.35	69.41	46.87	47.14	10.98	0.23	-0.44	1.32
39028	31	28.17	67.17	46.83	46.55	9.21	0.20	-0.12	-0.21
43005	34	9.86	56.81	40.58	38.79	9.68	0.25	-0.68	1.12
43006	34	20.66	56.57	36.36	37.11	7.174	0.19	0.49	0.94
48010	30	14.33	51.26	25.4	26.02	8.64	0.33	1.24	2.16
51001	31	8.638	40.46	21.82	23.15	7.54	0.33	0.57	-0.02
55016	30	1.12	22.12	5.29	7.67	5.83	0.76	1.33	0.79
55026	61	2.13	31.65	10.54	12.03	6.27	0.52	1.17	1.53
60002	38	2.98	29.53	11.46	13.02	7.03	0.54	0.70	-0.19
72004	40	5.15	38.24	11.3	14.99	8.36	0.56	0.96	0.14

Table A4.1.4: Characteristics of Annual Minima, where D = 60

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
9001	40	19.10	104.70	32.78	39.25	16.46	0.42	2.00	5.52
9002	38	15.55	92.32	27.53	33.45	14.96	0.45	1.93	5.36
14001	32	19.05	54.07	32.35	32.15	8.56	0.27	0.76	0.12
19002	38	13.12	61.82	23.08	26.32	10.73	0.41	1.40	2.11
19004	40	17.60	83.44	33.21	34.60	13.67	0.40	1.58	3.35
20001	39	16.87	63.23	26.80	30.64	11.10	0.36	1.06	0.89
20003	35	14.79	66.60	26.17	29.25	11.55	0.39	1.38	2.07
20005	35	14.69	64.24	24.88	28.21	11.26	0.40	1.66	3.17
21006	38	11.71	72.26	26.02	29.23	11.95	0.41	1.55	3.49
21012	36	8.51	67.36	18.56	22.63	13.67	0.60	1.49	2.26
21013	35	10.13	68.79	20.93	25.23	14.38	0.57	1.55	1.93
21015	33	10.22	65.07	17.58	22.48	13.14	0.58	1.88	3.51

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
21017	34	5.95	70.54	21.64	24.81	14.69	0.59	1.12	1.47
28031	31	11.38	67.67	27.34	29.05	13.83	0.48	1.18	1.57
34003	40	43.26	87.69	64.74	63.85	12.25	0.19	0.10	-0.71
39016	38	15.41	73.57	49.86	50.39	11.84	0.24	-0.36	0.90
39028	31	28.52	73.97	48.42	48.66	10.22	0.21	0.00	0.13
43005									
43006	34	21.88	64.67	39.19	40.10	8.202	0.20	0.60	1.56
48010	30	16.14	67.27	27.76	29.37	11.16	0.38	1.71	3.94
51001	31	9.29	45.98	23.73	25.39	8.45	0.33	0.59	0.11
55016	30	1.77	40.77	9.06	12.42	10.45	0.84	1.50	1.71
55026	61	4.37	62.91	18.75	20.48	11.24	0.55	1.46	3.08
60002	38	4.01	51.36	16.92	19.47	11.01	0.57	0.90	0.71
72004	40	6.66	59.41	19.64	22.46	12.72	0.57	0.99	0.50

Table A4.1.5: Characteristics of Annual Minima, where D = 90

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
9001	40	21.06	111.70	41.31	45.33	18.68	0.41	1.56	3.27
9002	36	16.92	106.9	33.55	39.67	19.54	0.49	1.69	3.4
14001	32	20.38	55.80	34.35	34.85	9.01	0.26	0.63	-0.21
19002	38	13.94	66.5	29.66	31.49	13.49	0.43	1.05	0.41
19004	40	19.88	93.67	36.57	39.69	15.78	0.40	1.31	2.30
20001	39	17.54	73.35	30.05	34.75	14.67	0.42	1.21	0.82
20003	35	16.43	76.76	28.58	33.45	14.86	0.44	1.40	1.63
20005	35	16.89	73.97	29.34	32.37	13.81	0.43	1.34	1.67
21006	38	14.11	77.06	32.37	33.69	12.60	0.37	1.23	2.51
21012	36	9.05	73.43	24.01	28.59	15.97	0.56	1.07	0.83
21013	35	13.07	73.09	24.15	29.33	15.58	0.53	1.29	1.06
21015	33	12.80	77.50	20.99	26.54	15.89	0.60	1.66	2.46
21017	34	6.95	78.58	28.86	31.69	16.29	0.51	1.00	1.03
28031	31	12.62	65.78	33.02	33.48	14.87	0.44	0.56	-0.36
34003	40	44.55	90.69	67.86	66.49	12.73	0.19	0.08	-0.70
39016	38	17.43	78.10	53.58	52.79	12.66	0.24	-0.23	0.54
39028	31	29.70	80.16	50.23	50.61	11.15	0.22	0.13	0.40
43005	34	13.53	65.06	46.20	43.68	11.09	0.25	-0.39	0.49
43006	34	23.13	64.77	41.56	42.65	8.88	0.21	0.47	0.63
48010	30	16.86	70.84	30.95	32.36	12.79	0.39	1.55	2.72
51001	30	10.71	52.11	26.15	27.19	8.80	0.32	0.73	1.07
55016	30	2.83	41.31	12.30	16.19	11.96	0.74	1.03	-0.19
55026	60	5.42	65.75	26.64	28.23	12.83	0.45	0.71	0.45
60002	38	4.43	58.87	25.92	27.42	14.34	0.52	0.46	-0.36
72004	38	11.51	69.51	27.56	31.61	15.06	0.48	0.77	-0.11

Table A4.1.6: Characteristics of Annual Minima, where D = 180

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
9001	40	29.55	115.10	56.88	60.15	19.94	0.33	0.73	0.12
9002	36	25.72	113.2	49.82	54.56	22.12	0.4	0.93	0.24
14001	32	21.67	94.74	46.06	48.58	15.95	0.33	1.15	1.70
19002	37	25.12	88.36	43.36	47.52	16.73	0.35	0.51	-0.61
19004	39	29.25	102.60	54.84	56.79	18.72	0.33	0.55	-0.15

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
20001	38	20.01	106.00	44.05	50.29	23.64	0.47	0.88	-0.30
20003	34	20.01	104.50	40.17	47.91	22.86	0.48	0.93	-0.21
20005	34	20.06	90.67	41.41	46.79	21.14	0.45	0.74	-0.67
21006	38	22.66	86.08	52.48	51.49	15.93	0.31	0.27	-0.58
21012	36	14.51	89.23	44.60	46.28	18.41	0.40	0.43	-0.56
21013	35	18.81	96.58	41.71	47.82	22.54	0.47	0.64	-0.67
21015	33	16.79	98.75	37.49	42.31	22.38	0.53	0.97	0.18
21017	34	17.64	92.32	50.72	53.31	16.26	0.31	0.29	0.20
28031	31	20.52	82.47	50.83	49.05	18.41	0.38	-0.02	-0.75
34003	40	52.86	104.30	74.04	73.87	13.56	0.18	0.28	-0.54
39016	38	24.07	105.40	63.05	63.16	16.60	0.26	0.08	0.11
39028	31	34.73	93.61	61.81	60.25	14.04	0.23	-0.02	-0.22
43005	34	19.23	82.82	55.96	54.57	14.74	0.27	-0.34	-0.33
43006	33	28.79	76.63	51.81	52.37	11.88	0.23	0.11	-0.44
48010	30	23.57	80.75	43.29	44.63	15.86	0.35	0.74	-0.001
51001	29	18.80	77.51	39.13	39.11	14.04	0.36	0.91	0.96
55016	29	7.29	63.97	26.52	31.34	16.38	0.52	0.33	-1.04
55026	59	14.65	83.26	43.38	45.17	15.10	0.33	0.29	-0.04
60002	38	8.39	75.93	42.12	44.02	17.56	0.40	-0.22	-0.60
72004	37	18.26	86.81	51.59	51.04	17.22	0.34	-0.02	-0.41

Table A4.1.7: Characteristics of Annual Minima, where D = 365

Station	No. Years	Min.	Max.	Median	Mean	Standard Deviation	C.V.	Skewness	Kurtosis
9001	39	40.58	122.90	82.69	82.12	20.97	0.26	-0.27	-0.66
9002	35	33.02	122.7	84.85	80.95	22.76	0.28	-0.38	-0.4
14001	31	36.60	124.30	86.01	82.98	21.88	0.26	-0.44	-0.42
19002	37	50.13	110.8	85.70	82.85	15.81	0.19	-0.33	-0.51
19004	39	44.82	104.90	92.13	85.55	17.08	0.20	-0.91	-0.18
20001	38	23.67	120.30	82.07	78.56	24.82	0.32	-0.54	-0.50
20003	34	24.72	117.80	79.96	78.23	24.28	0.31	-0.46	-0.43
20005	34	28.20	124.00	80.36	78.47	22.32	0.28	-0.50	0.00
21006	37	47.57	104.10	90.99	84.36	15.42	0.18	-0.78	-0.35
21012	35	44.49	102.90	86.49	82.67	15.67	0.19	-0.77	-0.28
21013	34	36.49	115.50	88.04	82.42	19.26	0.23	-0.76	0.17
21015	32	28.47	107.70	82.86	77.95	22.33	0.29	-0.67	-0.38
21017	33	52.58	104.50	88.63	84.72	14.78	0.17	-0.49	-0.95
28031	31	48.18	117.90	84.93	84.52	18.75	0.22	-0.23	-0.52
34003	40	65.28	121.20	90.53	88.55	14.93	0.17	0.28	-0.85
39016	37	36.31	122.00	88.94	84.30	23.24	0.28	-0.45	-0.84
39028	31	42.94	119.10	93.23	85.26	22.80	0.27	-0.40	-1.15
43005	34	28.11	126.00	88.46	83.91	24.71	0.30	-0.54	-0.35
43006	33	35.75	127.00	88.14	84.32	22.75	0.27	-0.32	-0.22
48010	29	58.27	118.10	82.7	84.83	17.51	0.21	0.33	-0.88
51001	29	28.09	127.90	78.42	75.50	23.75	0.31	-0.05	0.05
55016									
55026	58	54.64	131.10	80.23	82.40	15.54	0.19	0.47	0.45
60002	36	51.80	109.10	81.14	80.21	16.22	0.20	-0.05	-1.10
72004	36	39.37	113.40	82.68	83.31	16.45	0.20	-0.66	0.79

APPENDIX 4.2:
REGRESSION TEST FOR STATIONARITY

Table A4.2.1: Results of Regression Tests

Station	Duration	Regression			Coefficients			
					Constant		Year	
		No. Years	R^2	Std. Error	Value	Std. Error	Value	Std. Error
9001	1	40	0	7.90	45.28	214.30	-0.09	0.11
9001	30	40	0.004	11.11	156.88	301.28	-0.06	0.15
9001	90	40	0.001	19.92	116.10	512.89	-0.04	0.26
9001	365	39	0.044	20.79	843.42	585.55	-0.38	0.30
9002	1	38	0.006	6.77	113.22	191.53	-0.05	0.10
9002	30	38	0.006	10.16	156.63	287.70	-0.07	0.15
9002	30	36	0	19.82	109.52	567.65	-0.04	0.29
9002	30	35	0.04	22.64	867.68	673.34	-0.40	0.34
14001	1	32	0.018	6.06	-143.1	229.0	0.09	0.12
14001	30	32	0.026	8.15	-249.8	309.5	0.14	0.16
14001	90	32	0.032	9.01	-306.1	342.1	0.17	0.17
14001	365	31	0.162	20.37	-1840.5	811.6	0.97	0.41
19002	1	38	0.104	4.43	278.2	129.9	-0.13	0.07
19002	30	38	0.104	6.57	414.2	192.4	-0.20	0.10
19002	90	38	0.073	13.17	680.5	385.8	-0.33	0.20
19002	365	37	0.001	16.02	190.9	488.7	-0.06	0.25
19004	1	40	0.028	6.63	210.2	179.7	-0.10	0.09
19004	30	40	0.005	11.11	160.8	301.2	-0.07	0.15
19004	90	40	0.001	15.97	133.6	433.1	-0.05	0.22
19004	365	40	0.045	16.92	-540.2	476.7	0.32	0.24
20001	1	39	0.027	6.71	-171.5	188.9	0.10	0.10
20001	30	39	0.001	8.93	-32.3	-251.7	0.03	0.13
20001	90	39	0.007	14.81	252.5	417.2	-0.11	0.21
20001	365	38	0	25.16	158.4	737.1	-0.04	0.37
20003	1	35	0.001	6.93	61.7	229.8	-0.02	0.12
20003	30	35	0.004	9.42	132.3	312.5	-0.05	0.16
20003	90	35	0	15.08	71.6	500.1	-0.02	0.25
20003	365	34	0.014	24.47	-503.3	848.1	0.29	0.43
20005	1	35	0.069	5.49	-266.4	182.2	0.14	0.09
20005	30	35	0.012	7.74	-140.0	256.7	0.08	0.13
20005	90	35	0.003	14.00	-125.4	464.3	0.08	0.23
20005	365	35	0.008	22.57	-328.0	782.3	0.21	0.40
21006	1	38	0.1	4.20	-230.2	123.1	0.12	0.06
21006	30	38	0.015	8.38	-161.6	245.5	0.09	0.12
21006	90	38	0	12.78	41.0	374.3	0.00	0.19
21006	395	37	0.098	14.89	-778.8	454.3	0.44	0.23
21012	1	36	0.029	3.63	126.8	115.2	-0.06	0.06
21012	30	36	0.028	7.39	246.4	235.0	-0.12	0.12
21012	90	36	0.022	16.02	476.7	509.3	-0.23	0.26
21012	365	35	0.108	15.02	-912.8	498.3	0.50	0.25
21013	1	35	0.01	5.19	121.80	172.27	-0.05	0.09
21013	30	35	0.00	9.59	133.67	318.22	-0.06	0.16
21013	90	35	0.01	15.73	343.66	521.82	-0.16	0.26
21013	365	34	0.03	19.25	-593.62	667.03	0.34	0.34
21015	1	33	0.03	5.56	-190.98	201.64	0.10	0.10
21015	30	33	0.03	9.10	-273.50	329.77	0.15	0.17

Station	Duration	Regression			Coefficients			
					Constant		Year	
		No. Years	R^2	Std. Error	Value	Std. Error	Value	Std. Error
21015	90	33	0.02	16.01	-393.91	580.24	0.21	0.29
21015	365	32	0.01	22.63	-275.09	859.53	0.18	0.43
21017	1	34	0.00	16.53	-105.88	572.70	0.07	0.29
21017	30	34	0.03	6.65	-209.94	230.51	0.11	0.12
21017	90	34	0.00	16.53	-105.88	572.70	0.07	0.29
21017	365	33	0.30	12.53	-1585.16	454.29	0.84	0.23
28031	1	31	0.002	6.55	75.47	260.78	-0.03	0.13
28031	30	31	0.001	9.67	-42.19	385.16	0.03	0.19
28031	90	31	0.001	15.12	-50.22	602.44	0.04	0.30
28031	365	31	0.004	19.03	341.72	758.17	-0.13	0.38
34003	1	40	0.002	10.70	138.01	290.02	-0.04	0.15
34003	30	40	0.003	11.25	167.81	305.06	-0.06	0.15
34003	90	40	0.001	12.90	120.73	349.66	-0.03	0.18
34003	365	40	0.001	15.12	169.25	409.99	-0.04	0.21
39016	1	38	0.068	9.78	507.44	286.64	-0.24	0.15
39016	30	38	0.062	10.78	534.29	315.95	-0.25	0.16
39016	90	38	0.037	12.60	484.62	369.03	-0.22	0.19
39016	365	37	0.004	23.52	336.72	717.55	-0.13	0.36
39028	1	31	0.072	7.73	505.84	307.90	-0.23	0.16
39028	30	31	0.095	8.91	667.47	355.05	-0.31	0.18
39028	90	31	0.099	10.76	816.27	478.80	-0.39	0.22
39028	365	31	0.039	22.74	1061.72	905.90	-0.49	0.46
43005	1	34	0.05	8.99	426.39	311.39	-0.20	0.16
43005	30	34	0.05	9.57	478.66	331.61	-0.22	0.17
43005	90	34	0.04	11.02	503.44	381.76	-0.23	0.19
43005	365	34	0.02	24.84	788.94	860.70	-0.36	0.43
43006	1	34	0.03	6.47	234.21	224.06	-0.10	0.11
43006	30	34	0.02	7.20	256.48	249.49	-0.11	0.13
43006	90	34	0.02	8.92	305.08	309.11	-0.13	0.16
43006	365	33	0.00	23.07	-207.21	836.47	0.15	0.42
48010	1	30	0.02	6.60	-199.91	276.39	0.11	0.14
48010	30	30	0.04	8.61	-371.95	360.45	0.20	0.18
48010	90	30	0.06	12.61	-687.66	527.69	0.36	0.27
48010	365	30	0.03	17.60	-572.36	775.38	0.33	0.39
50001	1	31	0.09	5.85	-339.88	212.01	0.18	0.11
50001	30	31	0.11	7.23	-478.42	262.14	0.25	0.13
50001	90	31	0.07	8.62	-461.50	327.32	0.25	0.17
50001	365	29	0.02	23.99	-553.16	955.59	0.32	0.48
55016	1	30	0.026	3.023	-103.31	123.56	0.05	0.06
55016	30	30	0.049	5.785	-273.20	236.49	0.14	0.12
55016	90	30	0.025	12.014	-402.52	491.07	0.21	0.26
55026	1	61	0.01	3.085	39.02	43.13	-0.02	0.02
55026	30	61	0.002	6.317	41.25	88.33	-0.02	0.05
55026	90	60	0.001	12.932	-5.41	181.03	0.02	0.09
55026	365	58	0.161	14.46	-610.28	211.12	0.35	0.11
60002	1	38	0.02	3.59	97.06	105.05	-0.05	0.05
60002	30	38	0.045	6.96	279.18	203.99	-0.13	0.10
60002	90	38	0.05	14.16	600.56	414.97	-0.29	0.21
60002	365	38	0.003	16.44	-75.06	522.29	0.08	0.26
72004	1	40	0.01	2.93	61.33	73.86	-0.03	0.04
72004	30	40	0.06	8.23	329.36	207.44	-0.16	0.11
72004	90	38	0.12	14.36	816.89	362.54	-0.40	0.18
72004	365	38	0.01	16.61	366.50	451.38	-0.13	0.23

APPENDIX 4.3:
SPEARMAN'S RANK TEST

Table A4.3.1: Spearman's Correlation Coefficients

	Correlation Coefficient between Annual Minima and Year						
Station	**D= 1**	**D= 7**	**D= 30**	**D= 60**	**D= 90**	**D= 180**	**D= 365**
9001	-0.169	-0.211	-0.208	-0.163	-0.107	-0.095	-0.255
9002	-0.185	-0.182	-0.199	-0.164	-0.120	-0.138	-0.217
14001	0.159	0.168	0.139	0.177	0.194	0.271	0.336
19002	-0.334	-0.322	-0.296	-0.263	-0.281	-0.285	-0.038
19004	-0.236	-0.247	-0.218	-0.156	-0.129	-0.127	0.257
20001	0.119	0.045	-0.025	-0.105	-0.079	-0.160	-0.035
20003	0.090	-0.001	-0.040	-0.043	0.048	0.010	0.075
20005	0.361	0.274	0.171	0.114	0.117	0.014	0.107
21006	0.285	0.251	0.082	-0.020	-0.086	-0.139	0.278
21012	-0.097	-0.029	-0.108	-0.161	-0.154	-0.108	0.262
21013	-0.158	-0.182	-0.174	-0.208	-0.154	-0.048	0.127
21015	0.169	0.140	0.043	-0.015	0.033	0.121	-0.024
21017	0.088	0.085	0.079	0.130	0.116	0.147	0.538
28031	-0.109	-0.111	-0.053	-0.022	-0.032	-0.057	-0.056
34003	-0.057	-0.098	-0.074	-0.047	-0.037	0.002	-0.036
39016	-0.294	-0.299	-0.261	-0.223	-0.210	-0.200	-0.073
39028	-0.322	-0.367	-0.330	-0.342	-0.320	-0.282	-0.166
43005	-0.295	-0.282	-0.235		-0.194	-0.138	-0.174
43006	-0.101	-0.061	-0.066	-0.059	-0.017	0.022	0.065
48010	0.147	0.206	0.185	0.183	0.192	0.247	0.143
51001	0.167	0.199	0.189	0.229	0.221	0.225	-0.004
55016	0.0002	0.047	0.069	0.068	0.103	0.215	
55026	-0.083	-0.073	-0.077	-0.022	0.006	0.084	0.390
60002	-0.149	-0.209	-0.194	-0.084	-0.185	-0.208	0.016
72004	-0.192	-0.203	-0.27	-0.249	-0.413	-0.484	-0.047

APPENDIX 4.4:
PARTIAL AUTO-CORRELATION PLOTS

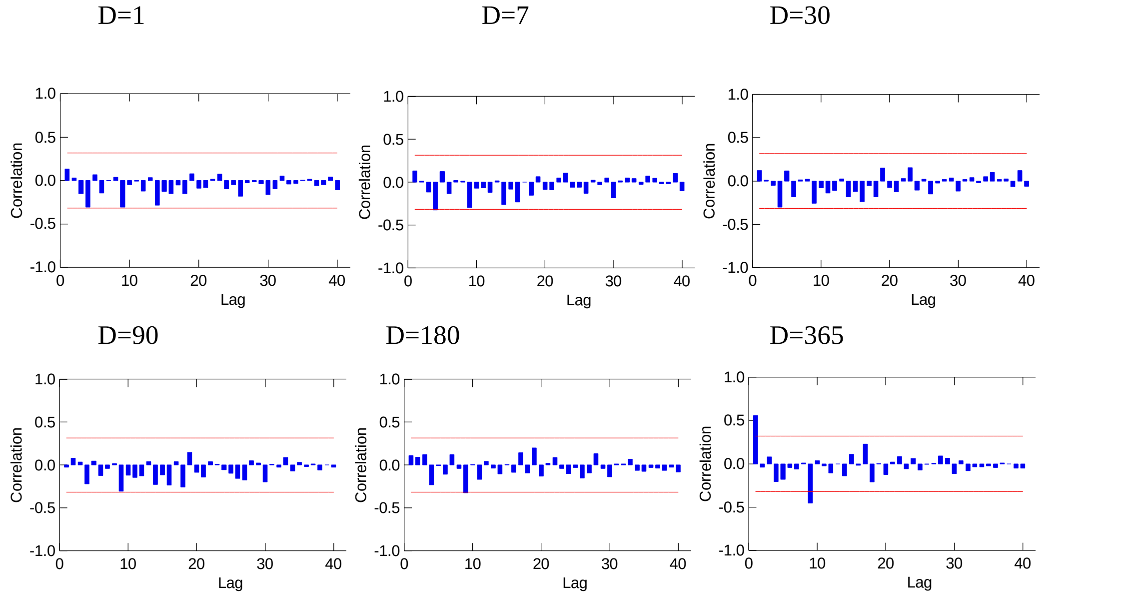


Figure A4.4.1: – Station 9001

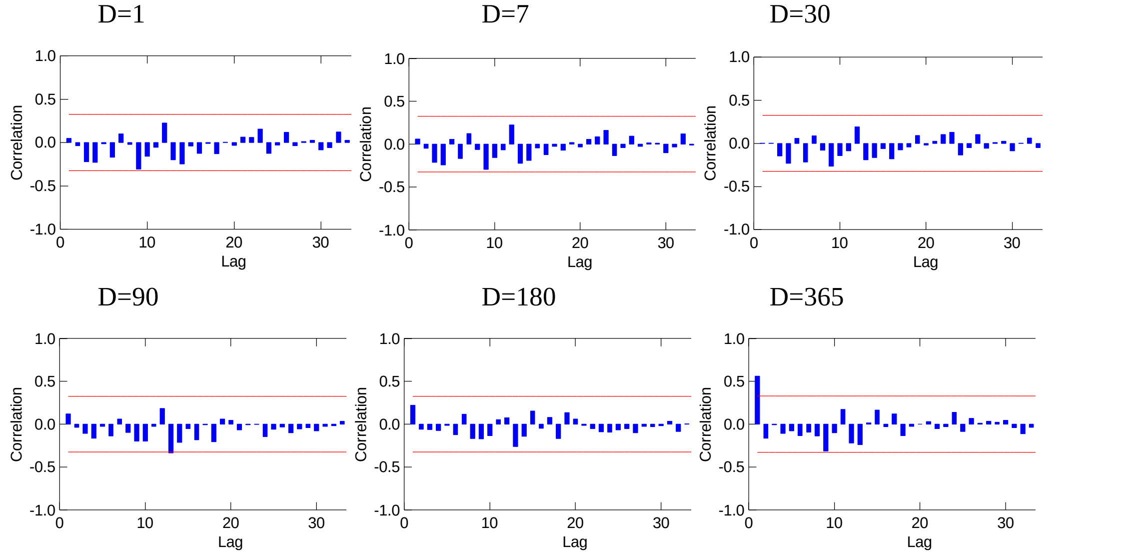


Figure A4.4.2: – Station 9002

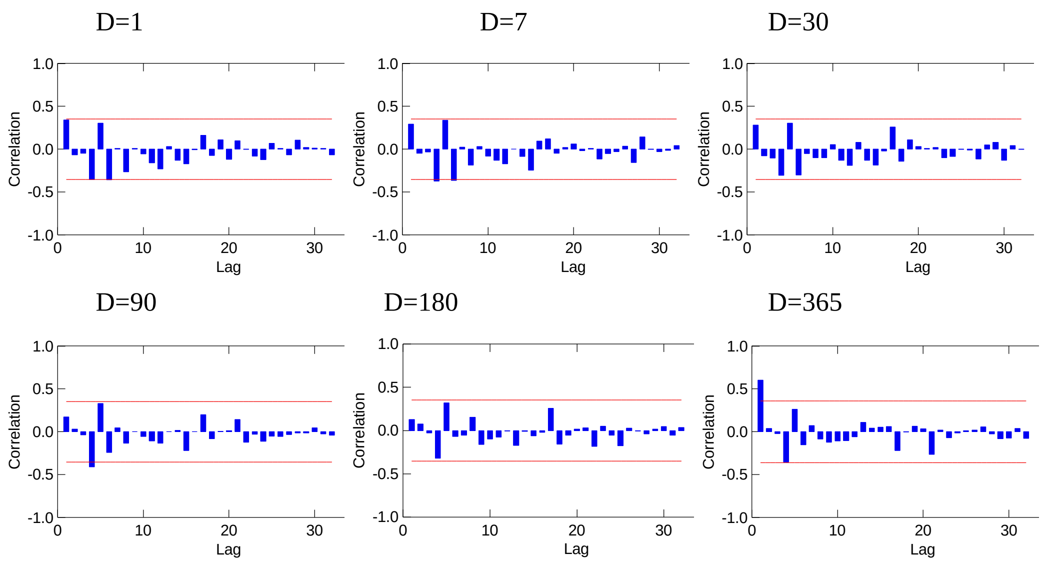


Figure A4.4.3: – Station 14002

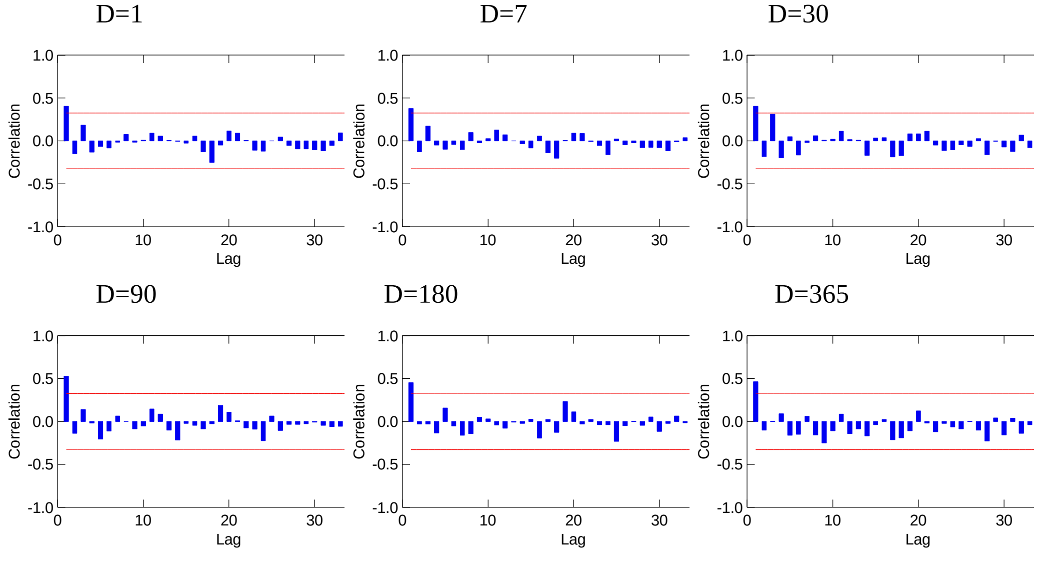


Figure A4.4.4: – Station 19002

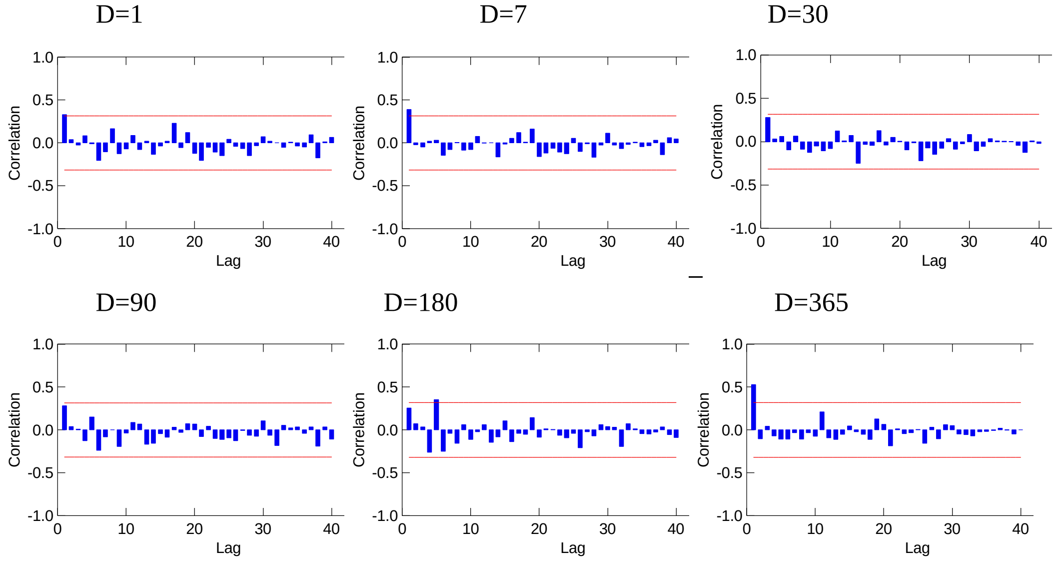


Figure A4.4.5: – Station 19004

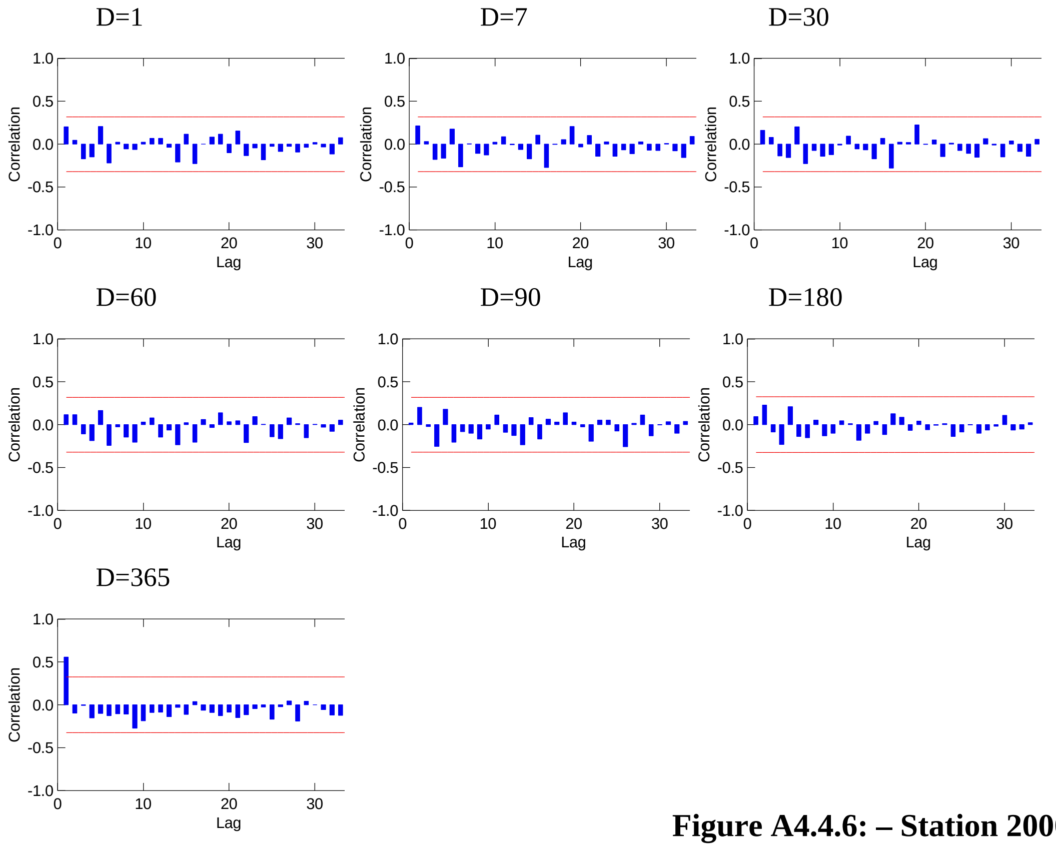


Figure A4.4.6: – Station 20001

D=1 D=60 D=30

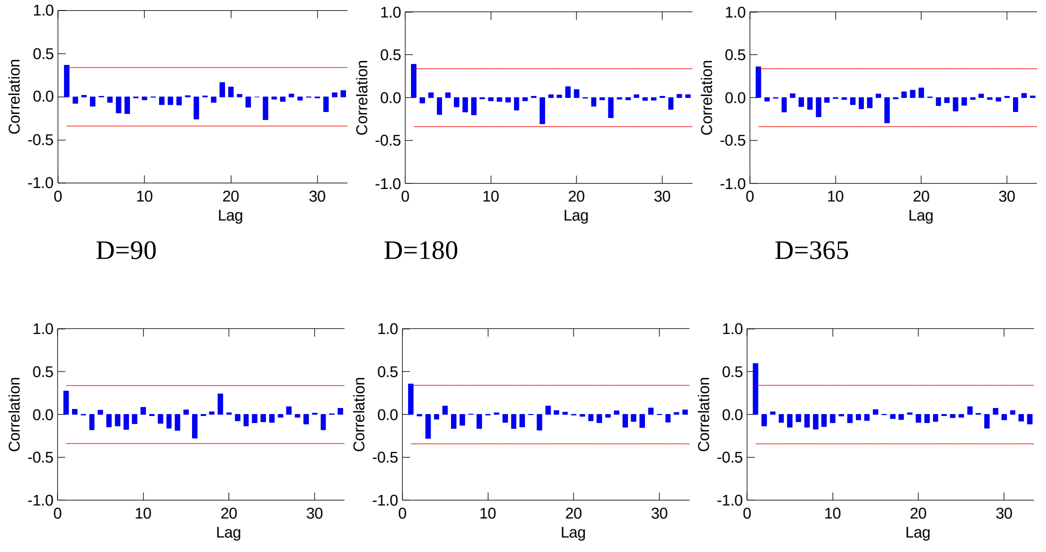


Figure A4.4.7: – Station 20003

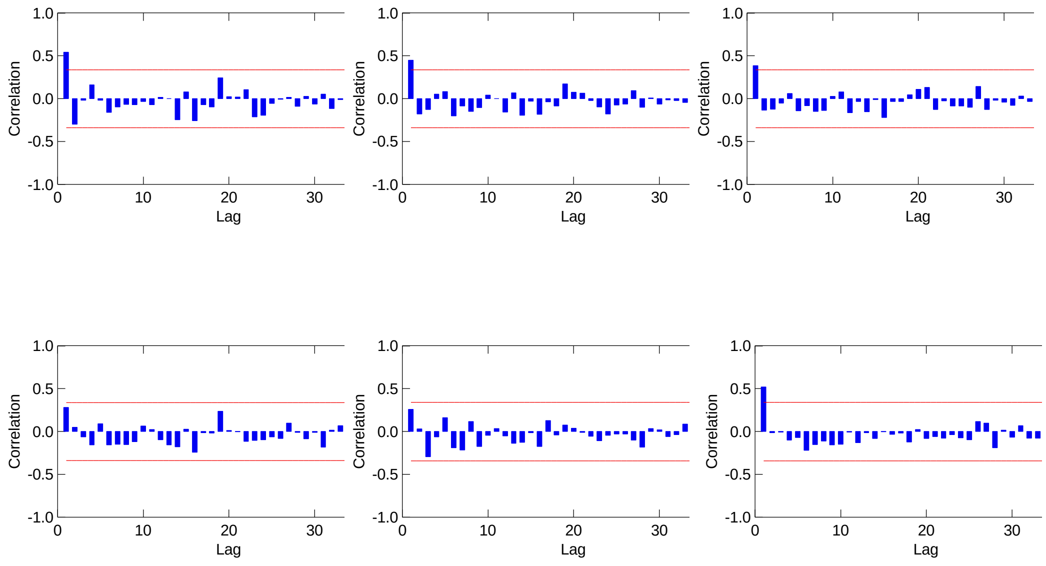


Figure A4.5.8: – Station 20005

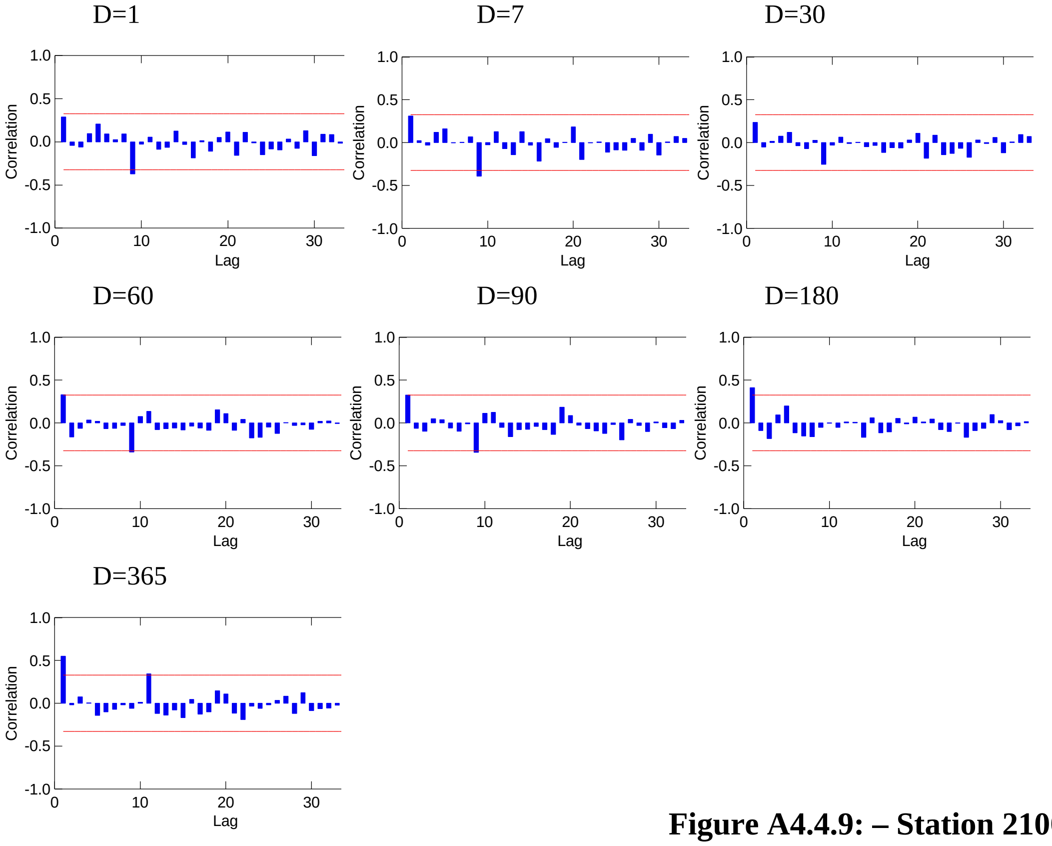


Figure A4.4.9: – Station 21006

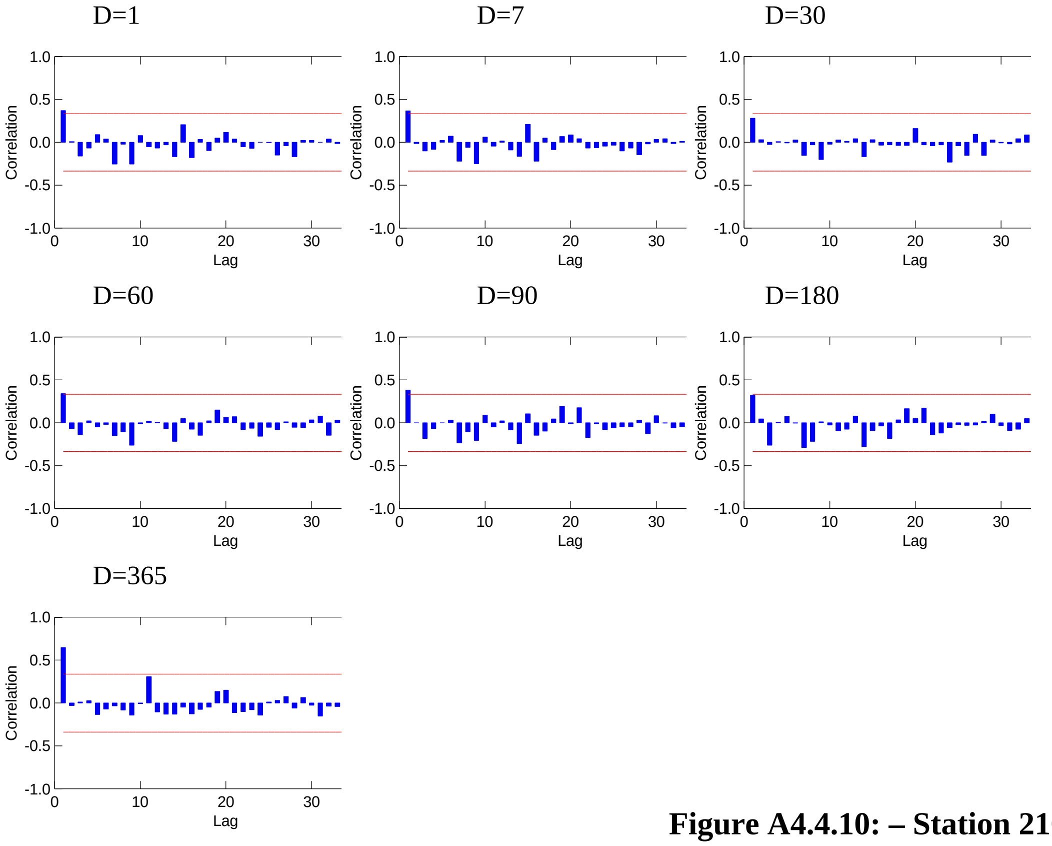


Figure A4.4.10: – Station 21012

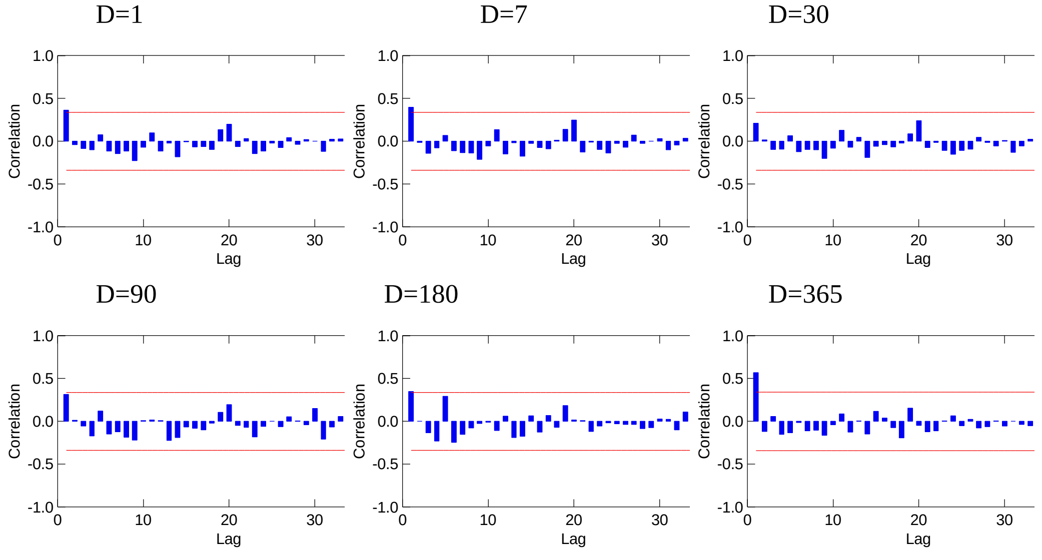


Figure A4.4.11: – Station 21013

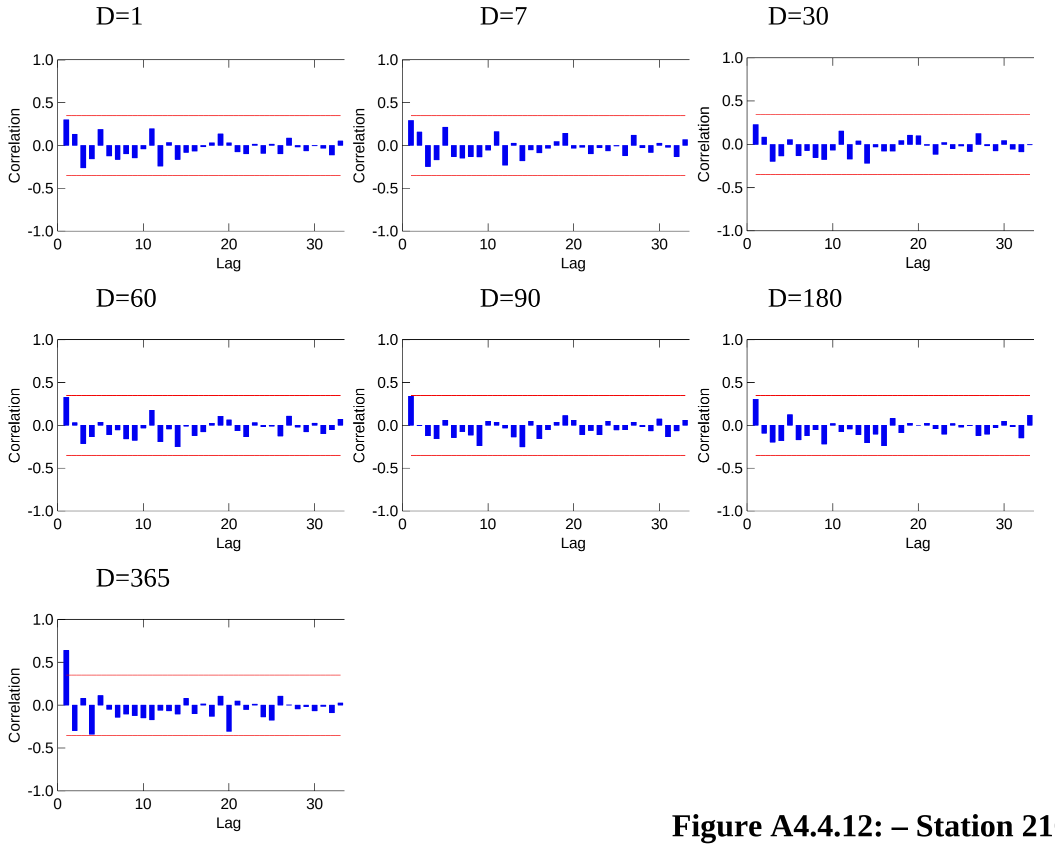


Figure A4.4.12: – Station 21015

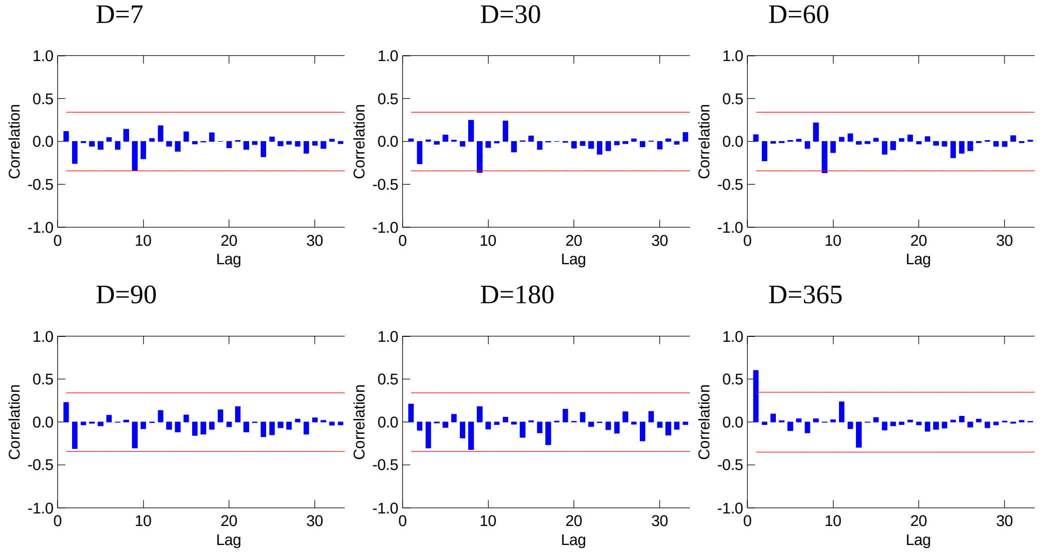


Figure A4.4.13: – Station 21017

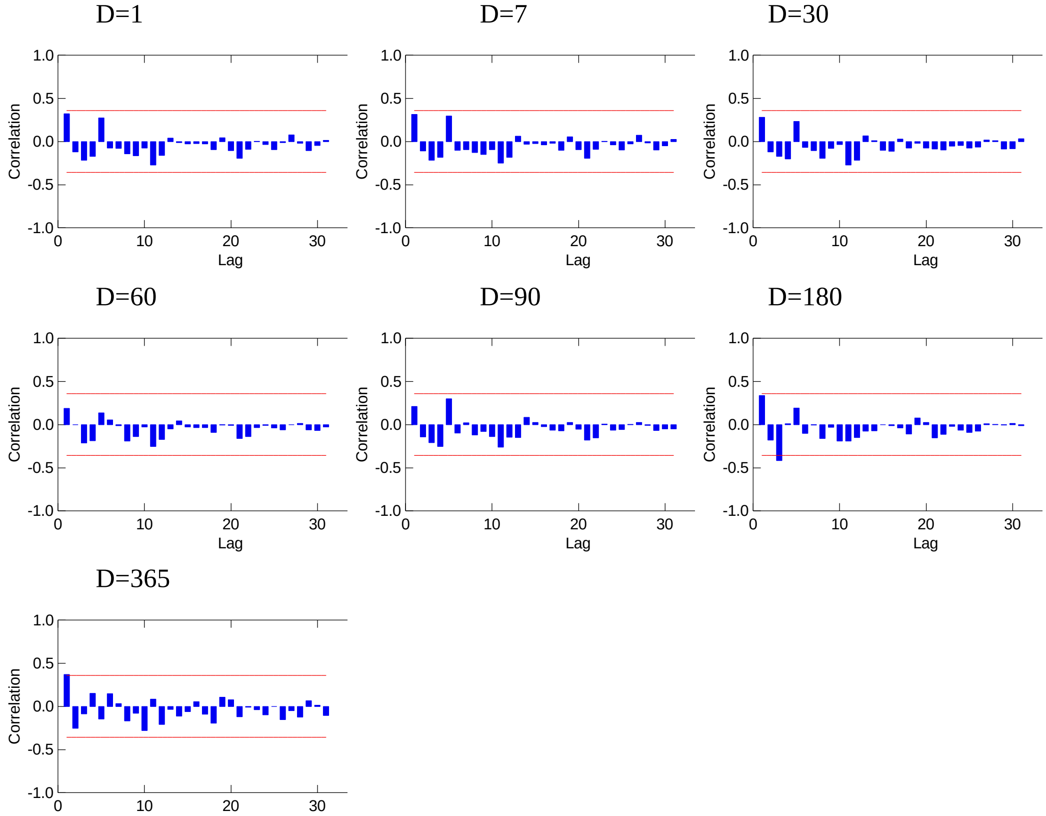


Figure A4.4.14: – Station 28031

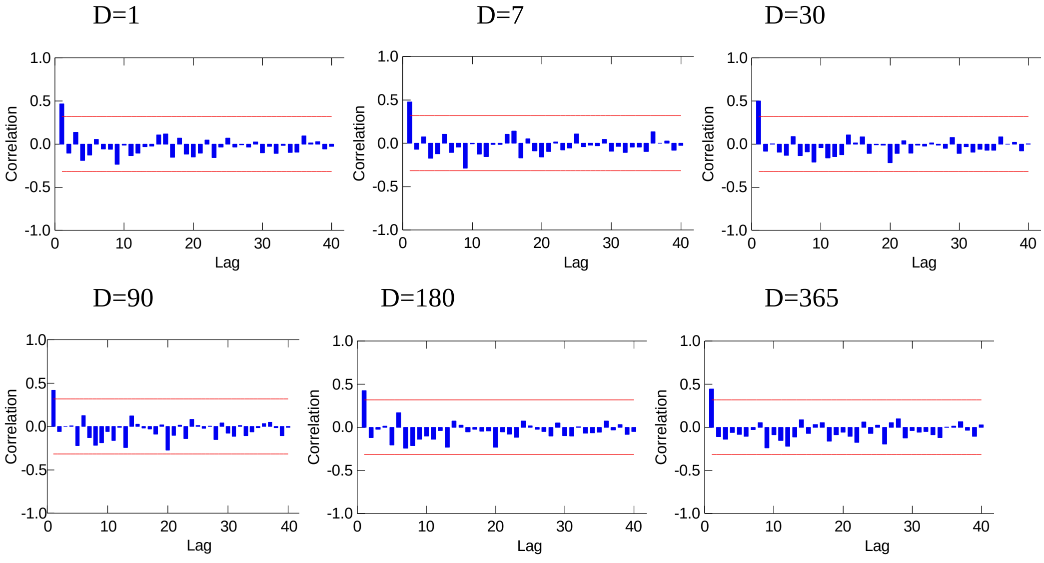


Figure A4.4.15: – Station 34003

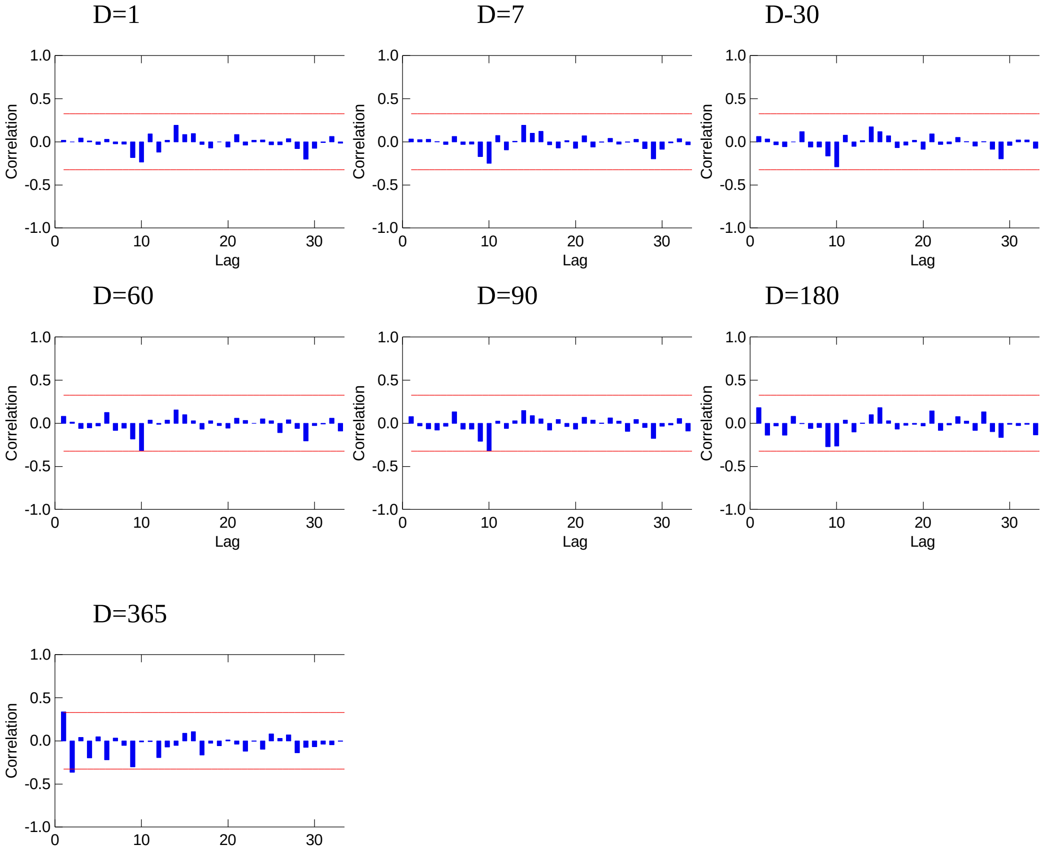


Figure A4.4.16: – Station 39016.

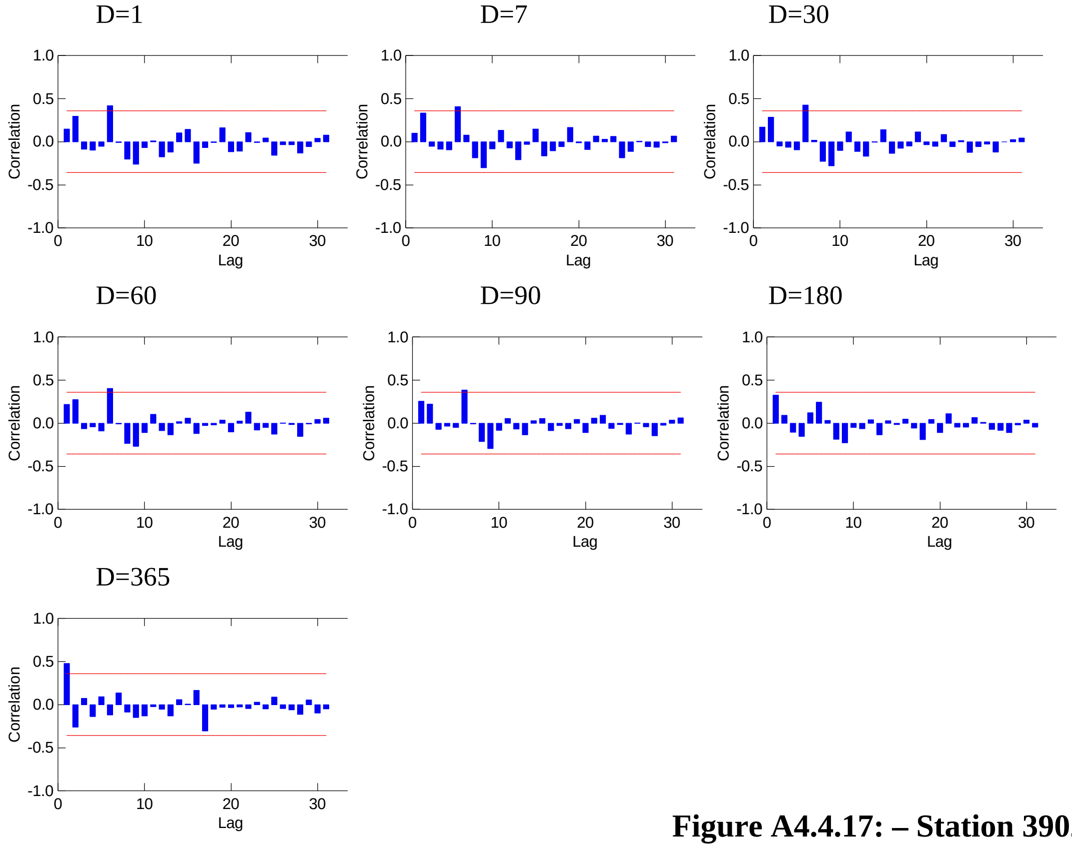


Figure A4.4.17: – Station 39028

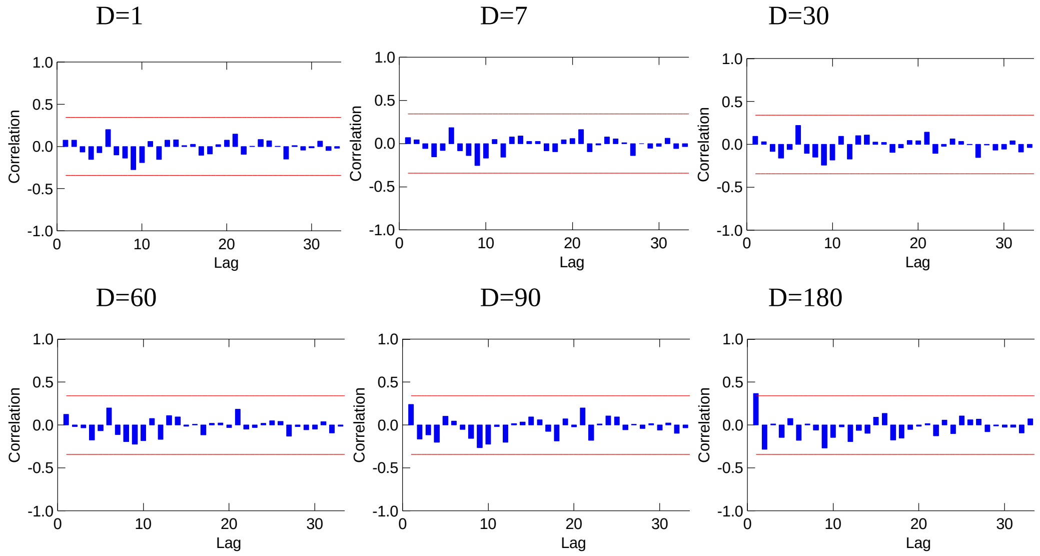


Figure A4.4.18: – Station 43005

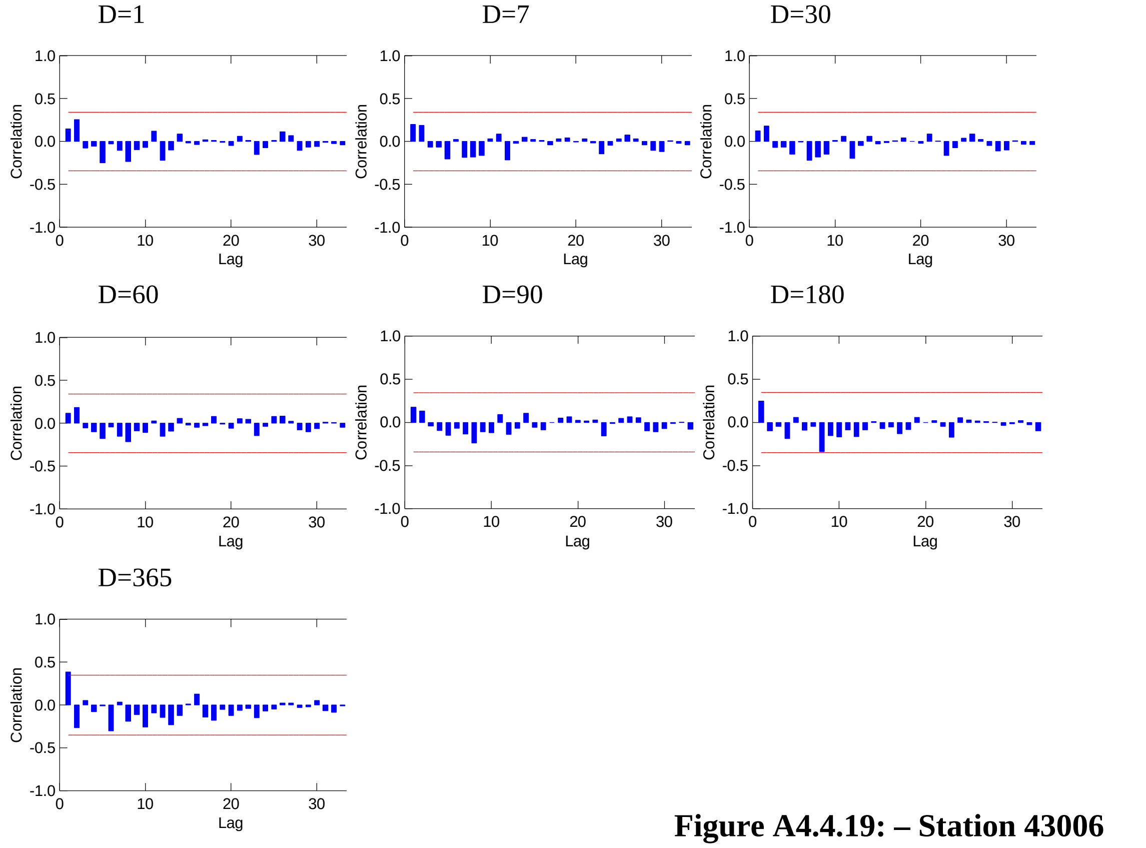


Figure A4.4.19: – Station 43006

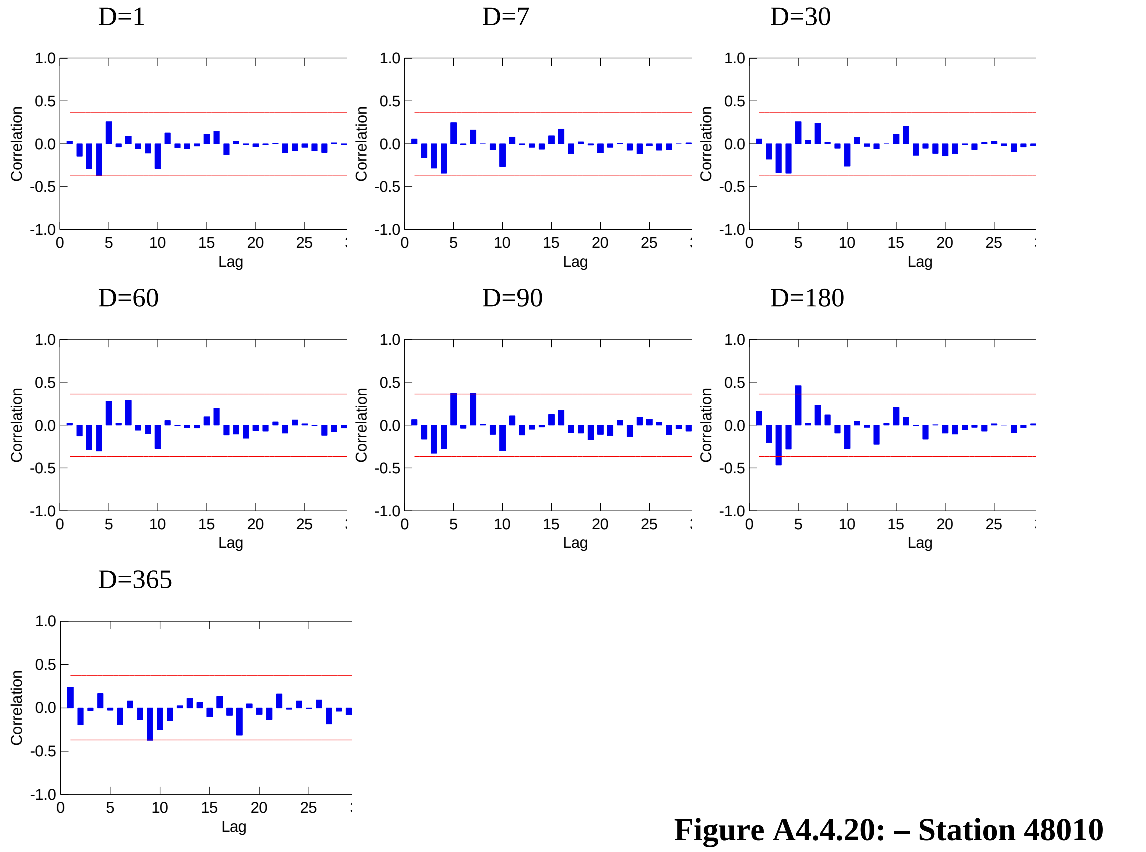


Figure A4.4.20: – Station 48010

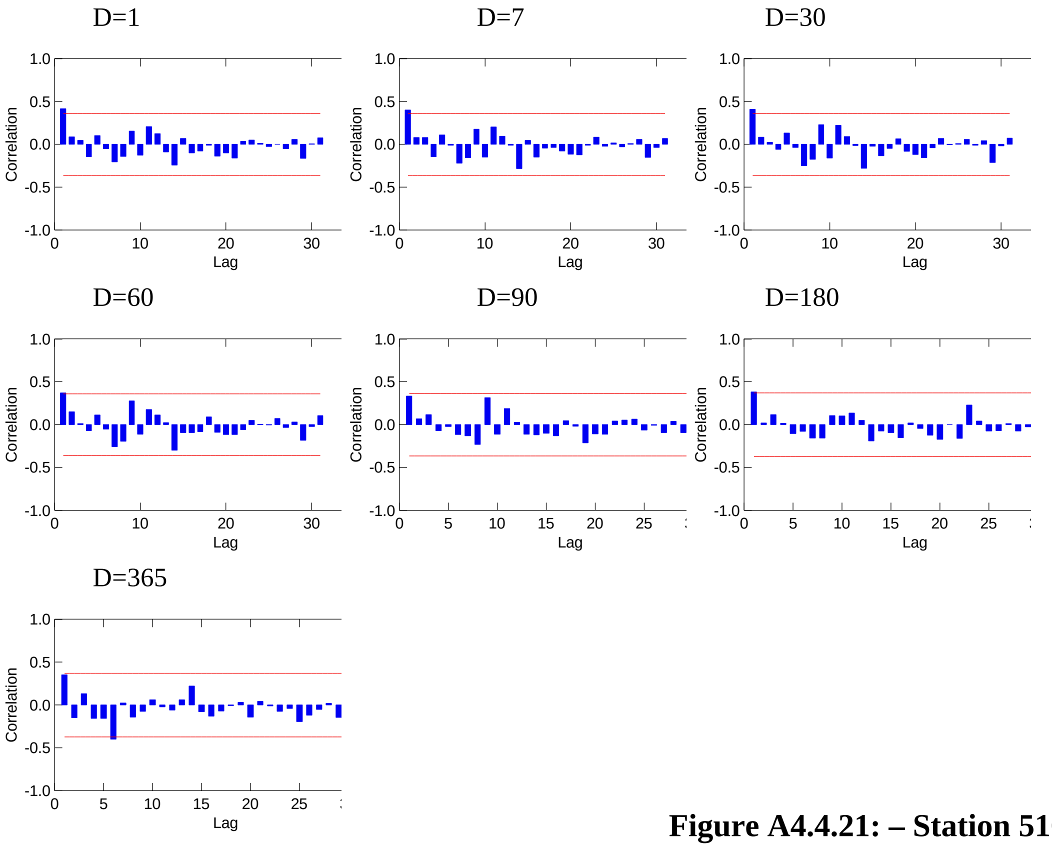


Figure A4.4.21: – Station 51001

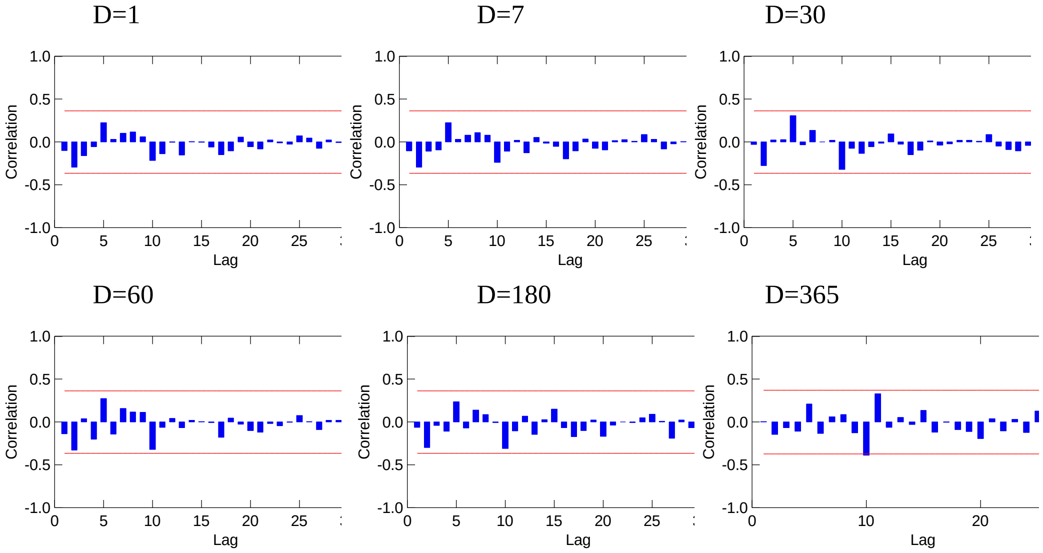


Figure A4.4.22: – Station 55016

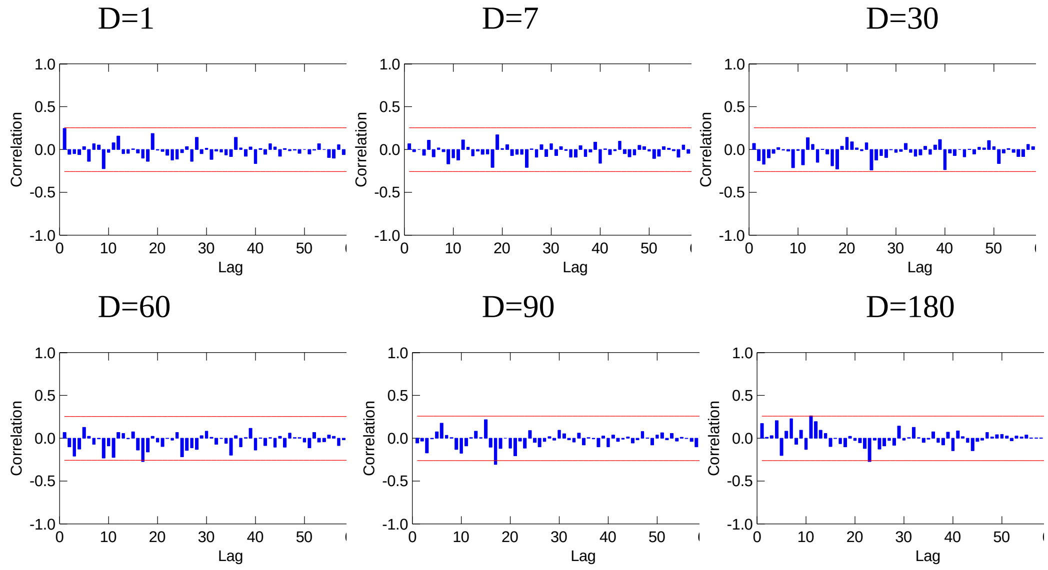


Figure A4.4.23: – Station 55026

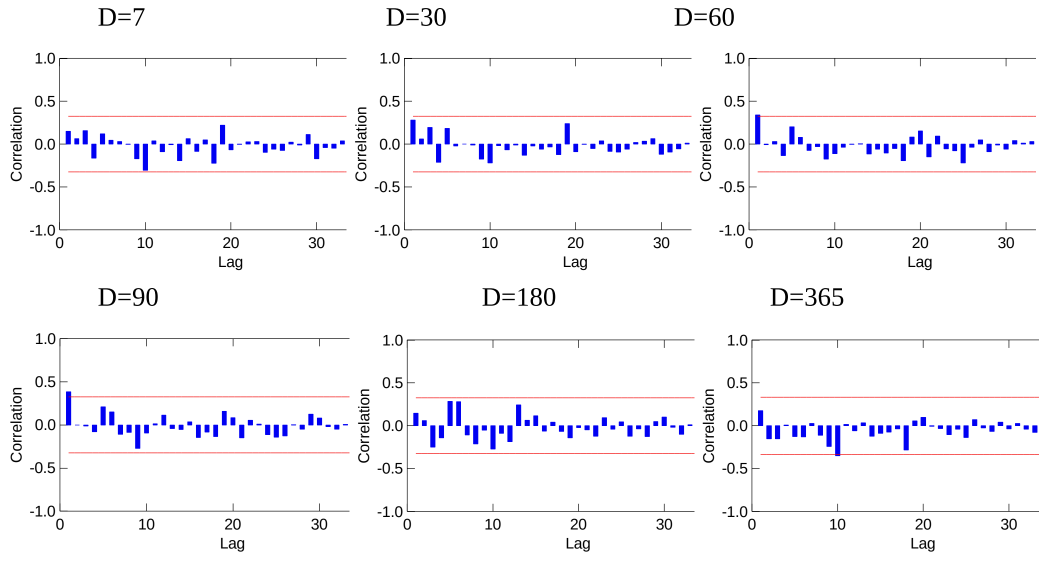


Figure A4.4.24: – Station 60002

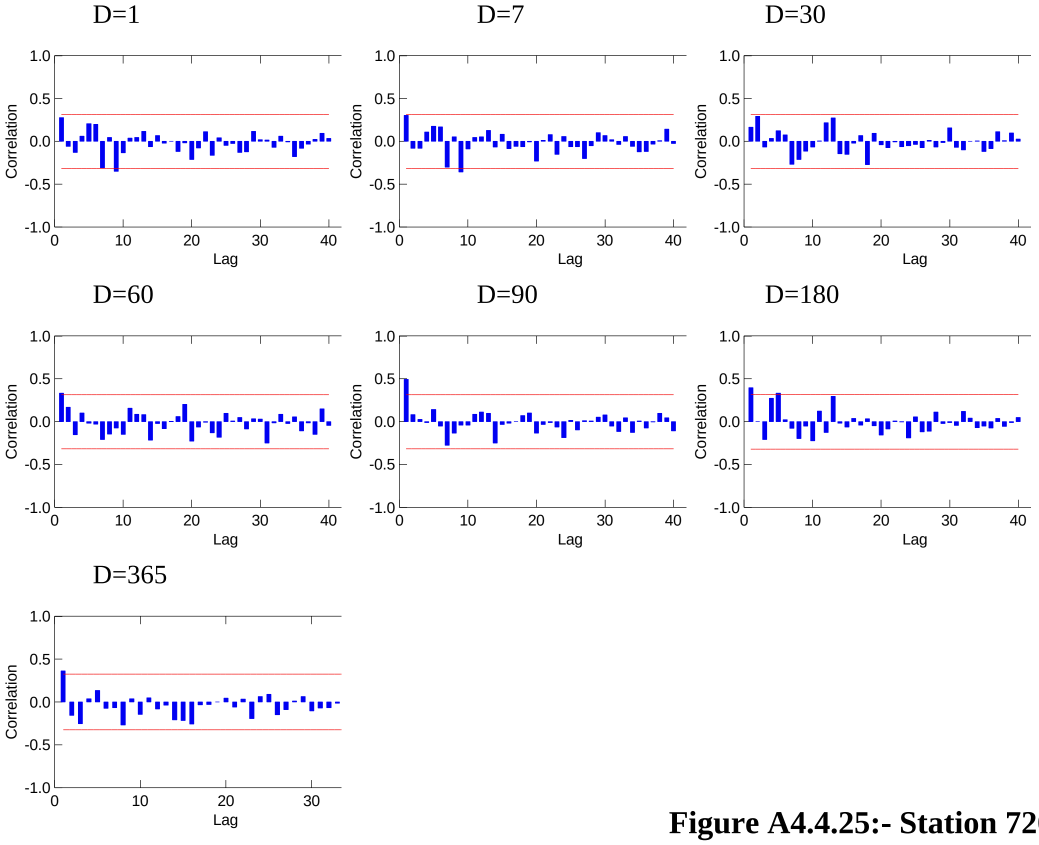


Figure A4.4.25:- Station 72004

APPENDIX 5:
ESTIMATION OF DISTRIBUTION PARAMETERS VIA THE L-MOMENTS METHOD

5.1 Generalised Extreme Value Distribution

<u>L-Moments:</u>

The four L-Moments, ?$_1$, ?$_{2,}$ t_3 and t_4 are given by the following relationships:

$$l_1 = x + a \frac{1 - \Gamma(1+k)}{k} \qquad \text{A5.1}$$

$$l_2 = a \, (1 - 2^{-k}) \frac{\Gamma(1+k)}{k} \qquad \text{A5.2}$$

$$t_3 = 2 \frac{(1 - 3^{-k})}{(1 - 2^{-k})} - 3 \qquad \text{A5.3}$$

$$t_4 = \frac{5\,(1 - 4^{-k}) - 10\,(1 - 3^{-k}) + 6\,(1 - 2^{-k})}{(1 - 2^{-k})} \qquad \text{A5.4}$$

where $\Gamma(.)$ denotes the gamma function:

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \qquad \text{A5.5}$$

In the particular case of a Gumbel distribution ($k = 0$), for example, the formulas are:

$$\begin{cases} l_1 = x + ag \\ l_2 = a \ \ln(2) \\ t_3 = 0.1699 = \dfrac{\ln(9/8)}{\ln(2)} \\ t_4 = 0.1504 = \dfrac{16 \ln(2) - 10 \ln(3)}{\ln(2)} \end{cases}$$

where g is Euler's constant and is equal to 0.5772

Parameters

To estimate k, equation A5.3, must be solved for k. No explicit solution is possible, but the following approximation, given by Hosking et al. (1985b), has accuracy better than $9*10^{-4}$ for $-0.5 \le t_3 \le 0.5$:

$$k \approx 7.8590\,c + 2.9554\,c^2\,, \quad \text{where } c = \frac{2}{3+t_3} - \frac{\log 2}{\log 3} \qquad \text{A5.6}$$

The other parameters, a and ? are then given by:

$$a = \frac{l_2 k}{(1-2^{-k})\,\Gamma(1+k)} \qquad \text{A5.7}$$

$$x = l_1 - a\,\frac{1-\Gamma(1+k)}{k} \qquad \text{A5.8}$$

5.2 Generalised Pareto Distribution

L-moments

The four L-Moments, ?$_1$, ?$_2$, t_3 and t_4 are given by the following relationships:

$$l_1 = x + \frac{a}{1+k} \qquad \text{A5.9}$$

$$l_2 = \frac{a}{(1+k)(2+k)} \qquad \text{A5.10}$$

$$t_3 = \frac{(1-k)}{(3+k)} \qquad \text{A5.11}$$

$$t_4 = \frac{(1-k)\,(2-k)}{(3+k)(4+k)} \qquad \text{A5.12}$$

Parameters

If the parameter ξ is known, then α and k are given by:

$$k = \frac{l_1 - x}{l_2} - 2 \qquad \text{A5.13}$$

$$a = (1+k)\,(l_1 - x) \qquad \text{A5.14}$$

If ξ is unknown, then the three parameters may be calculated from the L-moments as follows

$$k = \frac{1-3t_3}{1+t_3} \qquad \text{A5.15}$$

$$a = (1+k)\,(2+k)\;l_2 \qquad \text{A5.16}$$

$$x = l_1 - (2+k)\;l_2 \qquad \text{A5.17}$$

5.3 Pearson Type-III Distribution

L-Moments

Expressions for the distribution’s L-Moments in terms of its parameters are simpler when using the standard parameters; to represent these results, we therefore use the standard parametrisation, assuming $g > 0$. The corresponding results for $g < 0$ are obtained by changing the signs of l_1, t_3 and x whenever they occur in the following expressions.

L-Moments are defined for all values of a, $0 < a < \infty$.

$$l_1 = x + ab \qquad \text{A5.18}$$

$$l_2 = \frac{1}{\sqrt{p}}\, b\; \Gamma\left(a + \frac{1}{2}\right) / \;\Gamma(a) \qquad \text{A5.19}$$

$$t_3 = 6\,\mathrm{I}_{1/3}(a, 2a) - 3 \qquad \text{A5.20}$$

Here $I_x(p,q)$ denotes the incomplete beta function ratio

$$I_x(p,q) = \frac{\Gamma(p+q)}{\Gamma(p)\,\Gamma(q)} \int_0^x t^{p-1}(1-t)^{q-1}\,dt \qquad \text{A5.21}$$

There is no simple expression for t_4. Rational-function approximation can be used to express t_3 and t_4 approximately as functions of a. The following approximations are accurate to 10^{-6}. If $a \geq 1$,

$$t_3 \approx \frac{1}{\sqrt{a}} \frac{A_0 + A_1 a^{-1} + A_2 a^{-2} + A_3 a^{-3}}{1 + B_1 a^{-1} + B_2 a^{-2}} \qquad \text{A5.22}$$

$$t_4 \approx \frac{C_0 + C_1 a^{-1} + C_2 a^{-2} + C_3 a^{-3}}{1 + D_1 a^{-1} + D_2 a^{-2}} \qquad \text{A5.23}$$

if $a < 1$,

$$t_3 \approx \frac{1 + E_1 a + E_2 a^2 + E_3 a^3}{1 + F_1 a + F_2 a^2 + F_3 a^3} \qquad \text{A5.24}$$

$$t_4 \approx \frac{1 + G_1 a + G_2 a^2 + G_3 a^3}{1 + H_1 a + H_2 a^2 + H_3 a^3} \qquad \text{A5.25}$$

Coefficients of the approximations are given in the following table

Table A5.1: *Coefficients for parameter estimation for the Pearson Type-III distribution*

A0 = 3.2573501*10-1	C0 = 1.2260172*10-1
A1 = 1.6869150*10-1	C1 = 5.3730130*10-2
A2 = 7.8327243*10-2	C2 = 4.3384378*10-2
A3 = - 2.9120539*10-3	C3 = 1.1101277*10-2
B1 = 4.6697102*10-1	D1 = 1.8324466*10-1
B2 = 2.4255406*10-1	D2 = 2.0166036*10-1
E1 = 2.3807576	G1 = 2.1235833
E2 = 1.5931792	G2 = 4.1670213
E3 = 1.1618371*101	G3 = 3.1925299
F1 = 5.1533299	H1 = 9.0551443
F2 = 7.1425260	H2 = 2.6649995*101
F3 = 1.9745056	H3 = 2.6193668*101

Parameters

To estimate a, the third equation of the previous system must be solved for a, replacing t_3 by $|t_3|$ to enable a solution to be obtained when t_3 is negative. The following approximation has relative accuracy better than 5×10^{-5} for all values of a .

If $0<|t_3|<\frac{1}{3}$, let $z=3p\ t_3^{2}$ and use:

$$a \approx \frac{1+0.2906\,z}{z+0.1882\,z^2+0.0442\,z^3} \qquad \text{A5.26}$$

if $\frac{1}{3}\le|t_3|<1$, let $z=1-|t_3|$ and use

$$a \approx \frac{0.36067\,z-0.59567\,z^2+0.25361\,z^3}{1-2.78861\,z+2.56096\,z^2-0.77045\,z^3} \qquad \text{A5.27}$$

Given a , the parameters of our preferred parametrisation may be found from

$$g=2a^{-1/2}\mathrm{sign}(t_3), \quad s=l_2\sqrt{p}\,\sqrt{a}\,\frac{\Gamma(a)}{\Gamma\left(a+\frac{1}{2}\right)}, \quad m=l_1 \qquad \text{A5.28}$$

APPENDIX 6:

APPENDIX 6.1:
PARAMETER VALUES

Table 6.1.1: Parameter Values for the GEV distribution

Station	Parameter	Duration, D (days) 1	7	30	60	90	180	360
9001	a	0.434	0.497	0.675	0.822	1.080	1.440	1.920
	?	2.030	2.150	2.460	2.730	3.170	4.440	6.600
	k	-0.137	-0.114	-0.067	-0.181	-0.097	0.062	0.447
9002	a	0.798	0.884	1.110	1.480	1.980	2.770	4.050
	?	3.250	3.410	3.890	4.430	5.100	7.390	12.600
	k	-0.063	-0.045	-0.121	-0.164	0.160	-0.036	0.509
14001	a	0.237	0.244	0.272	0.276	0.301	0.463	0.915
	?	0.900	0.933	1.010	1.100	1.700	1.610	3.030
	k	0.342	0.233	0.111	0.072	0.101	-0.001	0.562
19002	a	0.029	0.035	0.046	0.066	0.090	0.138	0.159
	?	0.100	0.121	0.159	0.200	0.237	0.380	0.741
	k	-0.086	-0.053	-0.049	-0.129	-0.075	0.109	0.493
19004	a	0.095	0.101	0.114	0.141	0.175	0.256	0.285
	?	0.282	0.315	0.373	0.437	0.499	0.754	1.300
	k	0.246	0.209	-0.067	-0.085	-0.047	0.135	0.858
20001	a	0.152	0.150	0.180	0.227	0.268	0.452	0.737
	?	0.503	0.504	0.622	0.699	0.761	1.060	2.030
	k	0.665	0.025	-0.338	-0.513	-0.138	-0.103	0.585
20003	a	0.072	0.080	0.091	0.106	0.106	0.222	0.359
	?	0.250	0.270	0.297	0.329	0.329	0.511	0.998
	k	-0.004	0.015	-0.069	-0.113	-0.164	-0.088	0.507
20005	a	0.050	0.047	0.053	0.067	0.086	0.154	0.226
	?	0.160	0.173	0.194	0.218	0.244	0.350	0.697
	k	0.207	0.100	-0.042	-0.140	-0.120	-0.177	0.559
21006	a	1.440	1.670	2.160	3.040	0.274	0.558	0.843
	?	5.400	5.990	7.210	8.820	0.648	1.100	2.560
	k	0.154	0.186	-0.035	-0.072	-0.288	-0.066	0.652
21012	a	0.262	0.295	0.426	0.733	1.020	1.450	1.510
	?	0.758	0.821	1.060	1.419	1.860	3.410	7.110
	k	0.459	0.007	-0.109	-0.173	-0.060	0.132	0.740
21013	a	0.126	0.144	0.193	0.289	0.356	0.668	0.761
	?	0.430	0.463	0.539	0.666	0.740	1.370	2.890
	k	-0.101	-0.084	-0.195	-0.234	-0.165	0.254	0.675
21015	a	0.104	0.116	0.151	0.215	0.274	0.558	0.843
	?	0.385	0.408	0.470	0.559	0.648	1.100	2.560
	k	-0.222	-0.212	-0.280	-0.305	-0.288	-0.066	0.652
21017	a	0.051	0.061	0.102	0.207	0.239	0.281	0.368
	?	0.137	0.151	0.218	0.343	0.463	0.892	1.550
	k	0.108	0.113	0.036	-0.031	0.007	0.194	0.639
28031	a	0.201	0.212	0.262	0.356	0.448	0.638	0.681
	?	0.586	0.608	0.690	0.792	0.942	1.490	2.740

Station	Param eter	Duration, D (days)						
		1	7	30	60	90	180	360
	k	0.186	0.165	0.059	-0.018	0.107	0.340	0.433
34003	a	0.116	0.117	0.121	0.134	0.140	0.142	0.157
	?	0.541	0.565	0.609	0.650	0.678	0.750	0.904
	k	0.322	0.348	0.315	0.311	0.318	0.238	0.234
39016	a	0.934	0.950	1.030	1.118	1.193	1.560	2.380
	?	3.710	3.850	4.152	4.430	4.622	5.463	7.510
39016	k	0.375	0.361	0.362	0.358	0.339	0.314	0.554
39028	a	0.059	0.063	0.069	0.075	0.081	0.103	0.176
	?	0.288	0.299	0.313	0.326	0.337	0.400	0.571
	k	0.447	0.432	0.441	0.426	0.410	0.418	0.557
43005	a	0.321	0.329	0.345	0.365	0.389	0.532	0.909
	?	1.143	1.174	1.246	1.319	1.386	1.726	2.685
	k	0.602	0.589	0.527	0.440	0.441	0.463	0.579
43006	a	0.167	0.179	0.184	0.210	0.224	0.333	0.682
	?	0.868	0.913	0.976	1.052	1.113	1.369	2.232
	k	0.208	0.222	0.196	0.204	0.171	0.290	0.490
48010	a	0.053	0.056	0.067	0.075	0.087	0.133	0.167
	?	0.194	0.203	0.228	0.248	0.271	0.382	0.789
	k	0.082	0.050	0.056	-0.064	-0.060	0.063	0.211
51001	a	0.059	0.063	0.069	0.077	0.080	0.122	0.257
	?	0.186	0.195	0.213	0.233	0.253	0.356	0.730
	k	0.185	0.171	0.101	0.109	0.132	0.089	0.392
55016	a	0.164	0.177	0.274	0.477	0.641	1.230	1.510
	?	0.221	0.257	0.400	0.607	0.861	2.040	6.140
	k	-0.119	-0.151	-0.214	-0.234	-0.142	0.160	0.388
55026	a	0.153	0.183	0.309	0.540	0.732	0.947	0.980
	?	0.319	0.372	0.617	1.040	1.530	2.650	5.140
	k	-0.027	-0.121	-0.040	-0.045	0.080	0.204	0.213
60002	a	0.362	0.418	0.668	1.018	1.483	2.099	1.939
	?	0.714	0.796	1.143	1.685	2.489	4.486	8.682
	k	0.140	0.107	0.039	0.026	0.135	0.394	0.367
72004	a	0.847	0.971	2.100	3.320	4.260	6.160	6.100
	?	2.360	2.600	3.910	5.860	8.810	16.100	28.200
	k	0.040	-0.020	-0.082	-0.058	0.011	0.328	0.511

Table 6.1.2: Parameter Values for the Generalised Logistic distribution

Station	Parameter	Duration, D (days)						
		1	7	30	60	90	180	360
9001	a	0.310	0.350	0.463	0.601	0.755	0.920	1.000
	?	2.210	2.340	2.720	3.070	3.600	4.990	7.230
	k	-0.261	-0.245	-0.213	0.292	-0.234	-0.131	0.087
9002	a	0.546	0.599	0.782	1.070	1.430	1.870	2.050
	?	3.570	3.750	4.330	5.030	5.890	8.470	13.900
	k	-0.211	-0.199	-0.250	-0.279	-0.277	-0.193	0.118
14001	a	0.130	0.142	0.169	0.175	0.188	0.306	0.451
	?	0.982	1.020	1.120	1.210	1.320	1.790	3.320

Station	Parameter	Duration, D (days)						
		1	7	30	60	90	180	360
	k	0.032	-0.029	-0.100	-0.125	-0.107	-0.171	0.145
19002	a	0.020	0.024	0.031	0.047	0.062	0.086	0.081
	?	0.112	0.135	0.177	0.226	0.272	0.432	0.793
	k	-0.226	-0.205	-0.202	-0.255	-0.219	-0.102	0.110
19004	a	0.055	0.059	0.078	0.098	0.119	0.157	0.124
	?	0.316	0.352	0.418	0.490	0.568	0.849	3.750
	k	-0.214	-0.422	-0.214	-0.226	-0.200	-0.863	0.277
20001	a	0.097	0.098	0.122	0.155	0.191	0.316	0.370
	?	0.560	0.612	0.692	0.789	0.868	1.240	2.260
	k	-0.127	-0.154	-0.192	-0.203	-0.262	-0.238	0.156
20003	a	0.048	0.053	0.063	0.075	0.093	0.154	0.182
	?	0.278	0.301	0.332	0.371	0.415	0.599	1.110
	k	-0.172	-0.160	-0.215	-0.245	-0.278	0.228	0.117
20005	a	0.029	0.029	0.036	0.048	0.061	0.103	0.112
	?	0.178	0.191	0.214	0.244	0.279	0.410	0.769
	k	0.043	-0.107	-0.197	-0.263	-0.249	-0.181	0.143
21006	a	0.871	0.997	1.450	2.090	0.213	0.383	0.399
	?	5.930	6.600	8.050	10.000	0.762	1.320	2.310
	k	-0.075	-0.056	-0.192	-0.217	-0.369	-0.213	0.187
21012	a	0.169	0.194	0.299	0.533	0.699	0.891	0.690
	?	0.858	0.936	1.290	1.720	2.260	3.943	7.549
	k	-0.141	-0.165	-0.243	-0.286	-0.209	-0.088	0.226
21013	a	0.088	0.100	0.142	0.218	0.258	0.435	0.357
	?	0.480	0.519	0.617	0.785	0.938	1.630	3.120
	k	-0.236	-0.225	-0.302	0.329	-0.280	-0.154	0.197
21015	a	0.078	0.087	0.117	0.163	0.213	0.383	0.399
	?	0.432	0.456	0.533	0.648	0.762	1.320	2.310
	k	-0.320	-0.314	-0.363	-0.382	-0.369	-0.213	0.187
21017	a	0.032	0.038	0.066	0.139	0.157	0.167	0.147
	?	0.156	0.174	0.257	0.424	0.555	0.994	1.650
	k	-0.103	-0.100	-0.147	-0.190	-0.165	-0.051	0.181
28031	a	0.120	0.128	0.167	0.238	0.279	0.351	0.358
	?	0.659	0.686	0.790	0.930	1.110	1.710	2.570
	k	-0.056	-0.068	0.133	-0.182	-0.103	0.031	0.080
34003	a	0.064	0.064	0.067	0.075	0.078	0.082	0.091
	?	0.582	0.606	0.651	0.697	0.726	0.801	0.960
	k	0.021	0.035	0.017	0.015	0.019	-0.026	-0.028
39016	a	0.505	0.517	0.561	0.610	0.656	0.870	1.180
	?	4.020	4.180	4.500	4.810	5.030	6.007	8.260
	k	0.050	0.042	0.043	0.040	0.031	0.017	0.140
39028	a	0.031	0.033	0.036	0.040	0.043	0.055	0.087
	?	0.307	0.320	0.336	0.351	0.365	0.435	0.627
	k	0.087	0.079	0.084	0.077	0.068	0.072	0.142
43005	a	0.156	0.160	0.173	0.191	0.203	0.275	0.445
	?	1.243	1.277	1.357	1.441	1.515	1.901	2.970
	k	0.164	0.157	0.127	0.084	0.084	0.095	0.153
43006	a	0.098	0.105	0.109	0.124	0.135	0.188	0.348
	?	0.929	0.978	1.043	1.128	1.195	1.486	2.454
	k	-0.118	-0.138	-0.135	-0.211	-0.209	-0.130	-0.041
48010	a	0.034	0.036	0.043	0.052	0.060	0.085	0.099
	?	0.214	0.225	0.253	0.278	0.305	0.432	0.850

Station	Parameter	Duration, D (days)						
		1	**7**	**30**	**60**	**90**	**180**	**360**
	k	-0.118	-0.138	-0.135	-0.211	-0.209	-0.130	-0.041
51001	a	0.035	0.038	0.043	0.048	0.049	0.077	0.138
	?	0.207	0.218	0.239	0.262	0.282	0.402	0.817
	k	-0.056	-0.065	-0.106	-0.102	-0.088	-0.114	0.058
55016	a	0.110	0.127	0.204	0.359	0.458	0.747	0.811
	?	0.286	0.328	0.513	0.802	1.120	2.500	6.650
	k	-0.249	-0.270	-0.315	-0.329	-0.264	-0.071	0.057
55026	a	0.103	0.121	0.208	0.366	0.463	0.559	0.576
	?	0.378	0.443	0.737	1.250	1.800	2.990	5.490
	k	-0.187	-0.178	-0.196	-0.199	-0.119	-0.045	-0.040
60002	a	0.222	0.261	0.432	0.663	0.909	1.124	1.052
	?	0.848	0.952	1.398	2.076	3.038	5.196	9.344
	k	-0.830	-0.103	-0.147	-0.154	-0.086	0.060	0.046
72004	a	0.547	0.649	1.460	2.270	2.800	3.410	3.080
	?	2.680	2.980	4.740	7.160	10.500	18.300	30.200
	k	-0.145	-0.183	-0.227	-0.208	-0.163	0.024	0.120

Table 6.1.3: Parameter Values for the Generalised Pareto distribution

Station	Parameter	Duration, D (days)						
		1	**7**	**30**	**60**	**90**	**180**	**360**
9001	a	0.884	1.040	1.490	1.600	2.300	3.690	8.180
	?	1.770	2.000	2.290	2.930	3.920	6.320	18.800
	k	0.871	0.775	0.739	0.716	0.908	1.100	4.590
9002	a	0.063	0.078	0.104	0.135	0.196	0.375	0.721
	?	0.306	0.310	0.251	0.305	0.395	0.718	2.260
	k	0.391	0.368	0.413	0.510	0.545	0.955	3.930
14001	a	0.171	0.194	0.200	0.222	0.254	0.477	0.166
	?	0.153	0.125	0.120	0.135	0.179	0.360	1.130
	k	4.120	5.000	4.930	6.660	0.477	1.230	4.840
19002	a	0.659	0.708	0.895	1.430	2.270	4.050	9.930
	?	0.266	0.311	0.369	0.532	0.703	1.640	4.520
	k	0.194	0.219	0.265	0.368	0.477	1.230	4.840
19004	a	0.138	0.167	0.254	0.476	0.575	0.850	1.740
	?	0.600	0.618	0.669	0.830	1.210	2.340	2.840
	k	0.413	0.434	0.428	0.472	0.497	0.453	0.498
20001	a	3.590	3.580	3.890	4.200	4.370	5.520	11.800
	?	0.251	0.262	0.289	0.309	0.326	0.422	0.880
	k	1.710	1.718	1.644	1.537	1.639	2.312	4.681
20003	a	0.513	0.562	0.558	0.642	0.656	1.142	3.082
	?	0.140	0.143	0.171	0.166	0.194	0.340	0.517
	k	0.177	0.177	0.185	0.210	0.224	0.324	1.012
20005	a	0.340	0.354	0.515	0.877	1.300	3.570	5.920
	?	0.353	0.428	0.701	1.220	1.920	2.900	3.030
	k	1.023	1.133	1.667	2.493	4.159	8.285	7.380
21006	a	2.110	2.260	4.520	7.390	10.300	22.200	28.400

Station	Parameter	Duration, D (days)						
		1	7	30	60	90	180	360
	?	1.600	1.640	1.750	1.930	2.060	2.790	3.660
	k	2.410	2.460	2.770	2.570	3.140	4.410	6.060
21012	a	0.567	0.616	0.692	0.785	0.848	1.100	1.500
	?	0.070	0.084	0.110	0.133	0.143	0.216	0.489
	k	0.158	0.187	0.253	0.290	0.312	0.444	0.674
21013	a	0.328	0.386	0.428	0.458	0.491	0.593	0.727
	?	0.171	0.181	0.201	0.220	0.237	0.281	0.421
	k	0.097	0.118	0.137	0.151	0.157	0.183	0.318
21015	a	3.640	3.890	4.890	5.640	3.970	0.509	1.030
	?	0.460	0.496	0.622	0.699	0.781	1.660	4.140
	k	0.301	0.313	0.352	0.392	0.443	0.626	1.490
21017	a	0.290	0.297	0.331	0.364	0.397	0.509	1.030
	?	0.076	0.079	0.103	0.120	0.198	0.537	1.000
	k	0.334	0.346	0.390	0.405	0.412	0.597	1.713
28031	a	0.381	0.400	0.442	0.466	0.485	0.566	0.701
	?	2.360	2.490	2.680	2.840	2.940	3.310	3.530
	k	0.198	0.204	0.209	0.213	0.217	0.246	0.276
34003	a	0.585	0.609	0.684	0.764	0.794	0.902	1.134
	?	0.656	0.682	0.744	0.786	0.836	0.918	1.153
	k	0.132	0.139	0.151	0.169	0.179	0.230	0.576
39016	a	0.112	0.112	0.132	0.142	0.156	0.213	0.354
	?	0.547	0.081	0.138	0.155	0.219	0.525	3.930
	k	0.153	0.173	0.286	0.461	0.675	1.440	3.890
39028	a	0.275	0.300	0.388	0.547	0.698	1.408	5.896
	?	1.400	1.550	1.730	2.340	4.100	7.550	18.400
	k	0.172	0.212	0.296	0.097	0.242	0.537	1.380
43005	a	0.304	0.336	0.199	0.126	0.131	0.352	1.540
	?	1.130	0.888	0.634	0.557	0.615	0.417	1.690
	k	0.262	0.320	0.328	0.186	0.281	0.631	1.496
43006	a	0.916	0.838	0.295	0.263	0.332	0.682	2.530
	?	0.546	0.437	0.356	0.324	0.170	0.231	0.174
	k	0.412	0.447	0.292	0.214	0.130	0.257	0.153
48010	a	0.834	0.613	0.341	0.167	0.201	0.386	1.670
	?	0.722	0.789	0.355	0.287	-0.079	0.297	1.920
	k	0.506	0.437	0.220	0.110	0.309	0.677	2.170
51001	a	0.236	0.265	0.073	0.009	0.124	0.467	1.980
	?	0.029	0.044	-0.066	-0.106	-0.079	0.297	1.920
	k	0.628	0.638	0.487	0.362	0.430	0.804	1.880
55016	a	0.789	0.744	0.532	0.385	0.626	1.130	2.350
	?	1.086	1.145	1.069	1.062	1.077	0.899	0.890
	k	1.210	1.180	1.180	1.170	1.130	1.070	1.650
55026	a	1.380	1.340	1.370	1.330	1.290	1.310	1.660
	?	1.782	1.747	1.584	1.365	1.368	1.421	1.719
	k	0.835	0.866	0.810	0.827	0.757	1.015	1.489
60002	a	0.577	0.514	0.526	0.302	0.309	0.541	0.841
	?	0.787	0.787	0.615	0.631	0.677	0.590	1.248
	k	0.203	0.149	0.042	0.009	0.164	0.734	1.240
72004	a	0.365	0.396	0.345	0.336	0.573	0.827	0.845
	?	0.694	0.627	0.495	0.466	0.684	1.253	1.191
	k	0.495	0.382	0.259	0.312	0.440	1.100	1.540

Table 6.1.4: Parameter Values for the Pearson Type III distribution

Station	Parameter	Duration, D (days)						
		1	7	30	60	90	180	360
9001	a	0.66	0.73	0.93	1.35	1.56	1.71	1.81
	?	2.35	1.49	2.90	3.39	3.91	5.19	1.81
	k	1.57	1.48	1.29	1.75	1.41	7.97	-0.54
9002	a	1.10	1.19	1.65	2.37	3.14	3.68	3.78
	?	3.77	3.96	4.68	4.68	5.57	0.66	9.09
	k	1.27	1.20	1.51	0.17	0.17	1.17	-0.72
14001	a	0.23	0.25	0.31	0.33	0.34	0.59	0.85
	?	0.99	1.03	1.14	1.24	1.35	1.88	3.21
	k	-0.20	0.18	0.62	0.76	0.65	1.04	-0.88
19002	a	0.04	0.05	0.06	0.10	1.56	1.71	1.81
	?	0.12	0.14	0.19	0.25	3.91	5.19	7.09
	k	1.36	1.24	1.22	1.54	1.41	0.18	-0.54
19004	a	0.10	0.11	0.16	0.20	0.24	0.29	0.27
	?	0.32	0.36	0.45	0.53	0.61	0.87	1.31
	k	0.13	0.26	1.29	0.14	1.21	0.53	-1.66
20001	a	0.18	0.18	0.24	0.31	0.41	0.66	0.70
	?	0.58	0.64	0.73	0.84	0.96	1.37	2.17
	k	0.78	0.94	1.16	1.23	1.58	0.14	-0.95
20003	a	0.09	0.10	0.13	0.16	0.20	0.32	0.34
	?	0.29	0.32	0.36	0.40	0.46	0.66	1.08
	k	1.04	0.98	1.30	1.47	1.67	0.14	-0.72
20005	a	0.05	0.05	0.07	0.10	0.13	0.20	0.21
	?	0.18	0.20	0.23	0.27	0.31	0.44	0.74
	k	0.27	0.66	1.19	1.58	1.50	1.10	-0.87
21006	a	1.57	1.78	2.85	4.23	0.55	0.77	0.78
	?	6.03	6.69	8.53	10.80	0.91	1.46	2.69
	k	0.46	0.34	1.16	1.31	2.22	1.29	-1.13
21012	a	0.32	0.37	0.63	1.19	1.40	1.62	1.41
	?	0.90	0.99	1.36	1.99	2.52	4.07	7.28
	k	0.86	1.00	1.46	1.72	1.26	0.54	-1.37
21013	a	0.18	0.20	0.33	0.52	0.57	0.82	0.71
	?	0.52	0.56	0.70	0.92	1.07	1.74	3.00
	k	1.42	1.36	1.81	1.98	1.68	0.34	-1.19
21015	a	0.18	0.20	0.30	0.45	0.55	0.77	0.78
	?	0.48	0.51	6.15	0.58	0.91	1.46	2.69
	k	1.92	1.88	2.18	2.30	2.22	1.29	-1.13
21017	a	0.06	0.07	0.13	0.27	0.30	0.30	0.29
	?	0.16	0.18	0.27	0.47	0.60	1.01	1.60
	k	0.63	0.61	0.90	1.15	1.00	0.32	-0.11
28031	a	0.21	0.23	0.31	0.46	0.51	0.62	0.65
	?	0.67	0.70	0.83	1.00	1.16	1.70	2.92
	k	0.34	0.42	0.81	1.10	0.63	-0.19	-0.49
34003	a	0.11	0.11	0.12	0.13	0.14	0.15	0.16
	?	0.58	0.60	0.65	0.70	0.72	0.80	0.96
	k	-0.13	-0.21	-0.10	-0.09	-0.12	0.16	0.17
39016	a	0.90	0.92	1.00	1.08	1.17	1.54	2.21
	?	3.98	4.14	4.46	4.77	5.00	5.98	7.98

Station	Parameter	Duration, D (days)						
		1	7	30	60	90	180	360
	k	-0.31	-0.26	-0.26	-0.25	-0.19	-0.10	-85.60
39028	a	0.06	0.06	0.06	0.07	0.08	0.10	0.16
	?	0.30	0.32	0.33	0.35	0.36	0.43	0.61
	k	-0.53	-0.49	-0.52	-0.47	-0.42	-0.44	-0.86
43005	a	0.30	0.30	0.32	0.35	0.37	0.50	0.84
	?	1.20	1.24	1.32	1.41	1.49	1.86	2.86
	k	-0.99	-0.96	-0.78	-0.51	-0.52	-0.58	-0.93
43006	a	0.18	0.19	0.19	0.22	0.24	0.33	0.64
	?	0.94	0.98	1.05	1.14	1.21	1.49	2.39
	k	0.26	0.21	0.31	0.28	0.40	-0.02	-0.67
48010	a	0.06	0.07	0.08	0.10	0.12	0.16	0.18
	?	0.22	0.23	0.26	0.30	0.33	0.45	0.86
	k	0.72	0.84	0.82	1.28	1.26	0.79	0.25
51001	a	0.06	0.07	0.08	0.09	0.09	0.14	0.25
	?	0.21	0.22	0.25	0.27	0.29	0.42	0.80
	k	0.35	0.40	0.65	0.62	0.54	0.70	-0.36
55016	a	0.24	0.28	0.48	0.86	0.99	1.34	1.45
	?	0.34	0.39	0.63	1.02	1.32	2.58	6.58
	k	1.50	1.63	1.89	1.96	1.59	0.44	-0.35
55026	a	0.20	0.24	0.41	0.73	0.85	1.00	1.03
	?	0.41	0.48	0.81	1.37	1.89	3.03	5.53
	k	1.13	1.08	1.18	1.20	0.73	0.28	0.25
60002	a	0.40	0.48	0.81	1.26	1.65	2.01	1.88
	?	0.88	1.00	1.50	2.25	3.17	5.09	9.27
	k	0.52	0.63	0.88	0.94	0.53	-0.37	-0.28
72004	a	1.03	1.26	2.99	4.53	5.34	6.05	5.69
	?	2.81	3.18	5.32	7.97	11.20	18.10	29.60
	k	0.88	1.11	1.37	1.26	0.99	-0.15	-0.73

APPENDIX 6.2:
GOODNESS OF FIT TEST RESULTS

Table 6.2.1: Chi-Square Test Results for Selected Stations

Station	Distn.	Chi-Square Error (?^{2})								
		D= 1	D = 7	D = 30	D = 60	D = 90	D = 180	D = 365	Average	Average (D= 1 to 90)
9001	GEV	1.74	1.82	3.84	4.36	2.41	3.13	5.12	3.20	2.83
	GPA	2.23	2.35	4.93	4.70	3.08	6.36	3.18	3.83	3.46
	GL	1.96	2.15	3.67	6.91	5.77	2.83	5.19	4.07	4.09
	PE	1.71	1.78	3.44	5.85	3.05	2.71	6.47	3.57	3.17
9002	GEV	1.14	1.51	4.20	5.87	6.59	5.60	5.90	4.40	3.86
	GPA	1.55	2.17	5.25	6.27	7.71	9.51	12.76	6.46	4.59
	GL	1.49	1.49	3.60	7.95	8.25	2.66	8.16	4.80	4.55
	PE	1.09	1.32	3.61	6.83	7.01	3.91	7.08	4.41	3.97
21006	GEV	0.78	0.96	3.11	1.89	2.60	4.87	4.56	2.68	1.87
	GPA	0.97	1.40	2.52	1.79	2.43	8.88	9.91	3.99	1.82
	GL	1.68	1.63	6.79	5.58	7.17	4.04	2.58	4.21	4.57
	PE	0.76	0.96	3.82	2.76	2.91	5.42	5.20	3.12	2.24
34003	GEV	2.29	2.47	2.13	2.24	2.46	2.73	2.47	2.40	2.32
	GPA	3.86	4.08	3.43	4.43	4.73	4.51	5.92	4.42	4.11
	GL	2.75	2.90	3.23	2.16	2.48	3.07	1.90	2.64	2.70
	PE	2.53	2.73	2.33	2.66	2.90	2.97	3.84	2.85	2.63
39016	GEV	12.80	12.80	8.70	6.66	4.98	2.99	13.80	8.96	9.19
	GPA	7.35	7.85	4.70	3.49	2.70	2.99	2.64	4.53	5.22
	GL	29.30	27.90	22.10	18.90	15.60	10.40	6.54	18.68	22.76
	PE	7.11	11.30	7.46	5.64	4.22	2.71	16.40	7.83	7.15
43005	GEV	13.30	13.50	4.57	4.42	3.48	1.26	6.51	6.72	7.85
	GPA	6.40	6.49	2.14	2.03	1.95	3.22	16.00	5.46	3.80
	GL	37.80	37.60	15.80	14.80	12.50	6.19	8.93	19.09	23.70
	PE	11.20	11.30	3.66	3.51	2.81	1.49	8.08	6.01	6.50
43006	GEV	1.60	1.19	1.28	1.67	1.74	0.74	5.01	1.89	1.50
	GPA	1.08	0.68	0.79	0.88	1.14	1.56	9.08	2.17	0.92
	GL	3.98	3.80	3.75	5.01	5.11	2.79	9.87	4.90	4.33
	PE	1.50	1.08	1.20	1.51	1.69	0.85	5.53	1.91	1.40
48010	GEV	2.12	1.80	4.90	5.02	6.51	4.03	2.23	3.80	4.07
	GPA	1.95	1.86	4.57	4.68	6.33	6.15	4.52	4.29	3.88
	GL	4.12	3.29	7.68	8.07	9.71	3.67	0.93	5.35	6.57
	PE	2.18	1.85	5.01	5.70	7.13	3.76	2.58	4.03	4.37
51001	GEV	0.76	0.86	13.60	1.56	2.21	5.64	5.46	4.30	3.80
	GPA	0.94	1.19	1.54	1.55	1.68	5.89	7.25	2.86	1.38
	GL	2.37	2.30	3.38	4.36	5.99	9.05	13.30	5.82	3.68
	PE	0.77	0.89	1.43	1.63	2.22	5.65	5.41	2.57	1.39
55026	GEV	0.79	0.71	1.95	4.07	1.82	4.50	2.63	2.35	1.87
	GPA	0.78	1.12	2.39	4.42	5.30	6.20	3.78	3.43	2.80
	GL	3.99	3.55	7.90	13.90	11.10	17.10	6.26	9.11	8.09
	PE	1.24	0.92	2.81	5.38	1.75	4.40	2.56	2.72	2.42
72004	GEV	0.96	1.36	6.52	5.02	5.28	4.36	4.08	3.94	3.83
	GPA	1.48	2.12	9.23	9.12	9.52	8.39	2.94	6.11	6.29
	GL	1.04	0.97	3.10	1.26	2.67	9.50	13.90	4.63	1.81
	PE	0.90	1.13	4.51	2.70	4.04	4.93	3.46	3.10	2.66

Table 6.2.2: Chi-Square Test Results for Selected Stations

Station	Distribution with lowest Chi-Square Error (?^{2})						
	D= 1	D = 7	D = 30	D = 60	D = 90	D = 180	D = 365
9001	gev/pe3	pe3	pe3	pe3	gev	pe3	gev
9002	gev/pe3	gpa	gpa	pe3	gev	gev	gpa
14001	gpa	gev	pe3	pe3	pe3	glo	gev
19002	gev	gev	gev	pe3	pe3	glo	gev
19004	glo	gev	glo	gev	pe3	gpa	gpa
20001	pe3	pe3	pe3	pe3	gpa	gpa	gpa
20003	gev	gev	gev	pe3	gpa	gpa	gev
20005	gev	gpa	pe3	gev	pe3	gpa	gev
21006	glo	glo	glo	glo	glo	gpa	glo
21012	pe3	pe3	gev	gpa	gpa	gev	gpa
21013	gev	gev	gev	gpa	pe3	gpa	gev
21015	gev	pe3	gev	pe3	pe3	gpa	gpa
21017	pe3	pe3	gpa	gpa	pe3	glo	gpa
28031	pe3	pe3	pe3/gev	gev	gpa	gpa	gev
34003	gev	gev	gev	gpa	gev	gev	gpa
39016	glo	glo	glo	glo	glo	pe3	gpa
39028	gev	gev	gev	gev	gev	gev	pe3
55016	pe3	pe3	pe3	gpa	gpa	gpa	pe3
60002	pe3	pe3	gpa	gev	gev	gev	gpa
72004	pe3	gpa	gpa	gpa	gpa	gev	pe3

Table 6.2.3: Root Mean Square Errors

Station	Distn.	Root Mean Square Error (RMSE)								
		D= 1	D = 7	D = 30	D = 60	D = 90	D = 180	D = 365	Average	Average (D= 1 to 90)
9001	GEV	1.120	1.130	1.960	2.560	2.120	1.960	2.700	1.936	1.778
	GPA	1.080	1.210	1.980	3.390	3.380	2.350	3.190	2.369	2.208
	GL	1.310	1.320	2.190	2.550	2.140	2.780	3.990	2.326	1.902
	PE	1.030	1.080	1.880	3.100	2.530	1.850	2.980	2.064	1.924
9002	GEV	0.822	0.986	1.949	2.763	3.216	3.057	3.103	2.271	1.947
	GPA	0.975	0.971	1.730	3.434	3.728	2.168	4.264	2.467	2.168
	GL	0.966	1.197	2.194	2.733	3.399	3.953	4.095	2.648	2.098
	PE	0.804	0.911	1.768	3.141	3.425	2.594	3.228	2.267	2.010
14001	GEV	0.033	0.039	0.044	0.046	0.038	0.094	0.117	0.059	0.040
	GPA	0.034	0.047	0.055	0.055	0.045	0.143	0.150	0.075	0.047
	GL	0.045	0.048	0.052	0.055	0.053	0.090	0.155	0.071	0.051
	PE	0.036	0.046	0.044	0.046	0.038	0.102	0.121	0.062	0.042
19002	GEV	0.096	0.097	0.165	0.221	0.183	0.169	0.233	0.166	0.153
	GPA	0.093	0.105	0.171	0.293	0.292	0.203	0.276	0.205	0.191
	GL	0.113	0.114	0.189	0.220	0.184	0.240	0.344	0.201	0.164
	PE	0.089	0.093	0.162	0.268	0.218	0.155	0.257	0.177	0.166
19004	GEV	0.020	0.021	0.032	0.031	0.028	0.036	0.481	0.093	0.026
	GPA	0.031	0.028	0.046	0.045	0.042	0.041	0.026	0.037	0.038
	GL	0.016	0.021	0.029	0.030	0.029	0.048	0.074	0.035	0.025

Station	Distn.	Root Mean Square Error (RMSE) D= 1	D = 7	D = 30	D = 60	D = 90	D = 180	D = 365	Average	Average (D= 1 to 90)
	PE	0.019	0.020	0.037	0.035	0.029	0.036	0.051	0.032	0.028
20001	GEV	0.039	0.024	0.030	0.038	0.080	0.148	0.093	0.064	0.042
	GPA	0.048	0.034	0.044	0.034	0.053	0.098	0.089	0.057	0.043
	GL	0.029	0.026	0.032	0.048	0.095	0.172	0.139	0.077	0.046
	PE	0.029	0.024	0.032	0.032	0.063	0.124	0.067	0.053	0.036
20003	GEV	0.012	0.013	0.013	0.016	0.030	0.071	0.036	0.027	0.017
	GPA	0.015	0.017	0.016	0.019	0.023	0.050	0.053	0.028	0.018
	GL	0.014	0.015	0.016	0.020	0.036	0.091	0.057	0.036	0.020
	PE	0.012	0.013	0.013	0.015	0.023	0.061	0.041	0.025	0.015
20005	GEV	0.006	0.008	0.011	0.017	0.016	0.048	0.031	0.019	0.011
	GPA	0.009	0.006	0.013	0.020	0.015	0.031	0.016	0.016	0.012
	GL	0.008	0.011	0.012	0.019	0.020	0.057	0.038	0.023	0.014
	PE	0.006	0.008	0.010	0.018	0.014	0.044	0.032	0.019	0.011
21006	GEV	0.628	0.678	1.780	1.560	1.970	2.340	2.850	1.687	1.323
	GPA	1.000	1.020	2.500	2.540	3.080	2.490	2.230	2.123	2.028
	GL	0.602	0.716	1.160	1.440	1.760	3.030	4.140	1.835	1.136
	PE	0.615	0.664	1.950	1.860	2.100	2.440	3.030	1.808	1.438
21012	GEV	0.040	0.055	0.108	0.184	0.173	0.185	0.219	0.138	0.112
	GPA	0.039	0.053	0.130	0.167	0.143	0.174	0.196	0.129	0.106
	GL	0.053	0.068	0.115	0.211	0.132	0.266	0.329	0.168	0.116
	PE	0.038	0.052	0.113	0.162	0.144	0.188	0.236	0.133	0.102
21013	GEV	0.024	0.023	0.066	0.099	0.097	0.155	0.100	0.081	0.062
	GPA	0.031	0.029	0.081	0.073	0.063	0.083	0.140	0.072	0.055
	GL	0.021	0.028	0.066	0.112	0.116	0.197	0.134	0.096	0.068
	PE	0.026	0.023	0.077	0.074	0.070	0.143	0.103	0.074	0.054
21015	GEV	0.043	0.045	0.052	0.098	0.099	0.130	0.103	0.081	0.067
	GPA	0.041	0.041	0.052	0.070	0.066	0.085	0.096	0.064	0.054
	GL	0.046	0.049	0.056	0.098	0.112	0.160	0.167	0.098	0.072
	PE	0.041	0.041	0.054	0.066	0.059	0.109	0.116	0.069	0.052
21017	GEV	0.006	0.008	0.017	0.033	0.033	0.039	0.052	0.027	0.019
	GPA	0.009	0.011	0.014	0.041	0.050	0.071	0.032	0.033	0.025
	GL	0.008	0.010	0.022	0.038	0.038	0.036	0.075	0.032	0.023
	PE	0.006	0.008	0.015		0.034	0.038	0.058	0.026	0.016
28031	GEV	0.042	0.042	0.066	0.080	0.073	0.138	0.081	0.075	0.061
	GPA	0.058	0.060	0.088	0.095	0.064	0.139	0.112	0.088	0.073
	GL	0.040	0.091	0.065	0.086	0.097	0.164	0.106	0.093	0.076
	PE	0.041	0.042	0.067	0.080	0.073	0.143	0.081	0.075	0.061
34003	GEV	1.630	1.730	1.740	1.730	1.850	2.130	2.600	1.916	1.736
	GPA	2.100	2.180	2.350	1.960	2.120	2.440	2.150	2.186	2.142
	GL	1.990	2.090	2.090	2.410	2.540	2.720	3.390	2.461	2.224
	PE	1.670	1.770	1.790	1.880	2.000	2.230	2.730	2.010	1.822
39016	GEV	2.120	2.230	2.020	1.930	1.850	2.270	4.490	2.416	2.030
	GPA	3.360	3.380	3.390	3.380	3.320	3.740	3.900	3.496	3.366
	GL	1.700	1.900	1.580	1.580	1.600	2.220	5.980	2.366	1.672
	PE	2.000	2.130	1.880	1.810	1.750	2.160	4.810	2.363	1.914
39028	GEV	0.008	0.008	0.009	0.013	0.016	0.018	0.034	0.015	0.011
	GPA	0.012	0.013	0.014	0.018	0.021	0.022	0.018	0.017	0.016
	GL	0.009	0.009	0.014	0.013	0.016	0.020	0.046	0.018	0.012
	PE	0.008	0.008	0.009	0.012	0.016	0.038	0.037	0.018	0.011

Station	Distn.	Root Mean Square Error (RMSE)								
		D= 1	D = 7	D = 30	D = 60	D = 90	D = 180	D = 365	Average	Average (D= 1 to 90)
43005	GEV	1.880	2.020	1.620	1.780	1.760	1.320	3.260	1.949	1.812
	GPA	3.000	3.190	2.800	3.120	3.060	2.510	4.560	3.177	3.034
	GL	1.520	1.590	1.290	1.330	1.490	1.890	4.490	1.943	1.444
	PE	1.760	1.890	1.480	1.610	1.630	1.380	3.410	1.880	1.674
43006	GEV	1.140	1.030	1.160	1.520	1.420	1.080	3.360	1.530	1.254
	GPA	1.790	1.840	1.960	2.490	2.390	2.040	5.130	2.520	2.094
	GL	0.968	0.810	0.927	1.140	1.220	1.500	3.840	1.486	1.013
	PE	1.110	0.983	1.120	1.450	1.400	1.150	3.330	1.506	1.213
48010	GEV	1.370	1.310	2.300	2.640	3.200	2.530	2.520	2.267	2.164
	GPA	1.900	1.760	2.920	3.360	3.800	2.540	1.730	2.573	2.748
	GL	1.290	1.310	2.180	2.530	3.180	3.070	3.580	2.449	2.098
	PE	1.390	1.340	2.340	2.840	3.340	2.480	2.720	2.350	2.250
51001	GEV	0.598	0.708	0.966	1.110	1.540	2.930	3.700	1.650	0.984
	GPA	1.030	1.040	1.320	1.660	2.360	3.860	5.930	2.457	1.482
	GL	0.723	0.895	1.140	1.210	1.370	2.870	3.540	1.678	1.068
	PE	0.610	0.730	0.985	1.130	1.540	2.960	3.500	1.636	0.999
55016	GEV	0.031	0.040	0.129	0.180	0.289	0.245	0.223	0.162	0.134
	GPA	0.033	0.048	0.140	0.135	0.220	0.149	0.371	0.157	0.115
	GL	0.223	0.048	0.140	0.201	0.318	0.317	0.219	0.209	0.186
	PE	0.029	0.034	0.108	0.136	0.248	0.251	0.215	0.146	0.111
55026	GEV	0.316	0.277	0.739	1.500	0.978	1.780	2.050	1.091	0.762
	GPA	0.608	0.595	1.150	2.580	2.070	3.300	3.360	1.952	1.401
	GL	0.322	0.355	0.859	1.350	1.370	1.950	2.130	1.191	0.851
	PE	0.369	0.312	0.790	1.750	0.981	1.750	1.960	1.130	0.840
60002	GEV	0.060	0.060	0.092	0.105	0.173	0.288	0.311	0.156	0.098
	GPA	0.095	0.091	0.069	0.091	0.215	0.385	0.232	0.168	0.112
	GL	0.136	0.069	0.135	0.339	0.256	0.381	0.438	0.251	0.187
	PE	0.059	0.060	0.084	0.101	0.177	0.307	0.347	0.162	0.096
72004	GEV	0.494	0.674	1.630	1.580	2.030	1.900	2.610	1.560	1.282
	GPA	0.413	0.459	1.110	0.982	1.490	3.250	4.780	1.783	0.891
	GL	0.619	0.832	1.920	2.110	2.720	2.500	2.210	1.844	1.640
72004	PE	0.473	0.605	1.370	1.210	1.810	1.970	2.410	1.407	1.094

ANNEX

GUIDELINES FOR BEST PRACTICE

Low Flow Frequency Analysis
Guidelines for Best Practice

MD Zaidman, V Keller & AR Young

Research Contractor
Centre for Ecology & Hydrology

Publishing Organisation
Environment Agency, Rio House, Waterside Drive, Aztec West, Bristol BS32 4UD
Tel: 01454 624400 Fax: 01454 624409

ISBN: 1 85705 996 4

This report is the result of a collaborative study jointly funded by the Environment Agency and the Natural Environment Research Council.

Dissemination Status
Internal: Released to Regions
External: Released to Public Domain

Statement of Use
This report provides guidance for frequency analysis of annual minimum D-day flows. It will be of use to Water Resources staff involved in operational and planning tasks.

Key Words
Low flow, annual minima, flow-duration-frequency methods, probability distributions.

Research Contractor
This document was produced under R&D Project W6-064 by:
Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB
Tel: 01491 838800 Fax: 01491 692424

Environment Agency Project Manager
The Agency's Project Manager for R&D Project W6-064 was:
Andy Wall, Environment Agency - Wales

ACKNOWLEDGEMENTS

The authors would like to thank Andrew Wall (Environment Agency, Wales) and Daniel Cadman (Environment Agency, Anglian Region) for their helpful guidance in preparation of this guidance note.

EXECUTIVE SUMMARY

This document is a technical annex to R&D project W6-064, 'Probability distributions for x-day daily mean flow events'. It provides guidance on best practice for frequency analysis of annual minimum flows.

The general issues to be considered when contemplating a frequency analysis are discussed with a strong emphasis on likely sources of error and uncertainty. In order to encourage a consistent methodology to be adopted in the UK a single parametric approach for estimating the probability of occurrence of low flow events is recommended. This approach is based on the use of L-moments with a Pearson Type III probability distribution to estimate the flow - return period relationship for annual minima of different durations.

CONTENTS

LIST OF SYMBOLS

x	Observed annual minimum flow
x_i	Annual minimum flow of rank i (in ascending order)
n	Number of sampled values
X	Random variable corresponding to annual minimum flow
D	Duration of low flow period (in days)
T	Return period or recurrence interval
p	Probability of non-exceedance
F(x)	Cumulative distribution function
Q(*p*)	Quantile function
k	Shape parameter
α	Scale parameter
?	Location parameter
G	Gamma function
G	Incomplete Gamma Function
$?_s$	Spearman Rank correlation coefficient
$?^2$	Chi-square goodness of fit parameter

LIST OF ABBREVIATIONS

p.d.f.	Cumulative probability distribution function
AM	Annual minima
MAM(D)	Mean annual minimum flow at duration D.
MF	Mean flow
LFFA	Low Flow Frequency Analysis
RMSE	Root mean square error
Q95	The one-day discharge exceeded 95% of the time

1. INTRODUCTION

1.1 Objectives and Applicability

This document provides guidance for frequency analysis of annual minimum flows. It is not intended as a 'rule book', rather its aim is to promote a consistent approach to low flow frequency analysis in the UK. Whilst a guideline methodology is described, no programs or software tools are provided for its implementation. However references to appropriate software packages are provided where required.

The guidelines given are specifically for good quality flow records longer than 25 years in length, and may give misleading and uncertain results if applied to shorter or incomplete flow records. For example, a semi-deterministic approach, relying on alternative sources information about the low flow regime, such as from an analogue catchment or rainfall-runoff model might be more suitable for sites with limited (less than 25 years) flow data. Furthermore, as the method has been derived and tested using measured flow series, it is not necessarily best practice for synthetic or modelled data, such as naturalised flow series. Similarly, it should be used with caution for series that are known to be subject to large hydrometric errors, especially where these are unquantifiable. The method is also generally inappropriate for frequency analysis of other hydrological variables, such as annual maxima, level, velocity and so on, or for annual minima from regimes unlike those found in the UK.

The document is pitched at those with a basic understanding of statistical principles, and a glossary is provided for clarification where necessary. For those unfamiliar with the subject, the main principles and assumptions of low flow frequency analysis are also summarised later in this chapter (Section 1.3). The main aim to increase awareness of some of the pitfalls and issues connected with frequency analysis, such as sources of uncertainty or bias. These points are discussed in Chapter 2, whilst a detailed step-by-step methodology is presented in Chapter 3. Finally a worked example is included in Chapter 4.

1.2 Project Details

The recommendations presented are based on the findings of R&D project (W6-064) entitled 'Probability Distributions for "x-day" Daily Mean Flow Events' (Zaidman *et al.*, 2002). The project reviewed the use of parametric estimation methods within low flow frequency analysis, and examined the ability of different candidate probability distributions to describe the occurrence of D-day annual minima flow events. The study was based on data for 25 UK rivers having long, stable and natural flow records. It should be noted that of these, 20 were located in upland areas, including three in Wales and 14 in Scotland, whilst only five were from the aquifer dominated regions to the south east of England. The range of durations examined included D=1, 7, 30, 60, 180 and 365 days. The study also examined different methods of deriving the minima and ensuring that these were both stationary and independent.

Three candidate distributions were considered in the study: the Generalised Extreme Value, the Pearson Type III and the Generalised Logistic. In each case the 3-parameter form was used. Extreme value theory suggests that the frequency behaviour of annual minima will follow that of a Generalised Extreme Value distribution, however in practice a number of other distributions have been shown to describe the observed equally well. Using the method of L-moments for parametric estimation, the parameters of each of the three candidate distributions were determined based on the flows and plotting positions probabilities of the observed data. A fourth distribution, the Generalised Pareto, was also considered. Although this distribution was not theoretically suitable for describing extreme events, it was included as a ‘control’, i.e. to see whether this expectation was born out in the results.

All of the four distributions were able to satisfactorily represent the form of the observed data points, to some degree or another. A method of ranking the candidates according to their goodness–of-fit criteria, and RMSE was used to identify which distribution best represented the observed frequency curve for each series. A number of trends became clear from this exercise. For annual minima of short duration the Generalised Extreme Value and Pearson Type III distributions performed well in responsive catchments found in upland areas. These distributions performed less well for series based on longer durations, where short-lived extremes become averaged out. The Generalised Logistic was best in the high-storage catchments that typify many lowland areas of south east England. In some cases the distributions were not able to provide physically reasonable estimates for annual minima-recurrence intervals much beyond the observed range. Where the prescribed flow was less than 10%MF ‘sensible’ estimates of recurrence interval were, in general, obtained only for annual minima of short duration for impermeable catchments. Where annual minima of longer duration or catchments of high permeability were considered recurrence interval estimates were realistic only for higher prescribed flows.

1.3 Low Flow Frequency Analysis

Low Flow Frequency Analysis (LFFA) is a stochastic approach for characterising low flow events. The pivotal aim is to quantify the likelihood that the flow at a particular site will persist below a particular level over a particular duration. LFFA is thus typically utilised where using a single statistic or index, such as the MAM7 or Q95, is insufficient to describe the low-flow regime. For example in water resource planning, where low flow events of different length and severity need to be considered within an historical context, LFFA provides a means to quantify the flow-duration-frequency behaviour of the site of interest. Individual low flow events can be delineated by considering periods where the flow falls below a threshold level (i.e. LFFA is applied to a partial duration series). However, unless the flow record is particularly short, a simplified approach, using some representative annual value to typify the overall character of the low flow season, is often favoured. Customarily the minimum D-day average discharge per year is considered.

Given a set of observed annual minima $\{X_i, i=1..n\}$, the goal is to estimate the probability of occurrence of some minima, x. Presuming that the observed values are independent and identically distributed (i.e. that observed minima are random realisations of a single population of annual minima), this can be achieved by finding the cumulative distribution function, denoted by F(x), which represents the probability, *p*, of any previous or future minima, X, being less than or equal to some given value, x.

$$p = F(x) \qquad (1)$$

Similarly the probability of x being exceeded by X, termed F'(x), is given by 1-*p*

$$1-p = F'(x) \qquad (2)$$

The corresponding quantile function, denoted by Q(p), defines the value x associated with the p^{th} quantile:

$$x = Q(p) \qquad (3)$$

and can be used to define the flow x_T, associated with a recurrence interval T, where T is the reciprocal of *p* (and vice versa), as follows:

$$x_T = Q(\frac{1}{T}) \qquad (4)$$

As a single probability distribution function is assumed to describe all annual minima occurring at the site of interest during the lifetime of the river, including those within the observed series, the sample data are used to define the form of F(x). For instance the statistical characteristics of the observed data, particularly aspects of the shape and density (e.g. skewness and kurtosis), are assumed to be valid for the population as a whole. However, with few rivers having flow records longer than 50 years in length, most annual minima series represent a relatively small fraction of the possible range for the site of interest (i.e. there is a large sampling basis). Moreover, the observed data provides little detail regarding the shape of the distribution function in its upper and lower tails (at extremely high and low probabilities. Thus the problem of estimating F(x) reliably for all possible values of x is very difficult. Ironically the interest is usually in minima that lie well outside the range of observed values. As a result frequency analyses predominantly rely on parametric estimation procedures.

In the parametric approach, *a priori* assumptions about the shape of the cumulative distribution function are used to select an appropriate hypothetical distribution family as the basis of the mathematical expression describing the distribution. The observed data set are then used to constrain the parameters of this distribution (i.e. select the most relevant family member) usually using fitting techniques such as L-moments (Hosking & Wallis, 1997) or maximum likelihood estimation (Cox & Hinkley, 1974). The danger of this approach lies in choosing an inappropriate hypothetic distribution that will produce a misleading quantile estimates. To minimise the chance of using an unsuitable distribution a number of 'candidate' distributions are considered. The uncertainties associated with the resulting models are then assessed using goodness of fit tests or by applying re-sampling techniques to produce confidence intervals (e.g. Takara & Stedinger, 1994). Unfortunately, as most flow records are short, a number of different distribution types may all fit the observed annual minima reasonably well and it may not be possible to discriminate between them on an objective basis. Thus a particular model may be favoured for practical reasons, such as computational convenience, or because it exhibits certain characteristics that the user believes a low flow distribution should have. For example, a distribution having a finite lower limit equal to zero (to represent the possibility of recording a zero, but not a negative, flow) is often considered preferable to one that does not. This lack of objectivity in LFFA, coupled with the unparsimonious use of assumptions, has brought criticism from a number of authors including Klemes (2000), and facilitated wider use of non-parametric function estimation routines.

Candidate distributions are generally chosen from the extensive 'library' of established distribution families - some of the more common families are summarised by Evans *et al*. (1993). 'Bespoke' distribution functions are also sometimes used, such as those of

Gottschalk *et al*. (1997) based on low flow recession behaviour. Extreme value distributions such as the EV1 (Gumbel) and EVIII (Weibull) have traditionally been considered the most applicable of the established distribution families. This goes back to the work of Fisher and Tippet (1928) who showed that the minima of a particular sample will theoretically tend to one of three extreme value forms, named the EVI, EVII and EVIII. The Weibull distribution, advocated by Gumbel (1963), has found particular favour in the UK (e.g. Institute of Hydrology, 1980). The Log Pearson Type III (Loganathan *et al*., 1986), the Log-Normal (Kroll & Vogel, 2002) and the Gamma (Bobée & Ashkar, 1991) are also commonly used. A summary of recent work is given by Tallaksen (1999) and by Zaidman *et al*. (2002).

2. GENERAL CONSIDERATIONS

2.1 Introduction

This chapter discusses the general considerations which must be made when undertaking a frequency analysis of annual minima data, outlining sources of error and uncertainty. The various considerations are presented in order of importance, with the need for a consistent and pragmatic approach outweighing some ‘scientific’ considerations.

Figure 1 shows the general stages that should be involved in a low flow analysis, from deriving and manipulating the annual minimum series, through parameterisation of a distribution selected *a-priori*, to the assessment of the resulting flow-recurrence relationship derived from its quantile function.

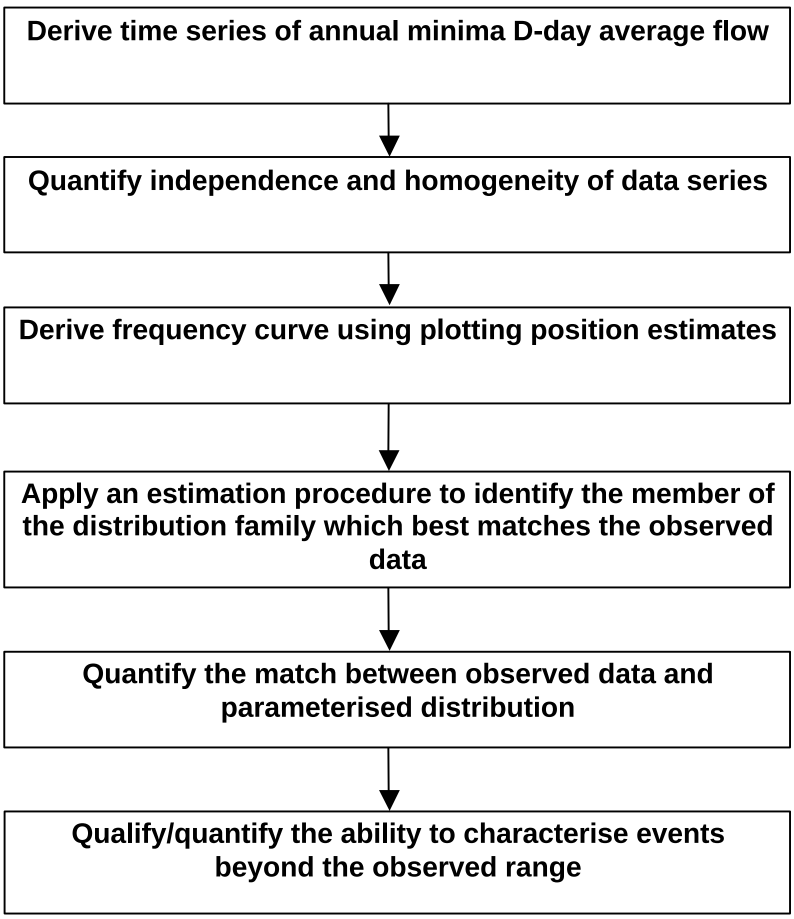


Figure 1: General stages of a low flow frequency analysis

2.2 Record Length

Record length is possibly the most influential factor in frequency analysis. Decreasing the sample size introduces sampling errors and increases the inherent uncertainty associated with the flow and recurrence interval relationship derived from the sampled data. This is because the assumption that the sampled minima are random representations of the true low-flow distribution breaks down where the record length is short. Whereas a large sample is likely to clearly exhibit the features (such as the level of skew and kurtosis) of the population of interest, a small sample is highly unlikely to be representative of the population. The level of sampling error is also influenced by the period over which the observations were made (e.g. a 30-year record from 1940 to 1970 may contain fewer extreme events than a 30-year record from 1970 to 2000). Furthermore, plotting position formulae are not robust for small sample sizes because they do not take into account the range and skew of the sampled flows, and thus cannot be thought of as unbiased estimators for short records, whilst many of the statistical tests used within the LFFA framework are unable to provide unequivocal answers where the sample size is small.

Ideally, to minimise sampling errors and increase the number of observed events in the extreme tails of the distribution, only records of several hundreds of years in length would be used for analysis. In practice, of course, most rivers in the UK have only been gauged since the 1960's, giving an average record length of around 40 years. Although it is difficult to quantify the length at which a record becomes 'short', an observed record length of 40 years is about the minimum record length that might give relatively reliable estimates for use in water resources application. **It is inadvisable to use LFFA with records less than 25 years in length.**

As a rule of thumb, a one in T year event requires a minimum record length of T/2 years, not counting missing or rejected years. If the aim is to characterise the 1 in 200 year annual minimum, then a longer record (100 years or more) is required than if the 1 in 50 year event was of interest (25 years or more). Similarly, a considerably higher uncertainty would be associated with a one in 100 year event based on a 30 year record, than a 1 in 100 year event estimated from a record of 60 years. Whilst uncertainties associated with individual components may be quantified (e.g. resampling methods may be used to provide some quantification of the sampling errors) the overall uncertainty associated with record length is not easily quantified.

2.3 Assumption of Statistical Independence and Stationarity

The assumptions that the data must be independent and originating from the same statistical population are central to the LFFA method. The latter assumption also implies that the data must be stationary (i.e. show no trends over time) and homogeneous, and there must be no outliers amongst the sample data. Short records are particularly vulnerable to the effects of short-term trends, which might average out over the long term.

Independence

Where sample data are not independent they cannot be thought as representing a population of random variables. Serial elements may manifest in annual minima due to catchment storage processes (e.g. carry-over of base flow conditions from one year to the next) and are more likely for longer duration minima, where two successive minima have a number of days in common. Statistical dependence has an important effect on the interpretation of the results from LFFA. Suppose for example that the analysis suggests that an annual minimum flow has a non-exceedance probability, p. The probability that annual minima in two consecutive years both do not exceed a particular flow will then be p^2, but if they are interdependent it would be closer to p. **Thus ignoring interdependence may lead to substantial underestimation of risks of sequences of years with low flows.** Non-independence can also influence the results of formal tests of fit and assessment of the uncertainty in parameter estimation used within the LFFA process.

Whilst a number of tests may be used to identify whether a data series is independent and originating from the same population (see Chapter 3), these tests are likely to perform poorly where the number of observations is small, and the user is advised to apply the most powerful tests as possible. It is possible to revise the LFFA procedure to take account of dependent data (e.g. Chung & Salas). However this will be beyond the capabilities of most non-statisticians! It is perhaps better to be aware of the effects of dependence on a qualitative level and take a view based on knowledge of the catchment and the drought events concerned. Generally, in impervious catchments where the soil moisture deficit is fully replenished during the winter, there is no serial correlation between annual minima induced by storage (indeed there may be independent events within the year). Storage from aquifers, reservoirs, soils that may not become saturated in winter and areas of persistent snow will all increase the potential for serial correlation.

Stationarity

Causes of non-stationarity include problems with the recording process (such as changes in rating equations, relocation of stations or changes in recording method), changes in the catchment (such as land use change), climatic variability or climate change. Trends are usually more evident in short records (the effects often average out over the longer term). If the user is confident that trend effects can be removed from the time series, the adjusted series may be used in LFFA.

Outliers

True outliers may be caused by 'one-off' artificial influences in the catchment or by measurement errors, and are less common for large durations, where the effects of abnormal daily flows are averaged out. However in a frequency analysis of extreme events it may be difficult to determine whether an outlying data point really is a true outlier or simply an extreme rare flow. In the latter case the data point is crucial as it will provide information that will help constrain the tail shape of the distribution function. It is therefore important not to remove apparent outliers arbitrarily. Where possible, outliers should be verified using circumstantial evidence, such as local knowledge, rainfall records, or by seeing if outlying events occurred on a regional basis. If an objective treatment is desired, a number of tests for outliers are well established in the statistical literature (e.g. see Barnett and Lewis, 1994).

2.4 Choice of Distribution

The aim of LFFA is to choose the theoretical distribution with the most appropriate shape for the data. However, the choice of distribution is rarely the major source of uncertainty in the end result. Whilst it is common practice to compare one or two different candidate distributions, there is little point in agonising over several alternatives giving similar results, particularly as record lengths for UK rivers are typically not long enough to ensure that a single 'best' distribution would be unequivocally identified in each case. We therefore **advocate the Pearson Type III distribution for use with all catchment types and all durations.** As well as giving the best performance in the study, the Pearson Type III has some physical attributes which make it a good choice for frequency analysis. In particular, the Pearson Type III is unlikely to predict negative flows for high recurrence intervals.

2.5 Fitting Technique

Within frequency analysis there is a profusion of competing estimation procedures that can give different results, and several studies have focussed on comparing the performance of these various techniques (e.g. Arora & Singh, 1987). However two parameter estimation methods dominate: likelihood based techniques (i.e. maximum likelihood), and moment based matching techniques, specifically the use of L-Moments. Both are well established techniques and there are many examples of their use in the frequency analysis literature.

L-Moments are linear combinations of the probability weighted moments and use the sample data to provide estimates of certain properties of the underlying population, which are then matched to a member of the chosen distribution family. A full description of the theory of the L-Moments method is beyond the scope of this document. Hosking and Wallis (1997) provide an extremely comprehensive review of the L-Moments methodology, although the same ideas are also discussed in a number of recent journal papers including Hosking (1990). The advantages of L-moments are that they have been shown to be unbiased, have relatively small sampling variance and are relatively insensitive to outliers (on the down side this latter point also means that large (or small) sample values reflecting important information in the tail of the parent distribution are given relatively little weight in the estimation procedure).

Likelihood based techniques are well established in statistical theory and practice. They are more versatile than L-moments and can be adapted to deal with a range of circumstances, such as dependence between observations. The estimators are known to have certain optimal properties in the sense that, one the sample size is large enough, no other estimators (including L-moments) have better properties. Unfortunately they do not perform as well where the sample size is small. A number of problems can arise in attempting to apply maximum likelihood in practise, such as the non-existence of a maxima, or existence of several local maxima, and so on.

Generally, the choice of fitting technique is irrelevant in the context of uncertainties generated by there being insufficient record length. Therefore from a pragmatic perspective, we suggest that **L-moments should be used for parameter estimation**. L-moments are widely applied in flood frequency estimation at present and are available in many statistical/ hydrological packages such as WINFAP and MIKE11.

2.6 Hydrometric Errors

It is assumed that the data sets to be used will be free from hydrometric errors. An analysis of errors carried out as part of the project showed that where the daily flow series are subject to random hydrometric errors of less than ± 10%, there will generally be little effect on the results of the frequency analysis, especially where longer duration are considered (the errors will tend to average out). However random hydrometric errors of up to ± 20% can result in more than a 20 year uncertainty in recurrence interval. Systematic hydrometric errors (bias) are likely to impose a systematic bias on the flow-recurrence interval curve.

3. RECOMMENDED METHOD

3.1 Introduction

Specific guidelines for low flow frequency analysis are presented in this chapter. As discussed earlier, the guidelines are based on gauged flows in catchments with minimal artificial influence and more than thirty years of record. The flow chart shown in Figure 2 gives an overview of the various stages in the recommended approach. The stages are discussed further in Sections 3.2 to 3.6.

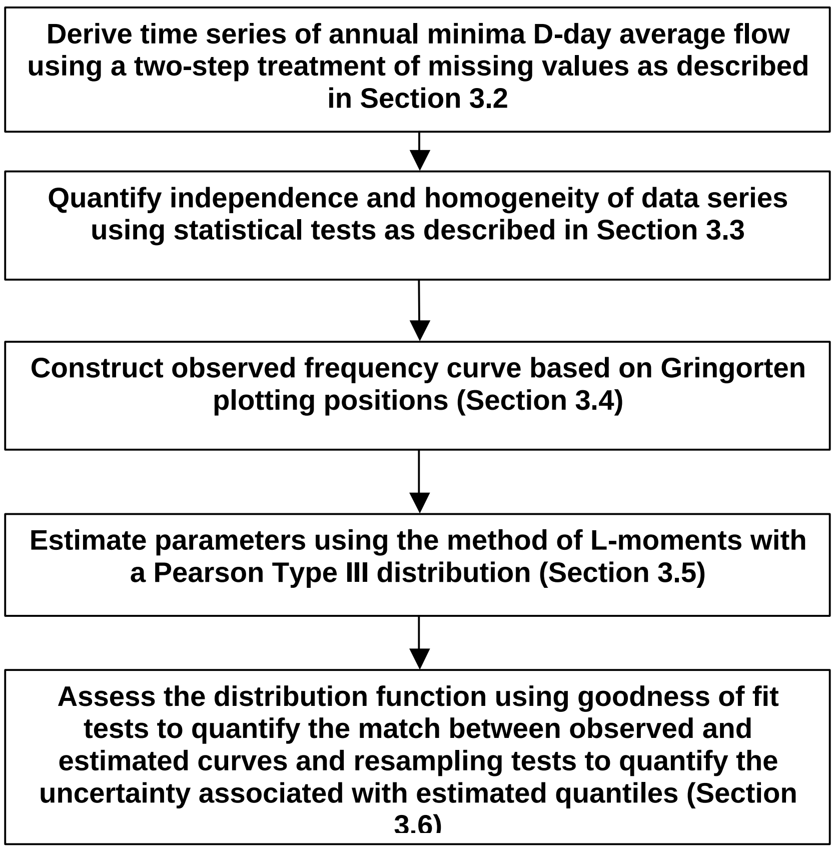


Figure 2. Overview of guidelines for best practice in LFFA

3.2 Computation of Minimum D-day Flow per Calendar Year

Here D-day average discharge is calculated on a running average basis for the entire period of record. This means that a window of duration D days is moved sequentially through the record using an increment of one-day, the mean flow over the duration being calculated at step. Each D-day mean is indexed to the middle day of the interval. For any interval of length D, the index day with be the m^{th} day where m is defined as:

$$m = \begin{cases} \left(\frac{D}{2}+1\right) & \text{, where D is even} \\ \left(\frac{D+1}{2}\right) & \text{, where D is odd} \end{cases} \quad (5)$$

For example the 7-day average determined using flow data from 3rd September 1983 to 9th September 1983 inclusive would be indexed to the 6th September 1983. As the running average method is utilised with an increment of one day, 365 running average values are determined for each calendar year, the smallest being used as the annual minimum.

If some daily flows are missing a particular year within the flow record, the level of uncertainty associated with the minimum D-day flow for that year increases, particularly where the missing values occur within the low flow period or where the number of missing values is relatively large. However, data missing from periods of relatively high-flow are unlikely to have much influence on the annual minima calculated for a particular year. Although missing values may be filled in by interpolation, this is inappropriate where several consecutive values are missing and an objective and consistent treatment of missing data is therefore required. To avoid filling in large gaps by interpolation whilst also avoiding rejecting years unnecessarily, the annual minima from data-poor years are rejected as a first step, whilst years with too much missing data within the low flow period are excluded in a second step as follows:

Step 1. The number of missing data allowed per calendar year is constrained to 30 days, i.e. an annual minimum is rejected outright if the year it represents contains 30 or more days with missing values.

Step 2. A ‘low flow period’ is delineated for each remaining year based on the shortest continuous period in which the lowest 20% (73 days) of daily flows are represented. For example if the 60 smallest flows occurred between 15th July and 15th September, and the next 13 lowest flows occurred between 10th and 22nd June the ‘low flow period’ is assumed to span the period from 10th June to 15th September. The year is discarded if the maximum number of consecutive missing values within the ‘low flow period’ exceeds 7 days. However if there are less than seven consecutive missing days, the year is only discarded if the aggregate number of missing data during the ‘low flow period’ is greater than 10 days, otherwise missing data can be filled by interpolation.

Although strictly not missing data, particular problems occur at the beginning and end of the flow record. The first (D-m) days at the start of a record and the last (m-1) values at the end of the flow record cannot be used as index days. For instance to derive a value for 1st January 1968 requires (m-1) extra days from the previous year 1967, yet if 1st January 1968 is the first day in the record, these extra data do not exist. A similar situation arises at the end of the flow record and results in a reduced number of running-averages being computed for years at the end and beginning of the record. Where these years are acceptable in Step 2, they can be included in the analysis provided that the value of m does not exceed 75 (tests showed that the value of the annual minima was changed only if more than 75 days were missing).

3.3 Validation of Independence and Stationarity

Sample data possessing serial elements cannot be thought of representing a population of random variables, therefore the independence, stationarity and homogeneity of the annual minimum series should be verified prior to beginning the frequency analysis procedures. In particular, non-independence may manifest as larger durations are considered.

The statistical literature gives examples of appropriate tests that may be applied to the data. For instance independence may be assessed using Anderson's Test, the Wald-Wolfowitz Test, the Durbin-Watson method or by determining the auto-correlation function of the data. Similarly the level of stationarity can be quantified by using tests such as the Spearman Rank Test and the Mann-Kendall Test, as well as via a number of non-linear methods. Tests for homogeneity include the Mann-Whitney Test. However it is important to note, that most of these tests do not perform well when the sample size is small (i.e. for many stations the record length will be insufficient to ensure the tests perform reliably).

3.4 Construction of the Observed Frequency Curve

Low flow frequency curves (probability plots) are used to depict the variation in observed annual minima with exceedance probability. The plotting positions along the probability axis are estimated using empirical formulae that evaluate the probability associated with a particular observation from its rank in the sample set (i.e. the plotting position is a distribution-free estimator of the probability). A number of plotting position formulae have been suggested over the years. Some are said to be optimised for particular distribution types, although these differences are really only evident for data points of the very lowest and highest ranks. Reviews of plotting position formulae are given by Cunnane (1978) and again by Cunnane (1989). Based on his recommendations, this guidance document advocates the use of the Gringorten formula for estimating plotting positions. The Gringorten formula is given by

$$p_i = (i-0.44) / (n +0.12) \qquad (6)$$

where p_i is the estimated **exceedence** probability for the data point of rank i, and n is the total number of observations. The rank, i, is calculated by reordering the set of data-values according to size, the largest value being assigned a rank of 1, and the smallest value a rank of n.

The probability plot is then constructed by plotting X_i (vertically) against p_i. The plotting position is often expressed as a recurrence interval (i.e. reciprocally) such that the curve describes the average interval between years in which the annual minimum D-day falls below a given discharge. If the user does not have the facility to plot on a probability scale a reduced variate scale may be used to 'linearise' the probability axis. In this approach the variate value corresponding to each plotting position or recurrence interval is determined, making use of the assumption that the data is likely to conform to an extreme-value type distribution. For ease, the Weibull reduced variate given by the following formula is recommended:

$$V = 4\left(1 - e^{-0.25(-\ln(-\ln p))}\right) \qquad (7)$$

where V is the reduced variate value, and p is the exceedance probability.

It is important to note that whilst plotting positions are fairly accurate estimators of the probability where n is very large (and the sampling errors are small), they are predominantly influenced by the number of observations in the sample set where n is small. This means that plotting positions will be equally spaced along the probability axis, regardless of how the magnitudes of the observed flows are distributed. For instance for a sample size of 35 years a recurrence interval of 62 years is always assigned to the smallest flow, regardless of its size or relation to the other data.

3.5 Parameter Estimation

As discussed earlier, the method of L-moments is suggested as the preferred estimation method. Parameter estimation is best attempted using existing software, rather than from a first principle approach; commercial software packages providing L-moments include

MIKE11, S-PLUS and WINFAP, the latter being specifically designed for regional flood estimation procedures (Institute of Hydrology, 1999). Fortran subroutines for L-moments are also available as freeware e.g. Hosking (2000).

The Pearson Type III distribution family is advocated as the best all-round choice of distribution for use in low flow frequency analysis. However where procedures for estimation based on the Pearson Type III are not available, the three parameter Generalised Extreme Value distribution is the next most preferred option.

When the three parameters have been determined, they can be used to define the form of F(x). Here we present the Pearson Type III formulation given by Hosking and Wallis (1997). The three parameters are shape, *k*, scale, α, and location, ξ. Where $k>0$ the appropriate cumulative distribution function is

$$F(x) = G\left(m, \frac{x-g}{b}\right) / \Gamma(m) \tag{8}$$

where $\mu = \frac{4}{k^2}$, ß= $\frac{1}{2} a |k|$ and ?= $x - \frac{2a}{k}$ (these do not hold true in where k=0, as in that case the distribution reduces to the normal distribution). The range of *x* is ? = *x* = 8 .

Similarly where $k<0$ the appropriate cumulative distribution function is

$$F(x) = G\left(m, \frac{g-x}{b}\right) / \Gamma(m) \tag{9}$$

and the range of *x* is ? = *x* = 8 .

G and ? represent the incomplete (integrated between 0 and *x*) and complete (integrated between 0 and infinity) Gamma Functions respectively, and may be estimated according to the following relationships:

$$?(x) = \int_o^\infty t^{x-1} e^{-t} dt \tag{10}$$

$$G(\mu,x) = \int_o^x t^{m-1} e^{-t} dt \tag{11}$$

3.6 Assessing the Estimation of F(x)

In assessing whether the fitted distribution is acceptable, the bias, variability and accuracy of the parameter estimates are taken into account. As the ‘true’ form of the distribution function for annual minima at the site of interest is always unknown, the estimation is usually assessed in relation to the observed data points. This usually involved some kind of comparison between the plotting position estimators of the observed minima and those inferred from the estimated form of the distribution function. Various goodness-of-fit tests may be used to quantify how well the distribution fits the observed data. Equally by-eye fitting tests are often used, but should be backed up with some quantitative justification. Resampling methods such as the jacknife or bootstrap may be used to quantify the uncertain associated with quantile estimates or to define confidence limits. These procedures are standard statistical methods, and are detailed in many different statistical texts. However a comprehensive description may be found in volumes 9 and 12 of the Handbook of Statistics (Rao, 1993; Patil & Rao, 1994).

4. WORKED EXAMPLE

This example is based on the North Esk gauged at Dalmore Weir (station 19004), located in the SEPA-East region. The annual minima of interest are those for durations of 1 day and 180 days. The upstream catchment area is 81.6 km^2 and the flow regime is moderately responsive - annual average rainfall is 951 mm and the Base Flow Index is 0.54. This record is of high quality and received a grade ‘A’ rating for hydrometric quality in the study by Gustard *et al*. (1992).

Daily Flow Record and Mean Flow

The daily flow record is available from 01/01/1960 to 31/12/1999. The period of record Mean Flow (MF) is 1.53 m^3s^{-1}.

Derivation of Annual Minima Series

Table 1 gives the minimum D-day flow derived for each year where D=1 and D=180, expressed both in absolute terms and as a percentage of the Mean Flow (%MF). For 1960, the annual minima where D=180 is rejected as no running average values could be calculated for the first 90 days of that year (i.e. there is no record for the latter months of 1959). For D=1 two or more years have the same annual minimum flow (e.g. 1961 and 1962 and 1977 and 1978). This is a feature of the data, but the same effect may occur due to rounding up errors and so on.

Validation tests for independence and homogeneity

Figure 3 shows the annual minima series for the North Esk derived for durations of D=1 and D=180. In both cases the annual minima varies from year to year, and on visual inspection no long term (linear) trends are apparent.

Table 1: Annual Minima (AM) Series for the North Esk at Dalmore Weir

	D=1		D=180	
Year	**AM (m^3s^{-1})**	**AM (%MF)**	**AM (m^3s^{-1})**	**AM (%MF)**
1960	0.323	20.85	N/A	N/A
1961	0.391	25.41	76.86	49.59
1962	0.391	25.41	113.82	73.43
1963	0.459	29.97	136.46	88.04
1964	0.416	27.36	89.2	57.55
1965	0.297	19.55	124.79	80.51
1966	0.388	25.41	111.54	71.96
1967	0.391	25.42	107.51	69.36
1968	0.549	35.83	153.74	99.19
1969	0.305	20.2	75.67	48.82
1970	0.315	20.3	93.93	60.6
1971	0.316	20.85	80.83	52.15
1972	0.233	14.98	45.34	29.25
1973	0.214	13.68	50.51	32.59
1974	0.236	15.63	51.82	33.43
1975	0.143	9.121	58.37	37.66
1976	0.24	15.64	58.64	37.83
1977	0.355	22.8	109.99	70.96
1978	0.355	22.8	92.26	59.52
1979	0.317	20.85	69.75	45
1980	0.361	23.45	84.09	54.25
1981	0.392	25.41	75.25	48.55
1982	0.298	19.55	71.63	46.21
1983	0.222	14.33	83.17	53.66
1984	0.138	9.12	49.72	32.08
1985	0.382	24.76	117.41	75.75
1986	0.29	18.89	103.2	66.58
1987	0.369	24.11	103.4	66.71
1988	0.234	14.98	106.07	68.43
1989	0.263	16.94	50.76	32.75
1990	0.276	18.24	85	54.84
1991	0.234	14.98	60.43	38.99
1992	0.1	6.515	93.16	60.1
1993	0.408	26.71	124.85	80.55
1994	0.27	17.59	57.02	36.79
1995	0.258	16.93	66.25	42.74
1996	0.263	16.94	53.38	34.44
1997	0.336	22.15	95.34	61.51
1998	0.611	39.74	159.03	102.6
1999	0.403	26.06	92.6	59.74

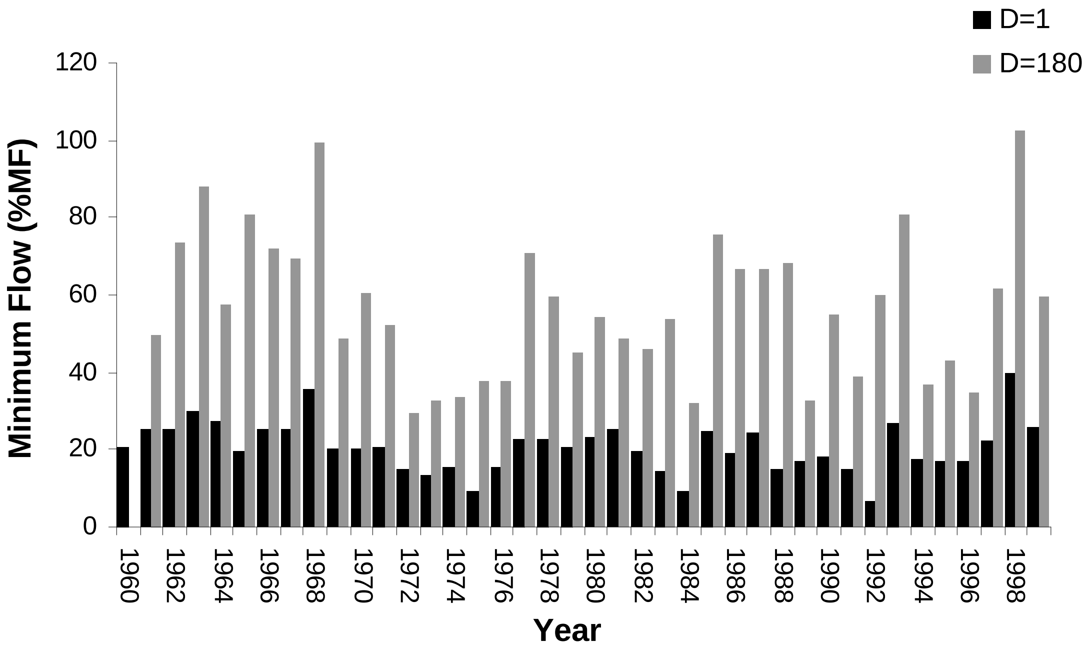


Figure 3: Annual Minimum Series for the North Esk at Dalmore Weir.

A number of validation test for independence, stationarity and homogeneity of the series are applied, in this case using the software package Systat 9 (©SPSS, 1998).

- Linear Regression Test

This shows that there is a poor linear relationship between the chronology and the annual minimum, giving a coefficient of (R^2) of 0.028 where D=1, and a value of 0.04 where D=180.

- Spearman Rank Correlation Test

The Spearman correlation coefficient between year and annual minima for D= 1 is equal to -0.236, and for D=180 is 0.127. As these values are much lower than the critical values for the samples, the hypothesis of rank order relationship between year and flow must be rejected.

- Autocorrelation test for independence

Figure 4 shows partial autocorrelation plots for a) D=1 and b) D=180. Autocorrelation measures the correlation of the series with itself shifted by a time lag. Autocorrelation can be calculated for a lag of any length, and if autocorrelation is present at one or more lags then the data is not independent. Partial autocorrelation plots show the relationship of points in a series to preceding points after ‘partialing’ out the influence of intervening points, and thus give a more conservative/ better perspective of autocorrelation.

a) b)

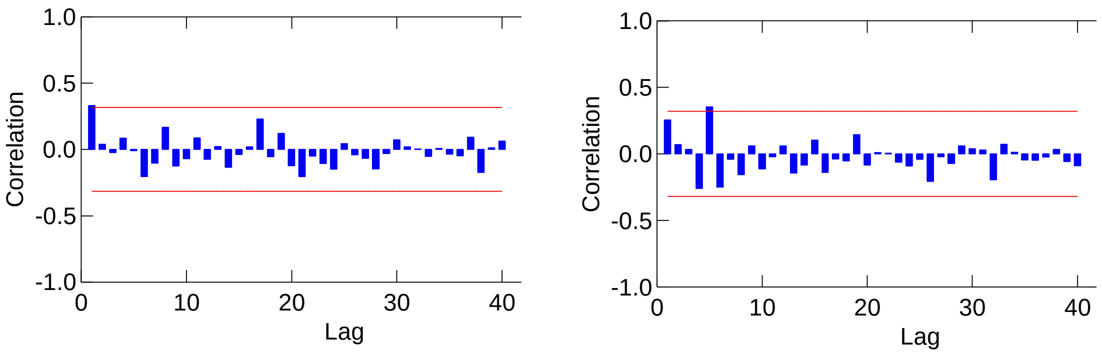


Figure 4: Partial autocorrelation plots for a) D=1 and b) D=180 for different lags (in days)

Values that lie outside the tramlines indicate that there is a significant correlation in the data at the lag (in years) indicated. Figure 4 shows that for D=1 there is little or no dependence within the data series, whilst for D=180 there seems to be some autocorrelation in the data at a lag of five years.

Construction of Observed Frequency Curve

The Gringorten plotting position is used to estimated the exceedance probability, p, using n = 40 where D = 1, and n = 39 where D = 180. Table 2 gives details of the values used, whilst Figure 5, shows the resulting probability plot.

Table 2: Derivation of plotting positions for D=1 and D= 180

	D=1, N= 40		D=180, N= 39	
I	**AM (%MF)**	**p**	**AM (%MF)**	**p**
1	6.515	0.014	29.25	0.014
2	9.12	0.039	32.08	0.040
3	9.121	0.064	32.59	0.065
4	13.68	0.089	32.75	0.091
5	14.33	0.114	33.43	0.117
:	:	:	:	:
35	26.06	0.861	80.51	0.883
36	26.71	0.886	80.55	0.909
37	27.36	0.911	88.04	0.935
38	29.97	0.936	99.19	0.960
39	35.83	0.961	102.6	0.986
40	39.74	0.986	N/A	N/A

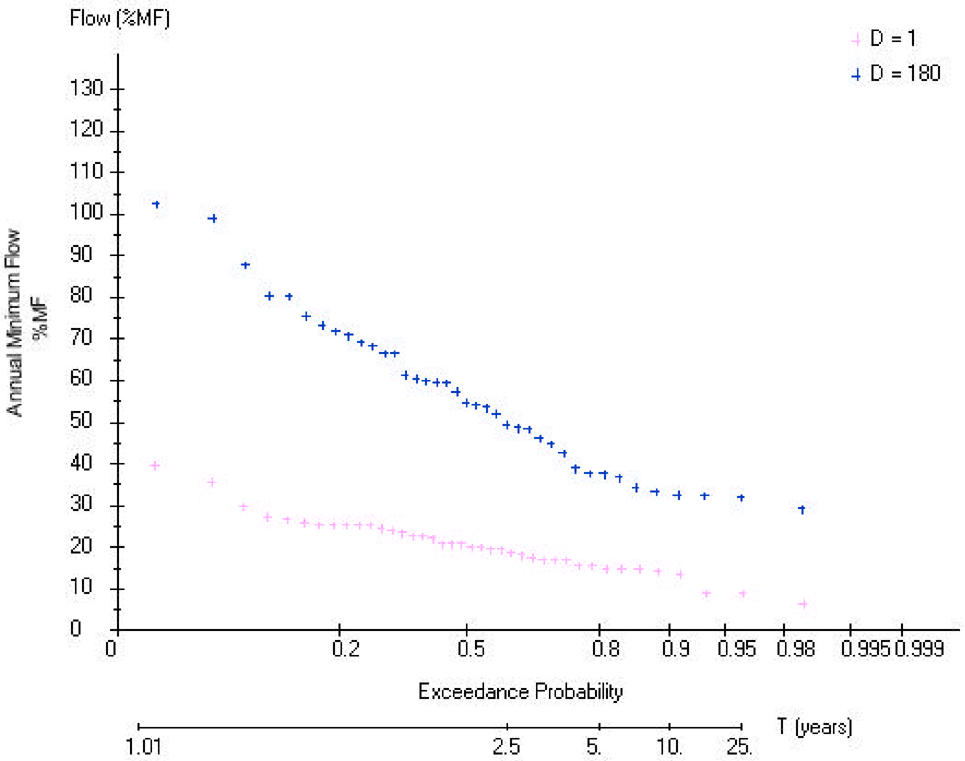


Figure 5: Frequency curves for D=1 and D=180

Fitting the Pearson Type III distribution

Parameter estimation via L-Moments is best attempted using existing software, rather than from a first principle approach. Several commercial software packages can provide the L-Moments estimation procedure with the Pearson Type III distribution, including MIKE11, S-PLUS, WINFAP. In this example the Fortran subroutines of Hosking (2000) were implemented yielding the parameters shown in Table 3.

Table 3: Parameters obtained via L-Moments for the Pearson Type III distribution

Parameter		**D=1**	**D=180**
a	scale	0.1	0.29
?	location	0.32	0.87
k	shape	0.13	0.53

As $k>0$ in both cases, the appropriate cumulative distribution function is

$$F(x) = G\left(m, \frac{x-g}{b}\right) / \Gamma(m) \quad (12)$$

For the case where D =1, the parameters are given by

$$\mu = \frac{4}{k^2} = 236.69 \quad (13)$$

$$ß = \frac{1}{2} a \left|k\right| = -0.0065 \quad (14)$$

$$?= x - \frac{2a}{k} = -1.22 \quad (15)$$

The range of x is -1.22 = x = 8. Substituting these parameters gives the following analytical solution for F(x):

$$F(x) = \frac{G\left(236.69, \frac{x+1.22}{0.0065}\right)}{\Gamma(236.69)} \quad (16)$$

F(x) can be evaluated by solving the incomplete Gamma functions for different values of x. Table 4 gives values of F(x) for different values of x, with corresponding recurrence intervals (T). The fitted curves are shown in Fig. 6.

Table 4: Estimation of F(x) for D =1

AM (%MF)	F(x)	T (years)
6.52	0.014	71.64
9.12	0.039	25.72
9.12	0.064	15.67
13.68	0.089	11.27
14.33	0.114	8.80
:	:	:
26.71	0.886	28.44
27.36	0.911	29.39
29.97	0.936	30.56
35.83	0.961	32.20
39.74	0.986	35.19

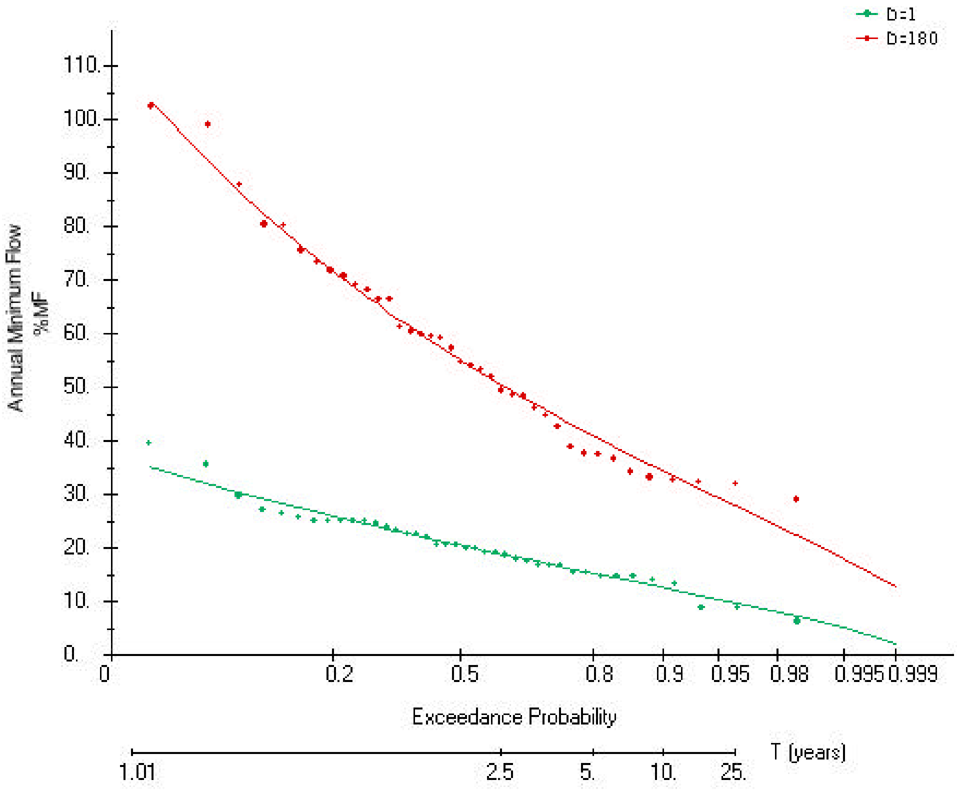


Figure 6: Fitted Curves for the North Esk for D=1 and D=180

Goodness of Fit Tests

The match between observed and predicted annual minima is quantified using root mean square error (RMSE) and chi square statistics, based on the residuals (Table 5).

Table 5: Comparison of observed and predicted flows, D= 180

T (years)	p	Observed AM (%MF)	Predicted AM (%MF)	residuals
69.86	0.014	29.25	22.51	-6.74
25.08	0.040	32.08	27.90	-4.18
15.28	0.065	32.59	31.11	-1.48
10.99	0.091	32.75	33.56	0.81
8.58	0.117	33.43	35.60	2.17
:	:	:	:	:
1.13	0.883	80.51	79.35	-1.16
1.10	0.909	80.55	82.56	2.01
1.07	0.935	88.04	86.64	-1.40
1.04	0.960	99.19	92.40	-6.79
1.02	0.986	102.60	103.40	0.80

Root mean square errors are 0.019 and 0.036 for D=1 and D=180 respectively, whilst the corresponding chi-square values are 1.23 and 2.31. At this point the user should refer to tables of critical values of the chi-square statistic, and based on the significance level of interest, determine whether they are exceeded by the observed values. The observed values must not exceed the critical values if the estimated curve is to be found acceptable.

Resampling methods would also be attempted, if desired, at this stage. A full description of resampling methods is beyond the scope of this guidance note. However there are plenty of introductory texts on the subject such as Good (1999).

Prediction of the Annual Minima – Recurrence Interval relationship

As the goodness-of-fit tests suggest that the estimated curve is satisfactory the distribution function, F(x), may be used to predict the probability (and thus the recurrence interval) associated with a particular annual minimum flow, x. Similarly if the annual minima associated with a particular recurrence interval was of interest, this could also be determined by calculating x based on F(x).

REFERENCES

Arora, K. & Singh, V.P.. 1987. An evaluation of seven methods for estimating parameters of the EV1 distribution. In: Singh, V.P. (Ed) Hydrologic Frequency Modelling, Reidel, Dordrecht, 383-394. A study comparing different estimation methods including likelihood techniques, L moments and maximum entropy for parameteric estimation. Focussing on the use of the Gumbel (EV1) distribution.

Barnettobée, B. & Ashkar, F. 1991. The Gamma family and derieved distributions applied in hydrology. Water Resources Publications, Colorado, USA, 205pp.
Provides a detailed description of the Pearson Family of Distributions, including the Pearson Type III distribution. Also provides a good summary of the logical steps included in frequency analysis, giving examples based on real data.

Bobée, B. & Ashkar, F. 1991. The Gamma family and derieved distributions applied in hydrology. Water Resources Publications, Colorado, USA, 205pp. Provides a detailed description of the Pearson Family of Distributions, including the Pearson Type III distribution. Also provides a good summary of the logical steps included in frequency analysis, giving examples based on real data.

Chung, C & Salas, J.D. 2000. Drought occurrence probabilities and risks of dependent hydrologic processes. J. Hydrologic. Eng., 5(3), 259-268. A treatment of dependent low flow events

Cox, D.R. & Hinkley, D.V. 1974. Theoretical Statistics. Chapman and Hall, London, UK.
A textbook on statistical procedure, giving a detailed treatment to likelihood based estimation procedures, including maximum likelihood.

Cunnane, C. 1978. Unbiased plotting positions – a review. J. Hydrol., 37, 205-222. A comprehensive critique of different plotting position formulae, including the Gringorten.

Cunnane, C. 1989. Statistical distributions for flood frequency analysis. Operational Hydrology Report No 33. World Meteorological Organisation, Geneva. A detailed report describing the use of different frequency distributions and plotting positions. Provides a good summary of the logical steps included in frequency analysis, but focusses primarily on annual maxima data.

Evans, M., Hastings, N.A.J. and Peacock, J.B. 1993. Statistical Distributions (2nd Edition). John Wiley, Chichester, UK. Paperback textbook summarising the forms of different frequency distributions.

Good, P.I. 1999. Resampling Methods: A Practical Guide to Data Analysis. Springer Verlag, 336pp. Introductory text describing the principles and use of resampling methods, including permutation tests and bootstrap tests.

Gustard, A., Bullock, A. & Dixon, J.M., 1992. Low Flow Estimation in the United Kingdom (IH Report No. 108). Institute of Hydrology, Wallingford, Oxon, OX10 8BB. 88pp. Manual on low flow estimation techniques in the UK. Includes sections on probability plot construction, and generalised estimation of frequency characteristics for UK rivers.

Gottschalk, L., Tallaksen, L.M. & Perzyna, G. 1997. Derivation of low flow distribution functions using recession curves. J. Hydrol., 194, 239-262. An example of the derivation and use of bespoke distributions for low flow frequency analysis.

Gumbel, E.J. 1963. Statistical forecast of Droughts. Bull. Int. Assoc. Sci. Hydrol., 8(1), 5-23. The classic paper on the use of the EV1 and EVIII distribution for describing annual minima data.

Hosking, J.R.M. 1990. L-Moments: analysis and estimation of distributions using linear combinations of order statistics. J. R. Statistic. Soc., B, 52(1), 105-124. Detailed methodology describing the properties and use of L-Moments.

Hosking, J.R.M. 2000. LMOMENTS: Fortran routines for use with the method of L-moments, version 3.03. IBM Research Division, Yorktown Heights, N.Y., U.S.A.. Freeware on the WEB from http://lib.stat.cmu.edu/general/lmoments

Hosking, J.R.M. & Wallis, J.R., 1997. Regional frequency analysis: an approach based on L moments. Cambridge University Press, Cambridge, UK, 221pp. A detailed, yet clear, account of the use of L-moments in frequency analysis by the leading authority on the subject. It summarises the main findings of Hoskings' previous papers on L-moments, in the framework of regional frequency analysis of annual maxima data. Mathematical expressions for the cumulative probability function and quantile functions for various distribution families are defined in the appendix.

Institute of Hydrology, 1980. Low Flow Studies. Institute of Hydrology, Wallingford, Oxon, OX10 8BB A four volume report describing a detailed regionalisation study of low flow duration and frequency, for use at ungauged sites in the UK.

Institute of Hydrology, 1999. Flood Estimation Handbook. Institute of Hydrology, Wallingford, Oxon, OX10 8BB A four-volume report detailing the recommended method for regional frequency analysis of annual maxima (flood) data for the UK.

Klemes, V., 2000. Tall Tales about Tails of Hydrological Distributions I. J. Hydrol. Eng., 5, 227-231. This set of two papers paper provides a clear and logical critique of the use of the parametric approach for extrapolating frequency curves, highlighting the uncertainties associated with extending the frequency curve beyond the observed range.

Kroll, C.N. & Vogel, R.M. 2002. Probability distribution of low streamflow series in the US. J. Hydrol. Eng., 7(2), 137-146. The study compared different hypothetical distributions to describe streamflow series in the US. They find different distributions for perenial and intermittent streams.

Loganathan, G.V., Mattejat, P., Kuo, C.Y. and Diskin, M.H. 1986. Frequency analysis of low flows: hypothetical distribution methods and a physically based approach. Nordic Hydrol., 17, 129-150. Study comparing the use of hypothetical distributions with a physically based model for predicting and characterising low flow events.

Patil, G. & Rao, C.R. Environmental Statistics. 1994. Handbook of Statistics Volume 12. Amsterdam, North-Holland, 1994. 927p.

Rao, C.R. Computational statistics. 1993 Handbook of Statistics Volume 9. Amsterdam, North Holland, 1993. 1045p. Statistical text book giving derivations and formulation for some tests and methods that are useful in assessing the fitted distributions.

Stedinger, J.R., Vogel, R.M. & Foufoula-Georgiou, E., 1992. Frequency analysis of extreme = events. In: Maidment, D.R. (Ed.). Handbook of Hydrology. McGraw Hill, New York, 18.1 18.66. A good background text describing frequency analysis. Although focussing primarily on flood frequency there is a section on low flows, which includes treatment of zero flows and regional frequency estimation

Tallaksen, L.M. 2000. Streamflow drought analysis. In: Drought and Drought mitigation in Europe. (ed by J.V. Vogt & Somma, F.) Kluwer Academic Publishers, Dorrdrecht, Netherlands, 103-117. Short review of the application of LFFA from a European perspective, describing recent work and the main principles involved.

Zaidman, M.D., Keller, V. & Young, A.R. 2002. Probability distributions for x-day daily mean flow events: Final Technical Report. Centre for Ecology and Hydrology, Wallingford, UK. 247pp. The final report for the study on which the recommendations presented in this document are based. It should be referred to for further detail or clarification.

GLOSSARY

Accuracy – In statistical estimation, accuracy refers to the deviation of an estimate from the true parameter value. In general, the term is used for the quality of a measurement that is both correct and precise.

Anderson–Darling test – A test procedure for testing the hypothesis that a given sample of observations comes from some specified theoretical population. It is particularly sensitive to deviations in the tails of the distribution. The combination of ease of computation and good power makes it an attractive procedure for a goodness-of-fit test.

Probability scale – Where a graph has uniform subdivisions for the x axis but the y axis is subdivided in such a way that a plot of the cumulative distribution appears as a straight line, it is said to be plotted using a probability scale. For example, the arithmetic probability scale describes a cumulative normal distribution.

Autocorrelation – In a time-series analysis autocorrelation is the internal correlation between observations often expressed as a function of the lag time between them.

Bias – A systematic error that may distort a statistical result in one direction. A biased estimator is one whose expected value does not equal the true value of the parameter being estimated.

Bootstrap – A nonparametric technique for estimating standard error of a statistic by repeated resampling (with replacement) from a sample. The technique treats a random sample of data as a substitute for the population and resamples from it a large number of times to produce sample bootstrap estimates and standard errors.

Chi-square test – A test of statistical significance based on the chi-square distribution. The chi-square statistic is obtained as the sum of all the quantities obtained by taking the difference between each observed and expected frequency, squaring the difference, and dividing this squared deviation by the expected frequency.

Correlation coefficient – An index used to measure correlation. It is also known as the Pearson product moment correlation coefficient. It is denoted by the letter r and its value ranges from -1 to +1. A value of +1 denotes that two sets are perfectly related in a positive sense and a value of -1 indicates that two sets are perfectly related in a negative sense. A value close to zero indicates that they are not linearly related.

Critical value – The theoretical value of a test statistic that leads to rejection of the null hypothesis at a given level of significance. Thus, in a statistical test, the critical value divides the rejection and the acceptance regions.

Distribution function – For any **random variable** *X*, the distribution function of *X*, denoted by *F*(*x*), is defined by $F(x) = P(X = x)$; that is, the distribution function is equal to the **probability** that a random variable assumes a value less than or equal to *x* for $-8 < x < 8$.

Durbin–Watson test – A procedure for testing independence of error terms in least squares regression against the alternative of autocorrelation or serial correlation. The test statistic *d* is a simple linear function of residual autocorrelations, and its value decreases as the autocorrelation increases.

Estimation – The process of using information from sample data in order to estimate the numerical values of unknown parameters in a population.

Gamma Function – A function generalizing the factorial expression for natural numbers, also known as Euler's Second Integral.

Goodness-of-fit test – A statistical procedure performed to test whether to accept or reject a hypothesized probability distribution describing the characteristics of a population. It is designed to ascertain how well the sample data conform to expected theoretical values. It involves testing the fit between an observed distribution of events and a hypothetical distribution based on a theoretical principle, research findings, or other evidence by means of a Pearson chi-square statistic or any other test statistic.

Independence – In probability theory, two events or observations are said to be independent when the occurrence of one event has no effect on the probability of occurrence of another event. Thus, two events are independent if the probability of occurrence of one is the same whether or not the other event has occurred.

Jackknife – A nonparametric technique for estimating standard error of a statistic. The procedure consists of taking repeated subsamples of the original sample of *n* independent observations by omitting a single observation at a time.

Kolmogorov–Smirnov tests – Nonparametric tests for testing significant differences between two cumulative distribution functions. The one-sample test is used to test whether the data are consistent with a given distribution function and the two-sample test is used to test the agreement between two observed cumulative distributions. The test is based on the maximum absolute difference between the two cumulative distribution functions.

Mann–Whitney *U* test – A nonparametric test for detecting differences between two location parameters based on the analysis of two independent samples. The test statistic is formed by counting all the bivariate pairs from the two samples in which one sample value is smaller than the other. It is equivalent to the Wilcoxon rank-sum test.

Maximum likelihood estimation – A method of estimation of one or more parameters of a population by maximizing the likelihood or log-likelihood function of the sample with respect to the parameter(s). The maximum likelihood estimators are functions of the sample observations that make the likelihood function greatest. The procedure consists of computing the probability that the particular sample statistic would have occurred if it were the true value of the parameter. Then for the estimate, we select the particular value for which the probability of the actual observed value is greatest. Maximum likelihood estimates are determined by using methods of calculus for maximization and minimization of a function.

Method of moments – A method of estimation of parameters by equating the sample moments to their respective population values. It is generally applicable and provides a fairly simple method for obtaining estimates in most cases. The method, however, yields estimators that, in certain cases, are less efficient than those obtained by the method of maximum likelihood.

Parametric methods – These are statistical procedures that are based on estimates of one or more population parameters obtained from the sample data. Parametric methods are used for estimating parameters or testing hypotheses about population parameters.

Partial autocorrelation – An autocorrelation between the two observations of a time series after controlling for the effects of intermediate observations.

Quantiles – A general term for the (n – 1) partitions that divide a frequency or probability distribution into n equal parts. In a probability distribution, the term is also used to indicate the value of the random variable that yields a particular probability.

Rank correlation – A nonparametric method for assessing association between two quantitative variables. A rank correlation is interpreted the same way as the Pearson product moment correlation coefficient. However, a rank correlation measures the association between the ranks rather than the original values. Two of the most commonly used methods of rank correlation are Kendall’s tau and Spearman’s rho.

Recurrence Interval (Return Period) – The average interval in years between two events of equal magnitude.

Resampling – The technique of selecting a sample many times and computing the statistic of interest with reweighted sample observations. Some commonly used resampling techniques include bootstrap, jackknife, and their variants.

Spearman’s rho (*?*) – A correlation coefficient between two random variables whose paired values have been replaced by their ranks within their respective samples or which are based on rank order measured on an ordinal scale. It provides a measure of the linear relationship between two variables. This measure is usually used for correlating variable(s) measured with rank-order scores.

Stochastic model – A mathematical model containing random or probabilistic elements.

Unbiased estimator – An estimator whose expected value or mean equals the true value of the parameter being estimated. Thus, an unbiased estimator on the average assumes a value equal to the true population parameter.

Uncertainty – A term denoting the lack of certainty inherent in a random phenomenon.

Wald–Wolfowitz run test – A nonparametric test for testing the null hypothesis that the distribution functions of two continuous populations are the same.