Title	Examination of Flood Estimation Techniques in the Irish Context
Author(s)	Das, Samiran
Publication Date	2010-07-07
Item record	http://hdl.handle.net/10379/1688

Examination of Flood Estimation Techniques in the Irish Context

Samiran Das

Department of Engineering Hydrology

National University of Ireland Galway

A thesis submitted for the degree of

PhilosophiæDoctor (PhD)

Supervisor: Professor C. Cunnane

-

Abstract

This thesis deals with the determination of the most suitable methods of finding the flood magnitude-return period relationship for Irish flood data using the annual maximum statistical model. Data from approximately 200 gauging stations in the Republic of Ireland, from the archives of OPW, EPA and ESB, were available. The research work is primarily based on the data of 115 of these gauging sites.

As a preliminary step to building statistical models for frequency analysis, flood flow characteristics are examined with a view to determining the regional behaviour of flood flow statistics, selecting appropriate statistical distributions to describe the flood data and examining the seasonal aspect of flooding. Most floods occur during the winter half of the year with some notable floods occurring during summer, especially in August. Because of Ireland's humid climate the range of variation of flow values from year to year, as measured by the coefficient of variation, is quite small by international standards. Likewise, the skewness of flow series is modest. While no single statistical distribution can be considered to be "best" at all locations in the context of at-site analysis it has been found that both the Extreme Value Type 1 (Gumbel) and the lognormal distributions provide reasonable models for the majority of stations.

Flood estimation (Q_T) by the Index Flood method using the Region of Influence (ROI) approach is investigated. The members of the pooling group in the ROI scheme are chosen with the help of a Euclidean distance or similarity measure, d_{ij}. Tests have been carried out on the effect on the estimated value of growth factor (X_T) of catchment area, peat coverage, lakiness (as measured by FARL), geographical location and period of record, where the periods fall between the 1950s and the 2000s. None of these attributes were

judged to be of sufficient influence to warrant special provision. Also, tests were carried out on the effectiveness of different combinations of catchment descriptors in the definition of d_{ij} and the most effective descriptors were found to be AREA, SAAR and BFI. The GEV distribution is recommended for use in the Index Flood method, except in cases where it implies an upper bound, where then EV1 is recommended.

Homogeneity is examined in the context of the estimation of X_T because a homogeneous pooling group of sites is required to minimise the error in estimating X_T. Tests based on Monte Carlo simulation were conducted to assess how successful a region of influence method of identifying pooling group membership is in selecting groups that qualify as being homogeneous. The results show that even with a carefully considered selection procedure, it is not certain the pooling groups identified are perfectly homogeneous. As a compromise it is recommended that a group containing more than 2 values of L-coefficient of variation outside the 95% confidence limits for that variable should not be considered as being homogeneous.

The standard errors (se) associated with estimates of X_T and Q_T are investigated. The standard error of Q_T estimated by the index flood method is dominated by $se(Qmed)$. The $se(Q_T)$, expressed as a percentage, is found to vary only slightly with T. When Qmed is estimated from at-site data and X_T is estimated from a pooling group containing approximately 500 station years of data then $se(Q_T)$ is of the order of 5% to 10% Q_T regardless of the magnitude of the return period. If Qmed is estimated from a catchment descriptor based formula alone and X_T is estimated from a pooling group containing approximately 500 station years of data then $se(Q_T)$ is of the order of 36% Q_T.

The performance of pooling group based estimation is also investigated. Experiments using Monte Carlo simulation show that the size of a pooling group is not a significant factor for estimating flood quantiles provided the number of station years is more than 350. An increased heterogeneity decreases the advantage of pooled estimation over that of at-site estimation.

A heterogeneity measure (H1) less than 4.0 can render the pooled estimation to be preferable to that of at-site estimation for estimating extreme quantiles. The at-site estimation is preferred when the record lengths at the site concerned exceed 50.

Guidance is provided on the estimation of the design flood of required annual exceedance probability at both gauged and ungauged locations. A number of examples is presented in Chapter 7 which cover a range of difficult cases that can arise in practice. In some of these cases, the user may be recommended to consider using an at-site based estimate in preference to the generally recommended regional pooling based method.

Acknowledgements

The author would like to thank his supervisor, Professor C. Cunnane, for his constant encouragement and invaluable guidance throughout the work.

He also wishes to thank Professor K. M. O'Connor for his kind support and ever available assistance.

Acknowledgements are due to the members of staff and technicians of the Department of Engineering Hydrology, especially to Dr. Monomoy Goswami, Mr. Dermot McDermot, Mr. Edward Kilcullen and Ms. Carmel King for their sincere cooperation.

The financial support in the form of a fellowship made available by the Irish Office of Public Works is gratefully acknowledged.

Finally, the author wishes to thank his wife, Nila, for her encouragement and invaluable support.

Contents

List of Figures

List of Tables

Glossary

A^2	Anderson-Darling goodness-of-fit statistic
D	Kolmogorov-Smirnov goodness-of-fit statistic
Q_T	the T year flood
W^2	Cramer-von Mises goodness-of-fit statistic
X_T	the T year growth factor
Γ	gamma function
α	scale parameter in extreme value distributions
$\bar{Q}$,Qmean	mean of AM flood series
β	scale parameter (growth curve)
χ^2	chi-squared goodness-of-fit statistic
γ	Euler's constant
λ_1	1^{st} L-moment
λ_2	2^{nd} L-moment
λ_3	3^{rd} L-moment
μ	mean
μ_3	3rd central moment
σ	standard deviation
ξ	location parameter in extreme value distributions
d_{ij}	similarity distance measure
g	coefficient of skewness
k	shape parameter of Generalized Extreme Value distribution
k	shape parameter of Generalized Logistic distribution
t_2	sample L-CV value
t_3	sample L-skewness value
t_4	sample L-kurtosis value
y	reduced variate
%PEAT	% of land area covered by peat bogs
A,AREA	catchment area

A1	gauging station grade A1
A2	gauging station grade A2
abs.bias	absolute bias
AM	annual maximum
ARTDRAIN2	% of the catchment river network included in the Drainage Schemes
B	gauging station grade B
BFI	baseflow index derived from soils data
BFIHOST	baseflow index derived from HOST soils data (U.K)
CV or C_v	coefficient of variation
D	discordancy measure
DRAIND	drainage density
E	expected value
EPA	Environment Protection Agency
ESB	Electricity Supply Board
EV1	Extreme Value Type 1
f(Q)	probability density function
F(Q), F	cumulative distribution function
FAI	flow attenuation index
FARL	flood attenuation by reservoir and lake
FEH	Flood Studies Handbook
FPEXT	Extent of flood plains in a catchment
FSE	factorial standard error
FSR	Flood Studies Report
FSU	Flood Studies Update
GEV	Generalized Extreme Value (distribution)
GLO	Generalized Logistic (distribution)
GNO	Generalized Normal (distribution)
H-skewness	Hazen corrected skewness
H1	heterogeneity measure (using L-CV)
H2	heterogeneity measure (using L-CV and L-skewness)
L-CV	L-coefficient of variation
L-kurtosis	L-coefficient of kurtosis

L-skewness	L-coefficient of skewness
LN	2 parameter Log Normal (distribution)
ln	natural logarithm
LN3	3 parameter Log Normal (distribution)
LO	2 parameter Logistic (distribution)
log10	10 base logarithm
M,m	number of sites
M100	1^{st} PWM
M110	2^{nd} PWM
M120	3^{rd} PWM
N,n	record length
OPW	Office of Public Works
PUM	Pooled Uncertainty Measure (as defined in FEH)
PWM	Probability Weighted Moment
Q	flood peak discharge
Qmax	maximum flood on record
Qmed	median of AM flood series
Qmin	minimum flood on record
rmse	root mean square error
S1085	10-85% stream slope (m/km)
SAAR	standard annual average rainfall (mm)
se	standard error
T	return period in years
USWRC	placecountry-regionUnited States Water Resources Council
Var	variance
W	weighting term

1

Introduction

1.1 Background

Flood frequency analysis is concerned with the assessment of flood magnitudes of stated frequency or stated degree of rarity for use as input to the process of flood risk assessment and management. Flood risk assessment is needed in the design of flood relief and protection works and in the assessment of the safety of existing and planned elements of infrastructure whether they be domestic houses, commercial or industrial buildings, bridges and roads, railways or other vital pieces of infrastructure such as hospitals, electrical stations, petrol filling stations or water and waste-water works. No project can be guaranteed to be immune from flooding during its projected life. Flood proofing every project would be prohibitively costly. As a result, projects for which the consequences and cost of occasional flooding is modest may be tolerated . Calculation of the flood risk associated with such projects requires that the flood magnitude which needs to be exceeded to cause flooding has to have its associated probability of occurrence determined. It is these probabilities that are calculated with the help of flood frequency analysis.

The estimation of these probabilities is, however, difficult and is a long standing problem in flood hydrology. Government agencies from around the world have been engaged in the development of frequency analyses for their own purposes or those of the public so that a reliable method of flood frequency analysis can be applied such that a consistent approach is adopted for design purposes at the national level.

With the same intention, a research component entitled 'Flood Frequency Analysis'

was included when the Irish Office of Public Works announced a fully fledged research programme, called the Flood Studies Update Programme, in 2005. This programme having many research components, arose out of the report of the Flood Policy Review Group (2004), aimed at providing improved hydrological information for input to decision making process. This thesis is based on the research work carried out on that 'Flood Frequency Analyses' research component.

1.2 Historical review of flood frequency procedure

The application of probability theory in flood estimation procedures was introduced by Fuller (1914) who calculated floods of different return periods for catchments in the U.S. The practical value of his work was soon recognized by his co-worker, Hazen (1914). Commenting on Fuller's paper, Hazen (1914, p. 626) wrote:

> This is a most important paper, because, as far as the writer knows, it is the first attempt to apply the principles of probabilities to the flood problem. The writer has followed the author's work in detail, and believes that these methods are sound. As time goes on, data covering longer periods and more streams may change some of the numerical values; but the underlying idea of treating the recurrence of floods as a matter of probabilities, to be determined by an examination of the records of many streams, will stand.

Empirical frequency curves were used in Fuller's (1914) study but the idea was soon extended to the use of statistical distribution functions having a specified theoretical form to describe the actual frequency distribution of the floods. Initially, attention was concentrated on tackling skewness which is inherent in the flood data. Foster (1924) introduced the Pearson type 3 (P3) and Hazen (1932) introduced the Log normal (LN) distribution for describing the flood data. The log transformed flood data often resulted in a skewness close to zero and thereby allowed the lognormal distribution to be considered.

Gumbel (1941) brought the basis of analysis to a new level by applying extreme value theory. Using the findings of Fisher and Tippett (1928), Gumbel (1941) introduced the Extreme Value Type I distribution (EV1) to flood frequency analysis. Fisher and Tippett (1928) had shown that the maximum value, selected from a sample of size n

chosen from a parent distribution, distributed as one of three limiting forms (EV1, EV2 or EV3) as the size of the samples increases. Gumbel (1941) postulated that, in considering floods as each year of record constitutes a sample of $n = 365$ flow values, the maximum flow in such a year should be distributed as an extreme value variate. The application of this concept later helped increase the popularity of the annual maximum approach to flood frequency analysis.

Since then, numerous methods for calculating the best estimates of flood frequency have evolved and are reviewed Benson (1968) and Cunnane (1989) by among others.

Benson (1968) reported on a study by a working group formed by the US Water Resources Council (USWRC) that compared the most commonly used methods of flood frequency analysis. This group recommended, based on probability plots, that the flood of given frequency be estimated by fitting the Log Pearson Type 3 (LP3) distribution to the series of annual maximum floods and that all U.S. government agencies adopt this as their base method in order to achieve a uniform procedure for estimating design floods.

In a report for WMO, Cunnane (1989) presented a detailed review on various issues related to flood frequency analysis up to that time including statistical properties of observed flood series, the modelling problem, methods of quantile estimation and methods of choosing between distributions. This report also summarized a world wide survey of flood frequency methods up to mid 1980's. It recommended that flood estimates be based on joint use of at-site and regional data using an Index Flood method of quantile estimation with model parameters estimated by probability weighted moments (PWMs). The report also revealed that conventional goodness of fit tests are of little value in the context of choosing between statistical distributions and that the EV1 and LN were the most commonly used distributions worldwide. However, the report discouraged the use of LP3 in general for flood frequency procedures.

Over the last two decades one of the major developments of flood frequency analysis can be regarded as the development of L-moments (see the book by Hosking and Wallis, 1997) which are defined as linear combinations of PWMs. Based on L-moments, Hosking and Wallis (1997) introduced the regional L-moment algorithm to obtain index-flood estimators. They also introduced homogeneity tests and an estimator for the goodness of fit applicable to regional flood frequency analysis based on L-moments. Their contribution to flood hydrology based on L-moments was summarised in their concise

and lucid book (Hosking and Wallis, 1997). Another major development in relation to regional flood frequency analysis can be regarded as the Region of Influence (ROI) approach, developed by Burn (1990). The ROI technique involves the identification of a region of influence, i.e. a separate group of stations, for each gauging station in a region. The regional approach based on L-moments and the formation of a group of stations using the ROI approach are now well established methods and have been applied in many recent flood studies (Zrinji and Burn, 1994; Zrinji and Burn, 1996; FEH, 1999; Castellarin et al., 2001; Cunderlik and Burn, 2002; Shu and Burn, 2004; Merz and Bloschl, 2005; Cunderlik and Burn, 2006b; Gaal et al., 2008).

1.3 Objective of the study

The main objective of this frequency analysis study is to determine suitable methods of finding the flood magnitude (Q)- return period (T) relationship, often called the $Q-T$ relationship, for Irish flood data using the annual maximum statistical model.

A preliminary step in building statistical models for frequency analysis requires knowledge of the flood flow characteristics of a particular region. This includes descriptive statistics, examination of probability plots, examination of suitability of candidate distributions and seasonal analysis. The first two characteristics essentially lead to the identification of appropriate probability distributions for describing the flood data in a region. Examination of suitability of candidate distributions can be made based on tests on descriptive statistics which mainly examine whether the assumed parent distribution is capable of producing random samples having the same statistical characteristics as the observed annual maximum (AM) series. Although Ireland experiences fairly uniform rainfall throughout the year, there is nevertheless a seasonal aspect of flooding. Systematic investigation of all these characteristics is necessary to draw inferences about the suitability of statistical models for describing flood data in Ireland.

Pooling of flood data is often used to provide a framework for estimating design floods by the Index Flood method. In pooling analysis, flood data are pooled from gauging stations that possess similar hydrological behaviour to the at-site subject station. Design flood estimation with this approach involves derivation of a growth curve which shows the relation between X_T and the return period T where $X_T = Q_T/Q_I$ and

Q_I is the index flood at the site of interest. The Region of Influence (ROI) approach is used as an objective way of forming a pooling group for a site of interest. The ROI selects stations that are hydrologically similar to the subject site in order to form the pooling group for that subject site. Similarity is commonly measured by a Euclidean distance measure in catchment descriptor space. However, the effective identification of a pooling group in the ROI approach is governed by two important criteria: the choice of appropriate site descriptors as pooling variables and the size of a group in terms of the number of sites and station years included. Careful consideration is necessary as to which forms of catchment characteristics are to be used and how many station years need to be included in a ROI method of pooling analysis.

It has been observed that a number of Irish catchments contain special types of features, such as peat bogs, catchment storage and in many water courses in Ireland, arterial drainage work. It is important therefore to know how the flood frequency growth curve behaves on such catchment types. The period of record can also have an effect on flood frequency curves and efforts should be made to identify its effect by splitting long records into several decades to determine if there are any changes over time.

The index flood method of pooling analysis works well under the assumption that data at different sites in a pooling group follow the same distribution except for scale, i.e. the performance of a pooled estimation method strongly depends on the grouping of sites into homogeneous pooling groups. Therefore, pooling groups formed by the ROI technique should be examined in the context of homogeneity.

In practical cases, the homogeneity assumption of the index flood procedure is often violated to some extent, i.e some degree of heterogeneity may exist. Hence the reliability of such estimates should be considered taking heterogeneity into account. It is therefore necessary to indicate, for Irish conditions, the magnitude of standard error that can be expected in those estimates. A detailed evaluation of pooled analysis is also required in this context.

This study has the following objectives:

1. Examination of the characteristics of Irish flood data

2. Investigation of the pooling group based estimate of Q_T

3. Investigation of the effect of catchment type and period of record on pooled growth curve estimates

4. Examination of the homogeneity of selected pooling groups

5. Evaluation of the error associated with design flood estimates and

6. Guidelines for engineering practice in flood estimation in the Irish context

1.4 Outline of the thesis

Chapter 1 gives a brief introduction of the whole thesis including a summary of the objectives of this study and a short historical review of flood frequency procedures. More detailed introductions, specific to each chapter, appear at the start of each individual chapter.

Chapter 2 is a review chapter which can be regarded as an elaboration of the review presented in Chapter 1 in the context of the study objectives. It covers some selected approaches of flood frequency procedures, which includes reviews on flood frequency modelling, regionalization techniques and standard error estimation.

Chapter 3 describes the selected characteristics of Irish flood flows which includes descriptive statistics, examination of probability plots, examination of suitability of candidate distributions and seasonal analysis. Necessary comparisons are also made in relation to some statistics that are reported in the 1975 Flood Studies Report (FSR) for the Irish context. This chapter also provides a good account of goodness-of-fit tests using chi-squared and empirical distribution function (EDF) statistics.

Chapter 4 covers the pooled based estimate of Q_T using the ROI technique. The first part of this chapter is concerned with investigating the choice of the number of catchment similarity variables on the effectiveness of the region of influence distance measures. The second part is concentrated on investigating the effect of catchment type and period of record on the pooled growth curve.

Chapter 5 presents a graphical approach based on L-moments of coefficient of variation (L-CV) to examine homogeneity of the pooling groups which are formed using the ROI technique described in the previous chapter. This chapter also includes a system for review of pooling groups using selected catchment characteristics.

Chapter 6 reports on the standard error of quantiles obtained in the Irish context. This chapter also gives a detailed account of simulation experiments done to test the effectiveness of the pooling analysis and includes explicit conclusions. The necessary comparisons are also made with the results of the studies of Hosking and Wallis (1997), Stedinger and Lu (1995) and Lettenmaier et al. (1987).

Chapter 7 provides the guidelines for engineering practice in flood estimation in Ireland.

Chapter 8 presents a summary, the conclusions and the recommendations of the study.

1.5 Data sets used in different parts of the study

While data are available for 202 stations, all data were not used for every part of the study. A summary of the data sets used in different parts of the study are summarised in Appendix A.

2

Literature Review

2.1 Introduction

The risk of failure, r of an engineering structure is expressed in flood frequency form as follows (FSR, 1975, I, p. 111).

$$r = 1 - (1 - 1/T)^L \tag{2.1}$$

where L is the design life of the structure and T is the design return period of a flood magnitude which may cause the structure to fail.

In flood frequency analysis (FFA), a relationship between a flood of magnitude Q and its return period T is developed by statistical modelling of a time series of past flows. The primary models used in flood frequency analysis are, as in (FSR, 1975, I, chap. 2):

Annual Maximum model (AM)

Peaks over Threshold model (POT)

Time Series model (TS)

In an AM model, the annual maximum flood series is used as input. An annual maximum flood series is a sequence of the largest peak discharge of each year of record. The $Q_T - T$ relationship in an AM model has the form,

$$1 - F(Q_T) = 1/T \tag{2.2}$$

where, $F()$ is the cumulative frequency distribution of flood magnitude, Q.

In a POT model, a flow data series containing floods exceeding a particular threshold (q_0) is fitted with a continuous probability distribution. The flood events are modelled

by a discrete probability distribution, such as Poisson distribution, and the model is of the form:

$$1 - F\left(Q_T / Q_T \geq q_0\right) = 1/\lambda T \tag{2.3}$$

where λ is the number of peaks per year included in the POT series. According to Cunnane (1989, p. 3), the AM model is statistically more efficient than the POT model when $\lambda < 1.65$.

In the Time Series model, the flow hydrograph is considered to be a time series in which the flows are represented by a series of ordinates at equally spaced intervals of time, preferably in days. The $Q - T$ relationship is obtained from a long data series that is generated by a time series model such as an auto-regressive model which is fitted to the daily flows. The basic difference between the TS model the other two models, namely, AM and POT models, is that the AM and POT models consider the floods as events and disregard their actual location on the time scale whereas in TS models the correlation structure of the events along the time scale is considered.

This thesis is concerned only with the AM model. Data from at-site or data from several sites in a homogeneous pooling group can be incorporated into the model to derive the $Q_T - T$ relationship. The former is known as at-site flood frequency analysis and the later as pooled or regional flood frequency analysis.

The research works in this study are concentrated on some aspects of these flood frequency analyses, although special focus is concentrated on the pooled flood frequency analysis and the estimation of standard errors of quantiles derived using these analyses.

In the present chapter, the pertinent literature on those topics is reviewed.

2.2 AM flood frequency procedure

The AM flood frequency procedure focuses on outlining an appropriate mathematical form to model the underlying distribution of flood data and then estimating the parameters of this underlying distribution. In the last four decades the research efforts in this area have concentrated mainly on two distinct fields: (1) finding the methods for fitting to the observed floods and (2) investigating the predictive ability of the $Q_T - T$ relationships estimated by these methods. In order to secure a better fit to the observed floods, a probability distribution need to be selected together with a method of parameter estimation.

The method of moments (MOM) is one of the simplest and conventional parameter estimation techniques used in statistical hydrology but is biased in estimating higher moments such as skewness. Maximum likelihood (ML) is statistically the most efficient method in the case of large samples, although the calculations are not simple. After the introduction of PWM by Greenwood et al. (1979) and subsequently the development of L-moments by Hosking (see Hosking and Wallis, 1997) based on PWMs, L-moments has become the method of choice among research hydrologists as it is almost free of bias, easy to use and generally unaffected by outliers. It has been also shown that the method is as good as the ML method in the case of very small samples (Hosking et al., 1985).

In general there has been no consensus on the use of any particular distribution function (see Cunnane, 1989, chap. 3). Therefore the choice of the probability distribution is, to some extent, arbitrary. However there is a number of procedures available for selecting the best possible probability distribution to describe the available data sample. These procedures are categorised as tests of descriptive ability (Cunnane, 1987). Subjective inspection of the data on a probability plot or goodness of fit indices (Benson, 1968; FSR, 1975) have been traditionally used to judge the suitability of any particular distribution. Log Pearson Type 3 (LP3) was recommended by USWRC in United States based on an index derived from probability plots for AM model used while the GEV was recommended for UK/Ireland in the FSR (1975) based on several goodness of fit indices such as Chi-square and Kolmogorov-Smirnov. The FSR also recommended using an EV1 for a record length of size less than 25. Although the goodness of fit indices, as a selection criterion have been used for a long time in flood frequency analysis, they have been criticized for not being able to adequately discriminate between statistical distributions.

Other descriptive ability tests, such as moment ratio diagrams, condition of separation and regional behaviour of statistics, are often used to identify a candidate distribution for a region in question. These tests mainly examine whether the assumed parent distribution is capable of producing random samples having the same statistical characteristics as the observed AM series. The condition of separation, as noted by Matalas et al. (1975), is a characteristic of regional flood samples which distinguishes them from simulated samples derived from distributions commonly used in hydrology. The method which is based on skewness got considerable attention in the late 70's and

in the 80's. In an extensive review of a large number of commonly used distributions, Cunnane (1989) concluded that random samples from a few thick-tailed distributions such as Wakeby (Houghton, 1978), 2-component Extreme Value, TCEV (Rossi et al., 1984), and Generalized Logistic, GLO (Ahmad et al., 1988b) distribution, can display the same skewness behaviour as do observed regional AM data sets. Although these distributions satisfied this important reproductive criterion, their use (particularly for Wakeby and TCEV) in single site analysis is limited because the associated parameter estimates often have large standard errors. The use of these distributions is mainly recommended in regional analysis. In more recent flood studies in the U.K., GLO has been selected for use with U.K. flood data (FEH, 1999).

In recent years, the most commonly used descriptive abitily test is that using moment ratio diagrams. Moment ratio diagrams based on conventional moment-based skewness and kurtosis are regarded as having very poor discriminating capability (Cunnane, 1989, App. 3) but L-moment ratio diagrams are considered to be more reliable as a diagnostic tool. L-moment ratio diagrams have been used as part of the distribution selection process in numerous published studies (Vogel et al., 1993; Onoz and Bayazit, 1995; PEEL et al., 2001; Castellarin et al., 2001)

The primary research area on the predictive ability of the annual maximum models is in investigating the robustness (e.g. Kuczera, 1982) of the assumed model. A procedure is said to be robust if it yields good estimations of quantiles when the procedure's assumptions depart even slightly from reality. The most commonly used procedure of investigation is the use of simulation experiments of random samples from a suitably assumed parent distribution. Appendix B describes the Monte Carlo technique of generating random samples. These experiments enable the comparison of the performance of Q_T estimates between different procedures of estimation. The performance of the Q_T estimates is assessed using indicators such as bias, se and rmse, the expressions for which are given below.

$$bias = E\left(\widehat{X} - X\right) \tag{2.4}$$

$$abs.bias = E\left(\left|\widehat{X} - X\right|\right) \tag{2.5}$$

$$se = \left\{E\left(\widehat{X} - E\left(X\right)\right)^2\right\}^{1/2} \tag{2.6}$$

$$rmse = \left\{E\left(\widehat{X} - X\right)^2\right\}^{1/2} \tag{2.7}$$

where, X is the population value and $\widehat{X}$ the estimated value of the statistic.

2.3 Regional flood frequency analysis

It is widely accepted that a short annual flood series is inadequate for the estimation of design floods of large return periods. The inadequacies are often associated with the identification of the appropriate statistical distribution for describing the data. The bias of quantile estimators from 3-or more parameter parameter distributions is generally small, but the standard error of such estimates may be unacceptably large. On the other hand, 2-parameter distributions (EV1, LN) lead to reasonably small standard errors of estimates, but may be highly biased i.e. they provide less flexibility for fitting the data. Regionalization provides a framework to overcome this problem by helping in the determination of the shape of parent distributions, allowing only a measure of scale to be estimated directly from the at-site data.

At present, the index flood method (Dalrymple, 1960) is the most widely used regional flood frequency procedure (FSR, 1975; Hosking and Wallis, 1997; FEH, 1999; Castellarin et al., 2001; Brath et al., 2001; Sveinsson et al., 2001; Grover et al., 2002; Sveinsson et al., 2003; López, 2004; Gaal et al., 2008). Cunnane (1988) reviewed twelve different methods of regional flood frequency analysis including well known methods such as the USWRC method , different variants of index flood methods, Station year methods, Bayesian methods and the two-component extreme value (TCEV) method and he rated the index flood using a regional algorithm based on PWMs as the best one. He also recommended using either the Wakeby or GEV distribution when floods are estimated by the index flood method.

The use of PWMs in the index flood method was introduced by Wallis in 1980 (see Greis and Wood, 1981, p. 1169). The technique calculates the PWMs at each site in a region from the standardized annual flood data and then the weighted regional average dimensionless PWMs are used to compute the dimensionless average growth curve. To obtain an estimate of the T-year flood at a specific site, the dimensionless T-year growth curve X_T is multiplied by the at-site mean flow information, i.e. by the index flood.

This technique was subsequently adopted as a viable way to estimate design floods and was further studied by, among others, Greis and Wood (1981, 1983); Lettenmaier

et al. (1987); Stedinger and Lu (1995) and Hosking and Wallis (1997).The early evaluation of the approach (EV1/PWM) was given by (Greis and Wood, 1981, 1983) although their approach was criticized for not considering the uncertainty arising from the at-site mean flow estimate (Lettenmaier et al., 1987, p. 316).

Addressing those issues of the Greis and Wood (1981, 1983) studies, Lettenmaier et al. (1987) explored the performance of index flood estimators with regions that exhibit various degrees of heterogeneity. Different variants of the GEV distribution with the PWM estimation procedure are used to obtain the index flood quantile. The GEV/PWM index flood quantile estimator performed well and gave the smallest mean squared errors in comparison with other at-site or regional quantile estimators for mildly heterogeneous regions.

Modifying the assumptions about the underlying frequency distribution based on real world evidence, Stedinger and Lu (1995) performed a similar kind of experiment to that designed by Lettenmaier et al. (1987) and concluded that index flood quantile estimators perform better in arid regions where CV values are usually large.

Hosking and Wallis (1997), using the regional L-moment algorithm, assessed the accuracy of index-flood estimators taking into account the possibility of heterogeneity in the region, mis-specification of the frequency distribution and statistical dependence between observations at different sites. They concluded that regional analysis is generally preferable to single site analysis even in regions with moderate amounts of heterogeneity, inter-site correlations, and mis-specification of the frequency distribution.

The regional approach based on L-moments is now a well established method and has frequently been used in numerous flood frequency studies (FEH, 1999; Castellarin et al., 2001; Kjeldsen et al., 2001; Ouarda et al., 2001; Cunderlik and Burn, 2002; Merz and Bloschl, 2005; Cunderlik and Ouarda, 2006; Gaal et al., 2008).

2.4 Formation of a homogeneous group

Formation of a homogeneous group is not guaranteed by casual selection of pooling group members. Geographical location has been used for a long time as a basis to form a region but has been criticised for being arbitrary. During the last two decades, several other methodologies, mainly involving the use of similarity measures to select a group of sites, have been introduced in hydrology. In this context, similarity between sites is

defined in a multidimensional space of catchment-related characteristics or statistical properties.

Cluster analysis for grouping sites into homogeneous pooling groups, using site characteristics is one of them. Nathan and McMahon (1990) presented a detailed regionalization methodology that addresses the problems associated with the selection of an appropriate clustering technique, selection of catchment variables, definition of homogeneous regions, and the prediction of group membership for new catchments whose group membership is otherwise unknown. The most suitable technique that they identified uses multiple regressions to select and weight the most appropriate variables and then uses cluster analysis to derive preliminary groupings, finally applying a multi-dimensional plotting technique to investigate further and refine the preliminary groupings.

A more appropriate and meaningful way of solving the formation issue is the Region of Influence (ROI) approach, developed by Burn (1990) and later studied and adopted by FEH (1999). The technique involves the identification of a region of influence i.e. a separate pooling group for each gauging station in a region. The identification of a pooling group consists of selecting stations that are hydrologically similar to the site of interest. Similarity is quantified by a Euclidean distance measure obtain for the catchment descriptor space. Selection of stations for inclusion in the pooling group is based on distance measure values, the stations having the smallest distance measure values relative to the subject site being chosen. A practical way of choosing an appropriate size of a pooling-group is proposed by FEH (1999). FEH investigated a range of pooling group sizes and decided on adoption of the $5T$ rule, namely that the total number of station years of data to be included when estimating the T year flood should be at least $5T$. An advantage of the region of influence method is that in the estimation of a regional growth curve, each site can be weighted according to its closeness to the site of interest.

Several other well known methods of forming regions, e.g. the partitioning technique (Pearson, 1991), hierarchical regions (Fiorentino et al., 1987; Gabriele and Arnell, 1991) etc are reviewed in detail by Hosking and Wallis (1997).

2.5 Homogeneity test

A homogeneity test is used to assess whether a proposed group of sites is homogeneous or not. A homogeneous group of sites leads to a reduction in the error of quantile estimators which is the main goal of a regional frequency analysis.

Dalrymple (1960) proposed such a test on homogeneity when he introduced the index flood approach of flood frequency analysis. The test, based on the assumption of an EV1 distribution, compares the variability of 10-year flood estimates, Q_{10}, from each site in the region with that expected supposing the differences between stations to be due to sampling error.

Wiltshire (1986) had doubts over the performance of the test and proposed two alternatives: one is based on the inter-site variability of CV and the other one is based on whether the exceedance probabilities associated with the samples have a uniform distribution as measured using his R- statistic. Of the two, the R- statistic test is that recommended by him and he further evaluated the power of the test using Monte Carlo simulations based on the assumption of the EV1 distribution . The test is seen to improve with larger numbers of sites and longer record lengths.

Lu and Stedinger (1992) formulated a significance test of homogeneity, based on the variability of normalized at-site GEV flood quantiles (X_{10}) estimated by L-moments and demonstrated that this test is more powerful than that of Wiltshire's R-statistic test.

Subsequently, Fill and Stedinger (1995) analyzed the relative performance of three homogeneity tests using Monte Carlo simulations in the case of Gumbel distributed flood flows. These tests are Dalrymple's test (1960), the MoM-CV test which similar to test by Lu and Stedinger (1992) but instead of using X_{10}, CV is used as a test statistic and finally the test by Lu and Stedinger (1992). They concluded that the test by Lu and Stedinger (1992) performed better than other two tests and that its superiority is due to its L-moment estimation procedure which yields a smaller sampling variance and a sampling distribution closer to normality than the use of classical product moments.

Hosking and Wallis (1993, 1997) proposed two more homogeneity tests based on L-moment ratios such as using L-CV alone (H1) and using L-CV & L-skewness jointly (H2). Both tests measure the sample variability of the L-moment ratios among the samples in the pooling group and compare it to the variation that would be expected in

a homogeneous pooling group. The variation is estimated through repeated simulations of homogeneous regions with samples drawn from a four parameter kappa distribution whose parameters are estimated from L-CV, L-skewness and L-kurtosis of the region's data. They recommended using the H1 over the H2 statistic as they found that the heterogeneity based on L-CV has better power to discriminate between homogeneous and heterogeneous regions.

This conclusion was further confirmed by Viglione et al. (2007) who compared four homogeneity tests through the determination of the power associated with the tests using Monte Carlo simulation experiment. The first two of these tests are those of Hosking and Wallis (1997), who proposed H1 and H2, and the other two, introduced by the authors, are based on the k sample Anderson-Darling test and Durbin & Knott test. Viglione et al. (2007) concluded that the H2 as a homogeneity test is lacking power. They further concluded that the H1 test should be preferred when skewness is low while the Anderson-Darling test should be used for more skewed regions, preferably those with L-skewness greater than 0.23.

The Homogeneity test based on L-CV (H1), proposed by Hosking and Wallis (1997), is nowadays routinely used in regional analysis. However, they recommended that although the heterogeneity statistic has the structure of a significance test, it should not be used in this way and they further stated that (Hosking and Wallis, 1997, p. 70):

> Significance levels obtained from such a test would be accurate only under special assumptions: that the data are independent both serially and between sites, and that the true regional distribution is kappa. We need to define a heterogeneity measure for regions that may not satisfy these assumptions, so we prefer not to use heterogeneity measure as a significance test. A significance test is of doubtful utility anyway, because even a moderately heterogeneous region can provide quantile estimates of sufficient accuracy for practical purposes. Thus a test of exact homogeneity is of little interest.

2.6 Standard error estimation

The standard error of a flood estimate indicates the reliability of that estimate. The concept of standard error and bias of statistical estimates of Q_T was introduced by

Kimball (1949) for the EV1 distribution and by Kaczmarek (1957) for EV1 and other distributions although this information filtered only slowly into hydrological practice (see FSR, 1975, I, p. 100-104). Simulation based studies of standard errors were first reported by Nash and Amorocho (1966),and by Lowery and Nash (1970) while in the mid-1970s extensive simulation studies on bias and standard error were reported by Matalas et al. (1975).

Since then, it has been a common practice to report standard error of an estimate. Several studies reported it either by providing theoretical expressions or by simulation.

For instance, FSR (1975) reported expressions of the standard error of Q_T in both the EV1 and GEV cases. The expression for the EV1 is moment based while that for the GEV is maximum likelihood based using the Taylor expansion approach. The additional shape parameter in the GEV provides complexity in developing the expression which resulted in inapplicable expressions for extreme quantiles and large positive values of the shape parameter.

Lu and Stedinger (1992) provided standard error expressions for both the EV1 and GEV cases using L-Moment (=PWM) based estimates of Q_T . While the expression for the EV1 is an analytical one derived using the Taylor series expansion approach, that for the GEV is an empirical one in which the coefficients for selected T values are obtained from Monte Carlo simulations. They also found that the Taylor expansion approach of deriving expression for the GEV provided unsatisfactory results for extreme quantiles.

After the introduction of the use of PWM to the index flood method, several studies concentrated on giving theoretical expressions for the standard error (*se*) of different variants of index flood estimators. The *se* variants resulted from the use of different distributions; different parameter estimation techniques; as well as the inclusion of heterogeneity and inter- site dependence among sites in a region.

For instance, Rosbjerg and Madsen (1995) presented analytical expressions of *se* of the T-year event considering heterogeneity among the sites in a pooling group based on the EV1 distribution with model parameters estimated by the method of moments, whereas Stedinger and Lu (1995) presented analogous results based on the GEV distribution with parameters estimated using PWMs. In both studies, the mean annual flood is used as the index flood.

De Michele and Rosso (2001) presented an asymptotic expression for the GEV distribution for evaluating the variance of X_T for a homogeneous region. They combined this variance with the variance of index flood estimators to obtain an uncertainty model for evaluating the variance of at site flood estimators.

Kjeldsen and Jones (2006) derived analytical approximations on the basis of asymptotic theory for the variance of the quantile estimates for both gauged and ungauged sites: they used the GLO distribution with parameters estimated using the method of L-moments and used the median flood as the index flood. Although their study considered the implication of intersite dependence, but it did not consider the implication of heterogeneity among sites in a pooling group.

A limitation to the theoretical approaches is that a number of assumptions are needed to formulate these expressions, particularly in the case of approaches based on L-moments where these assumptions might not necessarily be fulfilled to a satisfactory degree when considering observed flood data.

For these reasons, Hosking and Wallis (1997) opted for an L-moment based parametric simulation procedure to assess the accuracy of index-flood estimators taking into account the possibility of heterogeneity in the pooling region and statistical dependence between observations at different sites. They found that moderate heterogeneity and moderate inter-site dependence have a very little effect on the variance of regional index-flood quantile estimators. However they showed that the main effect of heterogeneity is to introduce bias into the estimated quantiles.

3

Examination of characteristics of Irish Flood Flows

3.1 Introduction

A detailed knowledge of the flood flow characteristics of a particular region is essential to understand the behaviour of the flood data which leads to building statistical models for flood frequency analysis. A previous comprehensive study on Irish flood data which was reported in FSR (1975) had been done in the early 1970s. 112 Irish stations with a total of some 1700 record years were used in that study. At present, many more data are available because i) OPW stations have longer records and ii) some new EPA stations are now accessible, and the analysis of these long records can be especially valuable in the understanding of the flood flow characteristics. Therefore, it is now opportune to conduct a comprehensive new study on Irish flood flows.

This chapter provides some of the characteristics of Irish flood flows including descriptive statistics, examination of probability plots, examination of suitability of candidate distributions and seasonal analyses. Descriptive statistics provide simple summaries about the flood samples and form a guide in deciding the next steps in the analyses. Summarizing the data in the form of a probability plot usually accompanies the calculation of descriptive statistics, both forming part of what is called 'Exploratory Data Analysis'. Selecting the most appropriate forms of distribution is a very important step in flood frequency analysis because it is often assumed for a certain region that flood data are coming from a population of a known distribution. Although Ire-

land experiences rainfall throughout the year there is nevertheless a seasonal aspect to flooding and this is also examined during the Exploratory Data Analysis.

3.2 Data

Annual Maximum flood data from 202 gauging stations have been obtained from the OPW. Based on rating curve and water level measurement reliability, as indicated by OPW, these data fall into the following categories:

- 45 Grade A1 stations
- 70 Grade A2 stations
- 70 Grade B stations and a further
- 16 stations (comprising 8 A1, 5 A2 with the remaining 3 classified as B) that all have pre-and post-drainage records.

The annual flood data from these 202 stations vary in record length from 8 to 65 years, having associated catchment areas varying from 9.2 to 7980.4 km^2. The mean and median record lengths are 37 and 35 years respectively. The conclusions of the study are based mainly on the results obtained from the Grade A1 and A2 stations. The B Grade stations are analysed and displayed but no general deductions are made from them. Among 115 A1 and A2 stations, 5 stations (Station no. 25001, 25002, 25003, 25004 and 25005) in the Mulkear catchment are not considered for the inference analysis because of embankments which contain 4 out of 5 annual maximum flows but which are overtopped by the larger flows causing damping of the larger flow magnitudes which leads to extremely convex downwards probability plots. The distribution of the remaining 110 A1 and A2 stations according to record length is shown in Table 3.1.

Table 3.1: Distribution of stations according to record length

sample size	number of stations
<11	0
11-20	13
21-30	34
31-40	18
41-50	25
51-60	20
60>	0

3.3 Descriptive statistics

Flood data are summarised by a number of statistics. These statistics are used as input for the calibration of statistical models. Such summary statistics, especially those expressed in dimensionless form, can also be used for comparison of flood frequency behaviour on different catchments and for drawing inferences about the suitability of probability distributions for describing the flood data. This section of the chapter describes these measures and their values as calculated from the available data series.

Descriptive statistics based on both conventional moments and the more recently developed L-moments (Hosking and Wallis, 1997) are used. L-moments, analogous to conventional moments, are defined as linear combinations of the probability weighted moments (PWMs) which were introduced by Greenwood et al. (1979).

In flood hydrology, the most useful statistics relate to

1. the overall magnitude of the flood data as conveyed by their mean, μ (equal to 1^{st} L-moment, λ_1) or median, Qmed

2. the scale or spread of the data as measured by standard deviation, σ or 2^{nd} L-moment, λ_2

3. the skewness or the lack of symmetry of the data, as measured by the 3^{rd} central moment μ_3 or by the 3^{rd} L-moment, λ_3

Dimensionless forms of (2) above are expressed as

- the coefficient of variation (CV), C_v

- L-CV, t_2

and of (3) above as

- Skewness, g
- L-skewness, t_3

These are defined in Table 3.2 in terms of the basic population quantities. Formulae for the calculation of these statistics from sample data are also provided in Table 3.2.

The Hazen skewness g_h is the ordinary skewness g multiplied by the factor $\{1 + 8.5/N\}$. This correction factor described by Hazen (1932) was shown by Wallis et al. (1974) to provide an almost unbiased estimator of skewness for lognormal and EV1 distributions and indeed for mildly skewed distributions in general.

For comparative studies these dimensionless measures of scale and asymmetry are useful for the detection of typical behaviours among the several series in the data set. Dimensionless measures of magnitude are not feasible although the 'specific flow value', Q/(catchment size), is sometimes used for comparison purposes. Other dimensionless ratios such as Qmax/Qmed, Qmax/Qmean or Qmed/Qmean can also be compared between catchments although it should be noted that the ratios Qmax/Qmean and Qmax/Qmed are very much dependent on the period of record from which they were derived and have high sampling variability.

The basic statistical descriptors, i.e. mean, median, maximum, CV, Hazen skewness, L-CV , L-skewness and L-kurtosis, as well as the ratios Qmax/Qmean, Qmax/Qmed, Qmax/Area and Qmed/Qmean, are shown in tabular form in Tables 3.7, 3.8 and 3.9 for A1, A2 and B grade stations respectively. The quantity specific median flood, i.e. Qmed/Area, for the A1, A2 and B stations, is mapped in Figure 3.1. The estimated CV and skewness values are also shown at location throughout Ireland in map format in Figures 3.2 and 3.3 respectively.

3.3.1 Regional behaviour of statistics

The values relating to CV and skewness are of particular interest because they convey information about the flood regime of a region which may be used for inter-comparison with other regions. The average values of these statistics, in addition to those of the ratios of L-CV/CV, L-skewness/H-skewness, Qmed/Qmean, Qmax/Qmean for the A1

Table 3.2: Calculation of moments and their dimensionless ratios

Definitions based on ordinary moments	
Population quantities	Sample estimates
μ = Mean	$\bar{Q} = \frac{1}{N}\sum_{i=1}^{N} Q_i$
σ = Standard deviation	$\hat{\sigma} = \sqrt{\sum_{i=1}^{N} \frac{(Q_i-\overline{Q})^2}{N-1}}$
μ_3 = 3rd Central Moment	$\hat{\mu}_3 = \frac{N}{(N-1)(N-2)}\sum_{i=1}^{N}(Q_i - \bar{Q})^3$
$CV = \frac{\sigma}{\mu}$ = Coefficient of variation	$\hat{C}v = \frac{\hat{\sigma}}{\bar{Q}}$
$g = \frac{\mu_3}{\sigma^3}$ = Coefficient of Skewness	$\hat{g} = \frac{\hat{\mu}_3}{\hat{\sigma}^3}$ $\hat{g}_H = \hat{g}\left\{1 + \frac{8.5}{N}\right\}$ =Hazen's Unbiased skewness
Definitions based on Probability Weighted Moments (PWMs)	
Population quantities	Sample estimates
M100 = 1st PWM	$\hat{M}_{100} = \bar{Q}$ = sample mean
M110 = 2nd PWM	$\hat{M}_{110} = \frac{1}{N}\sum_{i=1}^{N}\frac{(i-1)}{(N-1)}Q_{(i)}$
M120 = 3rd PWM	$\hat{M}_{120} = \frac{1}{N}\sum_{i=1}^{N}\frac{(i-1)(i-2)}{(N-1)(N-2)}Q_{(i)}$
M130 =4th PWM	$\hat{M}_{130} = \frac{1}{N}\sum_{i=1}^{N}\frac{(i-1)(i-2)(i-3)}{(N-1)(N-2)(N-3)}Q_{(i)}$ where Q(i) is the ith smallest value
λ_1 = 1st L-Moment	$L1 = \hat{M}_{100}$
λ_2 = 2nd L-Moment	$L2 = 2\hat{M}_{110} - \hat{M}_{100}$
λ_3 = 3rd L-Moment	$L3 = 6\hat{M}_{120} - 6\hat{M}_{110} + \hat{M}_{100}$
λ_4 = 4th L-Moment	$L4 = 20\hat{M}_{130} - 30\hat{M}_{120} + 12\hat{M}_{110} - \hat{M}_{100}$
τ_2 =L-CV	$t_2 = \frac{L2}{L1}$
τ_3 =L-skewness	$t_3 = \frac{L3}{L2}$
τ_4 =L-kurtosis	$t_4 = \frac{L4}{L2}$

and A2 stations are shown in Table 3.3. Arithmetic average values, such as defined by eq (3.1) are used here. Weighted average values relating to the sites' record lengths, as defined by eq (3.3), are often used in regional behaviour analysis but are not considered in this study because they are not needed for the comparisons being made.

The sampling distributions of the CV and skewness statistics are shown in graphical form also, in Figures 3.4 and 3.5 while Figures 3.6 and 3.7 show the relationships between CV and L-CV and also between L-skewness and H-skewness. While CV and skewness values vary between gauging stations, some of this variation is due to sampling variability the remaining may be due to true differences between the gauging stations themselves.

Table 3.3: Average values of some statistics for 43 A1 and 67 A2 gauging stations

St. Category	CV	L-CV	g_h	t_3	$\frac{L-CV}{CV}$	$\frac{g_h}{t_3}$	$\frac{Qmed}{Qmean}$	$\frac{Qmax}{Qmean}$
A1	0.31	0.16	1.15	0.17	0.52	0.15	0.94	1.85
A2	0.25	0.14	0.84	0.11	0.56	0.13	0.98	1.65
A1& A2	0.28	0.15	1	0.14	0.54	0.14	0.96	1.75

From Table 3.3, it is seen from the combined average A1 and A2 values that the ratios of L-CV/CV and L-skewness(t_3)/H-skewness(g_h) are 0.54 and 0.14 respectively. These values compare very well with the corresponding values obtained by simulation which are shown in Table 3.4. The simulation uses EV1 as the parent distributions for the population L-CV values in the range 0.1 to 0.2; this being the range relevant for Ireland and record lengths in the range 20-50. Alternative values of these ratios, namely 0.52 and 0.13, can be obtained from the slopes of the lines on Figures 3.6 and 3.7 respectively, where individual values of L-CV are plotted against individual values of CV and likewise for L-skewness and H-skewness.

The average values of Qmed/Qmean are 0.94 for the A1 stations and 0.98 for the A2 stations. These are just below the values expected in data from normal distributions whereas if the data were from an EV1 distribution this ratio would have a value around 0.96 as shown in Table 3.4. The inverse of 0.96 $\approx$ 1.05, which is similar to the average value of Qmean/Qmed reported in FSR (1975, I, p. 132).

The average values of Qmax/Qmean are 1.85 for the A1 stations and 1.65 for the A2 stations. These are extremely low by European standards. The largest value of

Table 3.4: Average values of some statistics derived using simulation based on EV1 distributions

L-CV	L-CV/CV			L-skewness/H-skewness			Qmed/Qmean			Qmax/Qmean		
	N=20	N=35	N=50	N=20	N=35	N=50	N=20	N=35	N=50	N=20	N=35	N=50
0.1	0.56	0.55	0.55	0.15	0.15	0.16	0.97	0.97	0.98	1.43	1.51	1.57
0.15	0.56	0.55	0.55	0.15	0.16	0.16	0.96	0.96	0.96	1.65	1.77	1.85
0.2	0.56	0.55	0.55	0.15	0.16	0.16	0.95	0.94	0.95	1.87	2.02	2.13

Qmax/Qmean among the A1 and A2 stations of this study was found to be 6.1 but this was from a short- record containing an unusual outlier. While similar values are found in UK data, there are only two other Irish stations where this statistic exceeds 3. The corresponding values derived from simulation are shown in Table 3.4. This shows that the statistic is sensitive to record length and population CV. Nevertheless, the values obtained for the A1 and A2 stations are comparable with the corresponding values obtained from the simulation experiments.

3.3.2 Comparison with the FSR study

In the FSR (1975), the regional estimates of CV and skewness reported were based on only ordinary moments. Three estimates, namely, arithmetic average, weighted average and pooled average (denoted here as FSR pooled average), are used to calculate regional value of CV and g. The equations for these estimates are shown below, as in (FSR, 1975, I, p. 123-125). The chosen regional values for CV and g in the FSR were based on eqs (3.5) and (3.6) respectively.

$$\bar{Cv} = \frac{1}{m}\sum_{j=1}^{m} Cv_j \tag{3.1}$$

$$\bar{g} = \frac{1}{m}\sum_{j=1}^{m} g_j \tag{3.2}$$

$$weighted \quad Cv = \sum_{j=1}^{m} (n_j - 1)\, Cv_j / \sum_{j=1}^{m} (n_j - 1) \tag{3.3}$$

$$weighted \quad g = \sum_{j=1}^{m} (n_j - 1)\, g_j / \sum_{j=1}^{m} (n_j - 1) \tag{3.4}$$

$$FSR \quad pooled \quad Cv = \left\{ \frac{\sum_{j=1}^{m} (n_j - 1) \, Cv_j^2}{\sum_{j=1}^{m} (n_j - 1)} \right\}^{1/2} \tag{3.5}$$

$$FSR \quad pooled \quad g = \frac{C3}{(pooledCv)}^{3} \tag{3.6}$$

$$in \quad which \quad C3 = \sum_{j=1}^{m} \left(\frac{n_j}{n_j - 2} \cdot S3_j \right) / \sum_{j=1}^{m} (n_j - 1) \tag{3.7}$$

$$and \qquad S3_j = \sum_{i=1}^{n_j} \left(\frac{Q_{ij} - \bar{Q}_{\cdot j}}{\bar{Q}_{\cdot j}} \right)^3 \tag{3.8}$$

Using these equations, the regional estimates of CV and g are also calculated based on the 110 A1 and A2 stations. The three quantities are tabulated in Table 3.5 alongside those reported in the FSR for Ireland. The reported values in FSR were based on 63 stations of record lengths > 15 years. The average record length of these 63 records was < 20 years whereas the average lengths used in the present study is 37 years.

Table 3.5 shows that the three measures of regional CV do not differ widely within either study and between studies but it is noted that the pooled average CV for this study is a little lower than the value reported in the FSR. On the other hand, the pooled average g in the present study is much higher than that in the FSR and it is higher than the other estimates in both studies. The estimate is obtained using the pooled average CV in its equation and it is extremely sensitive to the estimate of the pooled CV. If a value of the pooled average CV of 0.30 is used, instead of the value 0.28 obtained in this study, the estimate would have a value of 1.93. The weighted average estimate of g is a bit higher than that of the FSR while the arithmetic average estimate was not reported in the FSR.

Table 3.5: Different estimators of regional estimate of Cv and g

Regional estimator	FSR		This Study	
	Cv	g	Cv	g
Arithmetic avg.	0.27	not reported	0.28	0.75
Weighted avg.	0.28	0.66	0.26	0.72
FSR Pooled avg.	0.3	1.63	0.28	2.48

Table 3.6: Relative RMSE, absolute bias and SE of three estimators of regional skewness

Estimator	Relative RMSE (%)	Relative Abs.Bias (%)	Relative SE (%)
Avg. Skewness	22	21	7
Weighted avg. Skewness	30	29	8
FSR Pooled avg. Skewness	43	40	17

Having noted that the FSR pooled average g value is very sensitive; a simulation is carried out to assess which estimate of regional skewness is the most robust. This has been done based on the 110 A1 and A2 stations using the EV1 distribution. Random samples are drawn from EV1 populations for each stations taking the sample size as being equal to the length of the observed historical record at the site and taking the parameters as these estimated from the L-moments at the site. The results of the simulation are reported in Table 3.6. The rmse, abs. bias and se of the pooled estimate are much higher than those of the arithmetic average and weighted average estimates. The arithmetic average estimate gives the lowest rmse, abs. bias and se among the three. Hence the regional estimate of g based on eq (3.6) may give misleading results and should not be used in regional analysis.

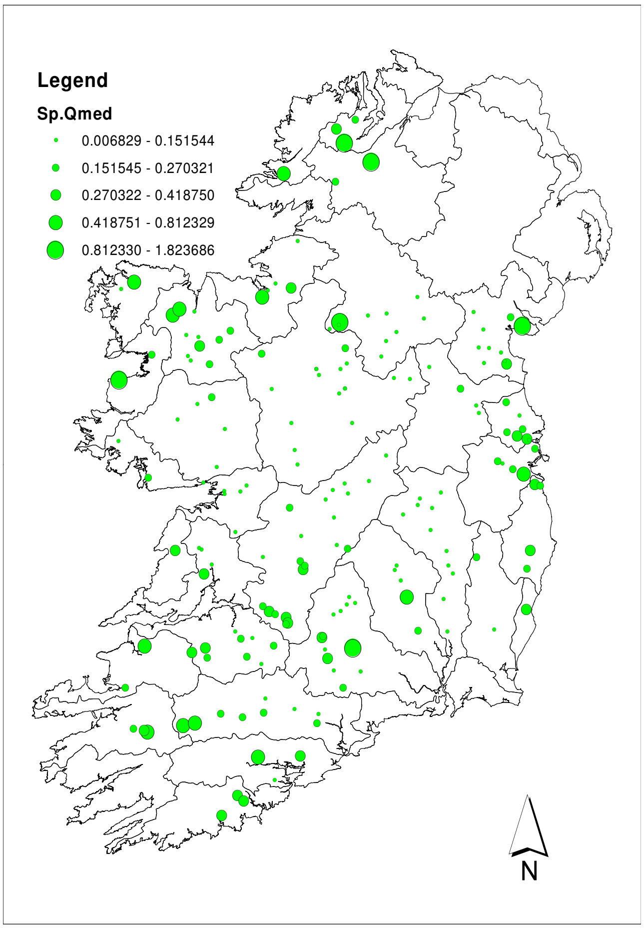


Figure 3.1: Specific Qmed ($m^3/s/km^2$) in map format - The figure shows Specific Qmed ($m^3/s/km^2$) at the A1 , A2 and B category station locations.

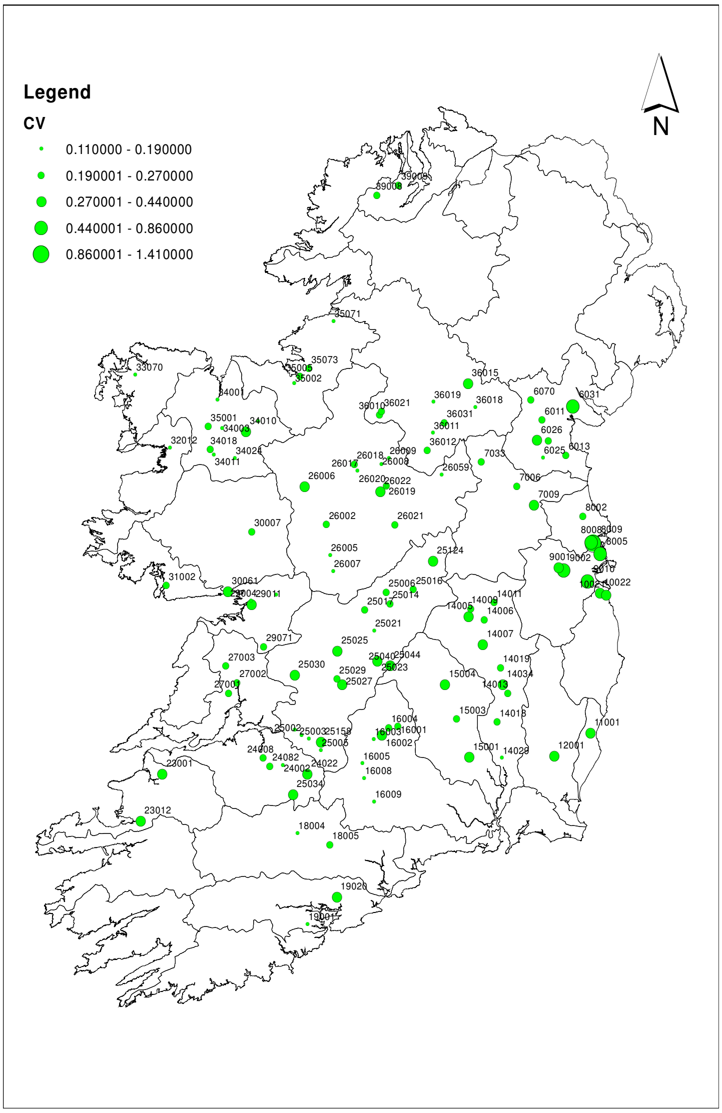


Figure 3.2: CV in map format - The figure shows CV at the 110 A1 and A2 category station locations.

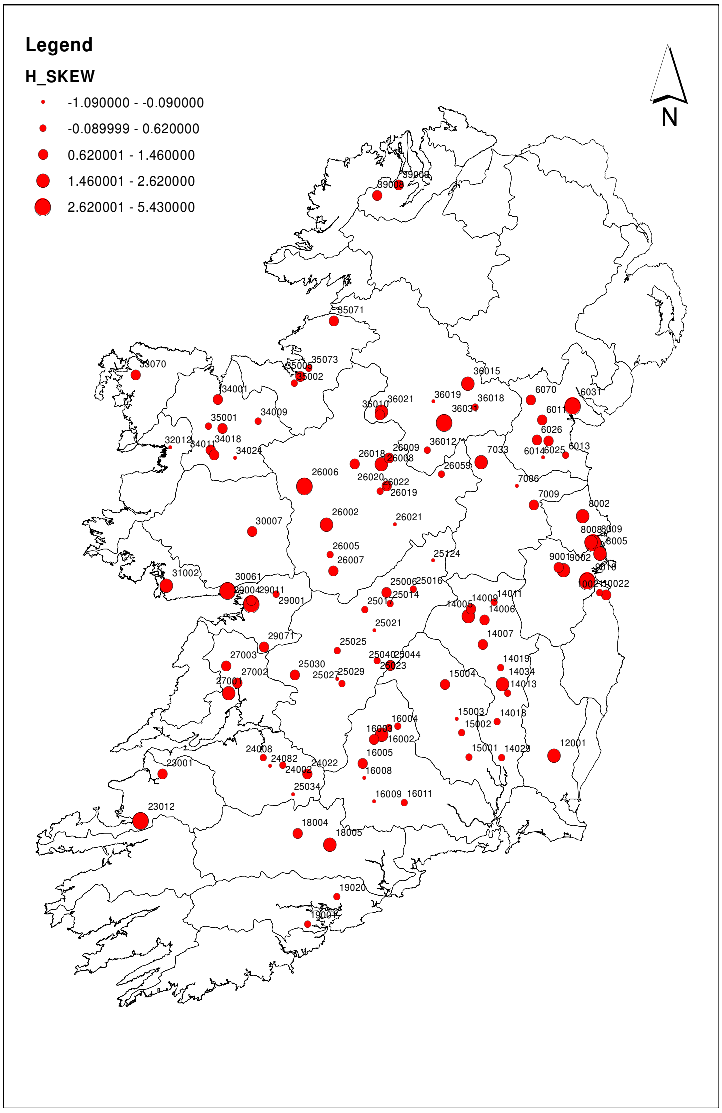


Figure 3.3: Hazen skewness in map format - The figure shows Hazen skewness at the A1 , A2 category station locations.

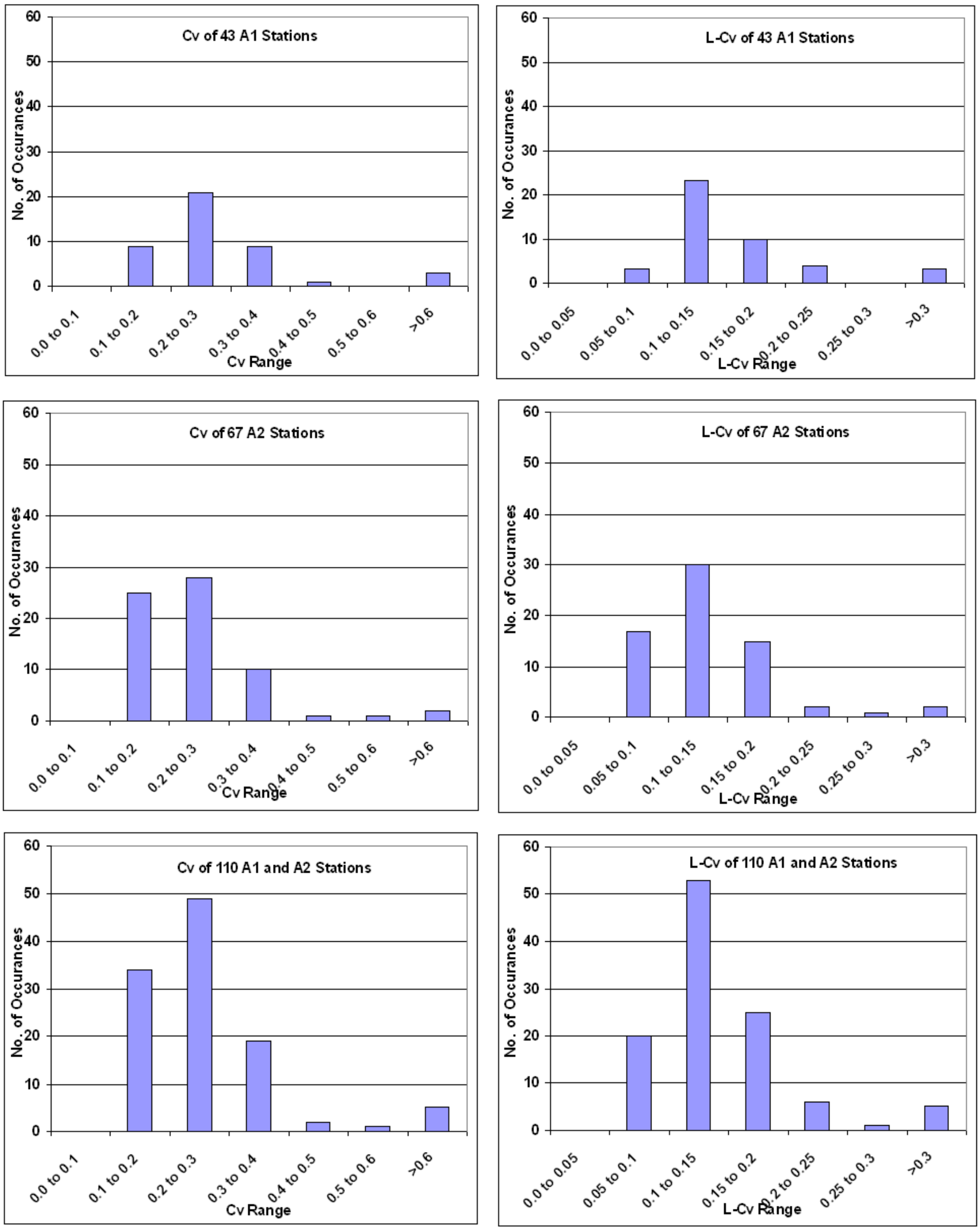


Figure 3.4: Histograms of CV and L-CV values - The figure shows the histograms of A1 , A2 and A1&A2 combined stations.

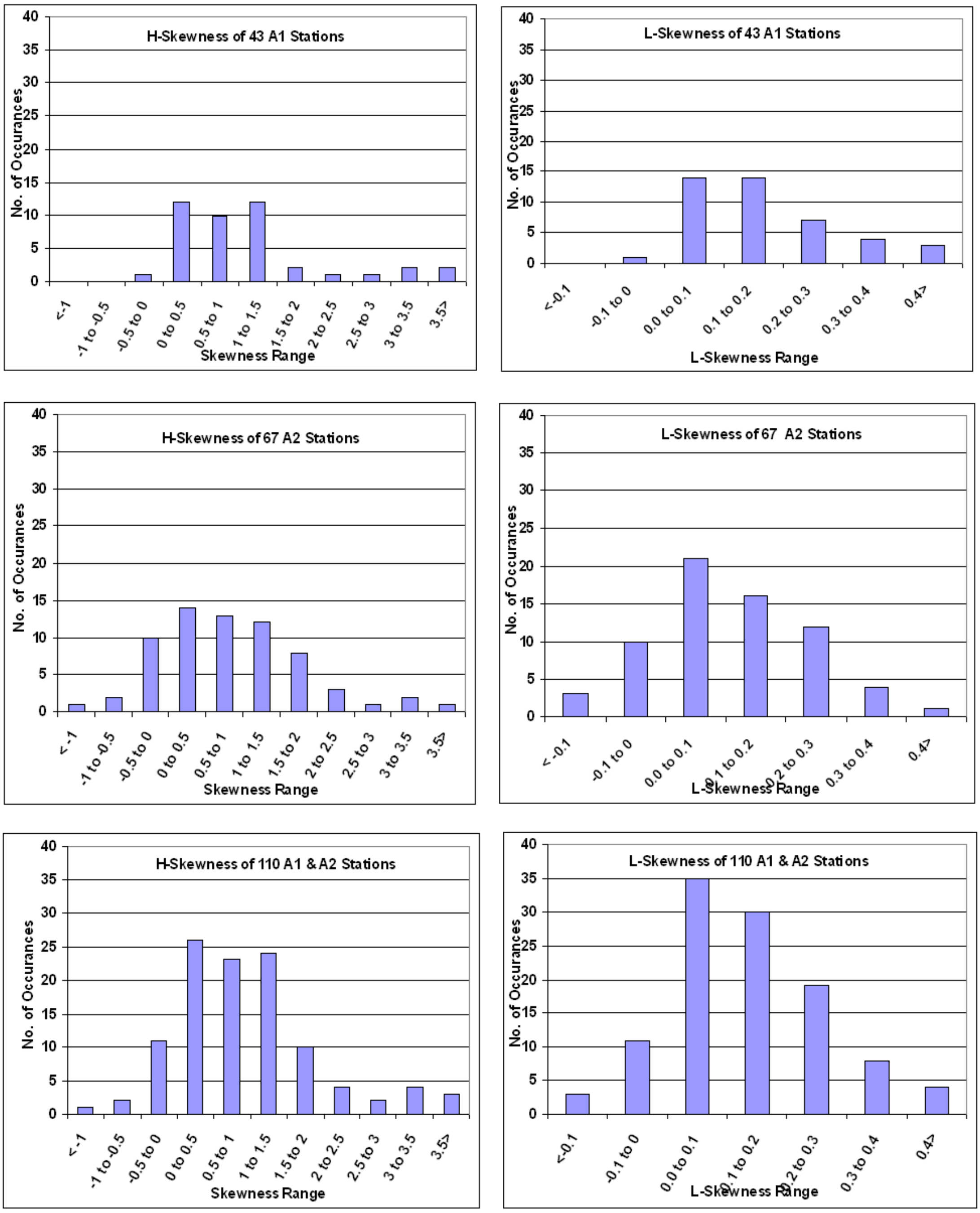


Figure 3.5: Histograms of Hazen skewness and L-skewness values - The figure shows the histograms of A1 , A2 and A1&A2 combined stations.

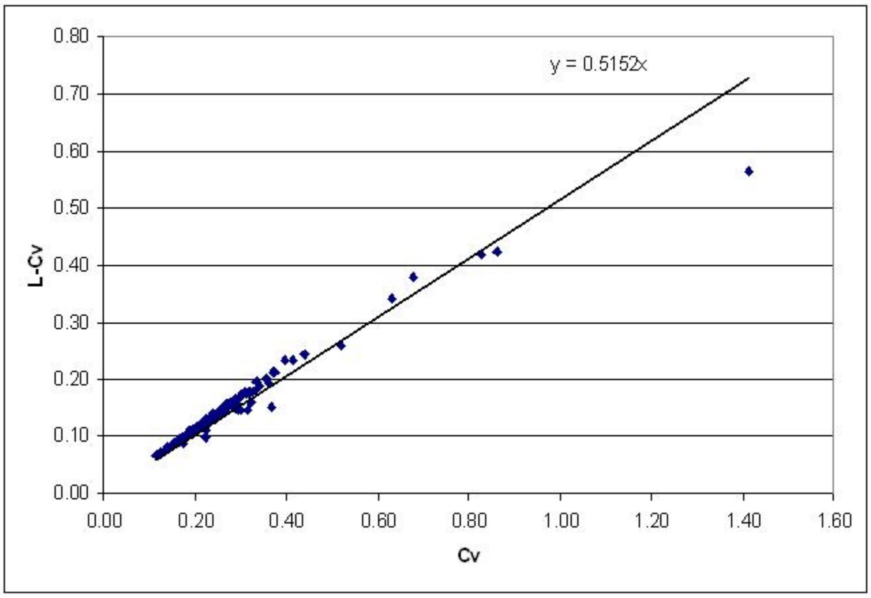


Figure 3.6: L-CV Vs CV for the A1 and A2 stations - Note if intercept is included the relevant trendlines are $y = 0.4426x + 0.0268$

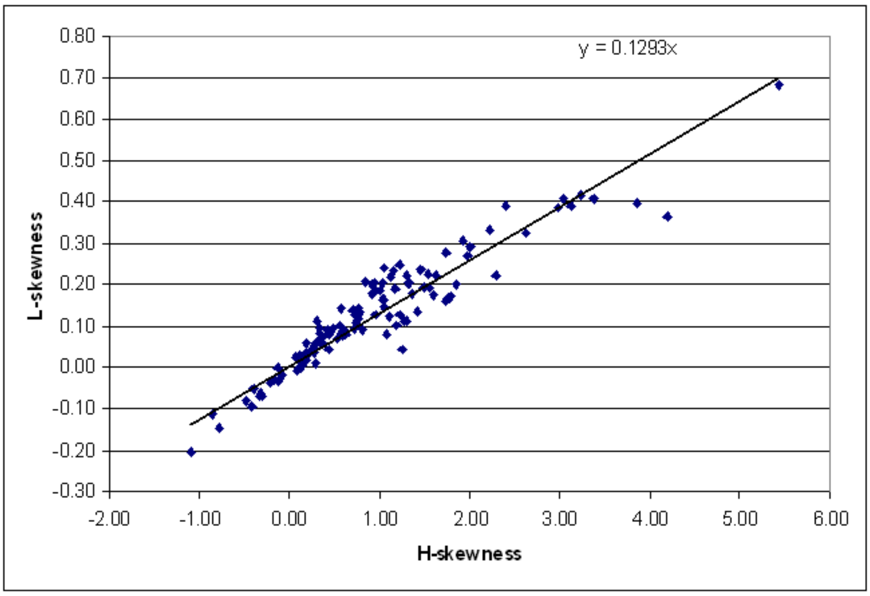


Figure 3.7: L-skewness Vs H-skewness for the A1 and A2 stations - Note if intercept is included the relevant trendlines are $y = 0.1225x + 0.0145$

Table 3.7: Summary statistics for the 45 A1 stations

Station No., River & Station Name	Area	N	Qmean	Qmed	Qmax	H-skew	CV	L-CV	L-skew	L-kur	Qmax/ Qmean	Qmax/ Qmed	Qmax/ Area	Qmed/ Qmean
06011 RIVER FANE @ MOYLES MILL	234	48	15.86	15.39	26.36	0.81	0.2	0.11	0.09	0.07	1.66	1.71	0.11	0.97
06013 RIVER DEE @ CHARLEVILLE WEIR	307	30	27.81	27.37	41.84	0.1	0.27	0.16	0.02	0.01	1.5	1.53	0.14	0.98
06014 RIVER GLYDE @ TALLANSTOWN	270	30	22.56	21.46	39.4	1.31	0.27	0.15	0.22	0.12	1.75	1.84	0.15	0.95
06025 RIVER DEE @ BURLEY BRIDGE	176	30	18.32	18.69	23.57	-0.33	0.16	0.09	-0.07	0.19	1.29	1.26	0.13	1.02
06026 RIVER GLYDE @ ACLINT BRIDGE	144	46	13.87	12.3	24.12	1.05	0.32	0.18	0.24	0.09	1.74	1.96	0.17	0.89
06070 RIVER MUCKNO L. @ MUCKNO	153.5	24	13.32	13.19	20.93	0.78	0.25	0.14	0.14	0.14	1.57	1.59	0.14	0.99
07009 RIVER BOYNE @ NAVAN WEIR	1610	29	162.64	134.8	297.6	0.93	0.38	0.21	0.21	0.11	1.83	2.21	0.18	0.83
08002 RIVER DELVIN @ NAUL	37	20	5.62	5.32	8.96	1.98	0.21	0.11	0.27	0.17	1.59	1.68	0.24	0.95
08009 RIVER WARD @ BALHEARY	62	11	10.38	6.59	53.6	5.43	1.41	0.56	0.68	0.65	5.17	8.13	0.86	0.64
09001 RIVER RYEWATER @ LEIXLIP	215	48	38.71	35.46	91.5	1.17	0.44	0.24	0.19	0.15	2.36	2.58	0.43	0.92
09002 RIVER GRIFFEEN @ LUCAN	37	24	7.24	5.4	23.7	2.4	0.83	0.42	0.39	0.25	3.28	4.39	0.64	0.75
09010 RIVER DODDER @ WALDRON'S BRIDGE	95.2	19	70.15	48	269	3.24	0.86	0.42	0.42	0.3	3.83	5.6	2.83	0.68
10021 RIVER SHANGANAGH @ COMMON'S ROAD	30.9	24	7.87	7.36	14.3	0.95	0.37	0.21	0.19	0.07	1.82	1.94	0.46	0.94
10022 RIVER CABINTEELY @ CARRICKMINES	10.4	18	3.84	3.85	6.89	0.36	0.4	0.23	0.06	0.06	1.79	1.79	0.66	1
14006 RIVER BARROW @ PASS BRIDGE	1096	51	83.76	80.52	137.38	1.46	0.2	0.11	0.24	0.21	1.64	1.71	0.13	0.96
14007 RIVER STRADBALLY @ DERRYBROCK	115	25	16.94	16.2	29.3	1.12	0.3	0.17	0.22	0.07	1.73	1.81	0.25	0.96
14011 RIVER SLATE @ RATHANGAN	163	26	12.07	12.3	18.7	0.19	0.25	0.14	0.02	0.16	1.55	1.52	0.11	1.02
14018 RIVER BARROW @ ROYAL OAK	2415	51	141.83	147.98	216.07	0.22	0.24	0.14	0.04	0.06	1.52	1.46	0.09	1.04
14019 RIVER BARROW @ LEVITSTOWN	1660	51	103.46	102.41	162.86	0.61	0.24	0.14	0.09	0.13	1.57	1.59	0.1	0.99
16011 RIVER SUIR @ CLONMEL	2173	52	234.52	223	389	0.42	0.3	0.17	0.09	0.05	1.66	1.74	0.18	0.95
25003 RIVER MULKEAR @ ABINGTON	397	51	69.45	68.98	92.79	0.05	0.15	0.08	0	0.11	1.34	1.35	0.23	0.99
25006 RIVER BROSNA @ FERBANE	1207	52	86.77	81.91	147.21	0.71	0.25	0.14	0.14	0.18	1.7	1.8	0.12	0.94
25014 RIVER SILVER @ MILLBROOK BRIDGE	165	54	17.67	17.25	27.03	0.57	0.23	0.13	0.1	0.13	1.53	1.57	0.16	0.98
25017 RIVER SHANNON @ BANAGHER	7980	55	413.25	407.68	596.51	0.18	0.2	0.12	0.04	0.08	1.44	1.46	0.07	0.99
25023 RIVER LITTLE BROSNA @ MILLTOWN	116	52	12.14	11.22	20.05	0.57	0.29	0.16	0.14	0.06	1.65	1.79	0.17	0.92
25025 RIVER BALLYFINBOY @ BALLYHOONEY	160	31	10.15	10.18	17.4	0.57	0.29	0.17	0.08	0.15	1.71	1.71	0.11	1
25027 RIVER OLLATRIM @ GOURDEEN BRIDGE	118	43	23.32	22.1	40.46	0.27	0.28	0.16	0.05	0.13	1.73	1.83	0.34	0.95
25030 RIVER GRANEY @ SCARRIFF BRIDGE	279	48	43.8	40.64	87.04	1.01	0.32	0.18	0.18	0.13	1.99	2.14	0.31	0.93
25158 RIVER BILBOA @ CAPPAMORE	116	18	47.66	43.88	75.09	0.17	0.29	0.17	0.05	0.13	1.58	1.71	0.65	0.92
26006 RIVER SUCK @ WILLSBROOK	182	53	26.57	24.23	70.06	3.87	0.37	0.15	0.4	0.41	2.64	2.89	0.38	0.91
26007 RIVER SUCK @ BELLAGILL BRIDGE	1184	53	91.75	88.15	147.84	1.05	0.19	0.11	0.16	0.16	1.61	1.68	0.12	0.96
26008 RIVER RINN @ JOHNSTON'S BRIDGE	292	49	23.68	22.94	41.02	1.49	0.19	0.1	0.19	0.19	1.73	1.79	0.14	0.97
26019 RIVER CAMLIN @ MULLAGH	260	51	22.34	21.18	37.03	1.17	0.25	0.14	0.23	0.12	1.66	1.75	0.14	0.95
26020 RIVER CAMLIN @ ARGAR BRIDGE	128	32	11.21	11.27	15.59	0.16	0.19	0.11	0.03	0.07	1.39	1.38	0.12	1.01
26059 RIVER INNY @ FINNEA BRIDGE	249	17	12.98	12.2	16.7	0.31	0.18	0.1	0.11	0.17	1.29	1.37	0.07	0.94
27002 RIVER FERGUS @ BALLYCOREY	562	51	34.22	32.6	59.76	1.37	0.23	0.12	0.18	0.22	1.75	1.83	0.11	0.95
29001 RIVER RAFORD @ RATHGORGIN	119	40	14.17	13.46	19.66	0.34	0.19	0.11	0.09	0.1	1.39	1.46	0.17	0.95
29011 RIVER DUNKELLIN @ KILCOLGAN BRIDGE	354	22	31.94	28.89	66.52	3.37	0.3	0.14	0.41	0.32	2.08	2.3	0.19	0.9
31002 RIVER CASHLA @ CASHLA	72	26	12.89	12.16	21.1	1.92	0.24	0.13	0.31	0.18	1.64	1.74	0.29	0.94
33070 RIVER CARROWMORE L. @ CARROWMORE	90	28	7.9	7.67	11.97	1.19	0.16	0.09	0.1	0.18	1.52	1.56	0.13	0.97
34018 RIVER CASTLEBAR @ TURLOUGH	93	27	11.5	11.28	17.33	0.93	0.2	0.11	0.18	0.03	1.51	1.54	0.19	0.98
36010 RIVER ANNALEE @ BUTLERS BRIDGE	774	50	66.56	66.8	106.62	1.05	0.22	0.12	0.15	0.22	1.6	1.6	0.14	1
36012 RIVER ERNE @ SALLAGHAN	263	47	14.22	14.12	21.63	0.21	0.22	0.13	0.03	0.1	1.52	1.53	0.08	0.99
36015 RIVER FINN @ ANLORE	175	33	23.14	22.08	49.99	2.62	0.32	0.16	0.32	0.3	2.16	2.26	0.29	0.95
36018 RIVER DRONMORE @ ASHFIELD BRIDGE	233	50	15.84	16.25	24.43	0.43	0.18	0.1	0.04	0.1	1.54	1.5	0.1	1.03

Table 3.8: Summary statistics for the 70 A2 stations

Station No., River & Station Name	Area	N	Qmean	Qmed	Qmax	H-skew	CV	L-CV	L-skew	L-Kur	Qmax/ Qmean	Qmax/ Qmed	Qmax/ Area	Qmed/ Qmean
06031 FLURRY @ CURRALHIR	45.3	18	13.58	11.7	35.8	3.12	0.52	0.26	0.39	0.32	2.64	3.06	0.79	0.86
07006 MOYNALTY @ FYANSTOWN	176	19	26.73	27.93	34.05	-1.09	0.21	0.12	-0.2	0.07	1.27	1.22	0.19	1.05
07033 BLACKWATER @ VIRGINIA	129	25	14.93	14.62	26.58	1.74	0.24	0.13	0.16	0.27	1.78	1.82	0.21	0.98
08005 SLUICE @ KINSALEY	10.1	18	3.04	2.5	7.81	1.54	0.68	0.38	0.23	0.19	2.57	3.13	0.77	0.82
08008 BROADMEADOW @ B.MEADOW	110	25	44.55	40.9	123.69	1.74	0.63	0.34	0.28	0.16	2.78	3.02	1.12	0.92
12001 SLANEY @ SCARAWALSH	1036	50	169.5	157	399	1.59	0.36	0.19	0.18	0.17	2.35	2.54	0.39	0.93
14005 BARROW @ PORTARLINGTON	398	48	40.81	38.27	80.42	2.01	0.29	0.15	0.29	0.22	1.97	2.1	0.2	0.94
14009 CUSHINA @ CUSHINA	68	25	6.69	6.79	11.19	1.42	0.23	0.13	0.14	0.24	1.67	1.65	0.16	1.02
14013 BURRIN @ BALLINACARRIG	154	50	16.54	16.05	26.32	0.37	0.26	0.15	0.07	0.06	1.59	1.64	0.17	0.97
14029 BARROW @ GRAIGUENAMANAGH	2762	47	162.54	160.74	206.21	0.18	0.14	0.08	0.06	0.07	1.27	1.28	0.07	0.99
14034 BARROW @ BESTFIELD	2060	14	137.3	125	247	2.23	0.32	0.17	0.33	0.17	1.8	1.98	0.12	0.91
15001 KINGS @ ANNAMULT	443	42	89.39	88.75	151	0.14	0.28	0.16	0	0.08	1.69	1.7	0.34	0.99
15002 NORE @ JOHN'S BR.	1605	35	211.98	197	393	0.59	0.31	0.18	0.08	0.09	1.85	1.99	0.24	0.93
15003 DINAN @ DINAN BR.	298	50	143.58	150.76	187.52	-0.78	0.2	0.11	-0.15	0.09	1.31	1.24	0.63	1.05
15004 NORE @ McMAHONS BR.	491	51	38.96	37.28	74.96	0.96	0.31	0.17	0.13	0.15	1.92	2.01	0.15	0.96
16001 DRISH @ ATHLUMMON	140	33	15.65	15.66	24.49	0.3	0.22	0.12	0.01	0.12	1.56	1.56	0.17	1
16002 SUIR @ BEAKSTOWN	512	51	55.4	52.66	123.88	1.79	0.3	0.16	0.17	0.19	2.24	2.35	0.24	0.95
16003 CLODIAGH @ RATHKENNAN	246	51	31.17	29.98	45.72	0.85	0.18	0.1	0.21	0.05	1.47	1.53	0.19	0.96
16004 SUIR @ THURLES	236	48	22.17	21.37	34.89	0.53	0.2	0.11	0.07	0.11	1.57	1.63	0.15	0.96
16005 MULTEEN @ AUGHNAGROSS	87	30	23.11	21.79	34.31	1.33	0.18	0.1	0.2	0.13	1.48	1.57	0.39	0.94
16008 SUIR @ NEW BRIDGE	1120	51	90.66	92.32	110.91	-0.32	0.13	0.07	-0.06	0.05	1.22	1.2	0.1	1.02
16009 SUIR @ CAHIR PARK	1602	52	159.29	162.21	206	-0.41	0.17	0.1	-0.1	0.05	1.29	1.27	0.13	1.02
18004 BALLYNAMONA @ AWBEG	324	46	30.96	31.2	52.7	1.25	0.17	0.09	0.04	0.33	1.7	1.69	0.16	1.01
18005 FUNSHION @ DOWNING BRIDGE	363	50	56.69	53.05	109.73	1.63	0.27	0.14	0.22	0.18	1.94	2.07	0.3	0.94
19001 OWENBOY @ BALLEA UPPER	106	48	15.87	15.42	22.03	0.48	0.17	0.09	0.09	0.12	1.39	1.43	0.21	0.97
19020 OWENNACURRA @ BALLYEDMOND	75	28	24.63	22.4	38.7	0.12	0.34	0.2	0.03	0.04	1.57	1.73	0.52	0.91
23001 GALEY @ INCH BRIDGE	196	45	97.39	99.05	210.07	1.22	0.33	0.18	0.13	0.18	2.16	2.12	1.07	1.02

23012 LEE @ BALLYMULLEN	60	18	16.87	15.66	31.74	2.98	0.29	0.15	0.39	0.33	1.88	2.03	0.53	0.93
24002 CAMOGUE @ GRAY'S BRIDGE	231	27	24.06	23.49	35.21	0.32	0.19	0.11	0.06	0.17	1.46	1.5	0.15	0.98
24008 MAIGUE @ CASTLEROBERTS	805	30	120.96	119.13	194.86	0.28	0.26	0.15	0.05	0.09	1.61	1.64	0.24	0.98
24022 MAHORE @ HOSPITAL	39.7	20	9.83	9.8	20.5	1.12	0.41	0.23	0.12	0.16	2.09	2.09	0.52	1
24082 MAIGUE @ ISLANDMORE	764	28	135.47	140.01	206.35	-0.22	0.26	0.15	-0.04	0.13	1.52	1.47	0.27	1.03
25001 MULKEAR @ ANNACOTTY	646	49	133.95	132.88	178.58	-0.19	0.16	0.09	-0.02	0.16	1.33	1.34	0.28	0.99
25002 NEWPORT @ BARRINGTONS BRIDGE	223	51	61.15	62.64	74.64	-0.65	0.16	0.09	-0.14	0.06	1.22	1.19	0.33	1.02
25005 DEAD @ SUNVILLE	190	46	28.73	29.63	33.42	-0.98	0.11	0.06	-0.19	0.1	1.16	1.13	0.18	1.03
25016 CLODIAGH @ RAHAN	274	42	23.04	22.57	36.14	0.62	0.22	0.13	0.08	0.16	1.57	1.6	0.13	0.98
25021 LITTLE BROSNA @ CROGHAN	493	44	28.03	28.58	35.8	-0.13	0.14	0.08	-0.03	0.07	1.28	1.25	0.07	1.02
25029 NENAGH @ CLARIANNA	301	33	54.12	56.48	74.06	-0.09	0.24	0.14	-0.02	-0.02	1.37	1.31	0.25	1.04
25034 L. ENNELL TRIB @ ROCHFORT	12	24	1.5	1.48	2.21	-0.48	0.29	0.17	-0.08	0.12	1.47	1.49	0.18	0.99
25040 BUNOW @ ROSCREA	30	20	3.78	3.59	6.29	1.32	0.27	0.15	0.2	0.18	1.67	1.75	0.21	0.95
25044 KILMASTULLA @ COOLE	98.9	33	25.38	22.7	45.75	1.23	0.34	0.19	0.25	0.15	1.8	2.02	0.46	0.89
25124 BROSNA @ BALLYNAGORE	254	18	12.79	13.65	22.5	-0.31	0.36	0.2	-0.07	0.23	1.76	1.65	0.09	1.07
26002 SUCK @ ROOKWOOD	626	53	56.98	56.56	104.87	2.29	0.22	0.11	0.22	0.29	1.84	1.85	0.17	0.99
26005 SUCK @ DERRYCAHILL	1050	51	92.8	93.21	135.94	0.27	0.18	0.1	0.03	0.14	1.46	1.46	0.13	1
26009 BLACK @ BELLANTRA BRIDGE	97	35	13.66	13.22	18.76	0.93	0.16	0.09	0.2	0.11	1.37	1.42	0.19	0.97
26018 OWENURE @ BELLAVAHAN	118	49	9.19	8.95	13.98	0.77	0.2	0.11	0.13	0.11	1.52	1.56	0.12	0.97
26021 INNY @ BALLYMAHON	1071	30	65.88	66.34	92.51	-0.86	0.25	0.14	-0.11	0.19	1.4	1.39	0.09	1.01
26022 FALLAN @ KILMORE	950	33	6.64	6.49	11.06	0.47	0.3	0.17	0.09	0.05	1.67	1.71	0.01	0.98
27001 CLAUREEN @ INCH BRIDGE	48	30	20.65	20.1	31.7	1.55	0.2	0.11	0.19	0.25	1.53	1.58	0.66	0.97
27003 FERGUS @ COROFIN	168	48	24.01	22.92	40.5	0.71	0.24	0.13	0.09	0.22	1.69	1.77	0.24	0.95
29004 CLARINBRIDGE @ CLARINBRIDGE	123	32	11.39	11.3	14.77	0.77	0.15	0.08	0.14	0.08	1.3	1.31	0.12	0.99
29071 L.CUTRA @ CUTRA	123.5	26	16	15.7	24.3	0.74	0.23	0.13	0.11	0.19	1.52	1.55	0.2	0.98
30007 CLARE @ BALLYGADDY	458	31	61.93	62.98	95.98	1.3	0.2	0.11	0.11	0.2	1.55	1.52	0.21	1.02
30061 CORRIB ESTUARY @ WOLFE TONE BRIDGE	3111	33	274.97	247.97	601.59	3.04	0.32	0.15	0.41	0.38	2.19	2.43	0.19	0.9
32012 NEWPORT @ NEWPORT WEIR	138.3	24	30.06	29.85	36.6	-0.12	0.12	0.07	0	0.18	1.22	1.23	0.26	0.99
34001 MOY @ RAHANS	1911	36	174.76	174.61	286.56	1.08	0.19	0.1	0.08	0.21	1.64	1.64	0.15	1
34003 MOY @ FOXFORD	1737	29	180.42	178	282	1.26	0.17	0.09	0.11	0.25	1.56	1.58	0.16	0.99

34009 OWENGARVE @ CURRAGHBONAUN	113	33	28.37	27.48	38.58	0.43	0.17	0.1	0.08	0.16	1.36	1.4	0.34	0.97
34011 MANULLA @ GNEEVE BRIDGE	144	30	18.8	18.73	26.05	0.78	0.16	0.09	0.1	0.23	1.39	1.39	0.18	1
34024 POLLAGH @ KILTIMAGH	128	28	20.7	20.8	24.7	-0.39	0.12	0.07	-0.05	0.14	1.19	1.19	0.19	1
35001 OWENMORE @ BALLYNACARROW	299	29	30.52	31.16	46.04	0.1	0.21	0.12	-0.01	0.19	1.51	1.48	0.15	1.02
35002 OWENBEG @ BILLA BRIDGE	90	34	51.78	50.48	69.37	0.07	0.17	0.1	0.03	0.08	1.34	1.37	0.77	0.97
35005 BALLYSADARE @ BALLYSADARE	642	55	77.78	75.42	132.71	1.04	0.26	0.14	0.2	0.14	1.71	1.76	0.21	0.97
35071 L.MELVIN @ LAREEN	247.2	30	26.95	26.29	37.91	0.76	0.18	0.1	0.12	0.2	1.41	1.44	0.15	0.98
35073 L.GILL @ LOUGH GILL	384	30	54.81	54.05	78.4	0.34	0.22	0.12	0.08	0.14	1.43	1.45	0.2	0.99
36019 ERNE @ BELTURBET	1501	47	89.6	89.95	119.43	-0.16	0.18	0.1	-0.03	0.06	1.33	1.33	0.08	1
36021 YELLOW @ KILTYBARDEN	23	27	24.96	23.37	43.57	1.85	0.22	0.12	0.2	0.22	1.75	1.86	1.89	0.94
36031 CAVAN @ LISDARN	52	30	6.85	6.45	13.7	4.19	0.22	0.1	0.37	0.39	2	2.12	0.26	0.94
39008 LEANNAN @ GARTAN BRIDGE	78	33	28.34	28.18	43.88	0.73	0.26	0.15	0.13	0.11	1.55	1.56	0.56	0.99
39009 FERN O/L @ AGHAWONEY	207	33	45.91	45.72	76.67	1.03	0.26	0.14	0.16	0.15	1.67	1.68	0.37	1

Table 3.9: Summary statistics for the 70 B stations

Station, River & St. Name	Area	N	Qmean	Qmed	Qmax	H-skew	CV	L-CV	L-skew	L-Kur	Qmax/ Qmean	Qmax/ Qmed	Qmax/ Area	Qmed/ Qmean
1041 DEELE @ SANDY MILLS	45.3	32	85.08	82.61	147.16	0.22	0.29	0.17	0.02	0.13	1.73	1.78	3.25	0.97
1055 MOURNE BEG @ M.BEG WEIR	10.8	9	2.92	2.7	4.72	1.03	0.38	0.23	0.16	0.02	1.61	1.75	0.44	0.92
6021 GLYDE @ MANFIELDSTOWN	321	50	21.54	21.5	33	0.45	0.24	0.13	0.09	0.09	1.53	1.53	0.1	1
6030 BIG @ BALLYGOLY	10.2	30	20.58	10.03	122	3.42	1.61	0.61	0.68	0.48	5.93	12.17	11.96	0.49
6033 DEE @ CONEYBURROW BR.	57.4	25	27.88	18.6	92.3	2.49	0.84	0.41	0.46	0.27	3.31	4.96	1.61	0.67
8003 B.DMEADOW @ FIELDSTOWN	76.2	18	26.88	22.55	110	4.21	0.87	0.39	0.36	0.34	4.09	4.88	1.44	0.84
8007 B.MEADOW @ ASHBOURNE	34	15	9.88	8.24	18.79	0.89	0.49	0.29	0.18	-0.01	1.9	2.28	0.55	0.83
8011 NANNY @ DULEEK ROAD BR.	181	23	31	32.22	45.41	-0.93	0.25	0.14	-0.15	0.2	1.46	1.41	0.25	1.04
8012 STREAM @ BALLYBOGHIL	22.1	19	4.21	4.35	8.15	-0.54	0.58	0.33	-0.09	0.11	1.93	1.87	0.37	1.03
9035 CAMMOCK @ KILLEEN ROAD	54.7	9	12.04	11.7	27.8	2.75	0.6	0.33	0.33	0.23	2.31	2.38	0.51	0.97
10002 AVONMORE @ RATHDRUM	233	47	88.19	83.49	266.64	2.85	0.43	0.21	0.27	0.31	3.02	3.19	1.14	0.95
10028 AUGHRIM @ KNOCKNAMOHILL	204.1	16	56.69	46.95	102	1.57	0.38	0.21	0.3	0.08	1.8	2.17	0.5	0.83
11001 OWENAVORRAGH @ BOLEANY	148	33	49.85	47.17	120.7	3.03	0.33	0.16	0.24	0.3	2.42	2.56	0.82	0.95
12013 SLANEY @ RATHVILLY	185	30	45.16	43.55	72.3	-0.01	0.27	0.15	0.03	0.12	1.6	1.66	0.39	0.96
14033 OWENASS @ MOUNTMELLICK	185	22	22.59	19.5	33	0.53	0.28	0.16	0.14	-0.07	1.46	1.69	0.18	0.86
15005 ERKINA @ DURROW FOOT BR.	387	50	28.47	27.44	61.85	1.78	0.34	0.18	0.19	0.21	2.17	2.25	0.16	0.96
15012 NORE @ BALLYRAGGET	945	16	77.16	77.11	133	0.9	0.3	0.17	0.08	0.21	1.72	1.72	0.14	1
16006 MULTEEN @ BALLINCLOGH BR.	75	33	30.37	27.87	58.07	0.34	0.39	0.23	0.06	0.03	1.91	2.08	0.77	0.92
16007 AHERLOW @ KILLARDRY	273	51	79.18	75.84	138.03	0.46	0.32	0.18	0.09	0.06	1.74	1.82	0.51	0.96
16011 SUIR @ CLONMEL	2173	52	234.52	223	389	0.42	0.3	0.17	0.09	0.05	1.66	1.74	0.18	0.95
16012 TAR @ TAR BRIDGE	228	36	55.2	54.57	92.2	0.39	0.28	0.16	0.06	0.08	1.67	1.69	0.4	0.99
16013 NIRE @ FOURMILEWATER	91	33	101.69	93.21	207.02	0.86	0.42	0.24	0.16	0.08	2.04	2.22	2.27	0.92
16051 ROSSESTOWN @ CLOBANNA	34.18	13	2.95	2.85	5.67	2.53	0.36	0.19	0.33	0.26	1.92	1.99	0.17	0.96
18001 BRIDE @ MOGEELY BRIDGE	335	48	71.07	71.49	96.93	-0.16	0.19	0.11	-0.03	0.1	1.36	1.36	0.29	1.01
18002 BALLYDUFF @ BALCKWATER	2338	49	353.65	344	479	0.24	0.16	0.09	0.06	0.1	1.35	1.39	0.2	0.97
18003 B.WATER @ KILLAVULLEN	1258	49	282.76	266.15	502.74	1.1	0.23	0.13	0.16	0.1	1.78	1.89	0.4	0.94
18006 B.WATER @ CSET MALLOW	1058	27	291.3	286	397	1.1	0.14	0.08	0.19	0.12	1.36	1.39	0.38	0.98

18016 B.WATER @ DUNCANNON	113	24	80.99	79.65	114.82	0.49	0.23	0.14	0.12	0	1.42	1.44	1.02	0.98
18048 B.WATER @ DROMCUMMER	881	23	222.77	220	269	1.11	0.08	0.05	0.15	0.22	1.21	1.22	0.31	0.99
18050 BLACKWATER @ DURRIGLE	244.6	24	121.96	124.5	175	0.12	0.2	0.11	-0.01	0.09	1.43	1.41	0.72	1.02
19014 LEE @ DROMCARRA (ESB)	N/A	20	79.69	71.89	157.21	1.67	0.38	0.21	0.3	0.12	1.97	1.4	0.69	0.9
19016 BRIDE @ OVENS BRIDGE (ESB)	N/A	8	28.74	29.58	34.87	-1.35	0.16	0.09	-0.14	0.22	1.21	1.65	0.6	1.03
19031 SULLANE @ MACROOM (ESB)	N/A	9	131.09	135.9	201.18	1.46	0.27	0.16	0.16	0.17	1.53	2.27	0.67	1.04
19046 MARTIN @ STATION ROAD	60.4	9	31.09	29.95	41.95	-0.34	0.26	0.16	-0.04	0.06	1.35	2.01	0.7	0.96
20002 BANDON @ CURRANURE	431	31	140.6	126.28	287.11	2.13	0.37	0.19	0.37	0.28	2.04	1.43	0.44	0.9
20006 ARGIDEEN @ CLONAKILTY	79.3	25	30.25	27.7	55.6	1.67	0.3	0.16	0.27	0.18	1.84	1.22	0.25	0.92
22006 FLESK @ FLESK BRIDGE	325	51	165.89	169.09	282.83	0.66	0.25	0.14	0.07	0.12	1.7	2.07	0.43	1.02
22009 DREENAGH @ WHITE BR.	37	24	11.91	11.47	16.35	1.17	0.16	0.09	0.18	0.19	1.37	1.64	0.63	0.96
22035 LAUNE @ LAUNE BRIDGE	559.65	14	112.81	116.4	141.48	-0.86	0.2	0.12	-0.15	0.03	1.25	1.29	0.4	1.03
24004 MAIGUE @ BRUREE	246	52	54.86	50.63	104.55	0.91	0.39	0.22	0.18	0.1	1.91	1.39	0.29	0.92
24011 DEEL @ DEEL BRIDGE	273	33	103.01	104.55	171.74	0.42	0.22	0.12	-0.02	0.24	1.67	1.41	0.48	1.01
24012 DEEL @ GRANGE BRIDGE	359	41	110.45	109.99	141.95	-0.02	0.16	0.09	0.01	0.13	1.29	2.37	0.16	1
24030 DEEL @ DANGANBEG	248	25	52.89	52	72.2	0.88	0.15	0.08	0.11	0.1	1.37	2.05	0.45	0.98
25004 BILBOA @ NEWBRIDGE	125	30	41.7	42.3	59.5	-0.16	0.23	0.13	-0.04	0.12	1.43	1.73	0.49	1.01
25011 BROSNA @ MOYSTOWN	1227	51	85.64	82.02	194.25	1.31	0.34	0.18	0.14	0.23	2.27	2.38	0.41	0.96
25020 KILLIMOR @ KILLEEN	197	35	46.6	43.65	89.55	0.93	0.33	0.19	0.16	0.11	1.92	1.66	0.32	0.94
25038 NENAGH @ TYONE	139	17	42.08	39.3	67.9	1.3	0.27	0.15	0.17	0.22	1.61	2.28	0.21	0.93
26010 CLOONE @ RIVERSTOWN	100	35	20.03	17.17	40.8	1.64	0.41	0.22	0.35	0.18	2.04	1.27	0.14	0.86
26014 LUNG @ BANADA BRIDGE	222	16	44.1	42.82	70.9	1.8	0.23	0.12	0.19	0.26	1.61	2.72	0.77	0.97
26058 INNY @ BALLINRINK BR.	59	24	5.98	5.35	12.2	1.6	0.38	0.21	0.28	0.18	2.04	1.62	0.15	0.89
26108 OWENURE @ BOYLE BR.	533	15	56.29	57.32	73.07	0.28	0.18	0.11	0.06	-0.06	1.3	1.23	0.14	1.02
28001 INAGH @ ENNISTIMON	168	17	52.69	47.58	129.28	4.89	0.4	0.16	0.47	0.61	2.45	2.23	0.44	0.9
29007 L.CULLAUN @ CRAUGHWELL	278	22	27.83	26.49	42.93	1.03	0.22	0.12	0.15	0.2	1.54	1.3	0.14	0.95
30012 CLARE @ CLAREGALWAY	1075.4	9	126.89	126	155	1	0.12	0.07	0.13	0.15	1.22	1.79	0.02	0.99
30021 ROBE @ CHRISTINA'S BR.	138	26	28.17	27.2	60.7	2.17	0.34	0.18	0.21	0.29	2.15	2.38	0.12	0.97
30031 CONG @ CONG WEIR	891	24	94.35	93.88	122.28	-0.62	0.18	0.1	-0.02	0.13	1.3	1.93	1.82	0.99
30037 ROBE @ CLOONCORMICK	210	21	1.8	1.79	3.19	-0.42	0.37	0.21	-0.09	0.18	1.77	1.96	1.59	0.99

31072 CONG @ CONG WEIR	891	26	49.08	43.2	103	2.04	0.38	0.2	0.31	0.2	2.1	1.94	0.57	0.88
32011 BUNOWEN @ LOUISBERG W.	68.6	26	74.88	64.87	125	0.75	0.3	0.17	0.17	0.06	1.67	2.35	1.28	0.87
33001 GLENAMOY @ GLENAMOY	73	25	62.11	59.3	116	1.65	0.28	0.15	0.2	0.19	1.87	1.07	0.53	0.95
34007 DEEL @ BALLYCARROON	156	53	90.37	84.48	198.91	0.96	0.36	0.2	0.14	0.1	2.2	1.25	0.12	0.93
34010 MOY @ CLOONACANNANA	471	12	123.29	113.72	193.07	1.31	0.3	0.17	0.21	0.11	1.57	1.7	0.41	0.92
35011 BONET @ DROMAHAIR	294	36	116.02	115.36	188.01	0.04	0.3	0.17	0.01	0.08	1.62	1.39	1.26	0.99
36011 ERNE @ BELLAHILLAN	318	49	17.91	18.23	23.6	-0.51	0.18	0.1	-0.1	0.11	1.32	1.29	0.07	1.02
36071 L.SCUR @ GOWLY	66	20	6.36	6.49	8.13	-0.06	0.15	0.09	-0.04	0.01	1.28	1.78	3.25	1.02
38001 OWENEA @ CLONCONWAL	109	33	70.02	70.63	113.38	1.78	0.16	0.09	0.08	0.26	1.62	1.75	0.44	1.01
39001 NEW MILLS @ SWILLY	49	30	44.88	44.25	61.5	0.06	0.2	0.12	0.01	0.08	1.37	1.53	0.1	0.99

3.4 Examination of Probability plots

Probability plots are useful tools for the graphical display and analysis of flood data, particularly to determine whether or not a particular sample is consistent with a particular population distribution. Probability plots use an inverse distribution scale so that a simple, usually 2 parameter, cumulative distribution function (cdf) plots as a straight line in a $x - y$ plot, i.e the points $(x(i), y(i); i = 1, 2, ..., n)$ are expected to be close to the line $y = a + b \times x$, where a and b are location and scale parameters respectively; conversely, strong deviation from this line is evidence that the distribution did not produce the data. It is acknowledged that what constitutes 'strong deviation' from such a line is usually judged subjectively. Nevertheless they are widely used in flood hydrology for data display if not for determining a final distribution choice.

A probability plot is constructed by plotting the ordered observations $Q\,(i)\,, i = 1, \ldots, n$ against the inverse of the cdf which is defined as

$$y_i = F^{-1}(\hat{F}(Q_i)) \tag{3.9}$$

where $\hat{F}(Q_{(i)})$, termed a plotting position, is an estimate of the cdf corresponding to the i^{th} ordered observation.

The use of probability plots was introduced in the literature on hydrology by Hazen (1914). Cunnane (1978) provided detailed reviews and discussion on this subject.

Three probability plots

- Gumbel (EV1)
- Logistic (LO) and
- Lognormal (LN)

have been chosen for the AM data of the 110 Irish stations. The EV1 and LN are included as they are the most commonly used distributions of all (Cunnane, 1989, App. 6). In addition, the EV1 was suggested by FSR (1975) as the best suited 2-parameter distribution for Irish flood data. The choice of logistic plot arises because the FEH (1999) suggested it for use with U.K. flood data.

Unbiased plotting formulae, i.e. the Gringorten formula for the Gumbel and Logistic plots and the Blom formula for the log-normal plot have been used in this study. Their expressions are as follows:

$$F_i = \frac{i - 0.44}{N + 0.12}, \quad i = 1, 2, ...N \tag{3.10}$$

$$F_i = \frac{i - 3/8}{N + 1/4}, \quad i = 1, 2, ...N \tag{3.11}$$

Flood data, plotted on Gumbel, Logistic and Log-normal plots for the 110 Irish stations, are shown in Appendix C. These plots were investigated from the standpoint of linearity and non-linearity. In the linearity investigation, a best-fit least squares straight line is drawn through the data and is given a score from 1 to 5 on the basis of visual examination where, 1 stands for a very poor fit and 5 stands for a very good fit. On the other hand in the non-linearity investigation, an eye-guided assessment of the plotted data is indicates a curve pattern such as linear, concave, convex or S-curve.

It is acknowledged that deductions drawn from probability plots have low statistical power i.e. there is a high probability that a conclusion could be adopted even though it is not true. To confirm this, this issue has been addressed informally by examining a selection of probability plots of samples, of size 20, 35 and 50, drawn randomly from EV1 populations which are shown in Figures 3.8 to 3.13. These plots were also investigated from the standpoint of linearity and non-linearity.

Linear Patterns

Each data series was examined from the point of view of linearity and subjectively scored from 1 to 5 by visual judgment,

where,

5= very good fit;

4= good;

3= medium;

2= poor; and

1= very poor.

These scores are listed individually for each station in Tables 3.12 and 3.13 for the A1 and A2 stations respectively and a summary of these scores is given in Table 3.10, as follows:

Table 3.10: Linear pattern statistics for 110 A1 and A2 stations and for simulated samples

Score Patterns	Linear score for 110 stations						Linear score for simulated data	
	EV1		LO		LN			
	Number	%	Number	%	Number	%	Number	%
5	10	9	7	6	16	15	19	32
4	49	45	23	21	36	33	29	48
3	25	23	41	37	35	32	12	20
2	21	19	28	25	17	15	0	0
1	5	5	11	10	6	5	0	0

The above statistics show that in the case of the EV1, 54% of samples provide 'good' to 'very good' fit to the data while in case of the LN and LO the numbers are 52 and 27 respectively. Thus EV1 and LN give almost identical results and they provide a better fit than LO. It can be deduced from the linear patterns that Irish flood data are more likely to be distributed as EV1 or LN rather than LO among 2-parameter distributions.

The scores for 60 random samples drawn from EV1 populations are also summarised in the table above. 80% of all samples provide 'good' to 'very good' fit to the data. In 20% of cases, departures from linear patterns are observed that could cause an EV1 hypothesis to be rejected.

Non Linear Patterns

Each of the 110 data series was investigated from the point of view of non-linearity and was again assigned a curve pattern by visual judgment. Four curve patterns, i.e. linear, concave, convex and elongated S are introduced and again each has been classified as degree one (see U1 below) and degree two (see U2 below) for greater discrimination. However, there is also an 'X' introduced (see below) if the patterns of the end points of the plot do not match the main body pattern well. In summary, non-linear patterns are classified as:

L1= perfectly straight line

L2= little deviation from straight line

L2X= body pattern is quite straight but end points disturb the linearity

U1= mild concave upwards

U2= severe concave upwards

U1X= mild concave with end disturbance

U2X= severe concave with end disturbance

D1= mild convex upwards

D2= severe convex upwards

D1X= mild convex upwards with end disturbance

D2X= severe convex upwards with end disturbance

S1= mild S-curve

S2= severe S-curve

S1X = mild S-curve with end disturbance

S2X= severe S-curve with end disturbance

These scores are listed individually for each station in Tables 3.12 and 3.13 for the A1 and A2 stations respectively and a summary of these scores is given in Table 3.11.

Table 3.11: Linear and non- linear curve pattern statistics for 110 A1 and A2 stations and for simulated samples

Curve Patterns	statistics for 110 stations						statistics for simulated data	
	EV1		LO		LN			
	Number	%	Number	%	Number	%	Number	%
L1	10	9	0	0	12	11	19	32
L2	41	37	28	25	38	35	29	48
D1	17	15	3	3	8	7	7	12
D2	7	6	1	1	2	2	0	0
U1	2	2	28	25	8	7	3	5
U2	6	5	9	8	2	2	1	2
S1	10	9	21	19	19	17	0	0
S2	4	4	8	7	6	5	0	0
D1x	0	0	0	0	1	1	0	0
D2x	1	1	2	2	1	1	0	0
U1x	0	0	2	2	0	0	0	0
U2x	1	1	1	1	2	2	0	0
S1x	2	2	1	1	3	3	0	0
S2x	3	3	2	2	1	1	0	0
L2x	6	5	4	4	7	6	1	2

In EV1 probability plots, 46% of the stations show linear patterns and a convex upwards pattern comes out as the second dominant pattern, occuring in 21% of the

stations. Interestingly, very few stations exhibit concave downwards pattern on EV1 probability plots. This is because in many cases the largest floods on record are close together in magnitude and in very few cases are there notably large flood outliers. In Logistic plots, most of the stations show concave and linear patterns and a significant number of stations show S shapes. There is no station with a perfect straight line and hardly any stations show a convex pattern. In the case of lognormal plots, 46% of the stations show linear patterns. A considerable number of stations also show S and convex patterns. In Tables 3.12 and 3.13, it is seen that in many cases EV1 and Lognormal results complement one another. It was also observed that, in many cases, those stations the data of which fit quite straight lines on EV1 plots show a concave fit when plotted on Logistic paper. The reason might be due to the range of the probability axis, because the probability axis of LO is relatively compressed compared to that of EV1. For the same reasons, it was also noticed that stations which fit as convex downwards on EV1 plots are found to be more linear on LO plots.

The simulated data series, previously referred to, were also investigated and were given a curve pattern score. They are summarized in Table 3.11 . In a number of cases (about 20%), departures from linear patterns are observed. 12% of samples show a convex trend and a concave trend comes out in 7% of samples. No sample exhibits an S-shape. The corresponding patterns on EV1 plot, based on the 110 stations, came out in 23%, 8%, and 17% of samples respectively. In about 55% of the overall cases departures from linear patterns are observed which are much higher than in the case of the simulated series. On the basis of the above, it would seem that the choice of the EV1 distribution as parent for all 110 Irish stations is of doubtful validity.

It must be remembered that above described experiment has been done for a limited number of samples (60). The main aim to carrying out this experiment is to demonstrate once again that departures from linear patterns can occur whether it be convex downward or concave upward even when the data are drawn from a single population. Therefore, deductions from probability plots have to be considered carefully.

Table 3.12: Probability plot linear and non-linear scores for the A1 category Stations

Station No, Location	N	Linear Score			Non linear Score		
		EV1	LO	LN	EV1	LO	LN
6011 FANE	48	4	2	3	L2	U1	L2
6013 DEE	30	2	3	4	D1	S1	L2
6014 GLYDE	30	4	2	3	L1	U1	L2
6025 DEE	30	2	4	3	D1	L2	S1
6026 GLYDE	46	3	1	2	S1	S2	S2
6070 MUCKNO	24	4	4	5	L2	L2	L1
7009 BOYNE	29	4	1	3	L2	D2X	L2
8002 DELVIN	20	4	1	2	L2	U1	U1
8009 BALHEARY	11	1	1	2	U2	U2	U1
9001 RYEWATER	48	5	2	5	L1	U1	L1
9002 GRIFFEEN	24	1	1	4	S2X	U2	L1
9010 DODDER	19	2	2	4	U2	U2	L2
10021 SHANGANAGH	24	4	2	3	L2	S1	L2
10022 CARRICKMINES	18	4	3	4	L2	L2X	L2
14006 BARROW	51	5	2	3	L1	U1	U1
14007 DERRYBROCK	25	4	2	3	L2	U1	S1
14011RATHANGAN	26	3	3	3	L2	L2	L2
14018 BARROW	51	3	3	4	L2	S1	S1
14019 BARROW	51	3	3	4	D1	S2	S1
16011 SUIR	52	3	2	3	U1	S1	S1
25006 BROSNA	52	4	3	4	L2	L2	S1
25014 SILVER	54	4	3	4	L2	U1X	L2
25017 SHANNON	55	3	4	5	D1	L2	L1
25023 LITTLE BROSNA	52	2	2	2	S2	S2	S1X
25025 BALLYFINBOY	31	5	4	5	L1	L2	L2
25027 OLLATRIM	43	4	5	3	D1	L2	D1
25030 GRANEY	48	4	2	4	S1X	U1	L2
26006 SUCK	53	1	1	1	S2X	S2X	S2X
26007 SUCK	53	5	3	5	L1	U1	L1
26008 RINN	49	4	3	3	L2	U1	L2
26019 CAMLIN	51	4	3	3	L2	S1	L2
26020 CAMLIN	32	3	3	2	L2	S1	L2
26059 INNY	17	2	2	1	D1	D1	D2X
27002 FERGUS	51	4	2	4	L2	U1	L2
29001 RAFORD	40	3	4	4	D1	S1	S1
29011 DUNKELLIN	22	2	2	2	U1	S1	S1
31002 CASHLA	26	2	1	1	U2X	U2X	U2X
33070 CARROWMORE	28	4	3	4	L2X	L2X	L2X
34018 CASTLEBAR	27	4	3	3	L2	U1	U1
36010 ANNALEE	50	4	3	4	S1X	S1X	S1X
36012 ERNE	47	3	4	3	D1	L2	L2
36015 FINN	33	2	2	2	U2	U2	U1
36018 DRONMORE	50	4	4	4	L2	L2	L2

Table 3.13: Probability plot linear and non-linear scores for A2 category Stations

Station No,Location	N	Linear Score			Non Linear Score		
		Ev1	LO	LN	EV1	LO	LN
6031 FLURRY	18	2	2	3	U2	U2	U2
7006 MOYNALTY	19	2	3	2	D2	D1	D2
7033 BLACKWATER	25	4	3	3	S1	U1	S1
8005 SLUICE	18	4	3	5	L2	U1	L1
8008 BROADMEADOW	25	4	2	4	L2	U1	L2
12001 RIVER SLANEY	50	4	2	3	L2	U1	L2
14005 BARROW	48	3	2	2	L2	U1	U1
14009 CUSHINA	25	4	3	4	S1	U1	S1
14013 BURRIN	50	4	3	4	S1	S2	S1
14029 BARROW	47	3	4	5	D1	S1	L1
14034 BARROW	14	4	2	2	L2X	U2X	U2X
15001 RIVER KINGS	42	2	2	3	D2	S2	S1
15002 NORE	35	4	3	4	L2	U1	L2
15003 DINAN	50	2	3	2	D2	S1	D2
15004 NORE	51	5	3	5	L2	U1	L2
16001 DRISH	33	5	4	5	L1	U1	L1
16002 SUIR	51	3	2	4	L2X	U1X	L2X
16003 CLODIAGH	51	4	3	3	L2	S1	S1
16004 SUIR	48	5	4	5	L1	U1	L1
16005 MULTEEN	30	5	2	2	L2	U1	U1
16008 SUIR	51	2	4	3	D2	L2	D1
16009 SUIR	52	2	3	3	D2X	D2X	D1X
18004 BALLYNAMONA	46	2	2	2	S2	S2	S2
18005 FUNSHION	50	4	2	3	L2X	U2	L2X
19001 OWENBOY	48	4	4	4	L2	L2	S1X
19020 OWENNACURRA	28	3	4	3	D2	S1	D1
23001 GALEY	45	4	2	3	L2	U1	L2
23012 LEE	18	2	1	1	U2	U2	U2
24002 CAMOGUE	27	4	5	4	L2	L2	L2X
24008 MAIGUE	30	3	5	4	L2	L2	L2X
24022 MAHORE	20	5	3	5	L1	U1	L2
24082 MAIGUE	28	2	5	3	D1	L2	D1
25016 CLODIAGH	42	4	4	5	L2	L2	L2
25021 BROSNA	44	3	4	5	D1	L2	L2
25029 NENAGH	33	3	3	4	S2	S2	S1
25034 L. ENNELLTRIB	24	2	4	3	D2	L2	D1
25040 BUNOW	20	4	3	4	L2	U1	L2
25044 KILMASTULLA	33	3	3	3	S1	S1	S2
25124 BROSNA	18	3	4	2	D1	L2	D1
26002 SUCK	53	2	1	2	S2X	S2X	L2X
26005 SUCK	51	4	4	5	L1	L2	L1
26009 BLACK	35	4	3	2	S1	S1	S2
26018 OWENURE	49	4	3	4	L2	S1	L2
26021 INNY	30	1	2	1	D2	D1	D1
26022 FALLAN	33	4	3	4	L2	S1	S1
27001 CLAUREEN	30	4	3	3	L2	U1	L2
27003 FERGUS	48	4	4	4	L2	L2	L2
29004 CLARINBRIDGE	32	4	3	3	L2	L2	S1
29071 L.CUTRA	26	4	3	4	L2	S1	L2
30007 CLARE	31	4	3	4	L2X	L2X	L2X
30061 CORRIB ESTUARY	33	1	1	2	U2	U2	U1
32012 NEWPORT	24	3	5	3	D1	L2	L2
34001 MOY	36	4	3	4	L2X	L2X	L2
34003 MOY	29	3	3	3	S1	S1	S2
34009 OWENGARVE	33	4	4	5	L2	L2	L1
34011 MANULLA	30	4	4	4	S1	L2	L2
34024 POLLAGH	28	3	4	3	D1	L2	L2
35001 OWENMORE	29	4	5	4	L2	L2	L2
35002 OWENBEG	34	3	5	4	D1	L2	L2
35005 BALLYSADARE	55	4	2	4	S1	S1	L2
35071 L.MELVIN	30	4	3	4	L2	U1	L2
35073 L.GILL	30	3	4	4	D1	L2	D1
36019 ERNE	47	2	3	3	D1	S1	S1
36021 YELLOW	27	3	3	3	L2	U2	S1
36031 CAVAN	30	2	1	1	S2	S2	S2
39008 GARTAN	33	3	2	3	S1	S1	S1
39009 FERN O/L	33	5	3	5	L1	U1	L1

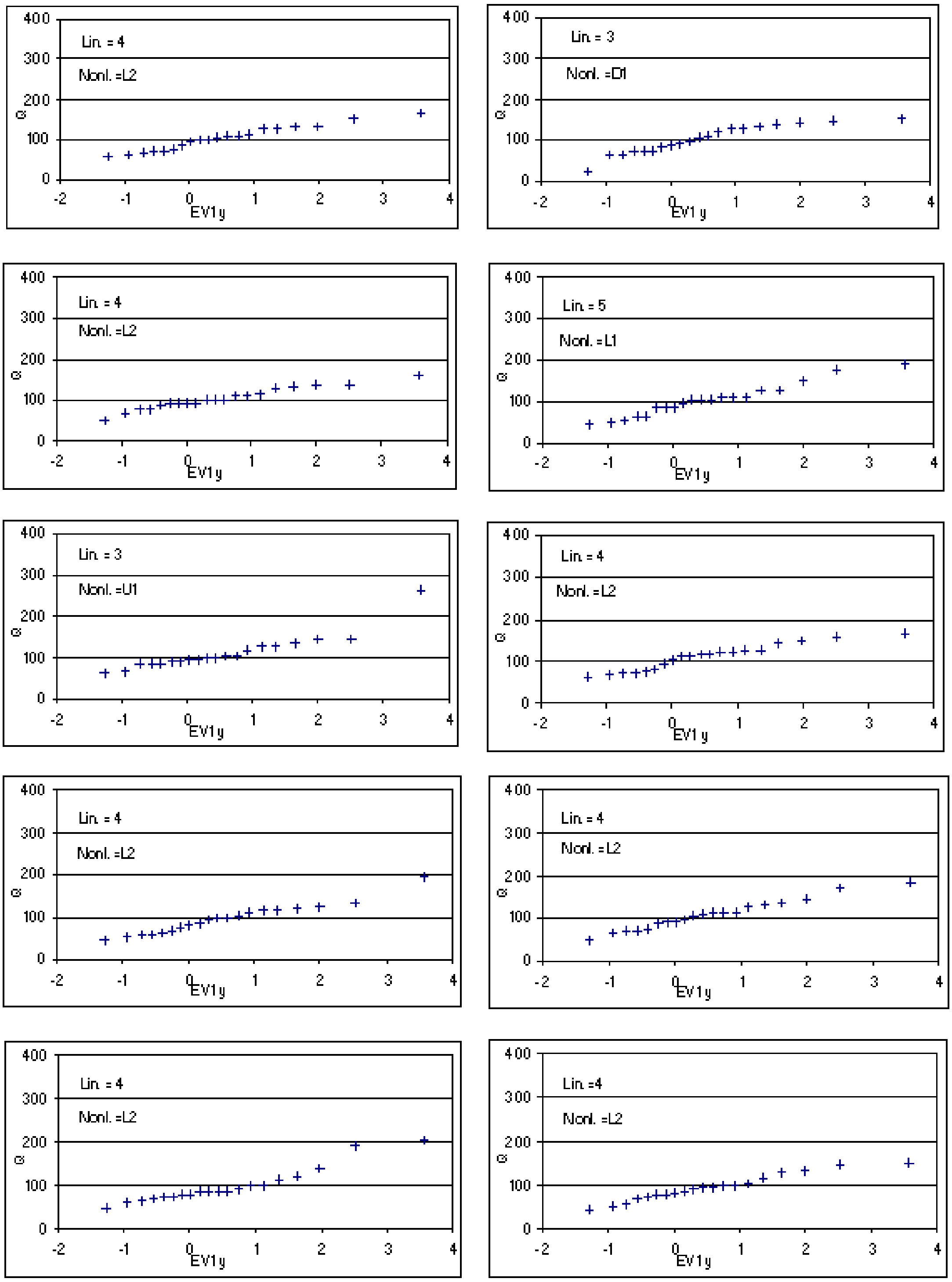


Figure 3.8: Random Samples of size 20 from an EV1 population (L-CV = 0.2)-first set -

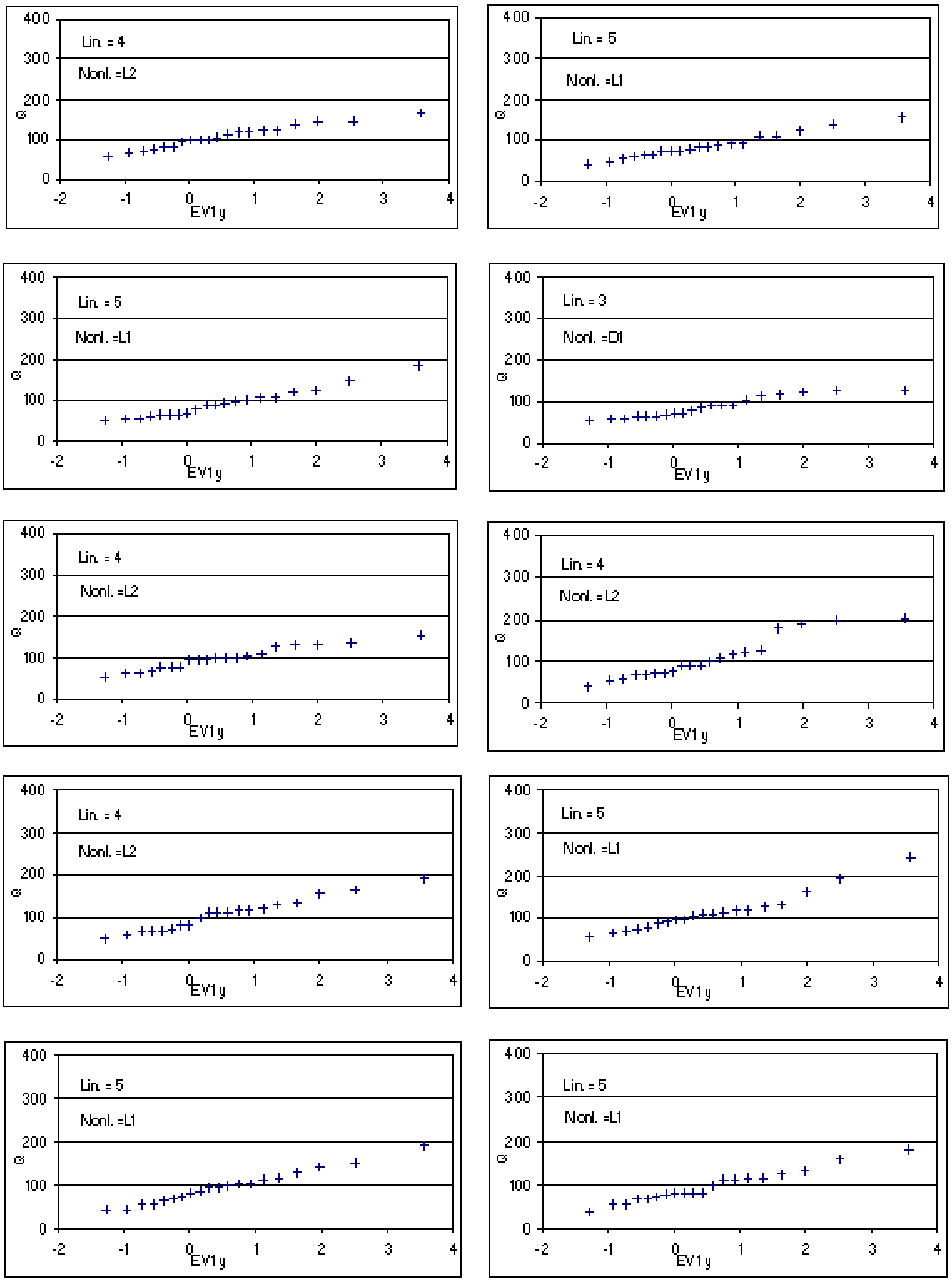


Figure 3.9: Random Samples of size 20 from an EV1 population (L-CV = 0.2)-second set -

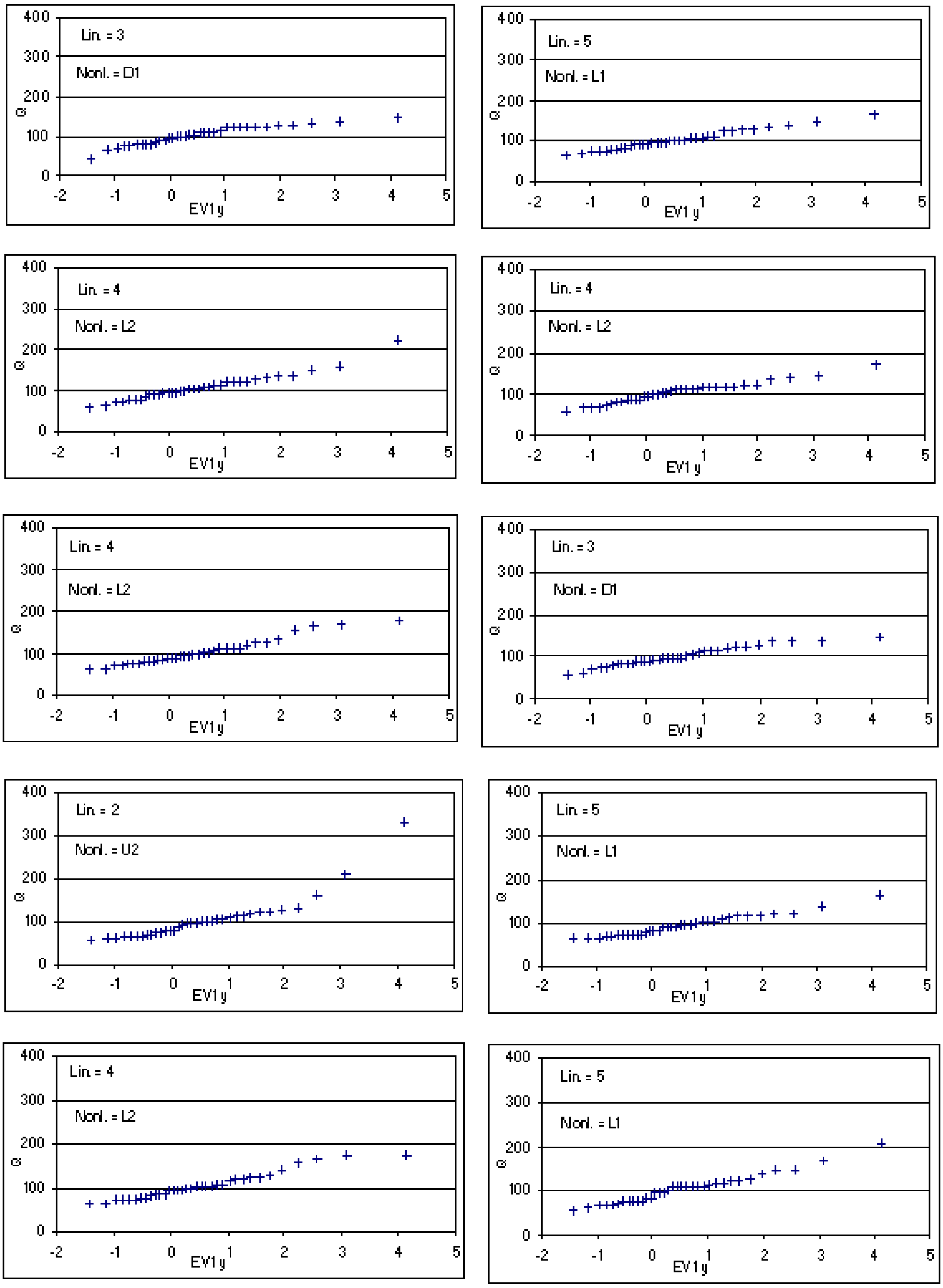


Figure 3.10: Random Samples of size 35 from an EV1 population (L-CV = 0.15)-first set -

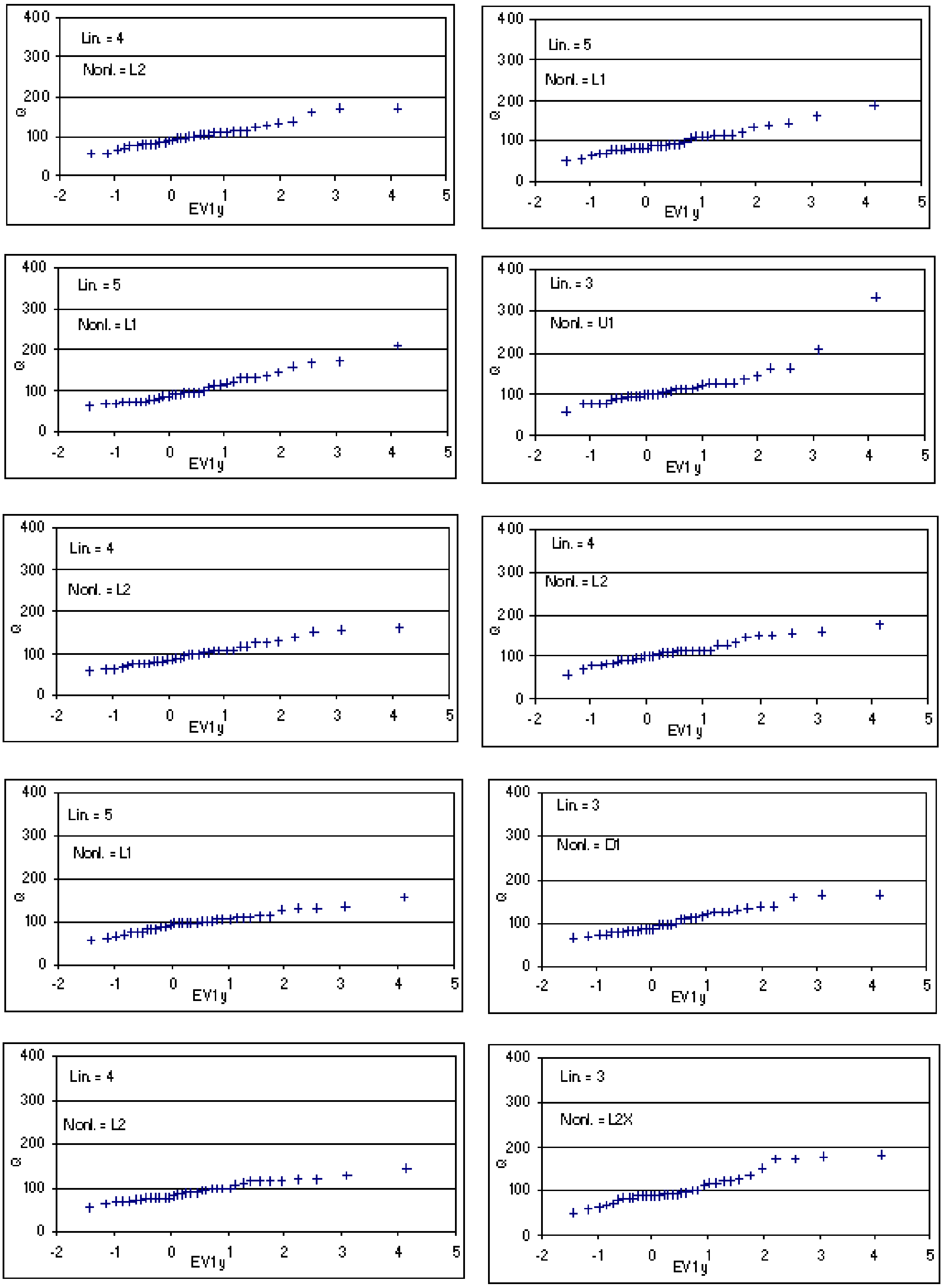


Figure 3.11: Random Samples of size 35 from an EV1 population (L-CV = 0.15)-second set -

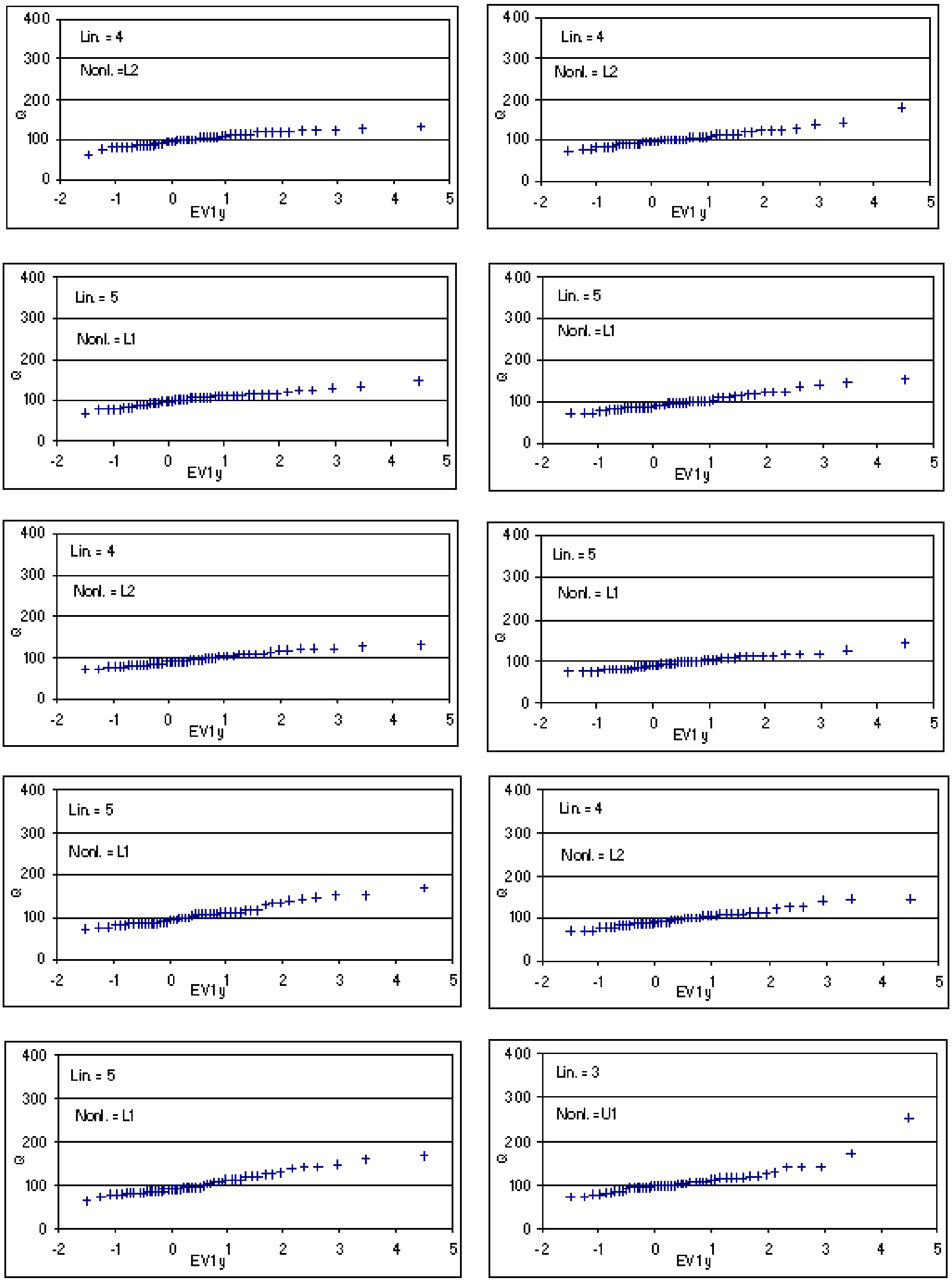


Figure 3.12: Random Samples of size 50 from an EV1 population (L-CV = 0.1)-first set -

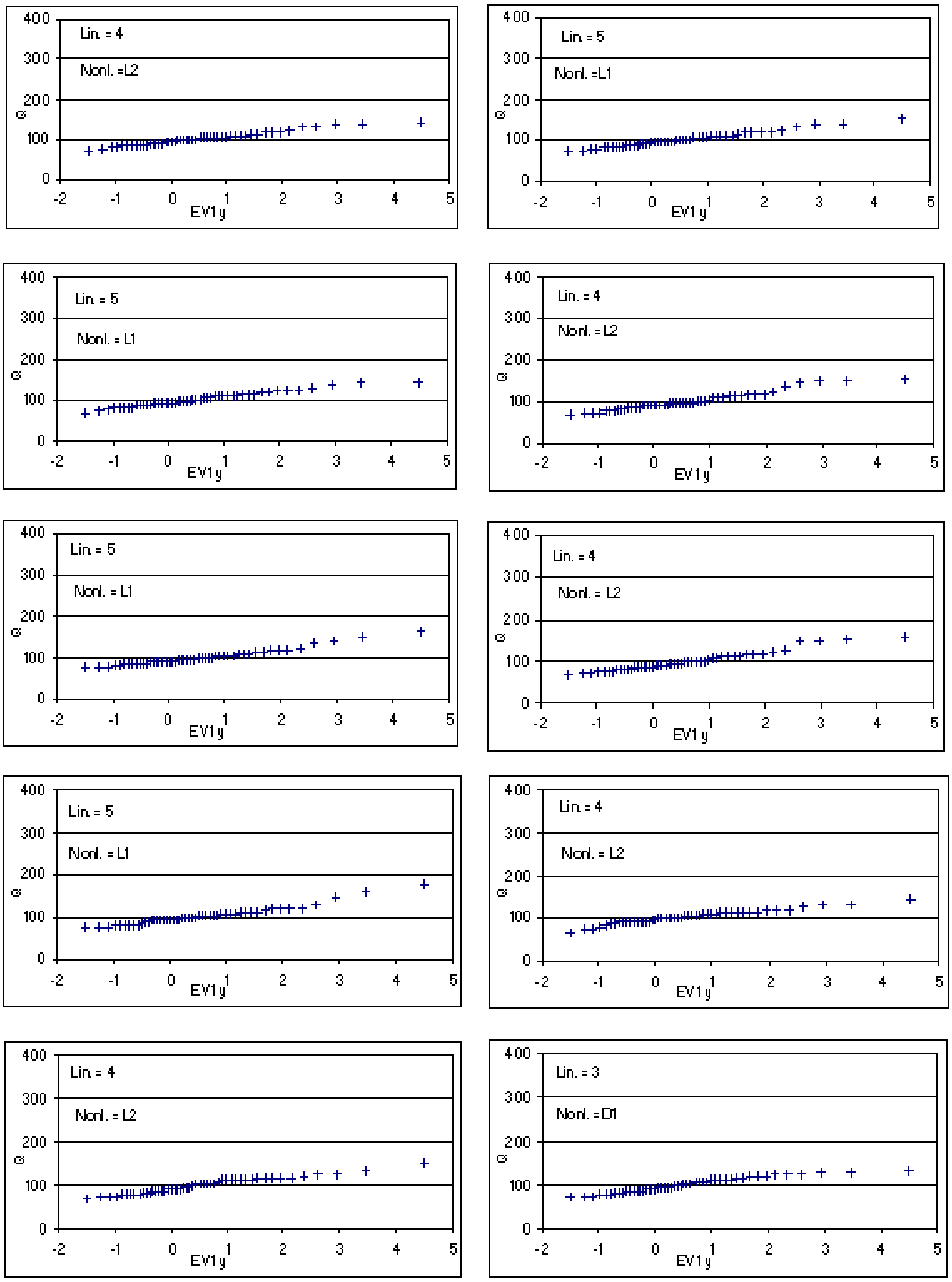


Figure 3.13: Random Samples of size 50 from an EV1 population (L-CV = 0.1)-second set -

3.5 Preliminary distribution choice

3.5.1 Goodness of fit tests

Probability plot techniques have a wide appeal in deciding if an observed AM flood series appears to come from a given distribution form but, as was shown in the last section, they are less objective and can be sensitive to random occurrences in the data. Hence, sole reliance on probability plots can lead to incorrect conclusions.

Objective techniques such as goodness-of fit indices are also used in this context. However, they have been often criticized for their inability to discriminate between statistical distributions for the same application (e.g. Cunnane, 1989). However goodness-of -fit tests can sensibly be used in FFA as the basis for rejecting some distributions but not to select a best population distribution.

A goodness of fit test calculates a test statistic or index which is used to assess the fit of an observed sample of AM flood to a population with a specified distribution. The population distribution may be completely specified (denoted here as Case 0) or it may result from fitting a distribution to the sample (denoted here as Case 1). If the observed index value lies in the tail of its sampling distribution, then it throws doubt on the hypothesis that the sample came from the specified distribution.

The most common goodness of fit tests that have been used in FFA are the chi-squared (χ^2) test (FSR, 1975), empirical distribution function (EDF) tests such as the Kolmogorov-Smirnov (KS) test (Chowdhury et al., 1991; FSR, 1975), the Anderson-Darling (AD) test (Ahmad et al., 1988a; Laio, 2004), the Cramer- von Mises (W^2) test (Ahmad et al., 1988a; Laio, 2004) and the probability plot correlation coefficient (PPCC) test (Chowdhury et al., 1991; Vogel and McMartin, 1991).

Laio (2004) discussed goodness of fit tests based on W^2 and AD indices. The power (the ability to reject the null hypothesis when it is false) of these tests was also assessed using Monte Carlo simulation and compared to the power of alternative approaches such as the χ^2 , KS and PPCC tests. The χ^2 test is the weaker among the considered statistics and even weaker when the sample size is small. Among EDF statistics, AD is slightly better than W^2, while the commonly used KS statistic tends to be rather weak in power. The PPCC test is on average more effective in detecting deviations from normality but the test has underperformed relative to the AD test for other distributions such as EV1, EV2, and LP3.

In this section, goodness-of-fit tests based on the chi-squared statistic and Empirical Distribution Function (EDF) statistics such as KS and AD are used for evaluating the suitability of different probability distributions for Irish conditions. Although chi-squared tests are generally less powerful than EDF tests, they are the most practical tests on the basis of flexibility and ease of use (see Moore, 1986).

3.5.1.1 $\chi^2 - test$

In this test, frequencies of the observed events in a number of class intervals are compared with their expected values for a selected distribution. The χ^2 goodness of fit is defined as

$$\chi^2 = \sum_{i=1}^{M} \frac{(O_i - E_i)^2}{E_i} \tag{3.12}$$

where M is the number of class intervals, O_i is the observed frequency for interval i and E_i is the expected frequency for interval i calculated as $E_i = N\,(P_i)$, in which P_i is the difference between upper and lower cumulative probability for the interval i and N is the sample size.

For a large sample (in which E_i>5 can be ensured), this quantity is distributed as χ^2 with $M - 1$ degrees of freedom. In Case 1, the quantity distributed as χ^2 with $(M - 1- \theta)$ degrees of freedom is commonly recommended (see Moore,1986). θ is the number of parameters to be estimated.

The choice of class intervals in a chi-squared test is an important issue. As there are no theoretical rules, subjectivity is often employed to choose class intervals. Since an objective procedure for choosing cells is desirable, this study adopts the choice of $M = 2n^{2/5}$ which is convenient (see Moore, 1986) .

3.5.1.2 EDF tests

EDF tests are based on the discrepancy between the theoretical and empirical distribution functions, $F_(Q)$ and $F_n(Q)$. The empirical distribution function is defined by the following these alternative forms

$$F_n(Q) = 0,\ Q < Q\,(1)$$

$$F_n(Q) = \frac{i}{N},\ Q_{(i)} \leq Q \leq Q_{(i+1)}$$

$$F_n(Q) = 1,\ Q_{(N)} \leq Q \tag{3.13}$$

The discrepancy between the two distributions $F_(Q)$ and $F_n(Q)$ can be measured using the KS test defined based on the statistic D defined as

$$D = \max_{1 \leq i \leq N} \left(\frac{i}{N} - F\left(Q,\theta\right), F\left(Q,\theta\right) - \frac{i-1}{N} \right) \tag{3.14}$$

or using quadratic statistics (see D' Agostino and Stephens, 1986, p. 100),

$$q^2 = n \int_{allQ} \left[F_n\left(Q\right) - F\left(Q,\theta\right)\right]^2 \psi\left(Q\right) dF\left(Q\right) \tag{3.15}$$

Use of different weights $\psi\left(Q\right)$ in the above equation results in different statistics: when $\psi\left(Q\right) = 1$, the statistic is called Cramer-von Mises statistic, W^2. When $\psi\left(Q\right) = \left[F\left(Q,\theta\right)\left(1 - F\left(Q,\theta\right)\right)\right]^{-1}$, the tails of the distribution are weighted more than the central part giving the Anderson-Darling statistic A^2 . The test formula for the Anderson-Darling statistic is defined as

$$A^2 = -N - \sum_{i=1}^{N} \left[(2i-1)\ln F\left(Q_i\right) + \left(2\left(n-i\right)+1\right)\ln\left(1 - F\left(Q_i\right)\right)\right] \tag{3.16}$$

A modified AD test applicable to FFA is proposed (Ahmad et al., 1988a) by assigning $\psi\left(Q\right) = \left[\left(1 - F\left(Q,\theta\right)\right)\right]^{-1}$ to the equation 3.16. In that case, a greater weight is imposed only to the upper tail.

The exact distributions of D and A^2 in Case 0 are different from those obtained in Case 1. Since the tests are performed in Case 1, appropriate percentage points should then be used; otherwise a serious error in significance level will result (see Stephens, 1986) . His results for Case 1 are available for a number of distributions, especially for 2-parameter distributions . He used Monte Carlo experiments to find percentage points of the test statistics for finite n.

The critical points of D for the GEV distribution are given by Chowdhury et al. (1991) while for GEV and GLO distributions the critical points of A^2 are given by Ahmad et al. (1988a).

In this study, the critical points of D and A^2 for the GNO distribution are calculated. These critical points are also applicable to the LN3 distribution provided that the sample data have been preliminarily log transformed. Monte Carlo simulation is used to obtain the sampling distribution of the test statistics D and A^2 , under the null hypothesis that the observations arise from a GNO distribution. Critical points are obtained as follows:

1. Random samples from GNO populations are drawn with L1= 1 and the following combinations of t_2 and t_3: $\{(t_2, t_3)$ = (0.1, 0.0), (0.1, 0.1), (0.2, 0.2), (0.3, 0.3) (0.4, 0.4)$\}$ for each sample of size n = 10, 15, 20, 25, 30, 35, 40, 45, 50, 100. A total of 5 sets of 10,000 samples were generated for each value of n. The variation of t_2 and t_3 is considered linear in the simulation study except for the first cell, bearing in mind that many Irish stations have skewness close to zero . A similar approach was used in a number of simulation studies by Hosking and Wallis (1997). The underlying premise of the approach is that it is very unlikely to have a larger t_3 for a smaller t_2 for a site unless there is a presence of an outlier. The theoretical expression for the GNO is given in Hosking and Wallis (1997).

2. For each of the previously generated samples the corresponding values of D and A^2 are obtained from eqs (3.14) and (3.16) respectively. This yielded 5 sets of 10,000 values of $D(i)$ and $A^2(i), i = 1, 2, \ldots$ 10,000 for each n and the five combination sets of t_2 and t_3.

3. Critical points of the distribution of D and A^2 are obtained using the empirical sampling procedure for $D_p = D_{10000 \cdot p}$ and $A^2_p = A^2_{10000 \cdot p}$, where $D_{10000 \cdot p}$th denotes the 10000$\cdot$ th largest D in the sequence of 10,000 generated values of D. The same is applicable to $A^2_{10000 \cdot p}$. The reported statistics D_p and A^2_p in Table 3.14 are the averages of the 5 values of D_p and A^2_p.

Table 3.14: Critical values of the Kolmogorov-Smirnov and Anderson-Darling tests for the GNO distribution in case 1

N	D_p			A_p^2		
	p = 0.90	p = 0.95	p = 0.99	p = 0.90	p = 0.95	p = 0.99
10	0.21	0.229	0.274	0.521	0.776	1.49
15	0.176	0.192	0.224	0.541	0.707	1.449
20	0.154	0.168	0.196	0.531	0.691	1.359
25	0.139	0.151	0.176	0.533	0.745	1.51
30	0.128	0.139	0.162	0.534	0.751	1.45
35	0.119	0.129	0.151	0.539	0.73	1.402
40	0.112	0.121	0.141	0.548	0.73	1.492
45	0.106	0.115	0.133	0.551	0.73	1.438
50	0.101	0.109	0.126	0.556	0.737	1.43
100	0.072	0.078	0.091	0.559	0.721	1.327

3.5.1.3 Results

In this section the results of calculating goodness of fit indices for records of annual maxima of 30 years and over on the hypothesis of different distributions are presented. There were 74 stations with 30 years or more of data.

Eight probability distributions are considered in this part of the study. Four of these have two parameters, i.e Normal, EV1, LN and LO and the remaining four have three parameters, i.e. GNO, GEV, LN3 and GLO. Parameters of these distributions are estimated by the L-moments. The theoretical expressions of these distributions and the formulae of parameter estimation using L-moments are given in (Hosking and Wallis, 1997).

χ^2, D and A^2 statistics at each site are computed for each of the eight distributions. The results are tabulated for all statistics in Table 3.18 . Of 74 stations, Table 3.15 summarizes the number of times out of 74 stations that distributions are rejected by these three tests at the 5% significance level.

Of 74 stations the chi-square goodness of fit index rejected the hypothesis of the Normal for 12, EV1 for 15, GEV for 10 and the remaining distributions for 11 stations also at the 5% significance level. Each distribution in question was rejected about an same number of times and hence indicated no outright single choice.

Of the same 74 stations, the Kolmogorov-Smirnov goodness of fit index rejected the two parameter Normal and LO distributions 22 and 32 times respectively but rejected a lesser number, 13 and 16, respectively for LN and EV1 at the 5% level. For the three parameter distributions, it did not reject the GEV at any station while for the GNO and LN3 it rejected 3 and 2 cases respectively. Similar results were obtained for the AD test for 2-parameter distributions but the AD rejected about 7 stations for the 3-parameter distributions.

Given that the AD test is more powerful than the other two, the results particularly for the 2-parameter distribution are in accord once with what was found earlier in section 3.4. Three parameter distributions are rejected a less number of times than the two parameter distributions but the discrimination between them is unclear in determining an outright single choice.

Table 3.15: Number of times out of 74 that distributions were rejected by the tests at the 5% significance level

Test	Normal	EV1	LN	L0	GNO	GEV	GLO	LN3
chi-square	12	15	11	11	11	10	11	11
K-S	22	16	13	32	3	0	N/A	2
AD	27	17	13	35	6	8	7	5

3.5.2 Skewness versus Record Length Plots

It should be possible to make some inference about a suitable choice of distribution from skewness values especially in relation to suitability of two fixed skewness distributions, both with low parent or inherent skewness, namely, Normal and EV1. However, the substantial standard error with which skewness is estimated from finite length samples limits the reliability of any conclusions drawn. As descriptive statistics show that Irish flood data have a low positive skewness, hence the analysis in this sub-section is limited to between a zero skewness distribution, such as Normal, and a low positive skewness distribution such as EV1 where shape is fixed and skewness, g, is equal to 1.14.

The estimated skewness values are tabulated in Tables 3.16 and 3.17 for the A1 stations and the A2 stations respectively. These values are presented in ranked (increasing)

order and the corresponding curve pattern of the data series was noted alongside for the case of the EV1 probability plots.

It is noticed that 14 stations have skewness values less than zero and they predominantly have convex upwards curve patterns. This kind of behaviour is also recognized for a number of stations in the B category stations. This may be related to floodplain attenuation effects. An investigation on this issue is addressed at the end of this section. However, when these stations are plotted on normal paper, the majority of them show straight lines on the plot. Figure 3.19 shows the normal probability plots for these stations.

The estimated skewness values are shown on Figures 3.14 and 3.15 which show L-skewness or Hazen skewness plotted versus length of record along with confidence intervals based on Normal and EV1 distribution assumptions. The confidence intervals in all cases are obtained by simulation.

From Figure 3.14, it is seen that among the 14 stations which have negative skewness, 13 of them fall below the 95% lower EV1 confidence interval band. At the upper end of the skewness range, there are 40 stations which fall above the 95% upper confidence band for the Normal distribution. There is a group of 57 stations which falls within both sets of confidence intervals (CI).

Because of the relatively low values of observed skewness at many gauging stations, it is appropriate to ask whether the data as a whole could be considered to have come from a Normal distribution. All but 1 of the lower values fall above the 95% lower Normal CI while 40 of the larger values lie above the upper 95% Normal CI. If the data as a whole were Normal, then only 2.5% of 110, say 3 values, would lie above the upper 95% Normal CI, so that as a whole the Normal hypothesis could not be accepted. The same question is considered in Figure 3.14(b) in relation to the EV1 distribution. Only 7 values lie above the upper 95% EV1 CI (≈ 3) while 13 values (> 3) lie below the lower 95% EV1 CI. Thus the hypothesis that all the samples have come from the EV1 family is more plausible than was the case for the Normal hypothesis but nevertheless cannot be accepted for all the stations.

The overall result is that, on the basis of L-skewness values, the stations with the larger values of skewness could be modelled by an EV1 distribution but with the reservation that there is a significant number (20 stations or 18% of all stations) which are not in keeping with this choice. Thus while 82 % of station skewness values are

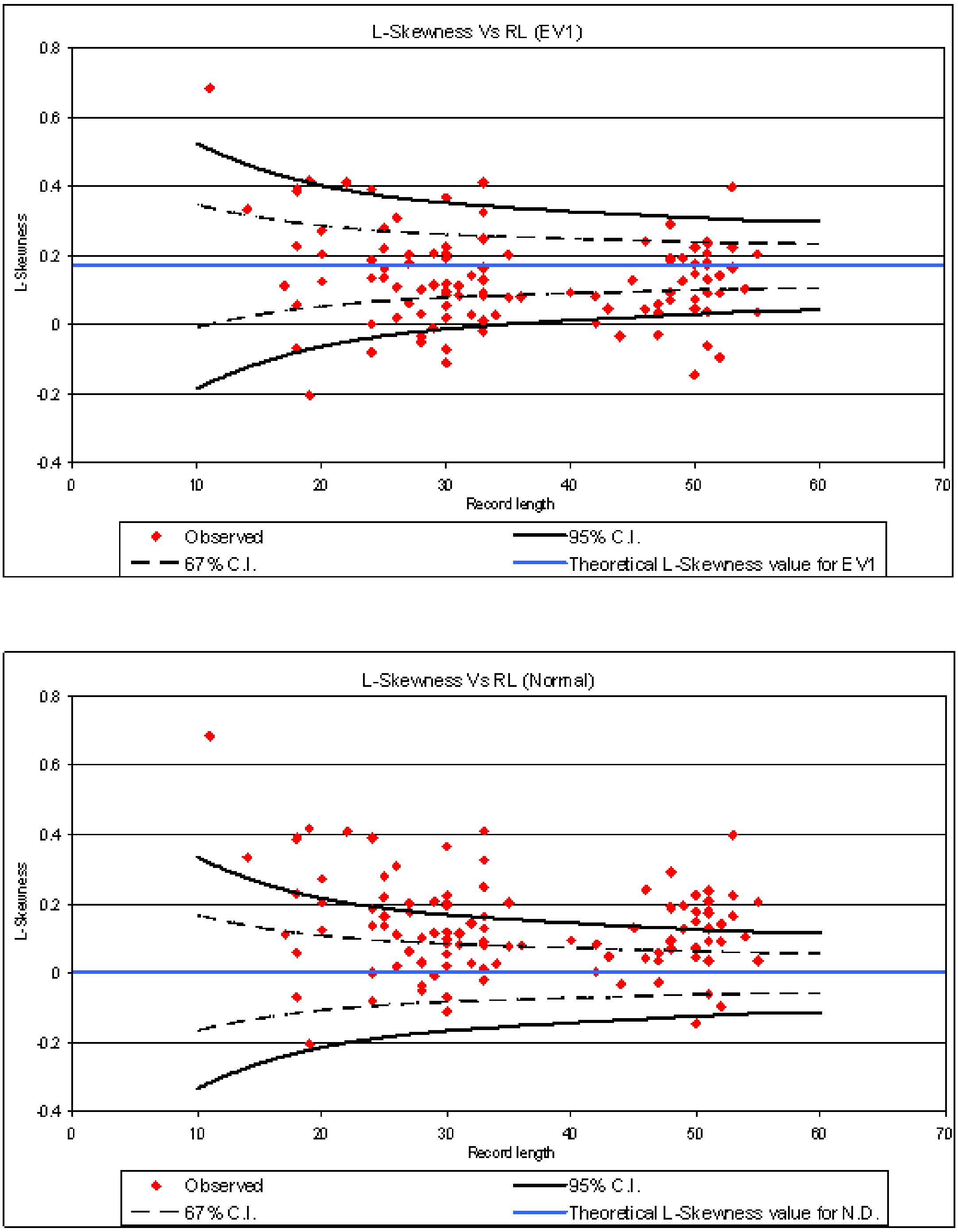
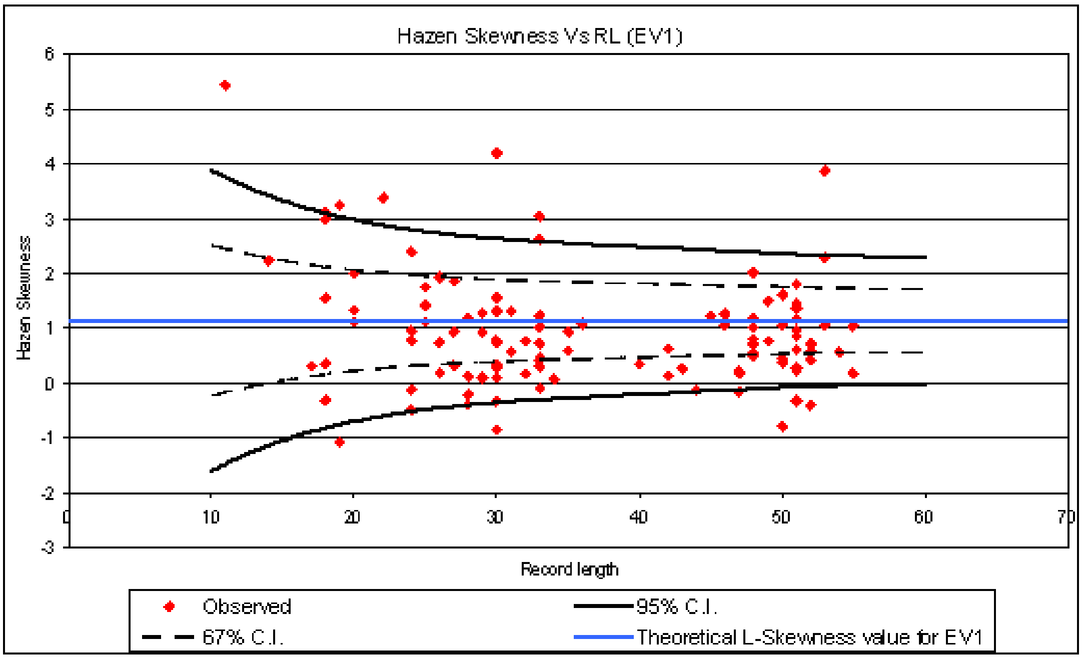


Figure 3.14: Observed L- skewness calculated at 110 A1 and A2 gauging stations plotted versus record length - Also shown are 67% and 95% Confidence Intervals for L-skewness calculated from (a) EV1 distribution samples and (b) Normal distribution samples. Samples are generated in both cases for a population of LCV =0.15.



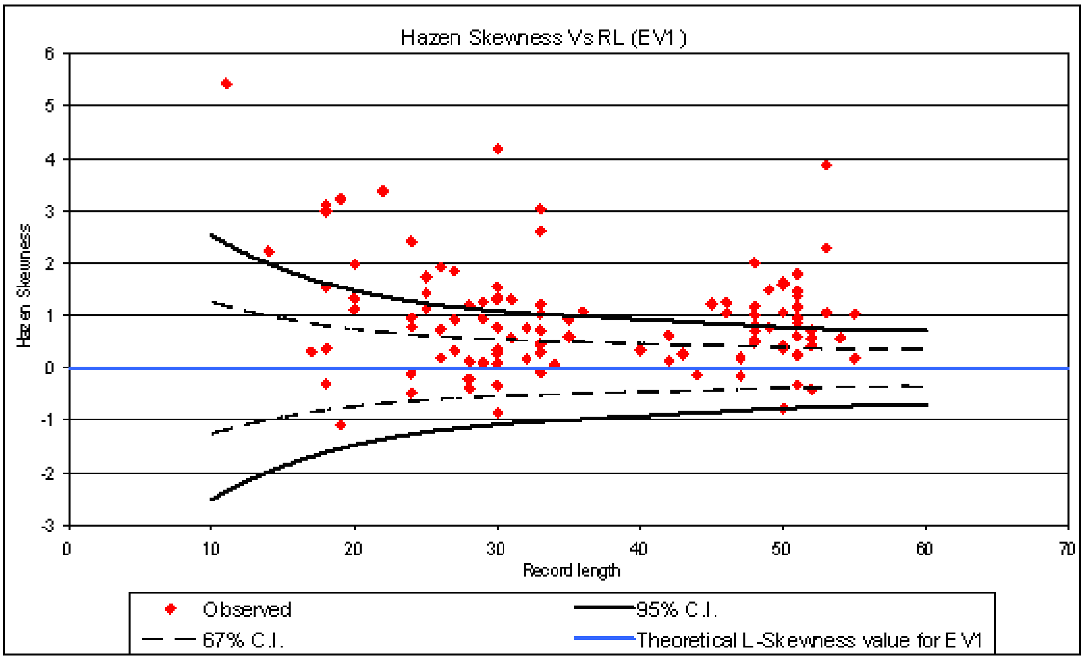


Figure 3.15: Observed Hazen skewness calculated at 110 A1 and A2 gauging stations plotted versus record length - Also shown are 67% and 95% Confidence Intervals for H-skewness calculated from (a) EV1 distribution samples and (b) Normal distribution samples. Samples are generated in both cases for a population of LCV =0.15.

consistent with EV1 there is a large proportion, 62%, which is also consistent with the normal distribution and 52% of all stations are consistent with both Normal and EV1 assumptions. Very similar findings can be obtained with the analogous plots of Hazen skewness as shown in Figure 3.15.

As noted earlier, the behaviour of convex upwards curve patterns on probability plots may be related to floodplain attenuation effect. As part of FSU Work Package 5.3, a catchment descriptor called Floodplain Attenuation Index, FAI, was developed for each gauged catchment in the study. The Index falls in the range 0 to 1 and it might be hoped that it could explain the behaviour described above. A plot (see Figure 3.16) of skewness , as an indicator of convex upwards behaviour , against FAI for those catchments whose skewness is less than 1 does not reveal any obvious correlation between these variables leading to the conclusion that such convex upwards behaviour cannot in fact be explained by the Flood Attenuation Index.

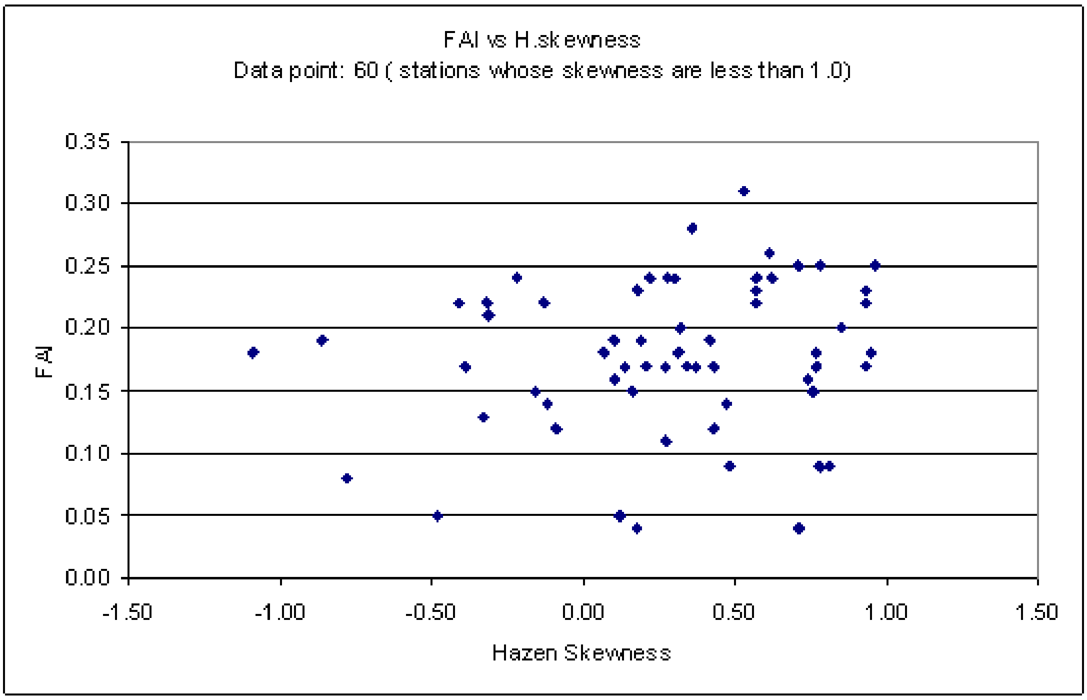


Figure 3.16: FAI plotted versus Hazen skewness. -

Thus it would seem that the goal of identifying of a single distribution for all stations in Ireland is of doubtful validity. Given the inherent difficulty involved in drawing conclusions from the above plots it seemed sensible to examine skewness in

another separate but complementary way. This is described in the next sub-section using moment ratio diagrams.

3.5.3 L-Moment Ratio diagrams

Another way to explore the suitability of different probability distributions with the help of dimensionless moments is to use moment ratio diagrams. At noted earlier, moment ratio diagrams based on conventional moment based skewness and kurtosis are regarded as having very poor discriminating capability (Cunnane, 1989, App. 3) but L-moment ratio diagrams can be more reliable as a diagnostic tool (Hosking and Wallis, 1997, p. 40). Nevertheless even L-moment ratio diagrams cannot be regarded as suitably reliable in the context of the available limited sample sizes.

The L-moment ratio diagram is a graph between L-kurtosis and L-skewness. Usually a 2-parameter distribution with a location and a scale parameter plots as a single point on such a diagram while a 3-parameter distribution with location, scale and shape plots as a line or curve on the diagram. Generally the distribution selection process involves plotting the sample L-moment ratios as a scatter plot and comparing them with theoretical L-moment ratio points or curves of candidate distributions.

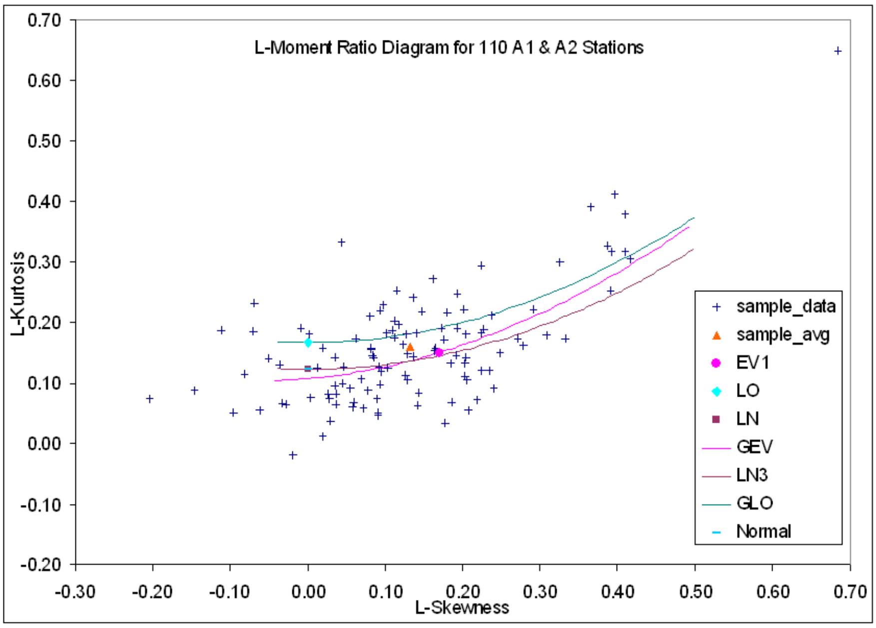


Figure 3.17: L-moment ratio diagram for annual maximum flow for 110 A1 and A2 stations. -

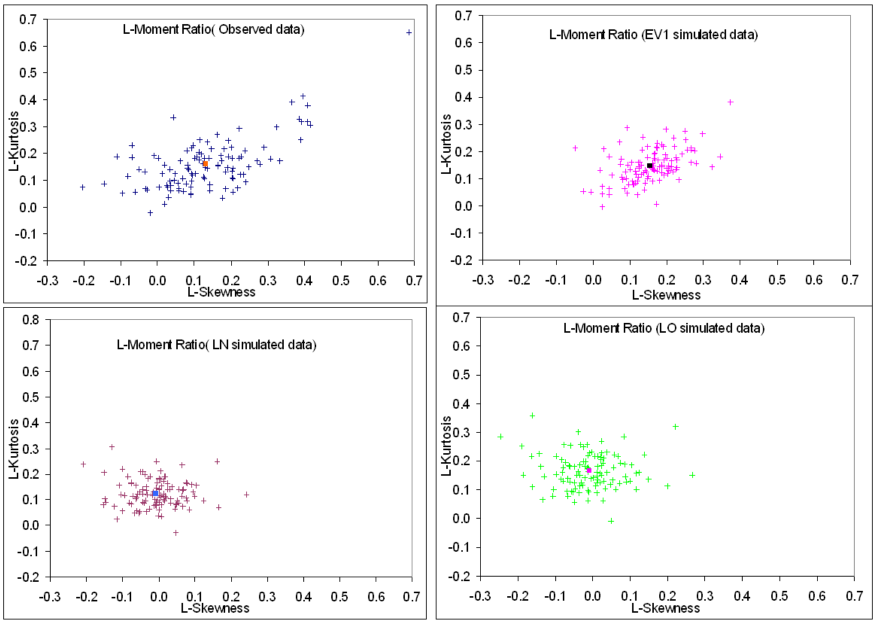


Figure 3.18: L-moment ratio diagram of a) observed data b) data simulated from EV1 distribution c) data simulated from LN distribution and d) data simulated from LO distribution. -

Plots of dimensionless L moment ratios are shown in Figure 3.17 which compares the observed and the theoretical relations between L-skewness and L-kurtosis for the annual maximum flood flows at the 110 A1 and A2 sites.

In the context of 2-parameter distributions alone, the data are on average closer to the population L-moments of an EV1 distribution rather than to those of the LO and LN distributions. Figure 3.18 shows sample L-moments of data simulated from EV1, LO and LN distributions each with the same record lengths and parameters as the actual records. In the case of the EV1 simulated data, the scatter of the points is narrower than in Figure 3.17 especially on the left hand side. There are many more negative historical values of L-skewness than there are among the simulated data. This may be related to the phenomenon called the flood plain attenuation effect which was mentioned in an earlier section. The LO and LN simulated L-moments cover the range of observed values in the negative L-skewness domain more adequately than the EV1 values but do not extend as far to the right as the EV1 simulated and observed L-skewness values. Four further realizations of this kind of simulation are shown in Figures 3.21 and 3.20.

While on the basis of average values, EV1 looks a more suitable candidate than either LO or LN, the occurrence of so many negative values of L-skewness among the observed data throws doubt on the suitability of EV1 as a universal 2 parameter model for Irish flood data. Note also that the average of the data points falls roughly half way between the GEV and GLO curves.

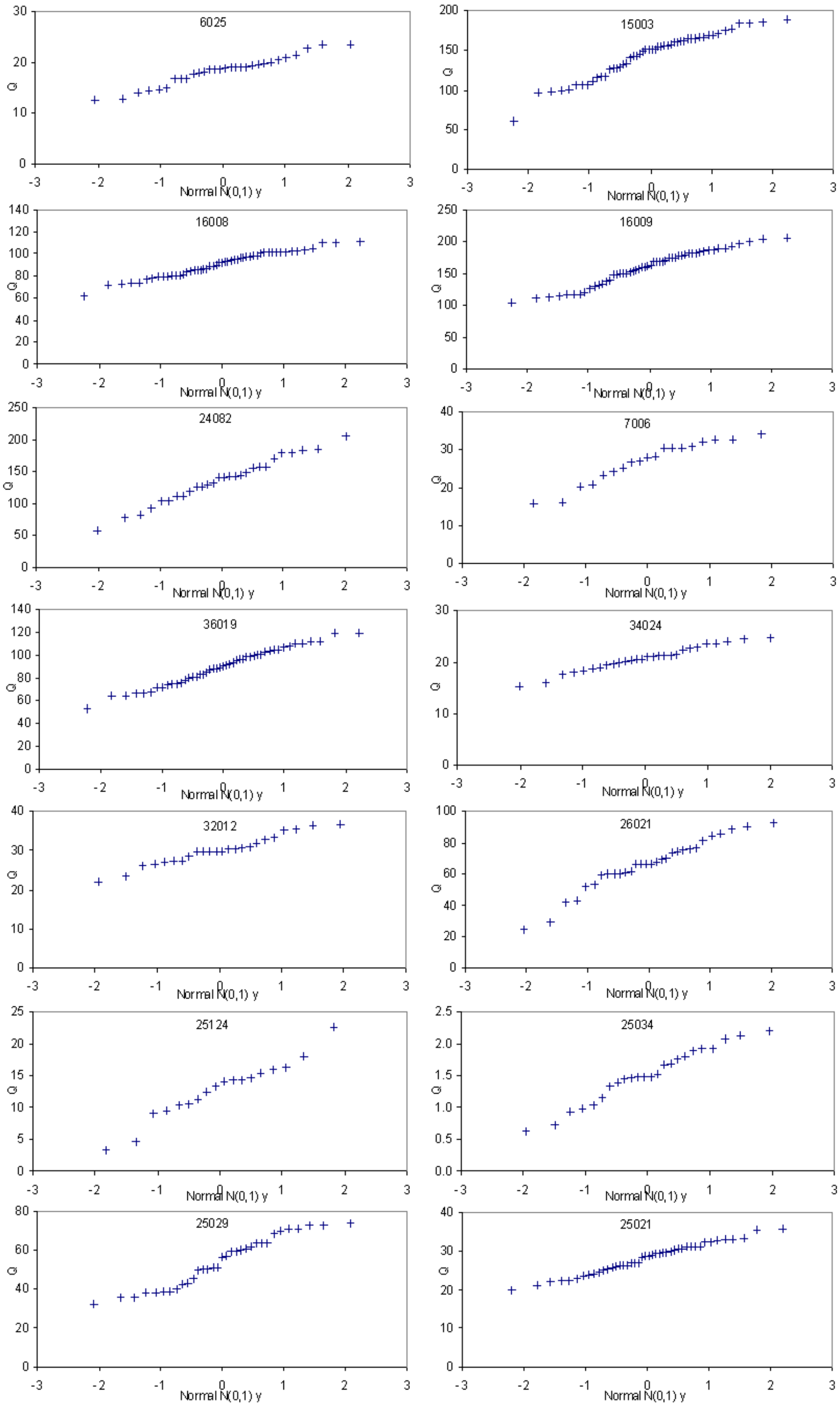


Figure 3.19: Normal probability plots for stations having g < 0. -

Table 3.16: A1 stations ranked by skewness

L-skewness	H-skewness	Station No, Location	N	Curve patterns on EV1
-0.07	-0.33	6025 DEE	30	convex
0.02	0.1	6013 DEE	30	convex
0.02	0.16	26020 CAMLIN	32	linear
0.03	0.18	25017 SHANNON	55	convex
0.03	0.19	14011Rathangan	26	linear
0.04	0.21	36012 ERNE	47	convex
0.04	0.22	14018 BARROW	51	linear
0.04	0.27	25027 OLLATRIM	43	convex
0.05	0.31	26059 INNY	17	convex
0.06	0.34	29001 RAFORD	40	convex
0.08	0.36	10022 CARRICKMINES	18	linear
0.09	0.42	16011 SUIR	52	convex
0.09	0.43	36018 DRONMORE	50	linear
0.09	0.57	25014 SILVER	54	linear
0.09	0.57	25023 LITTLE BROSNA	52	linear
0.1	0.57	25025 BALLYFINBOY	31	linear
0.1	0.61	14019 BARROW	51	convex
0.11	0.71	25006 BROSNA	52	linear
0.14	0.78	6070 MUCKNO	24	linear
0.14	0.81	6011 FANE	48	linear
0.14	0.93	34018 CASTLEBAR	27	linear
0.15	0.93	7009 BOYNE	29	linear
0.16	0.95	10021 SHANGANAGH	24	linear
0.18	1.01	25030 GRANEY	48	s-curve
0.18	1.05	26007 SUCK	53	linear
0.18	1.05	6026 GLYDE	46	s-curve
0.19	1.05	36010 ANNALEE	50	s-curve
0.19	1.12	14007 Derrybrock	25	linear
0.19	1.17	26019 CAMLIN	51	linear
0.21	1.17	9001 RYEWATER	48	linear
0.22	1.19	33070 CARROWMORE	28	linear
0.22	1.31	6014 GLYDE	30	linear
0.23	1.37	27002 FERGUS	51	linear
0.24	1.46	14006 BARROW	51	linear
0.24	1.49	26008 RINN	49	linear
0.27	1.92	31002 CASHLA	26	concave
0.31	1.98	8002 DELVIN	20	linear
0.32	2.4	9002 GRIFFEEN	24	s-curve
0.39	2.62	36015 FINN	33	concave
0.4	3.24	9010 DODDER	19	concave
0.41	3.37	29011 DUNKELLIN	22	concave
0.42	3.87	26006 SUCK	53	s-curve
0.68	5.43	8009 BALHEARY	11	concave

Table 3.17: A2 stations ranked by skewness

L-skewness	H-skewness	Station No, Location	N	Curve patterns on EV1
-0.2	-1.09	7006 MOYNALTY	19	convex
-0.15	-0.86	26021 INNY	30	convex
-0.11	-0.78	15003 DINAN	50	convex
-0.1	-0.48	25034 L. ENNELLTRIB	24	convex
-0.08	-0.41	16009 SUIR	52	convex
-0.07	-0.39	34024 POLLAGH	28	convex
-0.06	-0.32	16008 SUIR	51	convex
-0.05	-0.31	25124 BROSNA	18	convex
-0.04	-0.22	24082 MAIGUE	28	convex
-0.03	-0.16	36019 ERNE	47	convex
-0.03	-0.13	25021 BROSNA	44	convex
-0.02	-0.12	32012 NEWPORT	24	convex
-0.01	-0.09	25029 NENAGH	33	s-curve
0	0.07	35002 OWENBEG	34	convex
0	0.1	35001 OWENMORE	29	linear
0.01	0.12	19020 OWENNACURRA	28	convex
0.03	0.14	15001 RIVER KINGS	42	convex
0.03	0.18	14029 BARROW	47	convex
0.03	0.27	26005 SUCK	51	linear
0.04	0.28	24008 MAIGUE	30	linear
0.05	0.3	16001 DRISH	33	linear
0.06	0.32	24002 CAMOGUE	27	linear
0.06	0.34	35073 L.GILL	30	convex
0.07	0.37	14013 BURRIN	50	s-curve
0.07	0.43	34009 OWENGARVE	33	linear
0.08	0.47	26022 FALLAN	33	linear
0.08	0.48	19001 OWENBOY	48	linear
0.08	0.53	16004 SUIR	48	linear
0.08	0.59	15002 NORE	35	linear
0.08	0.62	25016 CLODIAGH	42	linear
0.09	0.71	27003 FERGUS	48	linear
0.09	0.73	39008 GARTAN	33	s-curve
0.09	0.74	29071 L.CUTRA	26	linear
0.1	0.76	35071 L.MELVIN	30	linear
0.11	0.77	26018 OWENURE	49	linear
0.11	0.77	29004 CLARINBRIDGE	32	linear
0.11	0.78	34011 MANULLA	30	s-curve
0.12	0.85	16003 CLODIAGH	51	linear
0.12	0.93	26009 BLACK	35	s-curve
0.13	0.96	15004 NORE	51	linear
0.13	1.03	39009 FERN O/L	33	linear
0.13	1.04	35005 BALLYSADARE	55	s-curve
0.13	1.08	34001 MOY	36	linear
0.14	1.12	24022 MAHORE	20	linear
0.14	1.22	23001 GALEY	45	linear
0.16	1.23	25044 KILMASTULLA	33	s-curve
0.16	1.25	18004 BALLYNAMONA	46	s-curve
0.17	1.26	34003 MOY	29	s-curve
0.18	1.3	30007 CLARE	31	linear
0.19	1.32	25040 BUNOW	20	linear
0.2	1.33	16005 MULTEEN	30	linear
0.2	1.42	14009 CUSHINA	25	s-curve
0.2	1.54	8005 SLUICE	18	linear
0.2	1.55	27001 CLAUREEN	30	linear
0.2	1.59	12001 RIVER SLANEY	50	linear
0.21	1.63	18005 FUNSHION	50	linear
0.22	1.74	8008 BROADMEADOW	25	linear
0.22	1.74	7033 BLACKWATER	25	s-curve
0.23	1.79	16002 SUIR	51	linear
0.25	1.85	36021 YELLOW	27	linear
0.28	2.01	14005 BARROW	48	linear
0.29	2.23	14034 BARROW	14	linear
0.33	2.29	26002 SUCK	53	s-curve
0.37	2.98	23012 LEE	18	concave
0.39	3.04	30061 CORRIB ESTUARY	33	concave
0.39	3.12	6031 FLURRY	18	concave
0.41	4.19	36031 CAVAN	30	s-curve

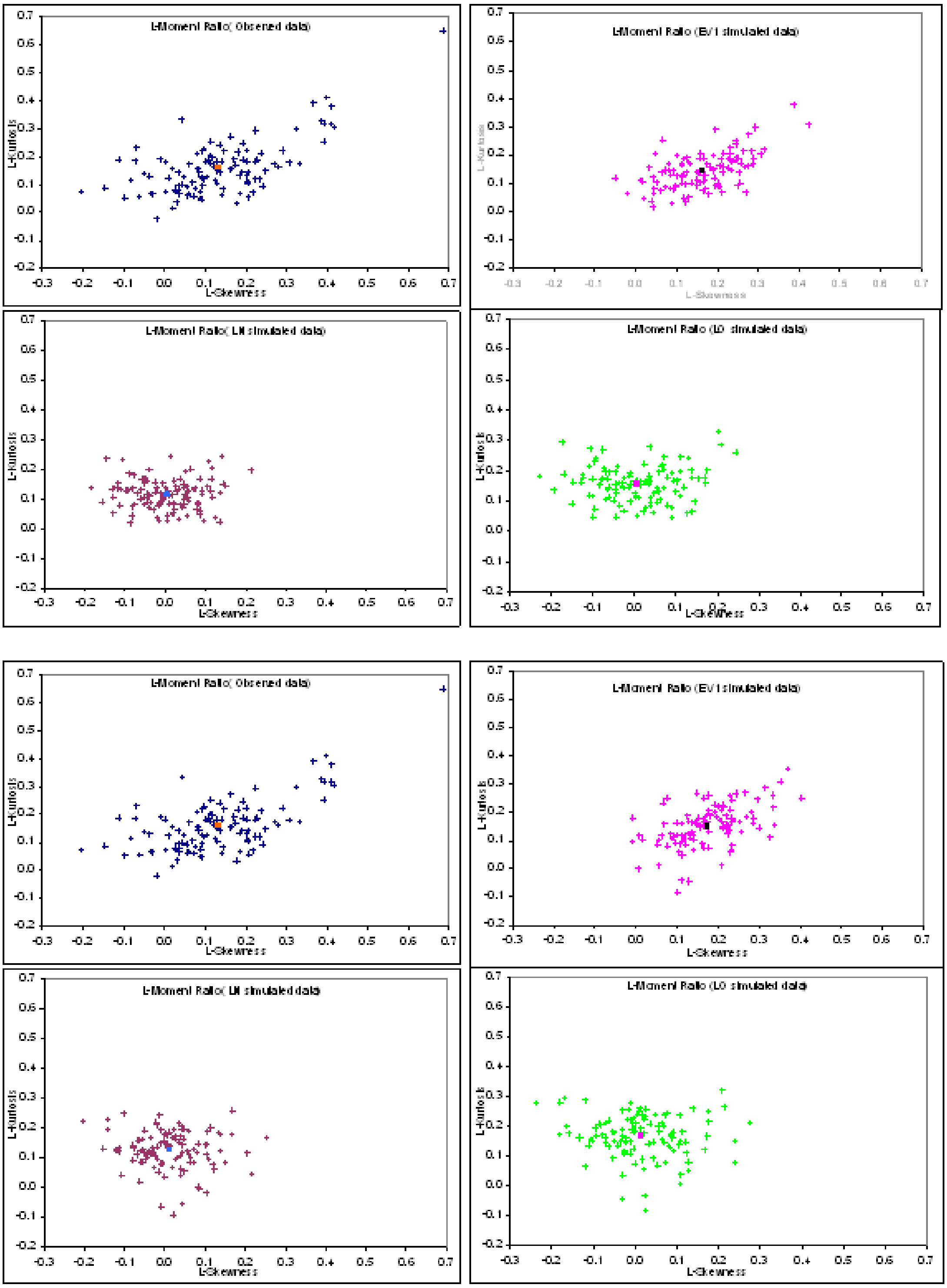


Figure 3.20: L-moment ratio diagrams of data simulated from EV1, LN and LO distributions - First set of two realizations .

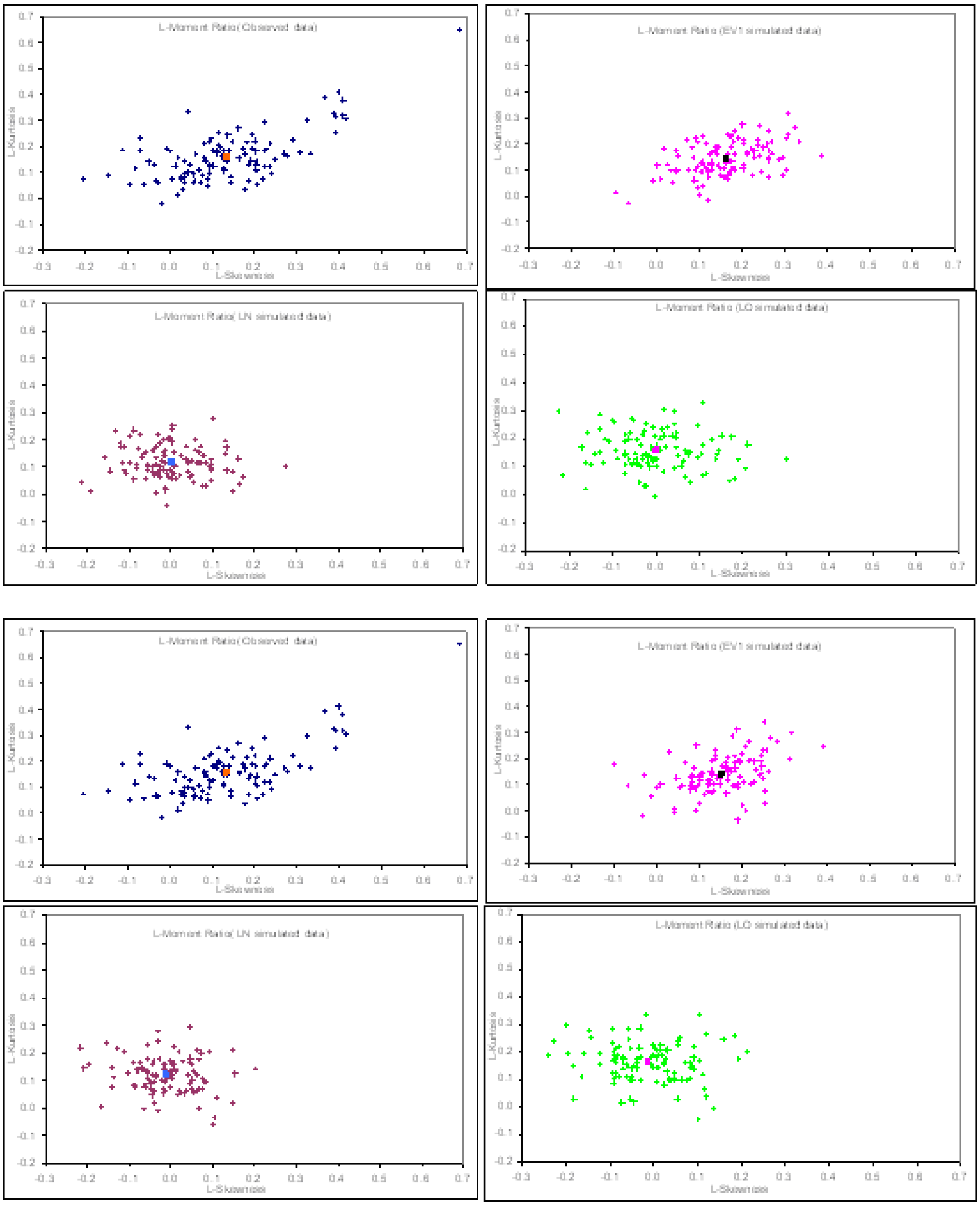


Figure 3.21: L-moment ratio diagrams of data simulated from EV1, LN and LO distributions - second set of two realizations.

Table 3.18: Chi-squared, SK and AD goodness of fit values. The asterisk sign indicate rejection of hypothesis at significance level of 5%

		Normal			EV1			LN			LO			GNO			GEV			GLO		LN3		
St. No	N	chi	SK	AD	chi	SK	AD	chi	SK	AD	chi	SK	AD	chi	SK	AD	chi	SK	AD	chi	AD	chi	SK	AD
6011	48	10.88	0.099	0.533	13.5*	0.103	0.478	8.25	0.08	0.378	11.25	0.101	0.665*	8.25	0.08	0.377	6.75	0.078	0.373	11.625*	0.533	8.25	0.08	0.377
6013	30	4.07	0.083	0.3	3.60	0.12	0.565	3.60	0.102	0.391	4.53	0.1	0.45	2.20	0.082	0.301	2.20	0.076	0.271	4.53	0.452	2.67	0.08	0.293
6014	30	4.53	0.131	0.778*	1.73	0.081	0.235	4.07	0.092	0.33	7.80	0.124	0.783*	0.80	0.067	0.178	3.13	0.067	0.191	2.20	0.252	1.73	0.064	0.179
6025	30	6.40	0.129	0.487	11.067*	0.198*	1.541*	13.4*	0.158	0.747*	5.93	0.132	0.453	5.93	0.1	0.419	5.93	0.107	0.486	4.07	0.361	6.87	0.103	0.439
6026	46	18.565*	0.175*	1.629*	5.26	0.109	0.599	6.83	0.121	0.711	20.13*	0.184*	1.756*	8.00	0.079	0.38	3.70	0.082	0.425	4.09	0.6	2.52	0.072	0.352
9001	48	9.00	0.119	0.844*	3.00	0.058	0.174	2.25	0.06	0.181	9.00	0.127*	0.848*	2.25	0.061	0.166	1.50	0.064	0.161	2.25	0.228	2.25	0.057	0.171
12001	50	7.60	0.137*	0.846*	9.40	0.067	0.244	9.40	0.069	0.245	5.44	0.139*	0.77*	9.40	0.067	0.249	6.52	0.066	0.243	6.52	0.253	9.40	0.07	0.245
14005	48	12.75*	0.222*	1.993*	9.38	0.152*	0.603	14.25*	0.168*	0.822*	8.63	0.224*	1.849*	4.88	0.102	0.335	4.13	0.101	0.295	4.50	0.29	4.13	0.104	0.3
14006	51	15.176*	0.17*	1.561*	5.65	0.099	0.398	9.88	0.131*	0.763*	13.412*	0.17*	1.444*	6.00	0.072	0.408	5.29	0.071	0.324	4.59	0.249	4.59	0.075	0.344
14013	50	6.88	0.075	0.388	7.60	0.111	0.533	4.00	0.094	0.346	6.16	0.08	0.546	6.16	0.083	0.306	7.60	0.081	0.289	5.44	0.477	6.16	0.085	0.303
14018	51	12.00	0.114	0.631	13.765*	0.182*	1.247*	17.294*	0.158*	0.833*	11.29	0.119*	0.812*	17.647*	0.128	0.653	18.353*	0.128	0.627*	19.765*	0.857*	18.353*	0.134*	0.668
14019	51	3.88	0.108	0.52	8.82	0.104	0.56	7.41	0.081	0.373	3.88	0.091	0.485	7.41	0.084	0.353	8.82	0.085	0.357	9.18	0.387	8.82	0.085	0.357
14029	47	6.04	0.098	0.372	4.51	0.09	0.459	6.04	0.081	0.284	9.11	0.115*	0.527	6.04	0.082	0.29	6.04	0.079	0.269	6.81	0.441	6.04	0.081	0.284
15001	42	4.86	0.069	0.236	5.62	0.103	0.765*	5.24	0.085	0.499	4.86	0.085	0.349	4.86	0.069	0.237	4.48	0.076	0.229	7.14	0.35	4.48	0.074	0.228
15002	35	6.83	0.119	0.411	5.00	0.101	0.393	4.54	0.089	0.331	8.20	0.124	0.509	5.46	0.088	0.306	6.37	0.087	0.299	6.37	0.403	6.37	0.087	0.295
15003	50	8.68	0.135*	0.847*	18.76*	0.2*	2.714*	10.48	0.168*	1.51*	10.48	0.145*	1.019*	5.44	0.079	0.344	4.36	0.075	0.322	3.64	0.508	4.36	0.078	0.322
15004	51	4.24	0.071	0.506	1.41	0.06	0.223	2.47	0.052	0.194	3.53	0.083	0.463	1.06	0.05	0.174	1.06	0.049	0.175	1.06	0.181	1.06	0.049	0.177
16001	33	1.18	0.08	0.203	8.94	0.139	0.688	4.09	0.111	0.363	7.00	0.092	0.218	4.09	0.083	0.206	4.09	0.081	0.231	7.00	0.224	5.06	0.082	0.211
16002	51	2.82	0.107	0.699	7.06	0.062	0.182	3.53	0.069	0.178	5.29	0.102	0.598	7.06	0.063	0.201	7.06	0.061	0.183	4.59	0.132	4.59	0.065	0.175
16003	51	23.647*	0.156*	1.582*	20.824*	0.116	0.889*	24*	0.139*	1.189*	24.706*	0.172*	1.848*	16.941*	0.104	0.815*	19.412*	0.109	0.856*	25.765*	1.12*	17.647*	0.104	0.813*
16004	48	8.25	0.083	0.302	6.00	0.092	0.4	10.50	0.077	0.21	11.63	0.091	0.397	13.125*	0.075	0.206	10.50	0.072	0.203	12.75*	0.303	10.50	0.074	0.204
16005	30	5.00	0.167*	0.783*	12.467*	0.116	0.317	3.13	0.137	0.49	6.87	0.174*	0.795*	9.667*	0.124	0.287	9.667*	0.127	0.299	8.733*	0.375	9.667*	0.126	0.292
16008	51	8.82	0.086	0.454	11.65	0.154*	1.413*	9.53	0.109	0.611	9.53	0.098	0.661*	9.18	0.078	0.387	9.18	0.068	0.334	8.82	0.595	9.18	0.076	0.374
16009	52	8.92	0.108	0.652	16.538*	0.167*	2.402*	11.00	0.132*	1.139*	7.88	0.121*	0.829*	6.15	0.074	0.352	3.73	0.069	0.298	6.50	0.544	5.12	0.077	0.358
16011	52	7.88	0.108	0.498	7.54	0.107	0.459	4.77	0.092	0.345	9.27	0.126*	0.687*	3.38	0.088	0.34	3.38	0.086	0.324	3.38	0.563	4.42	0.083	0.32
18004	46	8.78	0.132*	1.178*	18.957*	0.174*	1.952*	18.957*	0.14*	1.268*	11.91	0.118*	0.778*	13.478*	0.121	1.15*	13.478*	0.122	1.274*	9.17	0.764*	8.78	0.129	1.248*

18005	50	9.76	0.125*	1.236*	2.92	0.074	0.301	4.00	0.082	0.437	10.48	0.133*	1.117*	3.64	0.068	0.258	3.64	0.064	0.236	4.00	0.263	3.64	0.066	0.24
19001	48	10.13	0.102	0.627	4.88	0.099	0.528	5.25	0.086	0.409	15.75∗	0.1	0.689*	5.25	0.088	0.4	5.63	0.089	0.394	12∗	0.439	5.25	0.088	0.397
23001	45	7.20	0.104	0.622	14.4∗	0.114	0.435	17.6∗	0.107	0.425	7.20	0.116*	0.532	17.6∗	0.096	0.375	17.6∗	0.097	0.377	14∗	0.351	12∗	0.088	0.379
24008	30	2.67	0.104	0.27	4.07	0.104	0.393	2.67	0.084	0.292	2.67	0.122	0.324	2.67	0.097	0.236	1.27	0.094	0.231	2.67	0.291	1.27	0.095	0.233
25006	52	12.39	0.13*	0.883*	4.77	0.089	0.432	4.77	0.085	0.442	11.69	0.139*	0.857*	6.85	0.087	0.438	5.81	0.086	0.409	7.88	0.333	5.81	0.085	0.44
25014	54	5.67	0.088	0.428	6.33	0.065	0.276	7.00	0.056	0.158	5.00	0.089	0.442	7.00	0.055	0.159	7.00	0.056	0.157	7.67	0.198	7.00	0.056	0.158
25016	42	0.67	0.08	0.34	6.38	0.107	0.392	4.86	0.081	0.23	1.43	0.064	0.278	2.19	0.072	0.214	2.19	0.074	0.224	2.57	0.201	2.19	0.069	0.217
25017	55	5.38	0.075	0.276	6.69	0.118*	0.655	2.76	0.088	0.299	4.07	0.092	0.416	2.76	0.066	0.243	4.73	0.066	0.224	3.42	0.391	2.44	0.067	0.234
25021	44	2.64	0.08	0.247	12.05	0.149*	1.018*	12.05	0.106	0.387	6.32	0.091	0.399	3.46	0.067	0.221	3.04	0.067	0.19	3.04	0.37	3.04	0.069	0.216
25023	52	15.154∗	0.144*	1.309*	8.92	0.112	0.587	6.50	0.108	0.632	14.115∗	0.147*	1.557*	7.88	0.11	0.632	6.85	0.11	0.625*	11.00	0.85*	6.85	0.107	0.579
25025	31	4.45	0.102	0.272	2.65	0.096	0.322	2.65	0.082	0.248	3.10	0.099	0.256	2.65	0.081	0.187	3.10	0.083	0.19	1.74	0.189	2.65	0.087	0.19
25027	43	14.977∗	0.1	0.345	13.721∗	0.107	0.627	12.884∗	0.118	0.49	8.28	0.1	0.41	13.721∗	0.095	0.311	13.721∗	0.098	0.312	10.37	0.364	13.721∗	0.098	0.317
25029	33	7.00	0.114	0.568	13.303∗	0.167*	1.065*	8.46	0.143	0.729	6.51	0.129	0.762*	3.61	0.108	0.555	5.06	0.104	0.502	5.54	0.748*	6.51	0.114	0.555
25030	48	12.00	0.162*	1.025*	6.75	0.095	0.293	6.75	0.104	0.325	12.75∗	0.171*	1.112*	6.75	0.089	0.288	6.75	0.089	0.278	8.25	0.353	6.75	0.089	0.28
25044	33	9.42	0.196*	1.321*	5.54	0.125	0.443	6.51	0.132	0.492	9.91	0.196*	1.35*	8.46	0.094	0.3	7.97	0.093	0.296	6.03	0.339	10.394∗	0.092	0.282
26002	53	14.415∗	0.165*	1.85*	7.96	0.101	0.654	4.23	0.121	0.843*	10.00	0.156*	1.361*	12.377∗	0.114	0.756*	9.66	0.114	0.642*	8.98	0.432	17.132∗	0.103	0.606
26005	51	7.06	0.077	0.272	5.29	0.128*	0.783*	5.65	0.093	0.315	4.59	0.071	0.272	2.82	0.072	0.246	4.59	0.074	0.257	3.88	0.248	2.82	0.071	0.249
26006	53	22.906∗	0.224*	3.968*	15.434∗	0.154*	1.98*	12.38	0.15*	1.887*	21.547∗	0.223*	3.754*	10.68	0.153*	3.07*	10.68	0.134	1.805*	7.62	1.36*	12.717∗	0.128	1.806*
26007	53	7.96	0.138*	0.739	3.55	0.071	0.183	3.55	0.104	0.287	8.64	0.146*	0.754*	2.19	0.073	0.186	3.21	0.073	0.18	4.57	0.206	2.19	0.075	0.181
26008	49	21.714∗	0.148*	1.099*	16.204∗	0.113	0.521	19.143∗	0.111	0.633	14∗	0.148*	1.066*	9.22	0.117	0.599	15.837∗	0.113	0.544	12.163∗	0.456	15.837∗	0.114	0.542
26009	35	10.03	0.142	1.054*	4.09	0.107	0.468	10.03	0.116	0.718	10.03	0.146*	1.141*	5.00	0.11	0.457	3.63	0.106	0.45	9.571∗	0.511	5.00	0.108	0.449
26018	49	8.49	0.088	0.656	4.82	0.087	0.32	8.49	0.062	0.324	9.59	0.099	0.655	7.02	0.071	0.289	7.02	0.071	0.291	7.02	0.389	7.02	0.072	0.289
26019	51	15.176∗	0.127*	1.524*	6.35	0.085	0.419	11.29	0.101	0.671	20.118∗	0.132*	1.505*	5.29	0.071	0.256	4.24	0.077	0.29	4.24	0.414	3.18	0.073	0.257
26020	32	3.00	0.07	0.182	3.00	0.126	0.528	3.50	0.094	0.248	3.00	0.079	0.252	3.00	0.072	0.184	3.00	0.074	0.175	5.00	0.257	3.00	0.075	0.187
26021	30	2.20	0.142	0.425	5.00	0.196*	1.768*	4.07	0.203*	1.234*	2.67	0.127	0.349	5.00	0.108	0.261	6.40	0.116	0.298	4.53	0.222	6.40	0.108	0.321
26022	33	6.51	0.104	0.386	7.00	0.114	0.371	6.03	0.101	0.302	8.46	0.122	0.515	4.09	0.094	0.3	4.09	0.092	0.29	5.54	0.45	4.09	0.094	0.287
27001	30	2.20	0.156	0.828*	1.73	0.091	0.297	1.27	0.119	0.407	2.20	0.148*	0.633	3.13	0.088	0.334	1.73	0.084	0.302	1.27	0.202	1.73	0.089	0.297
27002	51	6.00	0.108	0.961*	6.00	0.082	0.285	6.71	0.067	0.351	4.94	0.106	0.762*	6.00	0.087	0.321	6.00	0.084	0.292	5.65	0.18	6.00	0.081	0.282
27003	48	6.75	0.102	0.65	10.13	0.104	0.761*	13.125∗	0.107	0.556	4.13	0.109	0.48	12.75∗	0.101	0.503	12.75∗	0.102	0.523	5.63	0.326	8.63	0.101	0.513
29001	40	6.00	0.115	0.507	3.60	0.087	0.352	3.60	0.082	0.31	9.60	0.126*	0.624	2.00	0.081	0.313	2.00	0.079	0.301	4.40	0.399	2.00	0.08	0.306

29004	32	3.50	0.13	0.665	6.00	0.082	0.307	5.00	0.112	0.434	3.50	0.144*	0.678*	6.00	0.087	0.308	6.00	0.088	0.31	6.00	0.388	6.00	0.087	0.304
30007	31	7.16	0.126	0.732	7.16	0.125	0.528	7.16	0.108	0.493	4.45	0.109	0.566	7.16	0.107	0.483	7.16	0.108	0.494	7.16	0.439	7.16	0.107	0.487
30061	33	17.182*	0.284*	2.952*	12.818*	0.215*	1.334*	11.848*	0.219*	1.557*	15.242*	0.28*	2.677*	4.58	0.146*	1.403*	7.97	0.129	0.799*	8.46	0.666	6.51	0.128	0.792*
34001	36	6.67	0.116	0.39	6.22	0.137	0.576	4.00	0.104	0.338	6.67	0.105	0.284	4.00	0.1	0.328	4.44	0.102	0.358	4.44	0.249	3.56	0.094	0.321
34009	33	5.54	0.109	0.37	11.848*	0.134	0.434	7.97	0.107	0.272	7.49	0.115	0.36	7.97	0.108	0.272	7.97	0.11	0.275	5.54	0.257	7.97	0.106	0.271
34011	30	3.60	0.135	0.504	4.07	0.131	0.464	3.60	0.112	0.347	3.60	0.12	0.366	5.47	0.111	0.346	5.47	0.112	0.356	5.47	0.235	3.60	0.11	0.344
35002	34	5.53	0.095	0.279	5.53	0.118	0.493	4.59	0.09	0.279	7.88	0.105	0.378	5.53	0.085	0.264	5.53	0.082	0.25	6.00	0.362	5.53	0.084	0.258
35005	55	15.855*	0.17*	1.464*	4.07	0.101	0.38	7.67	0.121*	0.571	21.418*	0.166*	1.48*	3.42	0.089	0.32	3.42	0.087	0.324	5.38	0.399	3.42	0.087	0.313
35071	30	2.67	0.124	0.467	5.47	0.112	0.32	2.67	0.097	0.27	3.13	0.111	0.39	2.67	0.093	0.263	2.67	0.094	0.262	2.67	0.194	2.67	0.09	0.258
35073	30	5.47	0.118	0.386	0.80	0.1	0.345	1.73	0.088	0.266	8.27	0.113	0.385	3.13	0.084	0.262	3.13	0.083	0.261	4.07	0.247	3.13	0.089	0.267
36010	50	10.84	0.118	1.118*	11.20	0.096	0.568	9.04	0.086	0.572	6.88	0.106	0.852	13.72*	0.086	0.557	13.72*	0.087	0.541	13.36*	0.377	16.24*	0.084	0.536
36012	47	4.13	0.073	0.158	5.66	0.099	0.536	6.81	0.072	0.235	2.98	0.084	0.248	2.98	0.06	0.138	2.98	0.057	0.128	1.83	0.232	2.98	0.058	0.133
36015	33	7.49	0.228*	1.775*	8.94	0.159*	0.602	6.03	0.167*	0.691	8.46	0.225*	1.509*	5.54	0.105	0.497	5.54	0.1	0.342	5.06	0.266	5.54	0.105	0.338
36018	50	3.64	0.095	0.334	9.04	0.163*	0.68	9.40	0.129*	0.352	11.92	0.101	0.445	9.40	0.113	0.312	9.40	0.113	0.311	10.84	0.441	9.40	0.114	0.311
36019	47	1.06	0.06	0.195	6.04	0.113	0.946*	4.13	0.08	0.39	4.51	0.077	0.344	1.45	0.059	0.177	1.45	0.054	0.143	1.83	0.326	1.45	0.058	0.166
36031	30	9.20	0.187*	1.76*	5.93	0.13	0.851*	10.6*	0.153	1.102*	9.20	0.195*	1.64*	9.667*	0.181*	2.651*	12*	0.168	0.786*	8.267*	0.682*	8.267*	0.162*	0.832*
39008	33	4.09	0.089	0.423	2.15	0.091	0.197	2.15	0.068	0.178	4.09	0.097	0.428	2.15	0.073	0.176	2.15	0.074	0.176	2.15	0.23	2.15	0.075	0.173
39009	33	8.46	0.119	0.611	5.54	0.081	0.216	5.06	0.077	0.246	8.46	0.111	0.572	4.58	0.078	0.217	5.54	0.079	0.215	5.54	0.235	4.58	0.076	0.215

3.6 Flood Seasonality

Annual maximum floods were examined, with respect to seasonality, for 202 stations (A1, A2 and B categories). A total of 6969 station years of data were involved. The following seasonal characteristics were found:

1. Floods are found to occur at all times of year, but most rivers register at least 65% in the winter time i.e. the October-March half year. In all, 6094 of the 6969 annual maxima occurred in winter. Table 3.19 lists all stations with the corresponding percentage of peak floods in the October-March period. There are 11 stations where no peak floods were observed in the summer time i.e. April-September.

2. Figure 3.22 shows the frequency of flood peaks in each of the 12 months. The months December and January are associated with the greatest number of AM flood events followed by November and February. In the summer time, considerable numbers of flood peaks were observed in August while July has the least number of AM floods.

3. Figure 3.23 illustrates the seasonal distribution of AM flood peaks for the 202 gauging stations. The radial position indicates season and the distance from the centre shows magnitude of the flood peak. As stated above, most of the big floods were found in the months between October and February. There were six occasions when more than 600 cumec discharges were observed, one of which occurred in August. However, all these discharges occurred at a single station, 23002 Feale at Listowel, with a drainage area 646 km^2.

4. As drainage area exercises a substantial control over flood peaks, specific AM (flood magnitude divided by catchment area) values are plotted on Figure 3.24, which is similar in appearance to Figure 3.23. A number of large specific floods were observed in August and September. In fact among the biggest five, four were from a single station Sandy Mills (1041) on the river Deele with drainage area 45.3 km^2, while the other one occurred in August during Hurricane Charlie in 1986 at the station Waldron's Bridge (9010) on the river Dodder. It is also known that very large floods occurred in other rivers along the east coast as a

result of Hurricane Charlie , especially on the river Dargle where Q/A may have exceeded 3 cumec/ km^2. However this and others are not included in Figures 3.23 and 3.24 as no gauged record exists.

5. Finally the months corresponding to occurrence of maximum flow in each AMF data series are shown graphically in Figure 3.25. Among the 202 stations examined, Table 3.20 shows that the largest value in the series occurred during the summer months for 20 of them. Of those 20, 8 occurred in August. The month of December provided the series maximum in 91 stations (45%).

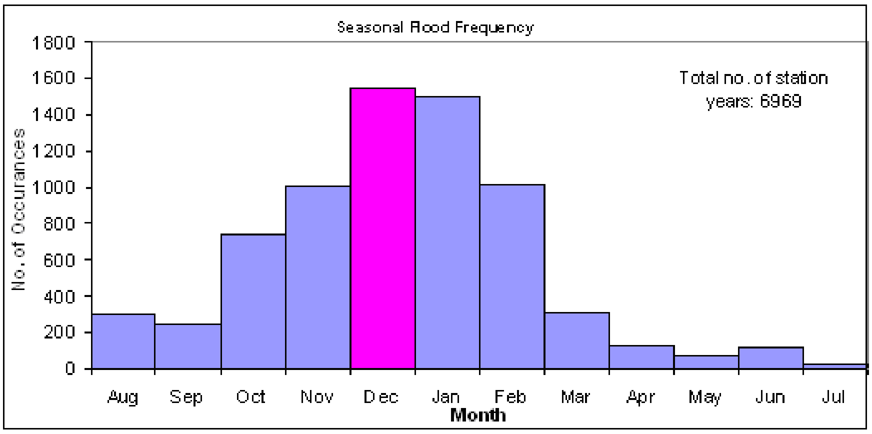


Figure 3.22: Seasonal occurrence of annual maximum flood values in 202 stations. -

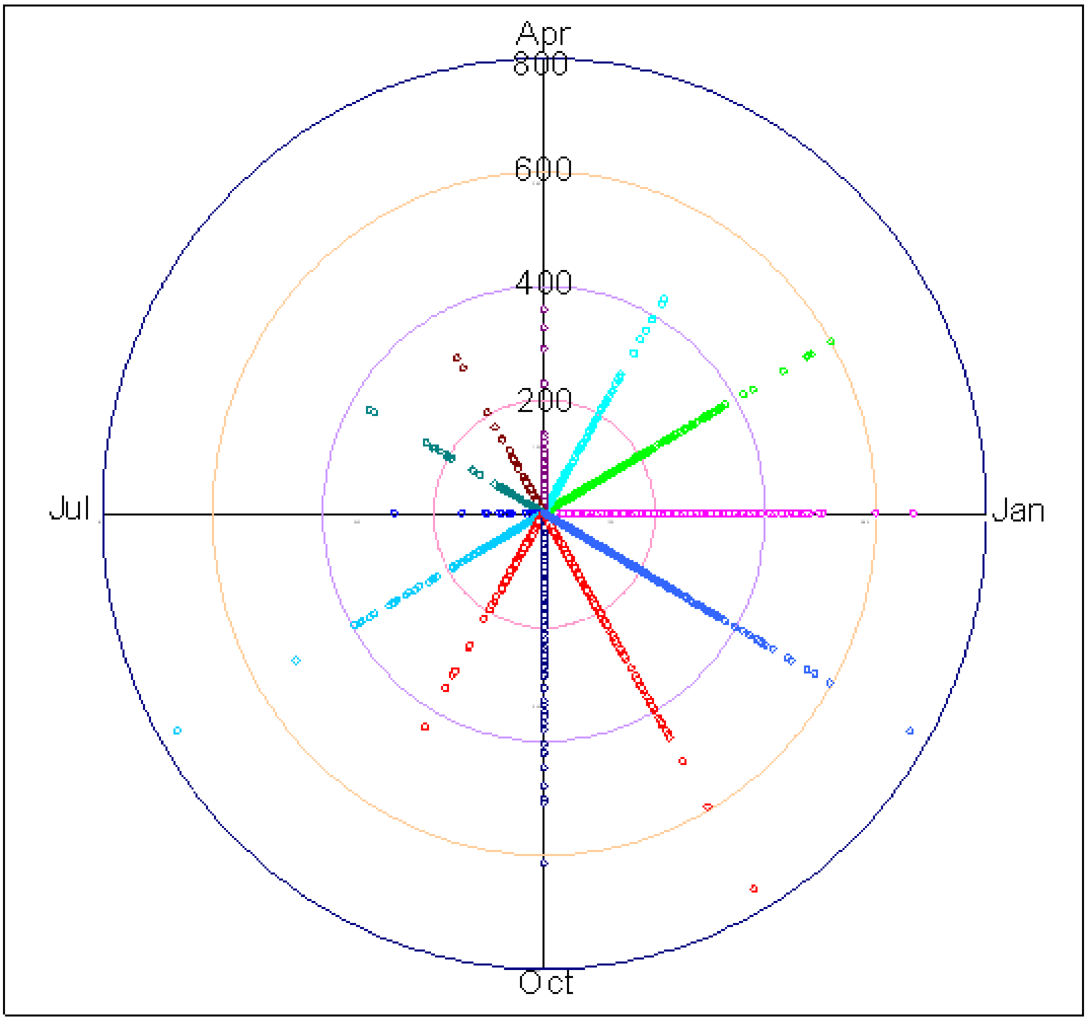


Figure 3.23: Flood seasonality showing actual magnitude in 202 Irish stations.
-

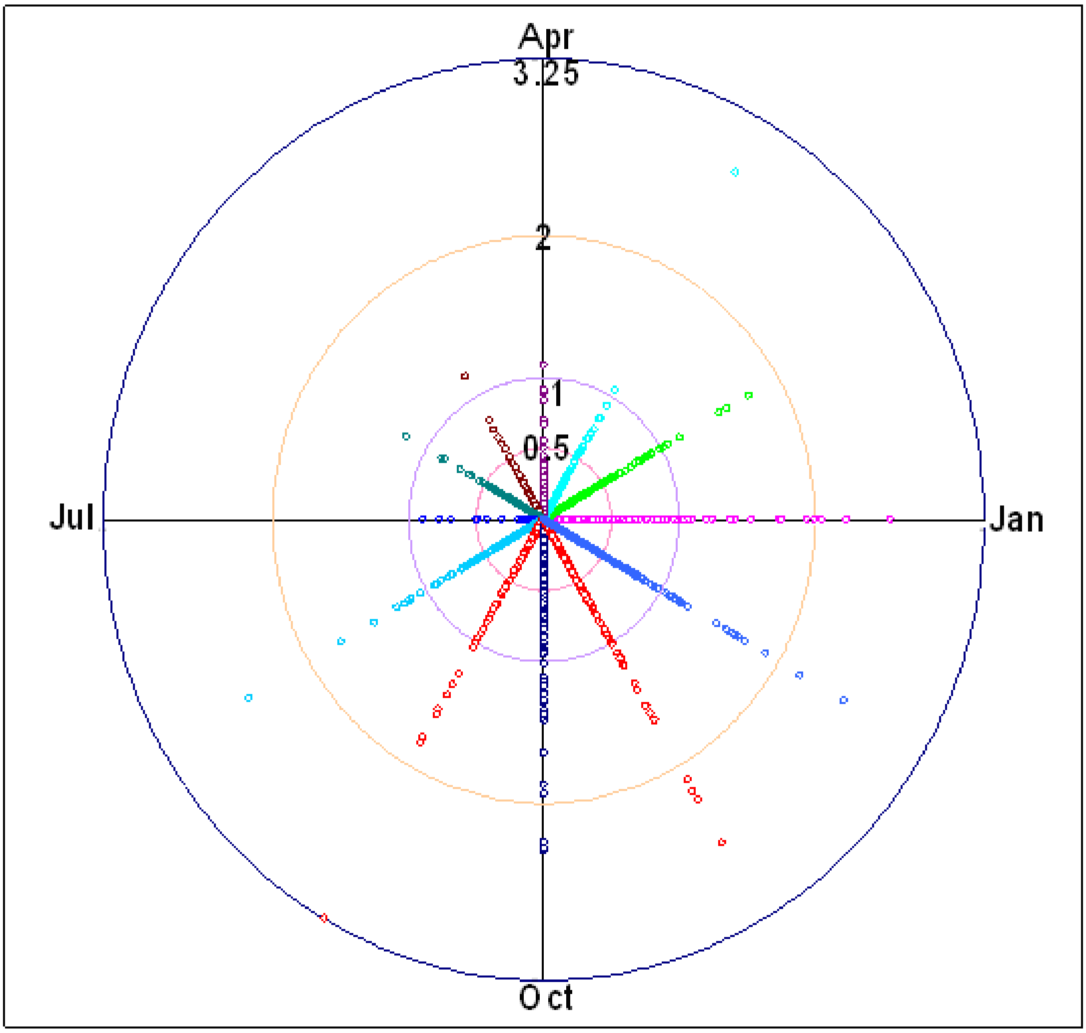


Figure 3.24: Flood seasonality with specific flood magnitude in 201 Irish stations. -

Table 3.19: Station Number with percent of peak floods in the October-March

Station No	% of events	Station No	% of events	Station No	% of events	Station No	% of events	Station No	% of events
1041	91	14006	88	20006	84	26009	89	34018	100
1055	89	14007	76	22006	86	26010	91	34024	90
3051	100	14009	80	22009	79	26012	98	34029	86
6011	100	14011	81	22035	100	26014	100	35001	90
6013	90	14013	82	23001	87	26017	94	35005	93
6014	97	14018	96	23002	80	26018	96	35011	89
6021	94	14019	92	23012	83	26019	88	35071	90
6025	80	14029	98	24001	88	26020	88	35073	87
6026	91	14033	82	24002	81	26021	90	36010	96
6030	63	14034	82	24004	87	26022	88	36011	96
6031	94	15001	88	24008	87	26058	83	36012	98
6033	80	15003	84	24011	85	26059	96	36015	94
6070	85	15004	92	24012	88	26108	100	36018	92
7002	96	15005	90	24013	87	27001	73	36019	96
7003	83	15012	75	24022	100	27002	96	36021	73
7004	98	16001	88	24030	88	27003	88	36027	93
7005	91	16002	90	24082	82	27070	79	36031	90
7006	84	16003	82	25001	73	28001	82	36071	75
7007	91	16004	90	25002	76	29001	85	38001	73
7009	90	16005	87	25003	78	29004	81	39001	83
7010	93	16006	82	25004	87	29007	82	39008	82
7011	98	16007	84	25005	87	29011	86	39009	88
7012	94	16008	90	25006	92	29071	83		
7033	84	16009	90	25011	78	30001	94		
7041	100	16011	90	25014	81	30004	97		
8002	76	16012	92	25016	93	30005	86		
8003	78	16013	88	25017	98	30007	94		
8005	67	16051	85	25020	91	30012	100		
8007	73	18001	90	25021	91	30021	92		
8008	84	18002	94	25023	71	30031	100		
8009	67	18003	96	25025	90	30037	57		
8011	87	18004	85	25027	88	30061	91		
8012	68	18005	84	25029	85	31002	81		
9001	77	18006	93	25030	92	31072	69		
9002	68	18016	83	25034	83	32011	65		
9010	63	18048	83	25038	82	32012	88		
9035	44	18050	92	25040	75	33001	68		
10002	83	19001	92	25044	83	33070	89		
10021	63	19014	90	25124	94	34001	97		
10022	60	19016	100	25158	72	34003	97		
10028	69	19020	89	26002	91	34004	92		
11001	91	19031	89	26005	94	34007	85		
12001	90	19046	67	26006	94	34009	91		
12013	73	20001	93	26007	94	34010	58		
14005	90	20002	94	26008	92	34011	97		

Table 3.20: Station number with month corresponding to maximum flow in an AM data series

Station No.	Month	Station No.	Month	Station No.	Month	Station No.	Month	Station No.	Month
1041	12	12001	11	19014	12	25034	12	30031	12
1055	12	12013	11	19016	1	25038	12	30037	11
3051	12	14005	12	19020	12	25038	2	30061	1
6011	12	14006	12	19031	2	25040	10	31002	10
6013	12	14007	2	19046	3	25044	12	31072	7
6014	12	14009	2	20001	10	25124	2	32011	9
6021	12	14011	2	20002	12	25158	1	32012	12
6025	11	14013	12	20006	12	26002	10	33001	9
6026	12	14018	2	22006	12	26005	12	33070	10
6030	11	14019	2	22009	12	26006	12	34001	10
6031	12	14029	12	22035	1	26007	11	34003	10
6033	12	14033	12	23001	12	26008	11	34004	10
7002	11	14034	2	23002	8	26009	10	34007	10
7003	12	15001	12	23012	8	26010	11	34009	11
7004	11	15003	8	24001	10	26012	2	34010	6
7005	12	15004	2	24002	2	26014	11	34011	10
7006	10	15005	12	24004	12	26017	12	34018	12
7007	12	15012	2	24008	12	26018	12	34024	11
7009	12	16001	12	24011	12	26019	10	34029	1
7010	11	16002	12	24012	12	26020	12	35001	10
7011	11	16003	12	24013	12	26021	12	35005	10
7012	12	16004	12	24022	12	26022	12	35011	10
7033	1	16005	9	24030	8	26058	1	35071	1
7041	1	16006	12	24082	12	26059	1	35073	12
8002	11	16007	9	25001	12	26108	1	36010	12
8003	8	16008	12	25002	10	27001	1	36011	12
8005	11	16009	12	25003	12	27002	12	36012	12
8007	8	16011	11	25004	12	27003	12	36015	10
8008	11	16012	11	25005	11	27070	12	36018	12
8009	6	16013	11	25006	12	28001	12	36019	12
8011	12	16051	1	25011	12	29001	1	36021	10
8012	8	18001	10	25014	12	29004	1	36027	1
9001	12	18002	10	25016	12	29007	1	36031	10
9002	12	18003	12	25017	12	29011	1	36071	4
9010	12	18004	1	25020	1	29071	12	38001	9
9035	10	18005	12	25021	12	30001	12	39001	9
10002	11	18006	11	25023	11	30004	11	39008	12
10021	5	18016	1	25025	1	30005	12	39009	12
10022	5	18048	12	25027	12	30007	2		
10028	10	18050	12	25029	2	30012	12		
11001	8	19001	1	25030	12	30021	1		

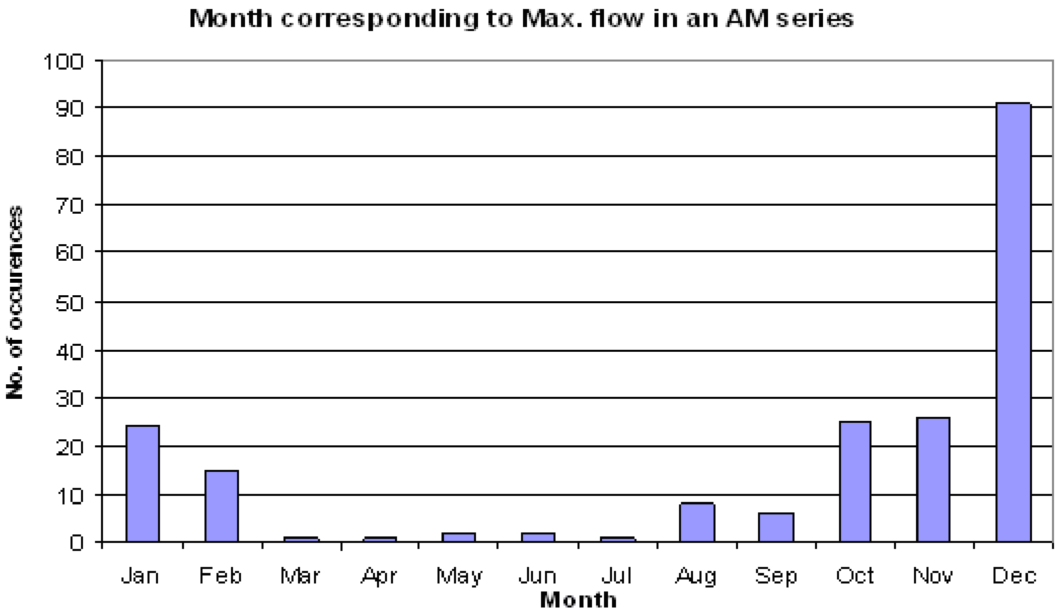


Figure 3.25: Month corresponding to maximum flow in each of 202 AM series.
-

3.7 Conclusion

The following conclusion can be drawn from the studies described in this chapter:

1. The descriptive statistics show that Irish annual maximum flood data have low CV and low skewness whether measured by traditional statistics or by L-moment statistics. Furthermore, the values are similar to some extent to what is observed for the Irish data in FSR (1975).

2. Examination of probability plots of observed data indicate that EV1 and LN distributions are most often the most suitable 2-parameter distributions but then only in about 50% of overall cases. The remaining cases show varying departures from linearity from concave downwards, to S-shaped, to convex upwards. In a good number of stations, the largest and 2^{nd} largest floods are not appreciably larger than the 3^{rd} largest flood leading to a flattening out of the probability plot. Examination of probability plots of random samples from EV1 parent distributions show that in about 20% of cases departures from linear patterns are observed. So another interpretation is that of the remaining 50% of stations

which have plots away from linear patterns on an EV1 plot, 40% of them can be attributed to randomness with the remaining 60% due to some additional reasons.

3. The adoption of one distribution on the basis of goodness-of -fit tests, particularly using the chi-square test, proved inconclusive. Based on SK and AD tests, LN and EV1 are the most suitable 2-parameter distribution in about 75% of overall cases. All of the four 3-parameter distributions (GEV, GLO, GNO and LN3) considered in this study are accepted in about 90% of cases based on the AD test.

4. Examination of suitability of candidate distributions for AM data using skewness values and L-moment ratio plots does not lead to entirely unambiguous results partly because of the preponderance of low or negative skewness values. In this context, even the Normal distribution, usually not regarded as an appropriate flood distribution, is not ruled out for some stations!

5. Seasonal analysis shows that two thirds of AM flows occur during the winter months of October to March and that at 11 stations no AM floods occurred during the summer months. July is the least likely month in which AM floods occur while August and September are the most likely summer months to have AM floods. At 20 out of the 202 stations examined, the largest flow on record occurred during summer, 8 of them in August, while 45% of stations had their largest flood on record during December.

).

4

Pooled based estimate of Q_T in the Irish context

4.1 Introduction

The traditional method of estimating a design flood, Q_T, is the at-site based method where flood data from the site of interest alone are used. A reasonable estimate of Q_T using the at-site approach requires a long record of AM data. If the record is not sufficiently long, the standard error of estimate of Q_T is large. An investigation in this context, which is summarized in Table 4.1, shows that a criterion of N>T is needed for a 2-parameter distribution model and a criterion of N>2T is needed for a 3-parameter distribution model in order to keep $se < 10\% Q_T$, where N is available record length and T is required return period. Irrespective of the selected distribution, the criterion of N≥ T is recommended by Hosking and Wallis (1997, p. 2).

Generally due to lack of at-site flood data, the above criterion of N≥ T, is not meet for the flood magnitudes of large return periods which are these of interest from a design point of view. Regional flood frequency analysis (FSR, 1975), i.e. pooling analysis (FEH, 1999), is therefore used to provide a framework for design floods. In pooling analysis, flood data are pooled from other gauging stations that possess similar hydrological behaviours to the at-site station. A very common way to implement the regional/pooling procedure is by using the index flood method proposed by Dalrymple (1960).

Flood estimation based on this approach involves derivation of a growth curve which

shows the relation between X_T and the return period T where $X_T = Q_T/Q_I$ and Q_I is the index flood at the site of interest. Generally the mean (FSR, 1975) or median (FEH, 1999) of the at-site AM flood series is taken as the index flood. It is assumed that the X_T-T relation is the same at all sites in a homogeneous pooling group.

The identification of a homogeneous pooling group is therefore important in pooling analysis. Hosking et al. (1985); Lettenmaier et al. (1987); Stedinger and Lu (1995) among other researchers have demonstrated that a successful pooling analysis requires a homogeneity criterion to be satisfied. They have also showed, however, that a small amount of heterogeneity causes very little deterioration in the quality of the flood estimate, from the point of view of overall root mean squared error. If the homogeneity criterion is met then the pooling method can also be used for the estimation of design values at un-gauged locations within the region apart from enhancing the reliability of design flood estimates at gauged sites (Hosking and Wallis, 1997; Zrinji and Burn, 1994).

Delineation of a homogeneous group is not guaranteed by casual selection of pooling group members. Geographical location has been used traditionally as a basis to form a region (FSR, 1975), but it has been often criticised as being arbitrary. Over the last couple of decades, several methods were evolved in which similarity between sites is defined in terms of similarity measures within the catchment descriptor space.

The Region of Influence (ROI) approach developed by Burn (1990) is one of such method. The ROI approach selects stations which are nearest to the subject site in the catchment descriptor space in order to form the pooling group for that subject site.

Ireland was treated as a single region when (FSR, 1975) used geographical regions as a basis to form pooling groups for U.K. and Ireland. Pooled growth curves for ten different regions in the U.K. and one for Ireland were estimated for design purposes. Later it was (FEH, 1999, 3, p. 177) found that such regions were less homogeneous than those formed by a region of influence approach. Unfortunately, no Irish flood data were used in the FEH studies. Therefore, it is now opportune to conduct a pooled flood frequency analysis for Ireland exploiting the ROI method using the longer records.

In this chapter, the author has investigated

1. How the choices and combinations of a number of catchment similarity variables affect the corresponding region of influence distance measures.

2. The effect of catchment type and period of record on pooled growth curve estimates .

Table 4.1: Investigation on the effect of record length

N	T	EV1 (k=0)		GEV (k=-0.1)	
		L-CV =0.15	L-CV =0.25	L-CV =0.15	L-CV =0.25
		$se(Q_T)/Q_T\%$	$se(Q_T)/Q_T\%$	$se(Q_T)/Q_T\%$	$se(Q_T)/Q_T\%$
T/2	50	12.3	16.1	27.5	34.8
	100	7.7	9.8	19.3	23.8
T	50	7.3	9.5	14.9	18.9
	100	5.5	7	11.6	14.3
2T	50	5.2	6.7	9.2	11.7
	100	3.8	4.8	8.4	10.3

4.2 ROI approach of pooling analysis

In pooling analysis, flood data are pooled from gauged sites that are hydrologically similar to the at-site station. Stations pooled for a particular site (denoted here as the subject site) are called, including the subject site, the pooling group. Data at the subject-site, if available, are used as part of the pooling group data set. The ROI approach provides an objective way of forming a pooling group for the site of interest. One of the main advantages of this approach over the traditional regional methods is that each site has its own unique set of similar stations which may even be geographically dispersed.

Developing a frequency model using the ROI approach involves the following series of steps

1. specification of a similarity distance measure

2. assembling the appropriate site descriptors (also known as pooling variables)

3. specification of the number of sites to include in a pooling group, i.e. its size

4. specification of a homogeneity check for the group

5. selection of an appropriate pooling distribution function of flood magnitudes and the selection of an estimation method of its parameters

Additionally, a data screening method, such as the discordancy measure of Hosking and Wallis (1997), is applied in advance to check for the appearance of unusual characteristics in the data.

4.2.1 Distance measure

The distance measure is used to define the closeness of the subject and pooled sites in an attribute space. The most commonly used distance measure in the ROI approach is the weighted Euclidean distance (Burn, 1990; FEH, 1999) . This weighted Euclidean distance, which is that adopted also in this study as the distance measure, is defined as

$$d_{ij} = \sqrt{\sum_{k=1}^{n} W_k \left(X_{k,i} - X_{k,j}\right)^2} \tag{4.1}$$

where n is the number of attribute variables, $X_{k,i}$ is the value of the k^{th} variable at the i^{th} site and W_k is the weight applied to attribute k reflecting the relative importance of that attribute. The subscript j applies to the subject site and the subscript i applies to the i^{th} pooled site.

In the above equation, the variables need to be standardized so that every variable can play its part equally. This is achieved by dividing each variable by the standard deviation of that variable as calculated from the values for all of the sites that are available in the study.

Redundancy of information can occur among the variables and this needs to be considered. Clearly, it is desirable to have the selected variables as independent as possible. However if the selected variables show a considerable amount of correlation an alternative distance measure, such as ' Mahalanobis' distance (Cunderlik and Burn, 2006a), can be applied to adequately account for the correlation among the variables.

Assigning weights to each variable relative to their importance is another point of discussion of the distance measure. The greater weights should be given in the distance measure to variables that show a greater influence on the form of the frequency distribution. Assigning appropriate weights to variables can most conveniently be tackled in

a subjective way as it is difficult to quantify in an objective manner. Hosking and Wallis (1997, p. 147) concede that choosing an appropriate weight is difficult and that the problem is analogous to that of deciding appropriate weights to assign to the variables used in a cluster analysis. The following quote from the U.S. National report (BOBEE and RASMUSSEN, 1995, p. 2) is quite appropriate in this context:

> The selection and weighting of variables is one of the problems where no strict mathematical solution is available, but use of common sense can lead to quite acceptable results.

4.2.2 Pooling variables

The choice of an appropriate set of site attributes as pooling variables is an important task. The attributes can be based on site characteristics (e.g. catchment area, soil type) or site flood statistics (e.g. C_v, g). The latter approach of selecting a pooling group should however be avoided as this might well result in groups consisting of sites that have experienced similar floods in recent history. Neither could such site flood statistics be used for ungauged catchments. Furthermore, hydrological processes are thought to be intuitively more related to catchment physiographic and climatic characteristics rather than site flood statistics.

Hence, FEH (1999); Hosking and Wallis (1997) recommended using site characteristics as pooling variables to form pooling groups and to test the homogeneity of the selected group on the basis of the site flood statistics. Another advantage of using site characteristics is that it offers a possible extension of pooling analysis to account for ungauged sites. Seasonality of the flood response (e.g. timing and regularity of flood events) has also recently been considered (Burn, 1997; Cunderlik and Burn, 2006a) as a similarity measure. Seasonality statistics are obtained from observed flood series. Therefore, a similarity measure based on these could not be used for ungauged sites, without additional assumptions.

In the case of availability of a wide database of site characteristics, a multivariate analysis might be needed to select a reduced set of site characteristics. Burn (1990) selected attributes which showed substantial correlation to the at-site estimate of the 100-year flood event. Furthermore, he selected weighting values corresponding to the observed correlation between the attributes and the 100-year event. The selection and

weighting of pooling variables in the Irish context is described in detail in section 4.4 of this chapter.

4.2.3 Size of a pooling group

In the ROI approach, a decision has to be taken about how many stations are to be included in the pooling group. Selection of stations for inclusion in the pooling group is based primarily on d_{ij} values (see eq 4.1), only those stations having the smaller d_{ij} values relative to the subject site i being chosen. Burn (1990) investigated a number of options to determine a threshold value based on the d_{ij} values to define a cut-off for the inclusion of stations in the ROI method for a target site.

However a more practical way of choosing an appropriate size of a pooling-group was presented by FEH (1999). FEH (1999) investigated a range of pooling group sizes based on the pooled uncertainty measure (PUM) and decided on adoption of the 5T rule, namely, that the total number of station years of data to be included when estimating the T year flood should be at least 5T. The adoption of such a rule was a compromise. If too few stations are included the precision of the Q_T estimate is sacrificed whereas if far too many stations are included then the assumption of homogeneity may be compromised. Hosking and Wallis (1997) however show that a small departure from homogeneity can be tolerated so that having too few stations included may be less desirable than having slightly too many. They also suggested not to use more than 20 sites in a group as little gain in the accuracy of quantile estimates is obtained by using more than about 20 sites in a group.

4.2.4 Homogeneity test

The homogeneity test is used to assess whether a proposed group of sites is homogeneous or not. This study adopts the homogeneity tests proposed by Hosking and Wallis (1997) based on L-moment ratios such as:

1. L-CV alone(the H1 statistic)

2. L-CV and L-skewness jointly (the H2 statistic)

These tests measure the sample variability of the L-moment ratios among the samples in the pooling group and compares it to the variation that would be expected

in a homogeneous pooling group. The sample variability of the L-moment ratios is measured as the standard deviation of the at-site sample L-moment ratios weighted proportionally to the sites' respective record lengths. The measure of the sample variability (which can also be called dispersion) based on L-CV alone, i.e. V_1 and L-CV & L-skewness jointly,i.e. V_2 are defined as

$$V_1 = \left[\sum_{i=1}^{M} n_i \left(t_2^i - t_2^R \right)^2 / \sum_{i=1}^{M} n_i \right]^{1/2} \tag{4.2}$$

$$V_2 = \sum_{i=1}^{M} n_i \left[\left(t_2^i - t_2^R \right)^2 + \left(t_3^i - t_3^R \right)^2 \right]^{1/2} / \sum_{i=1}^{M} n_i \tag{4.3}$$

where t_2^R and t_3^R are the group average of L-CV and L-skewness respectively; t_2^i , t_3^i and n_i are the values of L-CV, L-skewness and the sample size for site i and M is the number of sites in the pooling group.

Simulation is used to establish what 'would be expected' of a homogeneous region. Some 500 homogeneous regions are generated using a four-parameter kappa distribution with L-moment ratio values equal to t_2^R, t_3^R, t_4^R and the at-site mean, $L1 = 1$, in order to obtain the expected mean value, μ_{V_j}, and the standard deviation, σ_{V_j}, of the dispersion measures for a homogeneous group.

The heterogeneity measures H_j are then evaluated from the departure of the $V_j$of the actual region to those of the simulated regions, where

$$Hj = \frac{\left(V_j - \mu_{V_j} \right)}{\sigma_{V_j}}, \quad for j = 1, 2 \tag{4.4}$$

Hosking and Wallis (1997) recommended using the H1 statistic over the H2 statistic as they found that the heterogeneity measure based on V1 has better power to discriminate between homogeneous and heterogeneous regions. They suggested that a region is considered to be "acceptably homogeneous" if $H1 < 1$, "possibly heterogeneous" if $1 < H1 < 2$, and "definitely heterogeneous" if $H1 > 2$.

However, FEH (1999) opted for the H2 statistic as the heterogeneity measure for testing the homogeneity of pooling groups as both the L-CV and L-skewness are required for fitting pooled growth curves with a Generalised Logistic(GLO) or Generalised Extreme Value distribution(GEV). They revised the heterogeneity criteria based on the

H2 statistic, suggesting that if $2 < H2 < 4$, a region could be considered as heterogeneous whereas if $H2 > 4$ it could be considered as strongly heterogeneous. As part of this study, an extensive investigation on homogeneity for Irish pooling groups is presented in chapter 5.

4.2.5 Estimation of the Pooled Growth Curve

The growth factor X_T is the factor which when multiplied by the index flood Q_I (Qmed being used in this study), gives the flood magnitude of return period T, Q_T, as in equation

$$Q_T = Q_I \times X_T \tag{4.5}$$

The relationship between X_T and T is often referred to as the growth curve. When a growth curve is obtained by pooling the information from sites of a pooling group, it is called the pooled growth curve.

In this study the pooled growth curve is obtained using the approach based on L-moments. With this approach the derivation of a growth curve in a pooling group involves the following key steps:

1. computation of at-site and pooled L-moment ratios
2. selection of a suitable form of distribution and a method of estimation of its parameters

L-moments are calculated using PWMs at each site and the dimensionless L-moment ratios t_2 and t_3 are calculated for each site. Pooled L-moment ratios for the target site, i, are then computed using the following equation:

$$t^{(i)R} = \frac{\sum_{j=1}^{N} w_{ij} t^{(j)}}{\sum_{j=1}^{N} w_{ij}} \tag{4.6}$$

where $t^{(j)}$ is the L-moment ratio for the j^{th} most similar site and w_{ij} is a weighing term.

A weight can be given to each site according to its closeness (in terms of ranked d_{ij} values) to the subject site which is deemed to reflect their importance in a ROI group.

Here the weighing is calculated based on the d_{ij} values as in (Kay et al., 2007) using the following equation

$$w_{ij} = 1 - S_{ij} \tag{4.7}$$

where,

$$S_{ij} = \begin{cases} 0 & ifd_{i,j}=0 \\ d_{i,j}/d_{i,\max} & ifd_{i,j}\neq 0 \end{cases} \tag{4.8}$$

In the above equation, $d_{i,max}$ is set 10% larger than the distance associated with the group memeber which just qualified as a member of the pooling group.

Another weighting term related to a site's record length can be incorporated and is often used in a traditional regional analysis but in this study it is not considered as it was not essential to the comparisons presented in section 3.3.2.

In connection with the distribution choice, a 3-parameter distribution is preferred over a 2-parameter distribution, a practice also recommended by Hosking and Wallis (1997). A 3-parameter distribution can be used for a pooling group because the extra data involved in the estimation ensures that the resulting standard error is smaller than in the at-site or single sample case where the se using a 3-parameter distribution can be prohibitively large (see Table 4.1). Furthermore, the 3-parameter distribution avoids any possible bias resulting from using a fixed shape 2-parameter distribution when in fact a 3-parameter distribution is more appropriate – see (d) in Figure 4.1.

In this study, the Generalised Extreme Value (GEV) and the Generalised Logistic (GLO) distributions are chosen as the pooled distribution functions. The results from the L-moment ratio diagrams in Chapter 3 show that the average of the data points falls roughly half way between the GEV and GLO curves. The pooled distribution can also be selected using the goodness-of -fit measure described by Hosking and Wallis (1997). The application of this method to select a suitable form of distribution for pooling groups for the Irish data is described later, in section 4.5.

In order to estimate the chosen distribution's parameters, the values t_2^R, t_3^R are equated to expressions for these quantities written in terms of the distribution's unknown parameters (when the distributions growth curve is expressed in dimensionless form) and the resulting equations are solved for the unknown parameter values.

The dimensionless GEV growth curve is defined by two parameters k and β:

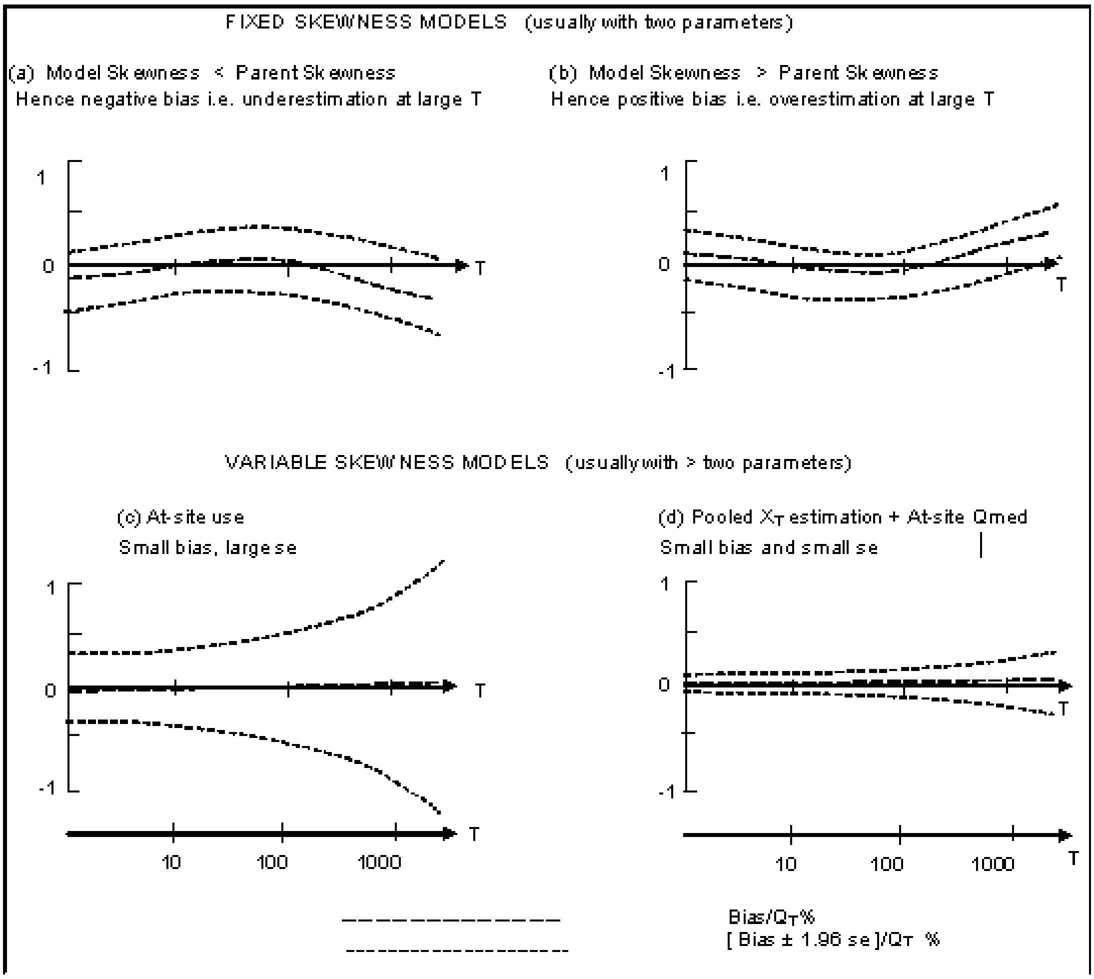


Figure 4.1: Simplified qualitative outline of simulation results for flood quantile estimates - (a) and (b) show the effect of choice of wrong distribution having fixed but incorrect skewness. (c) and (d) show the effect of using a flexible distribution i.e. low bias and large standard error when used in at-site mode but with much reduced se as well as low bias when us in at-site +pooling group mode. (Adapted from Figure 5.1 in WMO (Cunnane, 1989)

$$X_T = 1 + \frac{\beta}{k}\left((\ln 2)^k - \left(\quad \ln \frac{T}{T-1}\right)^k\right) \tag{4.9}$$

where T is the return period.

The two parameters k and β are estimated from the sample L-CV, t_2, and sample L-skewness, t_3, as follows (Hosking and Wallis, 1997)

$$k = 7.8590c + 2.9554c^2 \tag{4.10}$$

$$in \quad which, \quad c = \frac{2}{3+t_3} - \frac{\ln 2}{\ln 3} \tag{4.11}$$

$$\beta = \frac{kt_2}{t_2\left(\Gamma\left(1+k\right) - (\ln 2)^k\right) + \Gamma\left(1+k\right)\left(1-2^{-k}\right)} \tag{4.12}$$

where Γ denotes the complete gamma function.

The dimensionless GLO growth curve is defined by a shape parameter k and a scale parameter β as follows:

$$X_T = 1 + \frac{\beta}{k}\left(1 - \{(1-F)/F\}^k\right) = 1 + \frac{\beta}{k}\left(1 - \{T-1\}^{-k}\right), \quad k \neq 0 \tag{4.13}$$

The parameters k and β are estimated from sample t_2 and sample t_3, as follows (Hosking and Wallis, 1997)

$$k = -t_3 \tag{4.14}$$

$$\beta = \frac{kt_2 \sin(\pi k)}{k\pi(k+t_2) - t_2 \sin(\pi k)} \tag{4.15}$$

4.2.6 Discordancy Measure

A discordancy measure is used to identify unusual sites in a group or in the set of all available gauging stations. The former is called 'group-discordant' where a site is discordant relative to the sites in a particular pooling group that contains it and the latter is called 'globally-discordant' where a site is discordant relative to the set of all available gauging stations (FEH, 1999, 3, p. 159). The discordancy measure is treated as a kind of data screening process.

The discordancy measure, D developed by Hosking and Wallis (1997) is adopted in this study to identify globally-discordant sites. This discordancy measure D is given as

$$D_i = \frac{1}{3} M \left(u_i - \bar{u}\right)^T A^{-1} \left(u_i - \bar{u}\right) \tag{4.16}$$

where,

u_i is a vector containing the L-CV , L-skewness and L-kurtosis for station i,
A is the covariance matrix, $A = \sum_{i=1}^{M} \left(u_i - \bar{u}\right) \left(u_i - \bar{u}\right)^T$
$\bar{u}$ is the group average, $\bar{u} = M^{-1} \sum_{i=1}^{N} u_i$
and M is the number of sites in the group

In general, a site is declared discordant if D>3.

4.3 Data used

Analyses in this study use

1. observed annual maximum flood peaks of 88 A1 and A2 category stations
2. additional AM peaks of 16 stations that have pre- and post- drainage (used only in section 4.6.2 to explore the effect of arterial drainage on the pooled growth curve)
3. catchment descriptors (such as Area, SAAR, BFI, FARL, percent peat)

The discordancy measure was used to screen the data of 88 stations. Table 4.2 lists the 88 stations with their D values, L-moment ratios and catchment characteristics. There are 3 cases where the D values are seen to be greater than the discordant cut-off value of 3. The data of these stations have been examined and possible reasons for their discordancy are listed in Table 4.3. The EV1 probability plots of these data are shown in Figure 4.2. These three stations were excluded from further analysis for this part of the study even though it can be argued that this reduces the natural variability within the data set being studied. Table 4.4 summarizes the data sets of 85 stations.

Table 4.2: Stations with site characteristics, L-moment ratios and discordancy measures.

StationNo	RL	AREA	SAAR	BFI	FARL	L-Cv	L-Skew	L-Kur	D
6011	48	229.19	1028.98	0.71	0.87	0.113	0.09	0.074	0.438
6013	30	309.15	873.08	0.62	0.97	0.157	0.019	0.012	0.971

6014	30	270.38	927.45	0.63	0.93	0.149	0.224	0.121	0.572
6026	46	148.48	940.87	0.66	0.92	0.177	0.24	0.092	1.019
6031	18	46.17	930.66	0.56	1	0.258	0.392	0.317	1.978
6070	24	162.02	1046.28	0.73	0.83	0.145	0.135	0.143	0.019
7006	19	177.45	936.67	0.55	0.99	0.118	-0.205	0.074	2.462
7009	29	1658.19	868.55	0.71	0.99	0.211	0.206	0.106	0.738
7033	25	124.94	1032.22	0.44	0.89	0.129	0.162	0.272	0.721
8002	20	33.43	791.12	0.6	1	0.112	0.271	0.175	0.859
8005	18	9.17	710.76	0.52	1	0.378	0.228	0.189	4.715
9001	48	209.63	783.26	0.51	1	0.243	0.191	0.146	0.826
9002	24	34.95	754.75	0.67	1	0.417	0.39	0.253	6.261
9010	19	94.26	955.04	0.56	0.96	0.423	0.416	0.305	6.855
10021	24	32.51	799.07	0.65	1	0.213	0.186	0.068	1.1
10022	18	12.94	821.92	0.6	1	0.233	0.057	0.061	1.275
12001	50	1030.75	1167.31	0.72	1	0.194	0.176	0.171	0.193
14005	48	405.48	1014.9	0.5	1	0.149	0.291	0.223	0.535
14006	51	1063.59	899.07	0.57	1	0.106	0.238	0.212	0.549
14007	25	118.59	814.07	0.64	1	0.17	0.219	0.072	1.125
14009	25	68.35	831.24	0.67	1	0.125	0.136	0.242	0.473
14011	26	162.3	806.97	0.6	1	0.143	0.02	0.157	0.423
14018	51	2419.4	857.46	0.67	1	0.138	0.036	0.064	0.389
14019	51	1697.28	861.46	0.62	1	0.136	0.092	0.126	0.052
14029	47	2778.15	876.5	0.69	1	0.078	0.058	0.067	0.733
15001	42	444.35	935.24	0.51	1	0.159	0.003	0.075	0.488
15003	50	299.17	933.86	0.38	1	0.111	-0.146	0.087	1.618
16001	33	135.06	916.42	0.61	1	0.124	0.012	0.125	0.294
16002	51	485.7	932.15	0.63	1	0.156	0.172	0.191	0.05
16003	51	243.2	1192.01	0.55	1	0.099	0.207	0.054	1.851
16004	48	228.74	941.36	0.58	1	0.111	0.069	0.108	0.19
16005	30	84	1153.57	0.56	1	0.098	0.204	0.133	0.731
16008	51	1090.25	1029.63	0.64	1	0.072	-0.062	0.055	0.939
16009	52	1582.69	1078.57	0.63	1	0.099	-0.096	0.052	1.014
16011	52	2143.67	1124.95	0.67	1	0.173	0.089	0.053	0.618
18004	46	310.3	985.41	0.68	1	0.087	0.043	0.333	2.891
18005	50	378.47	1190.37	0.71	1	0.142	0.224	0.182	0.213
19001	48	103.28	1175.68	0.64	1	0.095	0.095	0.119	0.284
19020	28	73.95	1179.07	0.66	1	0.196	0.029	0.037	0.987
23001	45	191.74	1084.01	0.32	1	0.18	0.128	0.181	0.181
23012	18	61.63	1264	0.46	1	0.146	0.386	0.326	1.65
24008	30	806.04	939.47	0.54	1	0.151	0.055	0.092	0.223
24022	20	41.21	942.31	0.53	1	0.233	0.123	0.164	0.809
24082	28	762.84	941.7	0.52	1	0.152	-0.036	0.13	0.804
25006	52	1162.76	931.99	0.71	0.96	0.137	0.14	0.183	0.04
25014	54	164.42	1007.65	0.67	1	0.128	0.103	0.125	0.071

25016	42	275.17	947.09	0.61	1	0.125	0.082	0.155	0.08
25023	52	113.86	922.49	0.65	1	0.163	0.142	0.062	0.679
25025	31	161.2	904.54	0.73	1	0.167	0.083	0.146	0.143
25027	43	118.87	1021.15	0.65	1	0.158	0.046	0.126	0.219
25029	33	292.67	1108.68	0.58	1	0.139	-0.02	-0.02	1.367
25030	48	280.02	1183.81	0.54	0.85	0.178	0.185	0.132	0.235
25034	24	10.77	968.6	0.76	1	0.166	-0.08	0.115	1.358
25040	20	28.02	989.64	0.64	1	0.154	0.204	0.181	0.1
25044	33	92.55	1186.86	0.58	1	0.188	0.248	0.149	0.484
25124	18	215.45	954.84	0.87	0.78	0.2	-0.069	0.231	2.954
26002	53	641.45	1067.03	0.61	0.98	0.111	0.224	0.293	0.968
26005	51	1085.38	1054.4	0.56	0.98	0.101	0.035	0.142	0.282
26006	53	184.76	1120.64	0.54	0.97	0.152	0.397	0.412	2.833
26007	53	1207.22	1045.62	0.65	0.98	0.107	0.164	0.157	0.244
26008	49	280.31	1035.47	0.61	0.86	0.104	0.193	0.19	0.347
26018	49	119.48	1043.9	0.72	0.76	0.114	0.126	0.113	0.248
26019	51	252.96	979.62	0.54	0.99	0.139	0.234	0.121	0.715
26021	30	1098.78	945.25	0.83	0.81	0.14	-0.112	0.187	2.475
26022	33	61.88	915.82	0.58	1	0.173	0.091	0.047	0.695
26059	17	256.64	976.44	0.91	0.73	0.099	0.111	0.175	0.218
27001	30	46.7	1476.89	0.28	0.99	0.105	0.193	0.248	0.557
27002	51	564.27	1336.35	0.7	0.84	0.122	0.181	0.216	0.223
29004	32	121.44	1107.47	0.52	0.99	0.083	0.144	0.085	0.942
29011	22	354.14	1079.37	0.63	0.98	0.145	0.41	0.318	1.811
30007	31	469.9	1115.06	0.65	0.99	0.107	0.112	0.203	0.281
30061	33	3136.08	1422.43	0.78	0.66	0.147	0.41	0.379	2.383
31002	26	71.35	1530.25	0.53	0.63	0.129	0.308	0.178	1.03
32012	24	146.16	1784.36	0.59	0.84	0.071	0.001	0.181	0.902
34001	36	1974.76	1322.66	0.78	0.83	0.104	0.08	0.211	0.481
34003	29	1802.38	1339.66	0.8	0.82	0.093	0.114	0.253	0.854
34009	33	117.11	1256.71	0.4	1	0.097	0.081	0.158	0.22
34018	27	95.4	1554.59	0.66	0.73	0.114	0.177	0.033	1.725
34024	28	127.23	1177.46	0.52	0.92	0.066	-0.051	0.14	1.013
35001	29	299.45	1172.84	0.6	0.92	0.119	-0.009	0.19	0.923
35002	34	88.82	1380.56	0.42	0.99	0.098	0.026	0.081	0.405
35005	55	639.66	1198.32	0.61	0.9	0.144	0.205	0.143	0.271
35071	30	247.21	1364.46	0.77	1	0.1	0.118	0.197	0.285
36015	33	153.06	1090.72	0.42	0.96	0.159	0.325	0.299	0.987
36018	50	234.4	950.12	0.69	0.85	0.104	0.045	0.1	0.264
36019	47	1491.76	971.21	0.79	0.76	0.104	-0.029	0.064	0.579
36021	27	23.41	1569.64	0.27	1	0.119	0.201	0.222	0.3
36031	30	63.77	910.43	0.48	0.96	0.098	0.365	0.391	2.885

Table 4.3: Stations having large discordancy values

Station	River Name	D Value	Observation on discordancy
8005	Sluice	6.16	Large Cv of 0.6
9002	Griffeen	6.26	Large Cv of 0.8
9010	Dodder	8.94	Outlier due to Hurricane Charlie

Table 4.4: Summary of AM data sets used in the study

Number of stations	85
Shortest record length	17
Longest record length	55
Mean record length	36.5
Number of AMF events	3213

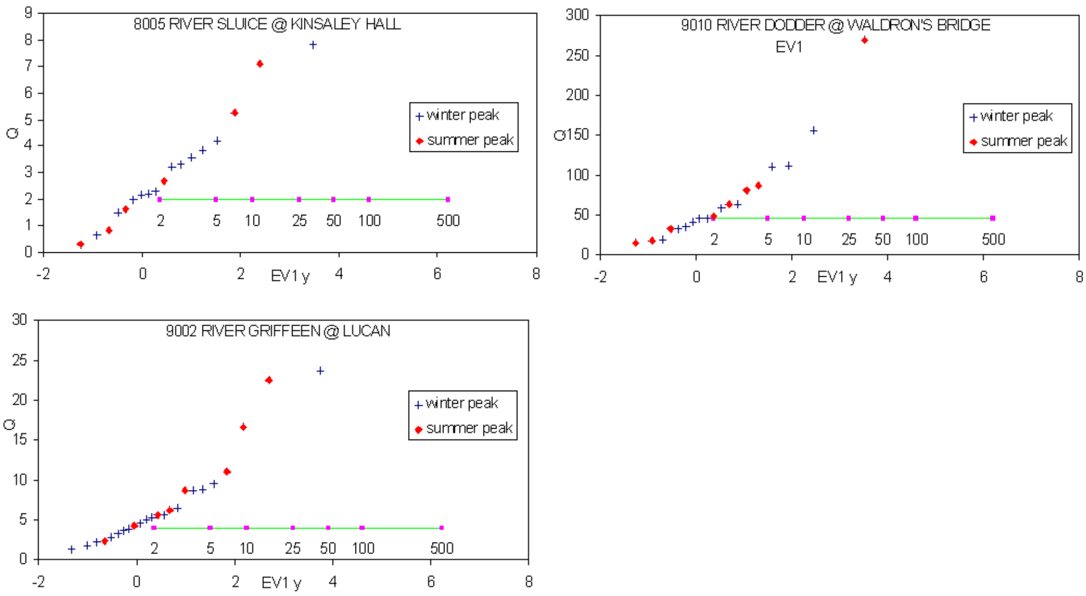


Figure 4.2: EV1 probability plots for the discordant stations -

4.4 Examination of choice of catchment descriptors on effectiveness of ROI distance measures

4.4.1 Selecting variables for pooling

For Irish conditions, four catchment descriptors have been selected as the potential pooling variables from a pool of nineteen catchment descriptors made available form the OPW. These are AREA, SAAR, BFI and FARL. These variables have been found to be effective in explaining the observed variation in Qmed values in the FSU WP2.3 (2009) regression studies.

The objective is to find which combinations of these descriptors lead to pooling groups which are most effective at exploiting the information about the flood distribution contained in the pooling groups. This is assessed by means of a Monte Carlo simulation procedure, which is described later in section 4.4.2.

The general form of the similarity measure used for selecting members of a pooling group was defined in eq (4.1). In choosing a distance measure d_{ij}, a decision has to be made about which catchment descriptors are to be included in the distance measure and what weightings are to be applied to them and whether logarithms or other transformations are to be employed. FEH (1999) gives a number of useful maxims for choosing a distance measure. For selecting the final set of pooling variables it uses the pooled uncertainty measure (PUM) which is a weighted average of the differences between each at-site growth factor and the pooled growth factor measured on a logarithmic scale. For UK data the distance measure is taken as:

$$d_{ij} = \sqrt{\frac{1}{2}\left(\frac{\ln A_i - \ln A_j}{\sigma_{\ln A}}\right)^2 + \left(\frac{\ln SAAR_i - \ln SAAR_j}{\sigma_{\ln SAAR}}\right)^2 + \left(\frac{BFIHOST_i - BFIHOST_j}{\sigma_{BFIHOST}}\right)^2} \tag{4.17}$$

The weight of 0.5 is given to AREA in the distance measure only to allow SAAR and BFIHOST to play a slightly more significant role in forming the pooling groups and to avoid dominance by AREA. Recently, a revised distance measure has been developed for UK data (Kjeldsen et al., 2008) where two new variables FARL and FPEXT (the extent of flood plains in a catchment) has been included and the variable BFIHOST has been omitted. This has the form

$$d_{ij} = \sqrt{3.2\left(\frac{\ln A_i - \ln A_j}{1.28}\right) + 0.5\left(\frac{\ln SAAR_i - \ln SAAR_j}{0.37}\right)^2 + 0.1\left(\frac{FARL_i - FARL_j}{0.05}\right)^2 + 0.2\left(\frac{FPEXT_i - FPEXT_j}{0.04}\right)^2} \quad (4.18)$$

4.4.2 Simulation Procedure

The GEV distribution has been chosen for the purpose of generating simulated data. Hosking and Wallis (1997, p. 93) suggested not to use the observed sample L-moment ratios as the population L-moment ratios of the simulated region because this would yield a simulated region that has much more heterogeneity than the actual data. Castellarin et al. (2001) addressed the issue by using a region of influence approach to estimate the at-site population values of t_2 and t_3. A similarity measure based on at-site flood statistics, as in eq (4.19), is used to form a group of sites for a subject site and its population values of t_2 and t_3 are considered as the corresponding pooled estimate of t_2 and t_3 for the group. Later, Gaal et al. (2008) adopted the above approach in their study. In this part of the study, a similar kind of approach is used based on a similarity measure defined below

$$d_{ij} = \sqrt{\left(\frac{t_{2,i} - t_{2,j}}{\sigma_{t_2}}\right)^2 + \left(\frac{t_{3,i} - t_{3,j}}{\sigma_{t_3}}\right)^2} \quad (4.19)$$

The estimated pooled values of t_2 and t_3 are then used as population values for each site in step 2 of the simulation procedure described below.

The steps of the simulation procedure to selecting variables are as follows.

1. The gauging stations in the subject site's pooling group are identified using the d_{ij} values of eq (4.1) for a set of catchment descriptors having a minimum of 5T station years of data in the pooling group.

2. Random samples are drawn from GEV populations for the subject site and for each site in the pooling group. For each site the sample size is taken as being equal to the length of the observed historical record at the site and the parameters are estimated from the at-site t_2 and t_3 values obtained using the procedure described above (e.g. using eq 4.19) .

3. The t_2 and t_3 values are obtained for each sample in the pooling group and the weighted average of these is calculated to represent the pooled t_2 and t_3 values.

4. The pooled t_2 and t_3 values are then used to determine the pooling group's GEV growth curve parameters k and β using eqs (4.10) and (4.12).

5. The subject site's $\hat{X}_T$ value is calculated for T = 50 and 100 years respectively using eq (4.9)

6. Steps 2 to 5 are repeated 10,000 times to provide 10,000 values of $\hat{X}_T$ and the $RMSE_T$ and $BIAS_T$ are calculated for the subject site by the following equations:

$$RMSE_T\,[\%] = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{X}_{i,r}^T - X_i^T}{X_i^T}\right)^2} \times 100 \tag{4.20}$$

$$BIAS_T\,[\%] = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{X}_{i,r}^T - X_i^T}{X_i^T}\right) \times 100 \tag{4.21}$$

where $\hat{X}_{i,r}^T$ is the estimated T-year growth factor at a site i at the r^{th} repetition; X_i^T is the assumed true T-year growth factor at site i; M is the number of sites in the pooling group and R is the number of repetitions.

4.4.3 Investigation of selected groups of pooling variables

In this part of the study, four catchment descriptors,namely, AREA, SAAR, BFI and FARL, have been considered in eight different combinations, as described below. $RMSE$ and $BIAS$ defined in the simulation procedure have been evaluated at the 50 and 100-year return periods for each site. The following combinations of the four variables have been tested based primarily) on $RMSE$.

lnAREA

lnAREA,lnSAAR

lnAREA,lnSAAR,BFI

lnAREA,lnSAAR,BFI,FARL

lnSAAR

BFI

lnAREA+BFI

lnAREA,lnSAAR,FARL

In all, 85 stations have been considered for the study. The data sets that have been used in the study are summarized in Tables 4.2 and 4.4. For each of these sites, a

pooling group was selected from the 85 stations. Initially, in the simulation procedure, all weights W_k in equation (4.1) were set to unity. Figure 4.3 shows, in box-plot form, the variation in the 100-year RMSE and BIAS values for different sets of catchment descriptors that are used in eq (4.1). Figure 4.4 summarises the corresponding RMSE and BIAS values for the 50-year return period. In Table 4.5, the corresponding mean variation of RMSE100 and RMSE50 values, for different sets of pooling variables, is summarised. Table 4.5 shows that the numerical measures of effectiveness vary by very little between rows. The set of two variables, lnAREA and lnSAAR, and the set of the single variable lnAREA performed best in terms of providing the lowest RMSE100 values. In terms of RMSE50, the set consisting of lnAREA and lnSAAR comes second best to the set consisting of, lnAREA on its own.

Table 4.5: Variation in the mean RMSEs corresponding to T=100 and 50 for different sets of pooling variables

Variables used in the model	RMSE100%	RMSE50%
lnAREA	15.13	12.47
lnAREA,lnSAAR	15.11	12.77
lnAREA,lnSAAR,BFI	15.52	13.48
lnAREA,lnSAAR,BFI,FARL	15.57	13.2
lnSAAR	15.27	13.23
BFI	15.97	13.83
lnAREA+BFI	16.21	13.44
lnAREA,lnSAAR,FARL	15.54	12.78

Overall, the set of variables comprised of lnAREA and lnSAAR may be considered as being the most suitable set of pooling variables for Irish conditions. However, if there is also a desire to incorporate another physical catchment effect then the BFI could be included with these two. While inclusion of just one or two catchment descriptors may indeed be best, there is an intuitive attraction in also representing some descriptor of catchment response even at the cost of a small apparent loss in effectiveness.

An extension to this investigation with varying values of weights W_k in equation (4.1) was also done, particularly for the set of variables of lnAREA, lnSAAR and BFI but the results of all variations examined are not reported in detail here. It was

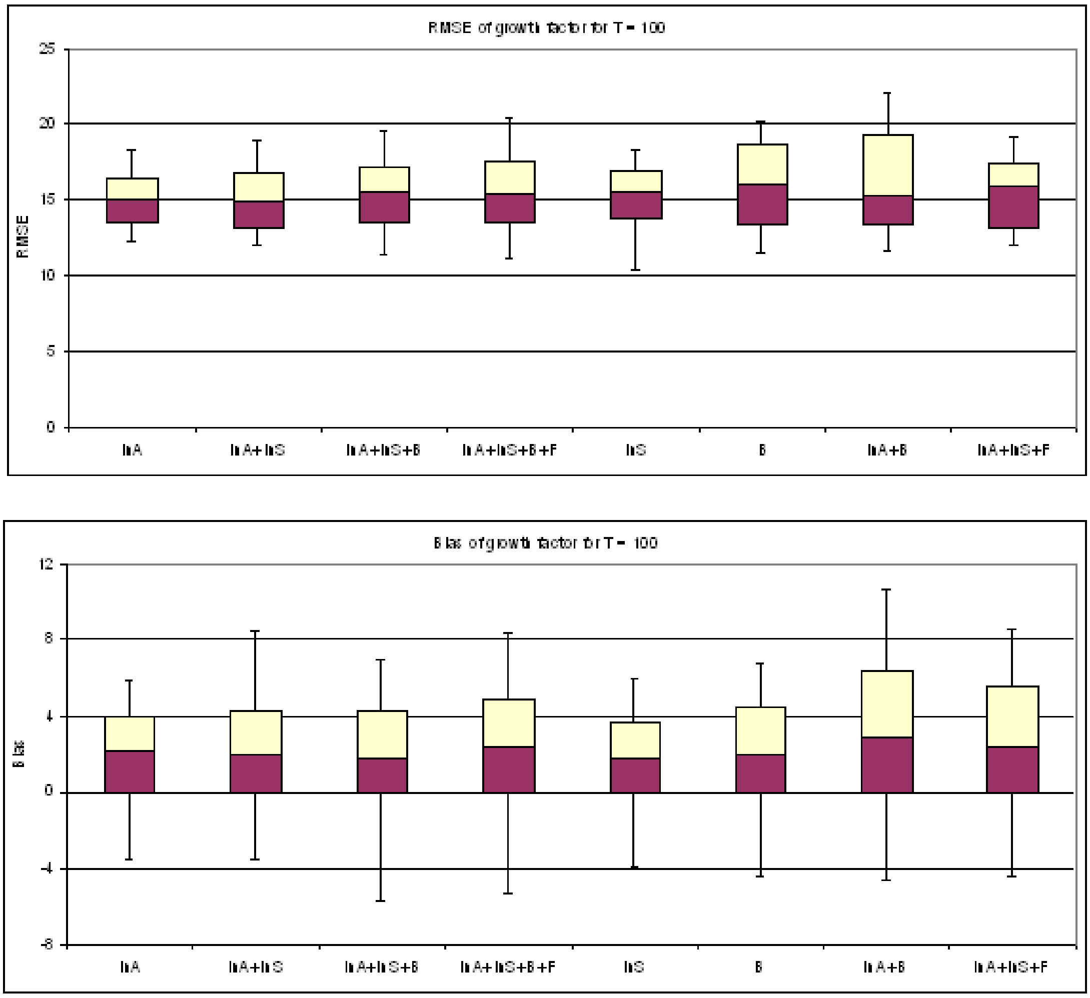


Figure 4.3: Box-plot of RMSE and BIAS of growth factors corresponding to 100 year return periods for different sets of catchment descriptors used in defining the distance measure d_{ij} -

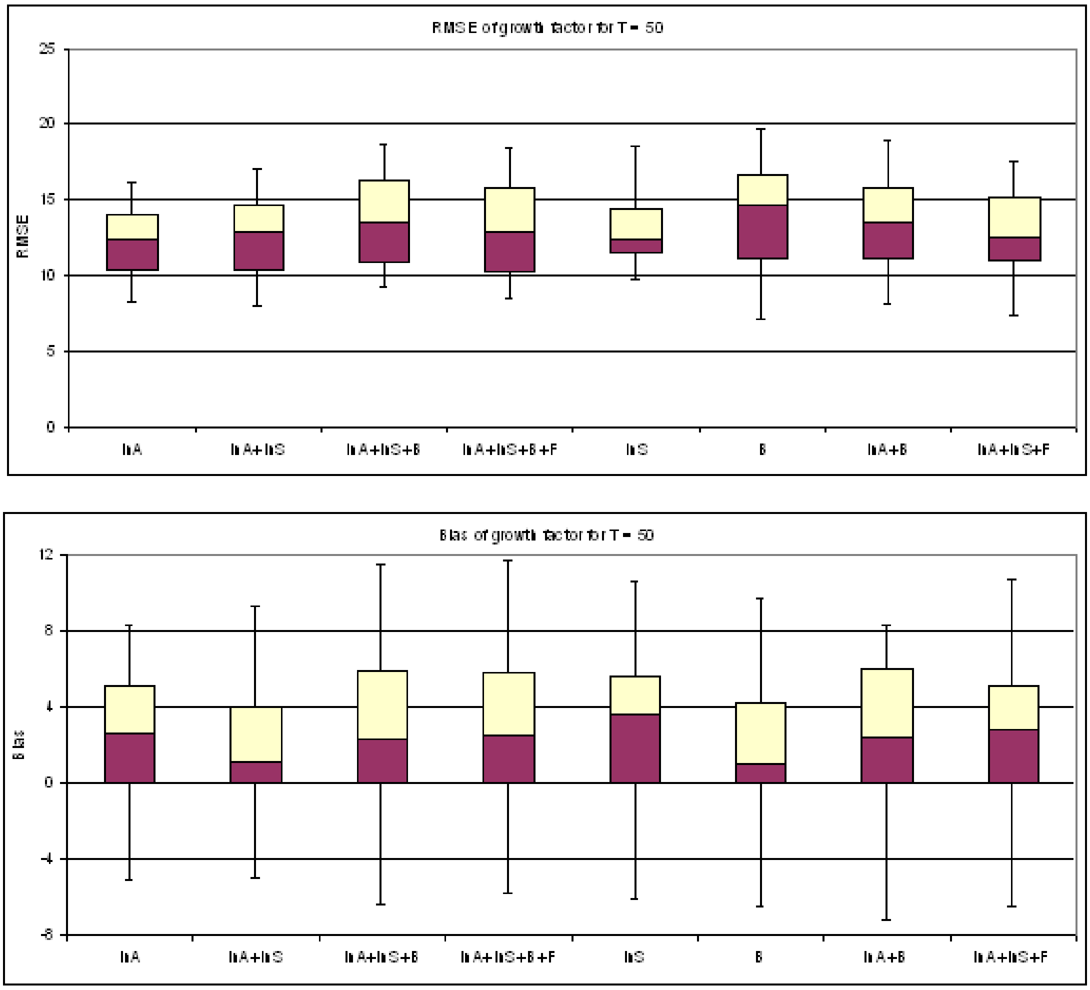


Figure 4.4: Box-plot of RMSE and BIAS of growth factors corresponding to 50 year return periods for different sets of catchment descriptors used in defining the distance measure d_{ij} -

found that the weights 1.5, 1.0 and 0.1 for lnAREA, lnSAAR and BFI respectively gave RMSE100 = 15.22 and RMSE50 = 12.81 which offer small improvements on the $W_k = 1.0$ values used in the calculations for the set of variables of lnAREA, lnSAAR and BFI.

Thus the following equations (4.22 and 4.23) are considered suitable for calculating the distance measure, d_{ij}.

$$d_{ij} = \sqrt{\left(\frac{\ln A_i - \ln A_j}{\sigma_{\ln A}}\right)^2 + \left(\frac{\ln SAAR_i - \ln SAAR_j}{\sigma_{SAAR}}\right)^2 + \left(\frac{BFI_i - BFI_j}{\sigma_{BFI}}\right)^2} \tag{4.22}$$

$$d_{ij} = \sqrt{1.5\left(\frac{\ln A_i - \ln A_j}{\sigma_{\ln A}}\right)^2 + \left(\frac{\ln SAAR_i - \ln SAAR_j}{\sigma_{SAAR}}\right)^2 + 0.1\left(\frac{BFI_i - BFI_j}{\sigma_{BFI}}\right)^2} \tag{4.23}$$

4.5 Goodness-of-fit measure

The goodness of fit test described by Hosking and Wallis (1997) is used for selecting a suitable distribution for a pooling group. The test compares the sample L-kurtosis with the population L-kurtosis for different distributions. The test statistic (Z^{DIST}) is given as

$$Z^{DIST} = \left(\tau_4^{DIST} - t_4^R + B_4\right)/\sigma_4 \tag{4.24}$$

where τ_4^{DIST}, L-kurtosis of the fitted distribution

t_4^R, Weighted regional average L-kurtosis

B_4, Bias of the t_4^R and

σ_4, Standard deviation of the t_4^R

The statistics B_4 and σ_4 are obtained from simulation. A 4-parameter kappa distribution is used for the simulation purpose. The distribution is fitted to t_2^R, t_3^R, t_4^R and the at-site mean/ 1^{st} L-moment, $L1 = 1$. In all, 500 pooling groups similar to the observed pooling group are simulated and from these B_4 and σ_4 statistics are estimated.

An absolute $Z^{DIST} \leq 1.64$ indicates a good fit to a given distribution model. If several distributions in question pass the above criterion, the preferred distribution is the one with minimum Z^{DIST} value.

The gauging stations in the subject site's pooling group are identified using the d_{ij} values of equation (4.22) having a minimum of 500 station years of data in the pooling group, which satisfies the 5T rule for the 100 year quantile. Table 4.7 summarises the values of Z^{DIST} for different 3-parameter distributions for the 85 Irish pooling groups considered here. Table 4.6 summarises the results of analyses to select a suitable form of distribution.

The GEV gives the best overall fit to the Irish AM data. It was the best distribution for 40% of cases and was acceptable in 79% of cases. The LN3 was acceptable for 75% of sites but was the best distribution in only 27% of cases. The GLO was the best distribution in 33% of cases but was acceptable in only 40% of cases. Overall, at least one acceptable distribution was found for 100% of sites and in 79% of these cases the GEV was accepted. The conclusion is therefore, that the GEV distribution provides the best fit among the three distributions.

Table 4.6: Results of the goodness-of-fit measure applied to Irish pooling groups

Criterion	GLO	GEV	LN3
acceptable	40%	79%	75%
best	33%	40%	27%

Table 4.7: Values of Goodness of fit measure as defined by eq 4.24 for different distributions

Station No	Number of site	Station years	GLO	GEV	LN3
6011	14	540	2.13	-0.74	-0.59
6013	15	503	4.27	1.48	1.5
6014	13	535	3.27	0.08	0.27
6026	13	522	2.15	-0.8	-0.62
6031	16	509	2.41	-0.47	-0.37
6070	13	516	1.95	-1.19	-0.99
7006	15	511	2.29	-0.39	-0.43
7009	11	519	4.31	0.4	1.04
7033	15	542	-0.8	-2.81	-3
8002	17	537	4.45	1.49	1.53
9001	17	531	2.74	-0.15	0
10021	16	505	4.64	1.62	1.71
10022	17	537	4.48	1.46	1.5
12001	11	520	0.59	-1.83	-1.88
14005	15	510	0.09	-1.9	-2.13

14006	11	513	1.17	-1.56	-1.53
14007	15	505	4.38	1.5	1.65
14009	16	518	3.14	-0.01	0.12
14011	17	542	3.69	0.67	0.76
14018	11	536	5.02	1.19	1.74
14019	11	523	4.29	0.56	1.06
14029	11	521	5.35	1.27	1.98
15001	15	515	1.34	-1.17	-1.15
15003	14	501	0.86	-1.63	-1.68
16001	16	534	4.13	1.06	1.21
16002	12	531	2.44	-0.46	-0.32
16003	14	529	1.61	-0.58	-0.96
16004	15	538	3.68	0.59	0.69
16005	14	535	2.35	0.01	-0.3
16008	11	525	0.94	-1.95	-1.76
16009	11	545	2.18	-0.94	-0.72
16011	12	548	4.78	0.84	1.49
18004	12	513	2.94	-0.1	-0.03
18005	13	515	1.5	-0.95	-1.05
19001	13	502	3.17	0.47	0.32
19020	14	547	3.09	0.01	0.05
23001	15	539	-1.12	-3.05	-3.19
23012	15	528	0.83	-1.19	-1.51
24008	12	514	0.96	-1.8	-1.73
24022	15	507	1.8	-0.73	-0.79
24082	12	514	1.11	-1.57	-1.51
25006	11	517	2.31	-0.95	-0.56
25014	12	511	3	0.01	0.07
25016	14	514	3.62	0.68	0.76
25023	14	547	2.57	-0.53	-0.38
25025	13	503	2.95	-0.01	0.14
25027	12	502	2.12	-0.63	-0.64
25029	14	540	0.64	-1.43	-1.72
25030	14	528	1.54	-0.46	-0.88
25034	13	505	2.58	-0.58	-0.27
25040	13	546	2.9	-0.43	-0.24
25044	14	539	2.23	-0.17	-0.46
25124	14	512	1.26	-1.86	-1.54
26002	12	513	-0.54	-3.12	-2.97
26005	11	514	1.72	-1.29	-1.05
26006	13	507	1.63	-0.35	-0.81
26007	11	550	1.32	-1.78	-1.61
26008	13	509	0.99	-1.51	-1.51
26018	13	537	2.66	-0.34	-0.17

26019	14	518	0.39	-1.62	-1.97
26021	14	520	1.19	-2.03	-1.52
26022	16	506	4.4	1.49	1.53
26059	14	534	1.95	-1.42	-0.98
27001	15	513	0.03	-1.72	-2.09
27002	13	505	2.13	-0.19	-0.44
29004	14	514	1.23	-0.72	-1.13
29011	14	546	0.53	-2.19	-2.02
30007	13	541	0.23	-2.26	-2.24
30061	12	530	2.95	-0.2	0.09
31002	15	535	2.87	0.09	-0.05
32012	15	531	2.1	-0.43	-0.68
34001	12	536	1.24	-1.94	-1.56
34003	12	514	1.22	-1.92	-1.52
34009	14	508	0.01	-1.71	-2.09
34018	13	500	2.79	0.41	0.2
34024	14	509	0.82	-1.1	-1.48
35001	13	510	1.82	-0.41	-0.75
35002	15	521	0.44	-1.49	-1.94
35005	13	511	2.3	-0.06	-0.3
35071	13	520	2.38	-0.38	-0.41
36015	14	536	-1.02	-3.04	-3.3
36018	13	544	2.81	-0.25	-0.16
36019	11	518	4.82	0.65	1.39
36021	15	507	-0.23	-1.93	-2.22
36031	15	530	1.27	-1.32	-1.32

4.6 Effect of Catchment Type and Period of Record on Pooled Estimates of Flood Magnitude

In this section, the effect of the following categories on pooled estimates of the pooled growth curve estimates are summarised.

1. Catchment type
2. Arterial drainage effect
3. Temporal effect on the pooled growth curve estimates obtained by considering data divided into decades

In all cases, the effect on either X_{100} or X_{50} has been assessed. For the first category above, X_{100} is used. In the second category, X_{50} is used because the limited number of data (only 16 stations having pre- and post- drainage data) does not allow to efficiently estimation beyond T=50 years. X_{50} is also used in the case of estimates from the decadal data of category three because of the obviously smaller amount of data available.

The growth curve ordinate, either X_{100} or X_{50}, has been estimated for each gauging station from its own pooling group based on the distance measure defined in equation (4.22) and applying the 5T rule to determine the number of data included in each. These estimates were obtained by both the GEV and the GLO estimating equations (eqs 4.9 and 4.13).

4.6.1 Effect of catchment type on pooled growth curve estimates

It can be argued that a meaningful number of catchment characteristics are already incorporated into the region of influence approach of pooling analysis to construct a growth curve for a particular site. Nevertheless, it would be informative to know how the flood frequency estimates, in terms of growth curve, depend on any particular catchment category. The following catchment features are considered here:

***Peat lands*:**

Peat-bogs are found in the midlands and along the west coast of Ireland. The Irish peat map (CORINE land cover data, 1990) shows that 13.2% of the national land area

is covered by peat bogs. In terms of flood hydrology, peat lands often tend to be sources for floods rather than sinks (see Holden et al., 2006).

Lakes and Reservoirs:

Many large lakes/reservoirs are found in the Shannon catchment which is located in the middle/west of Ireland. Catchments containing lakes or reservoirs are expected to attenuate flows. The presence of lakes and reservoirs in a catchment is measured by the FARL (Flood Attenuation by Reservoir and Lake) index. The value of FARL ranges between 0 and 1, the value of 0 indicating extreme storage effect and the value of 1 indicating no storage effect.

Catchment Area:

Catchment area plays a big role in the formation of a pooling group in the pooling analysis. Though it is already incorporated in the distance measure, it would be interesting to see whether the growth curve varies with the size of the catchment.

Geographical Location:

Geographical location can be thought of as surrogates of climatic behaviour such as SAAR. The west of Ireland gets about twice as much rainfall (on average 1200 mm per year) than the east part (on average 700 mm per year) and that may affect the growth curve.

Catchments have been divided into subgroups on the basis of

- catchment storage, as measured by the FARL value
- peat content of catchment, as measured by the % Peat value
- size of catchment, as measured by the area
- location of catchments, as regards its geographical region (see in Figure 4.5)

In all cases, the growth curve has been estimated for each station in a subgroup by forming a pooling group from stations that are available in that particular subgroup. An example is the case of location-based subgroups such as west of Ireland, where the growth curve has been estimated for each of the 22 stations by forming a pooling group

from among the available 22 stations. It can be noted here that, as a result of only 22 stations in the data set, there is a certain possibility to have overlaps between the memberships of the pooling groups.

The effects of catchment storage and of percentage peat are shown in Figures 4.6 and 4.7 in box plot form. Each box plot in Figure 4.6 summarise values of X_{100} while Figure 4.7 summarise the corresponding Hosking- Wallis homogeneity statistics, H1 and H2.

The leftmost box plot represents the growth curves of all 85 stations in the study. The median X_{100} values differ only slightly with variation in peat percentage and without a definite pattern. The median X_{100} values show a slight monotonically decreasing trend with increasing amount of storage i.e. smaller FARL values. The results are substantially the same for both GEV and GLO estimation. In nearly all cases, the median H1 values are greater than 2 while nearly all the median H2 values are less than 2 and those few that do exceed 2 do so by a very small amount.

The effects of geographical region and of catchment area are shown in Figures 4.8 and 4.9. In terms of geographical region the South-West has the lowest median X_{100} value of approximately 1.52 while the East has the highest value of approximately 1.9. However, catchments of all sizes, degree of FARL and of percentage peat are mixed together in these subjectively selected geographical groupings. The median X_{100} values show a significant dependance on area. They show a monotonic decrease with increase in catchment area, AREA. The results are substantially the same for both GEV and GLO estimation. Figure 4.9 displays H1 and H2 values that are slightly more dispersed than those in Figure 4.7, with some cases showing median values of H2 greater than 2.

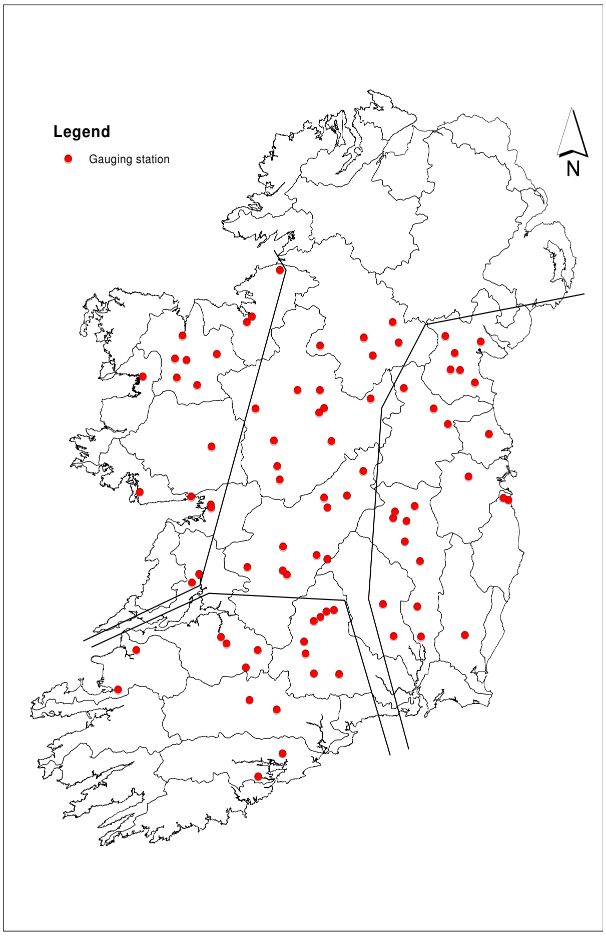


Figure 4.5: Location of the 85 stations and the four regions -

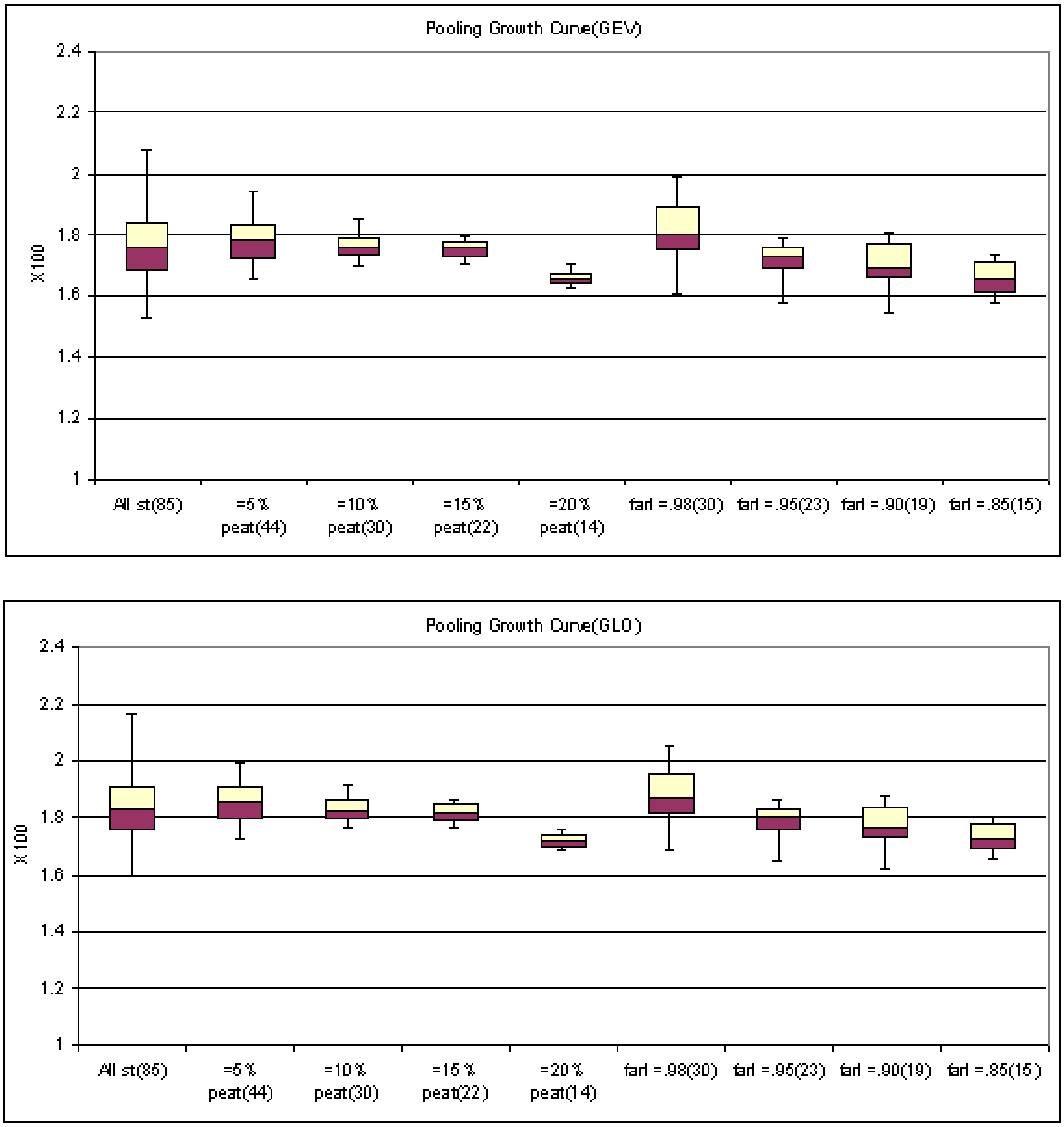
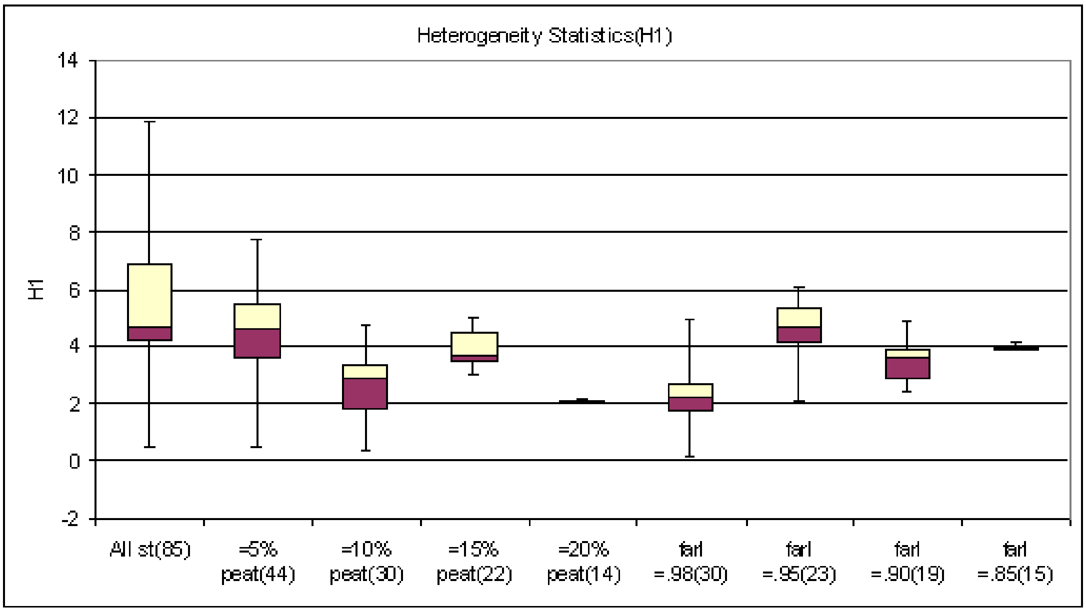


Figure 4.6: Box plots of X100 values from the region of influence pooling estimates of X_{100} obtained by (a) GEV, (b) GLO estimating equations showing the effects of %peat and FARL - The numbers in brackets (85, 44, 30...) indicate the number of X_{100} values contributing to each box plot



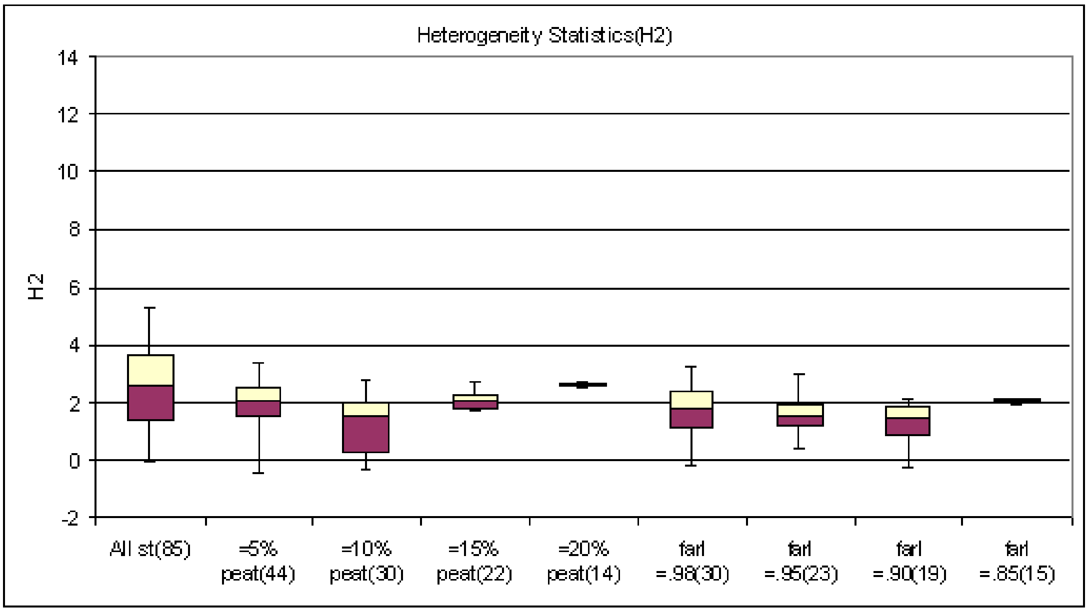


Figure 4.7: Values of (a) H1 (b) H2 for the analyses summarised in Figure 4.6.

-

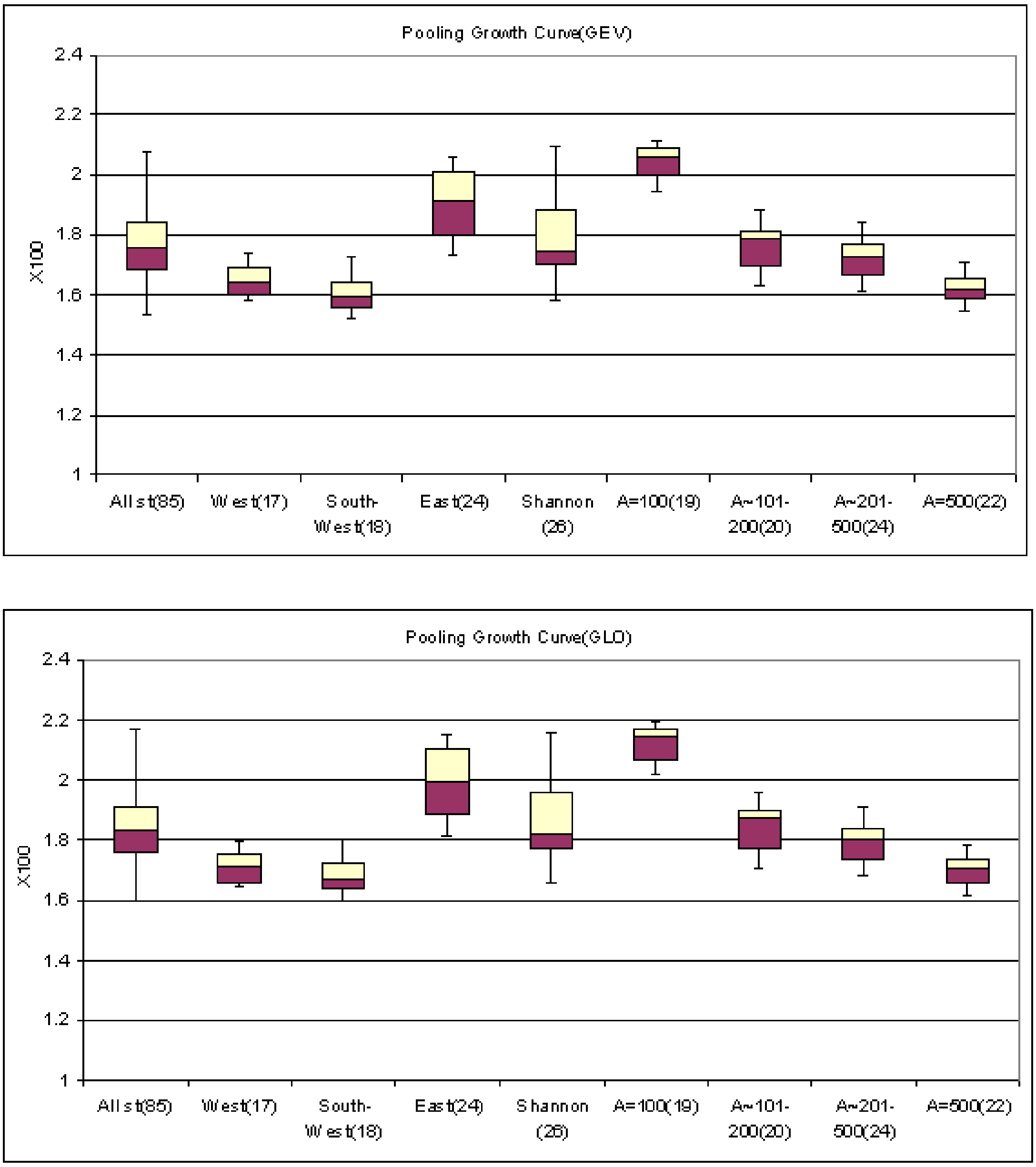


Figure 4.8: Box plots of (a) X_{100} (GEV) (b) X_{100} (GLO) showing the effects of geographical area and catchment size -

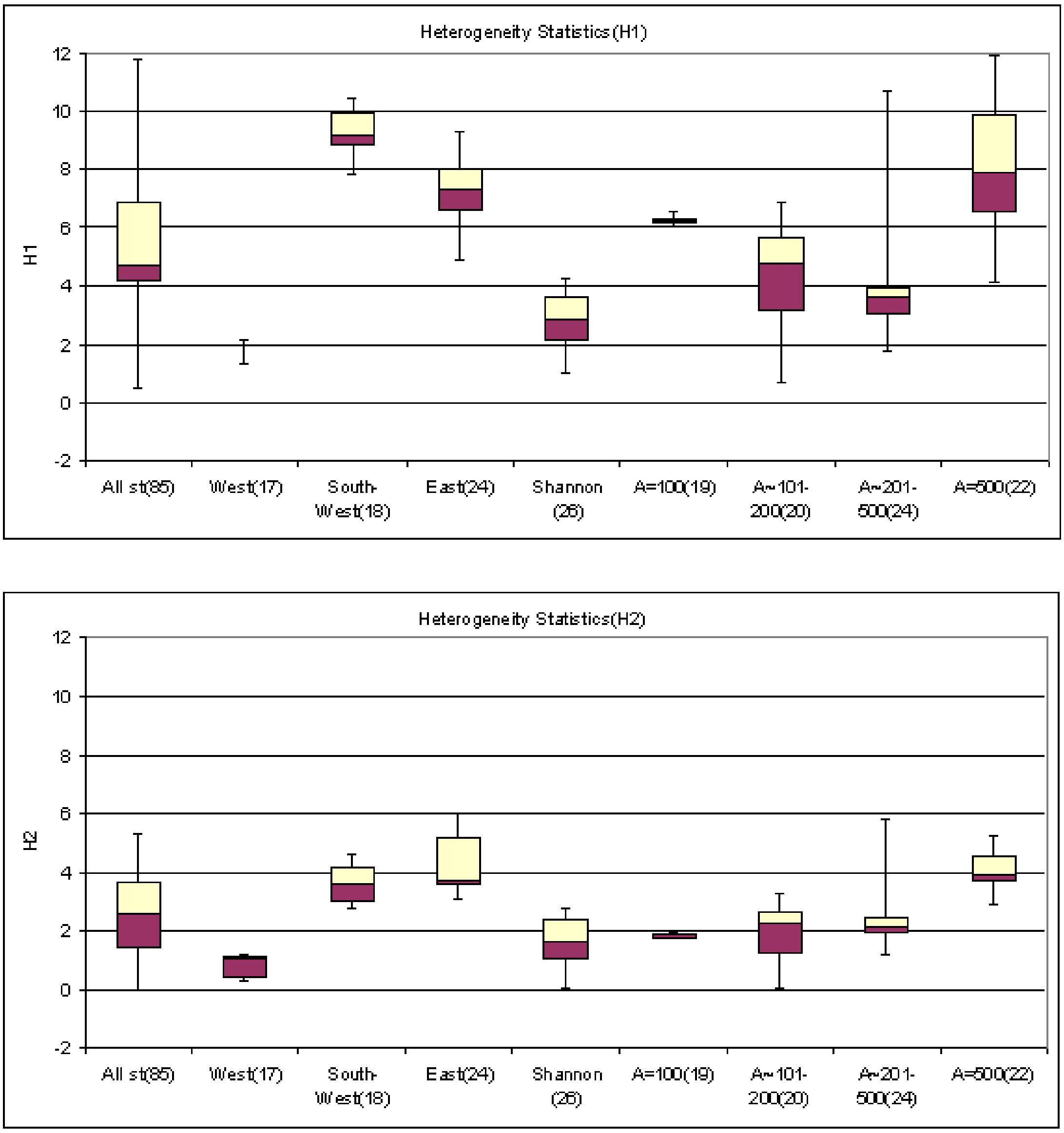


Figure 4.9: Values of (a) H1, (b) H2 for the analyses summarised in Figure 4.8. -

4.6.2 Arterial drainage effect on pooled growth curve estimates

Arterial drainage work (such as widening and deepening of the river channels) has been carried out in many watercourses in Ireland. This work was mainly intended to improve land drainage and to reduce the frequency and extent of overland flooding.

The Hydrology unit of the Research and Development Division at the OPW (1981) reported a detailed investigation based on 13 drained Irish catchments which identified a change of the $Q - T$ relation attributable to drainage and provided methods that could be used to predict post-drainage flood behaviour from a pre-drainage record for a catchment for which arterial drainage is proposed. Relationships established between pre and post drainage show that on average the post-drainage Qmean exceeds its pre-drainage value by about 60%. In the flood estimation procedure, the regional growth curve for Ireland in the FSR (1975) is used for both pre and post-drainage conditions largely because of the evidence of general agreement between the two sets of growth curves (see OPW, 1981, fig. 7). Later Bhattarai and O'Connor (2004) found that the change of the $Q - T$ relation may not be permanent in the Irish context.

In this part of the study pooled growth curves have been obtained from the pre and post drainage records of 16 gauging stations, listed in Table 4.8, 8 of which are located in the Boyne catchment. These 16 stations are those that have their pre- and post-drainage annual maximum series displayed on probability plots in Appendix C (Drainage). Although the descriptor BFI is needed in the pooling procedure, it was not available for three of these stations. As two of these are on the Boyne catchment (7007 and 7011), the BFI values at neighboring gauging stations on the same river have been used instead, i.e. 7003 for 7007 and 7004 for 7011. The third station lacking a BFI value, namely 30004, was assigned the BFI value for gauging station 30005 even though this is not on the same stream. A visual check was made on the appearance of the yearly hydrographs of these two gauging stations to check if it is reasonable assign that BFI value. The visual agreement is reasonable but not especially good. The probability plots in Appendix B (Drainage) show a clear increase in magnitude at all return periods for only 9 gauging stations. These are 3051, 7003, 7004, 7010, 7011, 7012, 24001, 24013 and 30004. One gauging station's plot, 30001, shows quite the opposite effect while the remaining six show a variety of cross-over effects, that is,

that the pre-drainage large floods generally exceed the post drainage large floods or else the lines fitted to each period of record might cross over.

The results of the pooled estimation of X_{50}, Qmed and X_{50} for these stations, from both pre-drainage and post-drainage data, are shown in Figure 4.10. These show that the pre-drainage growth factor exceeds the post-drainage growth factor in every case, whereas the Qmed values are greater in the post-drainage period than in the pre-drainage period in every case except for the above mentioned station 30001. The estimated X_{50} values for the post-drainage period exceed the pre-drainage values in 10 cases, 9 of them being those noted above in relation to their probability plot behaviour, the 10^{th} one being 26012 which shows just a small increase.

Table 4.8: List of 16 stations that have pre and post- drainage records

Station No	Station Name	River Name	Catchment	N	Timespan	Drainage works	Missing year
3051	Faukland	Blackwater	Blackwater	26	1975-2004	86-'89	1986-89
7002	Killyon	Deel	Boyne	46	1953-2004	3/74- 4/79	1973-78
7003	Castlerickard	Blackwater	Boyne	46	1953-2004	1970- 1975	1956,70-74
7004	Stramatt	Blackwater	Boyne	48	1956-2004	1973-1981	1973-81
7005	Trim	Boyne	Boyne	47	1953-2004	1971- 1974	1971-74
7007	Boyne Aqueduct	Boyne	Boyne	45	1953-2004	73-78	1973-78
7010	Liscartan	Blackwater	Boyne	46	1953-2004	1982- 1986	1970,71,82-85
7011	O'Dalys Bridge	Blackwater	Boyne	44	1958-2004	80- 83	1980-82
7012	Slane Castle	Boyne	Boyne	65	1940-2004	69-86	None
23002	Listowel	Feale	Feale	59	1946-2004	51-59	None
24001	Croom	Maigue	Maigue	51	1953-2003	75- 76	None
24013	Rathkeale	Deel	Maigue	46	1953-2004	1963-1968	1963-68
26012	Tinacurra	Boyle	Boyle	48	1957-2004	82-91	None
30001	Cartronbower	Aille	Cara	48	1952-2001	70-71	1970-71
30004	Corrofin	Clare	Cara	35	1952-2003	58-64	1958-73,79,86
30005	Foxhill	Robe	Cara	45	1953-2004	73-78	1961,73-78

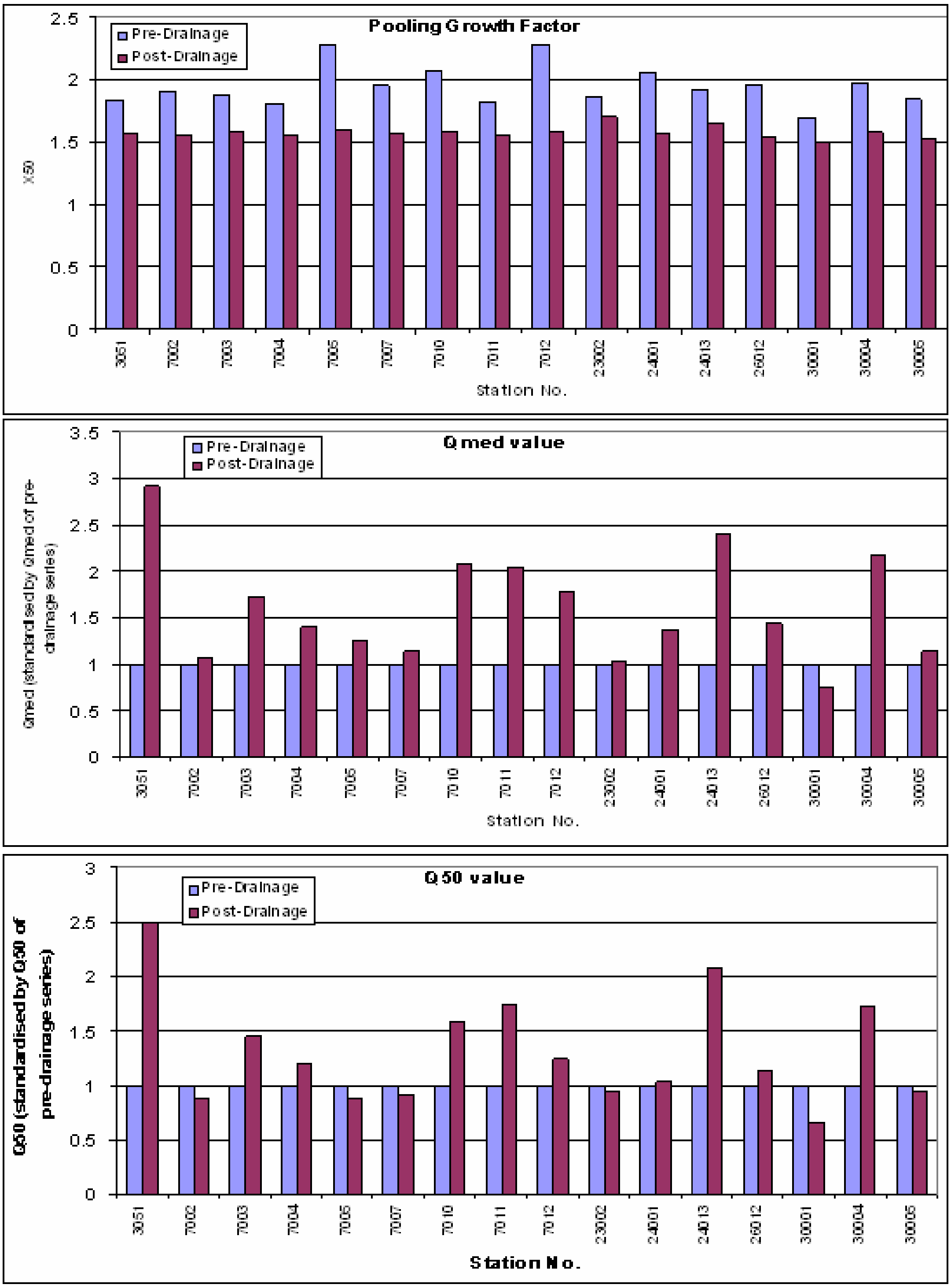


Figure 4.10: (a) Column chart of X_{50} (GEV) values (b) Column chart of Qmed values (c)Column chart of X_{50} values for the pre- and post-drainage data of 16 stations -

4.6.3 Temporal effect on pooled growth curve estimates

Pooled growth curves have been obtained from different periods of record as follows:

1. distinguishing between early and late halves of each record and
2. distinguishing by decades, 1972 - 81, 1982 - 91 and 1992 - 2001 (for which data are available for 50 stations) and
3. distinguishing by decades, 1952 - 61, 1962 - 71, 1972 - 81, 1982 - 91 and 1992 - 2001 (for which data are available for 26 stations).
4. distinguishing by decades 1957-66, 1967-76, 1977-86, 1987-96 and 1997-2006 where 2006 is the most recent year of record in the data set. There are 34 records available for all 5 decades.

This part of the study had already been completed using a total of 90 gauging stations before the data provider indicated some of these are unsuitable for detailed statistical analysis. Hence the number 90 is greater than 85 mentioned above in section 4.3. It is believed that the overall nature of the findings would not be greatly altered if the study had been repeated with just 85 stations. This would be especially true if the five omitted stations had reasonable water level records and were rejected on the basis of doubts about the reliability of rating curves.

In the first case, each record of 90 stations has been divided into two equal parts and a pooled growth curve has been estimated for each of the 90 stations for each part.

In the case of the decadal effect, three different cases are dealt as mentioned above. In all cases, only stations which have a minimum of 8 years of data available in a decade are included for that particular decade. Table 4.9 lists the number of stations that are available for each decade.

The effects of period of record (1^{st} half of each record compared with 2^{nd} half) and of successive decades are shown in Figures 4.11 and 4.12. H1 and H2 values are shown in Figure 4.12. The median X_{100} value shows a decrease between the earlier and later halves of the records while the decade based values of X_{50} show no particular pattern. Indeed the 3^{rd} decade value of median X_{100} is not much different to that of the 1^{st} decade median X_{100} value.

Table 4.9: Number of stations available for each decade

Decade Period	No. of St. available	Decade Period	No. of St. available
1952-61	29	1957-66	37
1962-71	42	1967-76	41
1972-81	52	1977-86	75
1982-91	83	1987-96	87
1992-2001	82	1997-2006	79

There are 50 stations for which decade estimates are available for all 3 decades. The values of X_{100} for each decade for each station are displayed in Figure 4.13(a). From this it can be seen that the largest X_{50} values occur in the middle decade, 1982-91, in every case. The corresponding Qmed values, standardized at each station by the 1972-81 value of Qmed, are shown in Figure 4.13(b). In most cases, the Qmed values for the 2^{nd} and 3rd decade exceed the 1st decade value and while the 3^{rd} decade displays a number of remarkable Qmed values, only about one third of the stations have their largest Qmed in the 3^{rd} decade. Figure 4.13(c) shows the estimated Q_{50} values, standardized at each station, by the 1972-81 value. The majority of Q_{50} values in the 2^{nd} and 3^{rd} decades exceed those of the 1^{st} decade, with more than half of these exceedances occurring during the 2^{nd} decade.

There are 26 stations for which decade estimates are available for 5 decades. The results of the analyses are examined in the same way as for the 3 decade case result above and presented in Figure 4.14. The 2^{nd} decade (1960s) value of X_{50} is larger than those of the other decades at every station with an average value of approximately 1.7. This contrasts strongly with the results of the last decade (1990s) where the average value is approximately 1.4. Figure 4.14(b) shows the decadal standardized Qmed values. This shows that the largest values tend to be from the 2nd (1960s) and the last (1990s) decades. The Q_{50} values in Figure 4.14(c) also conveys the impression that the largest Q_{50} values occur during the 2^{nd} (1960s) and last (1990s) decades.

The results for the decades ending in 1966, 1976, 1986, 1996 and 2006 show results that are similar to those described in the last paragraph. Results shown in Figure 4.15 are analogous to those of Figure 4.14.

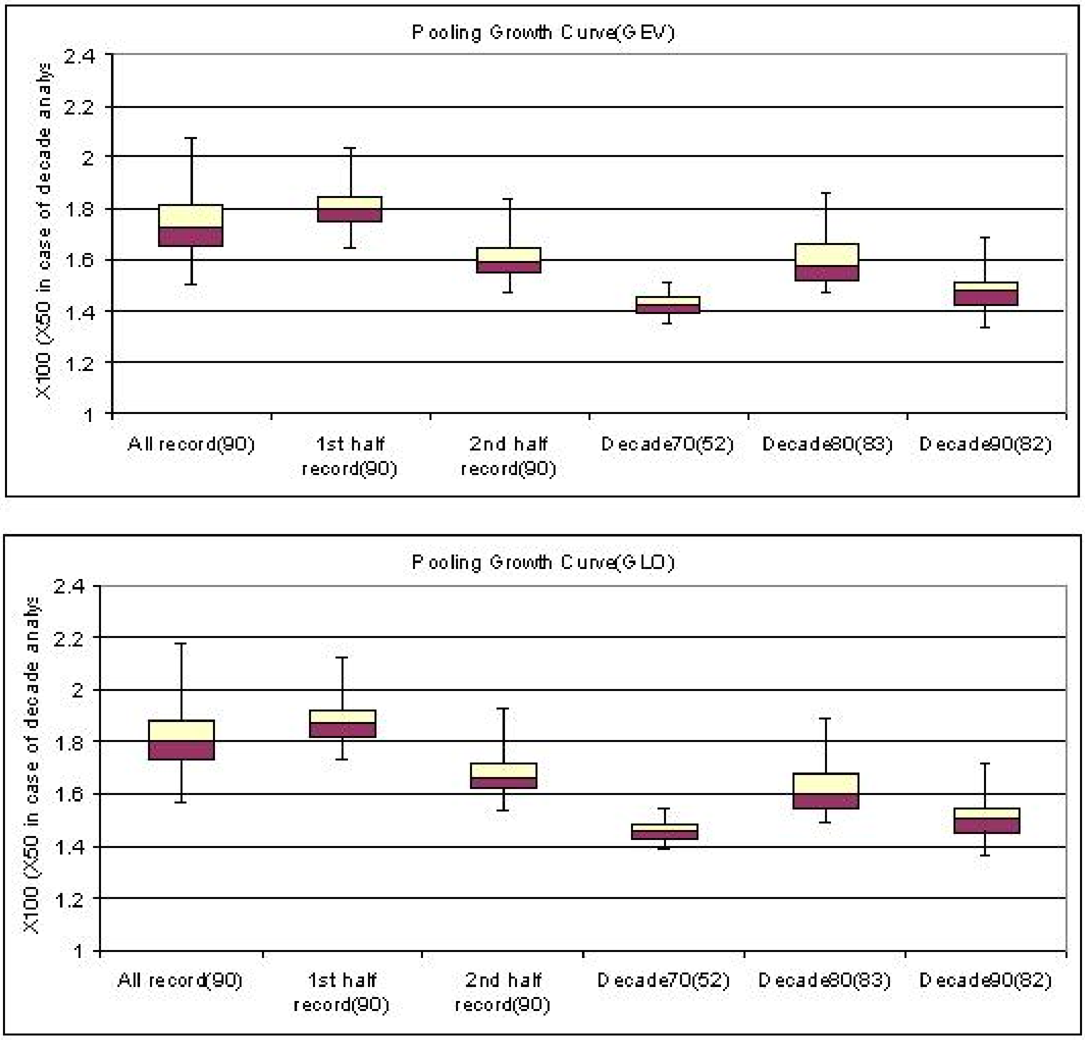


Figure 4.11: Variation of Q_{100} with period of record (first half versus second half of each record) and variation of Q_{50} with decades, as estimated by (a) GEV, (b) GLO from each individual stations pooling group using the 5T rule -

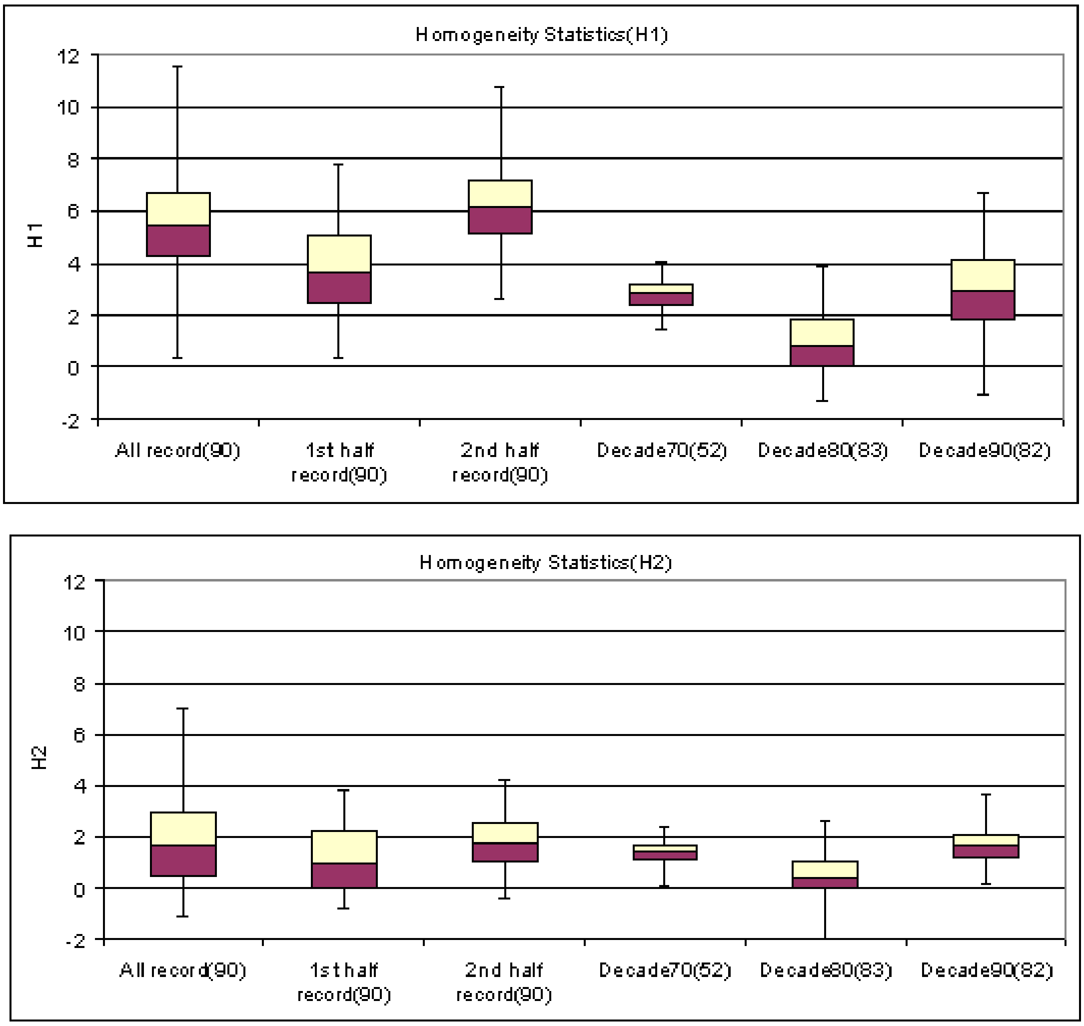


Figure 4.12: Values of (a) H1, (b) H2 for the analyses summarised in Figure 4.11 -

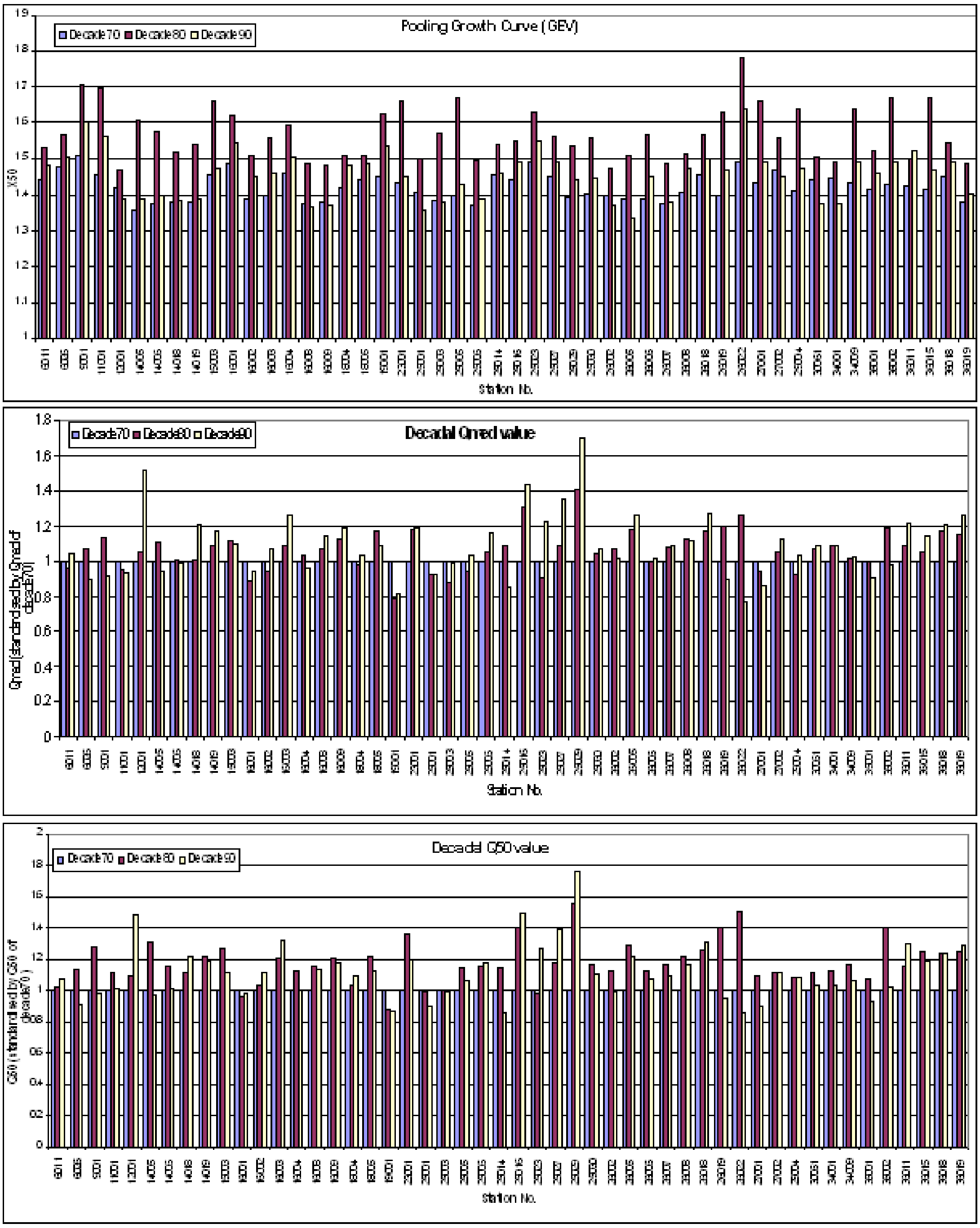


Figure 4.13: (a) Column chart of X_{50} (GEV) values (b) Column chart of Qmed values (c) Column chart of Q_{50} values for the decadal data ('70, '80 and '90) of 50 stations -

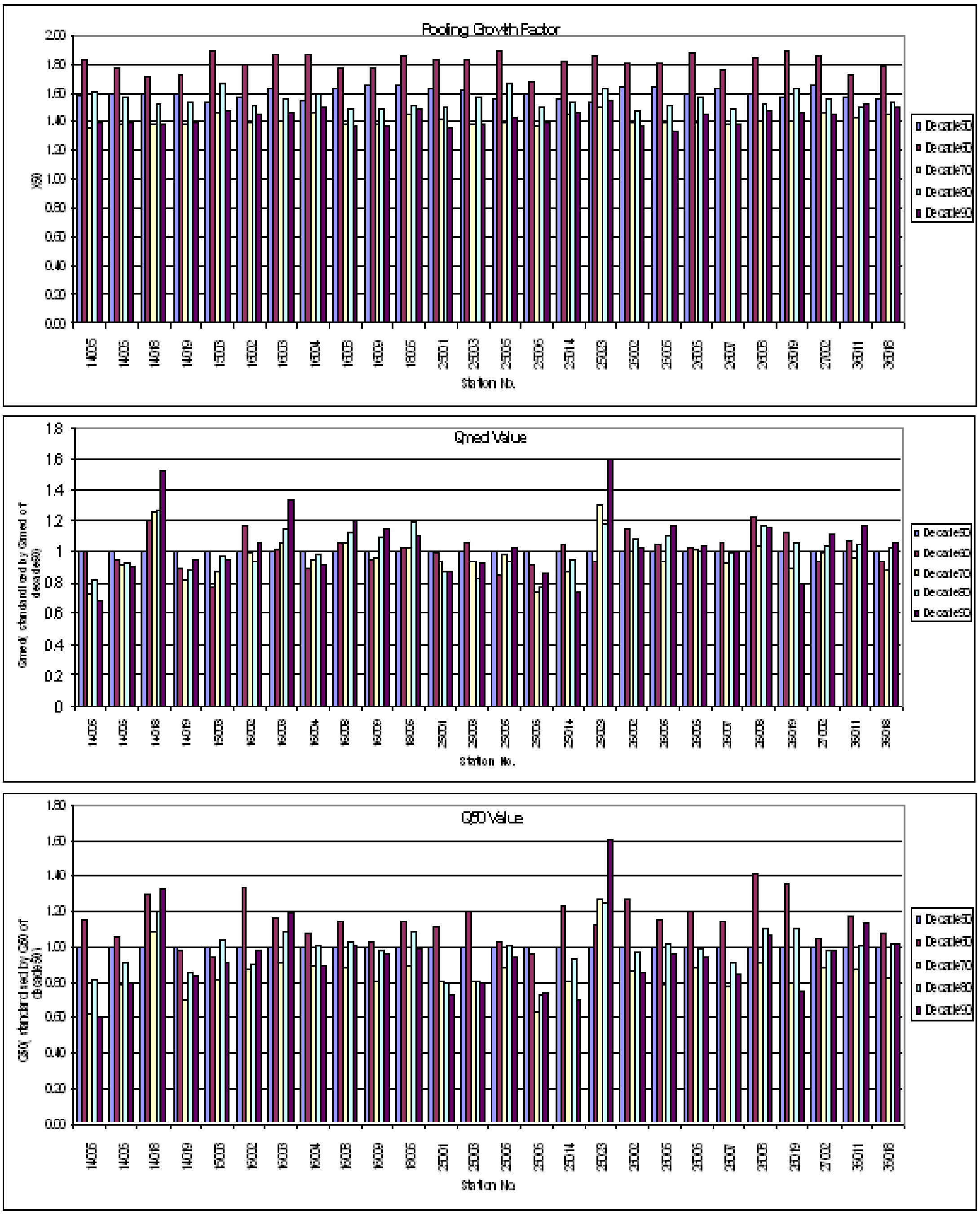


Figure 4.14: (a) Column chart ofX_{50} (GEV) values (b)Column chart of Qmed values (c) Column chart of Q_{50} values for the decadal data ('50, '60 '70, '80 and '90) of 26 stations -

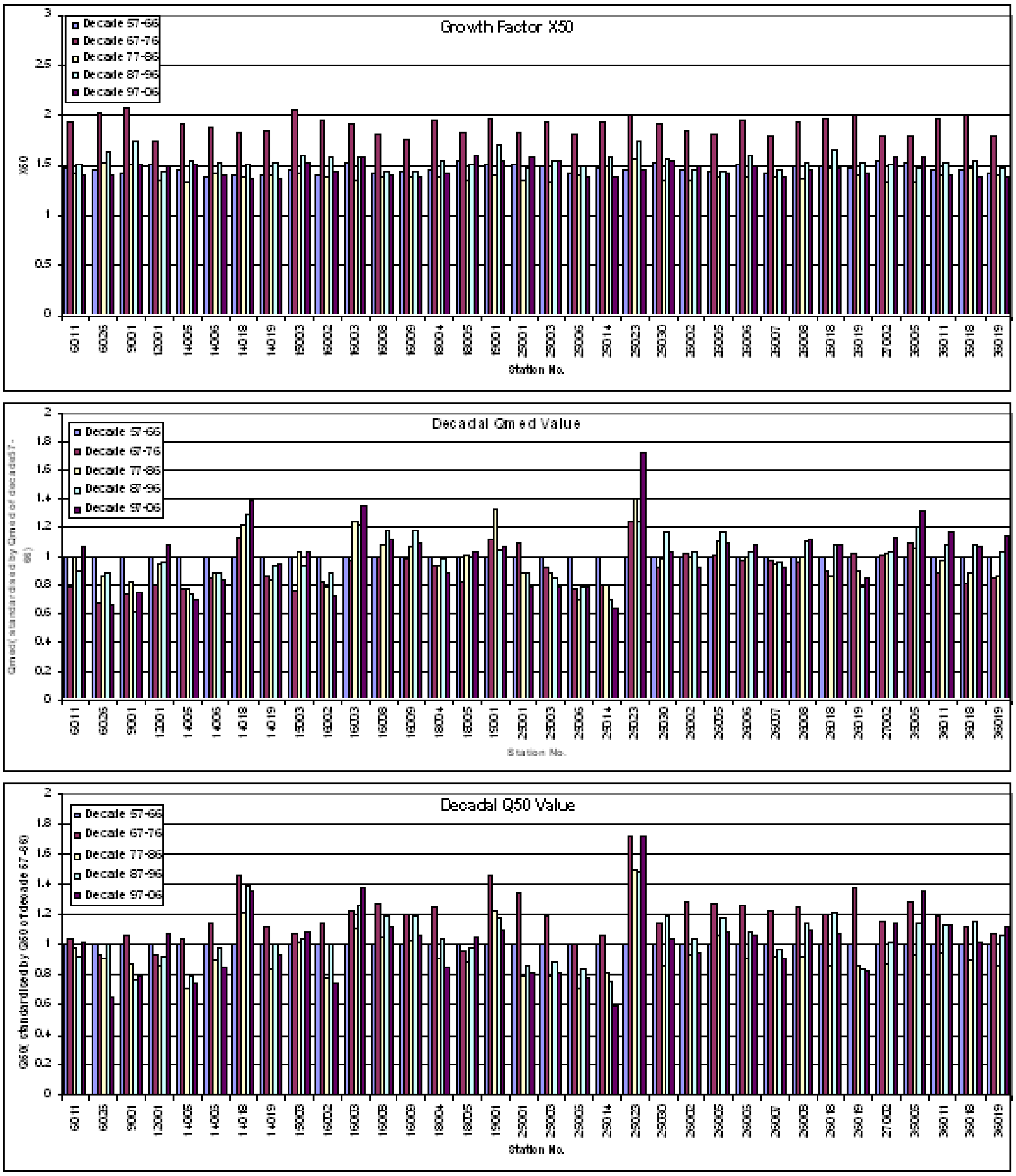


Figure 4.15: (a) Column chart of X_{50} (GEV) values (b)Column chart of Qmed values (c) Column chart of Q_{50} values for the decadal data ('50, '60 '70, '80 and '90) of 26 stations -

4.6.4 Summary of the above effects

The FARL value, catchment area and geographical location appear to have an effect on the X_{100} growth factor with the % peat having virtually no noticeable effect. In temporal terms, the middle of the 3 decades between 1972 and 2001 produces the largest X_{50} values as a whole, while the same decade displays many, but not all, of the largest Q_{50} values. The 1972-81 decade displays the smaller values of X_{50}, Qmed and Q_{50} with just a small number of exceptions. When the smaller number of cases for which 5 decades of data are available the impression is that the 2nd (1960s) and last (1990s) decades produced the largest floods.

4.7 Conclusion

The following conclusions are obtained from the above studies

1. The composition of the distance measure d_{ij} for expressing similarity between catchments has been investigated. Various combinations of the catchment descriptors AREA, SAAR, BFI and FARL have been assessed using root mean square error. Different combinations of weights applied to these descriptors were also tested. Equations (4.22) and (4.23) are recommended as being suitable for use but because of its simiplicity eq (4.22) is used in the remaining chapters.

2. The results of the Hosking and Wallis goodness-of-fit measure show that the GEV distribution provides the best fit among the three distributions considered for Irish pooling groups.

3. The effects of catchment size, % peat content, catchment storage (measured by FARL) and geographical position on the magnitude of the growth factor X_T have also been examined. It was found that catchment area, FARL and geographical location did have an effect on X_T, although the statistical significance of the effect has not been tested.

4. The effect of period of record e.g. 1950s versus 1960s on X_T was also investigated. While there were small effects present indicating higher X_T values when estimated from data of the 1960s and 1990s, it is felt not to be worthwhile to exclude the data of any decade from the flood estimation process.

5

Examination of the homogeneity of selected Irish pooling groups

5.1 Introduction

One important stage in the forming of a homogeneous pooling group of gauging sites is establishing the proof of its homogeneity. The homogeneity criterion implies that the $X_T - T$ relation is the same at all sites in a pooling group. A homogeneous pooling group of sites leads to a minimisation in the error of quantile estimators which is the main aim of a pooled analysis. The examination of homogeneity is, therefore, an essential element of a pooled flood frequency analysis. In this chapter, the homogeneity of selected groups identified by the Region of Influence (ROI) approach is examined.

5.2 Approach used in the study

A test of homogeneity is normally used to assess whether a proposed group of sites is homogeneous or not. Tests for the homogeneity of regions/pooling groups are usually based on a statistic that relates to the formulation of a frequency distribution model, e.g. the coefficient of variation, C_v (Wiltshire, 1986; Fill and Stedinger, 1995) and/or skew coefficient C_s, their L-moment equivalents (Chowdhury et al., 1991; Hosking and Wallis, 1997) or of dimensionless quantiles such as the 10-yr event, X10 (Dalrymple, 1960; Lu and Stedinger, 1992).

Hosking and Wallis (1993, 1997) proposed homogeneity tests based on L-moment ratios such as L-CV alone (H1) and L-CV & L-skewness jointly (H2) which are widely

used in flood frequency analysis although the former one is recommended by those authors for having better power to discriminate between homogeneous and heterogeneous regions. Very recently, a similar conclusion has been drawn by Viglione et al. (2007) when they compared several homogeneity tests. They stated that the H1 test is ahead of all others when the L-skewness is lower than 0.23. They further concluded that the H2 as a homogeneity test lacks power. These findings certainly indicate that the heterogeneity among the sites in a group is mainly due to variations in the sample L-CVs.

However, one of the main assumptions of these tests is that the true regional distribution is kappa. For that reason and others, Hosking and Wallis (1997) recommended that though the heterogeneity statistic is constructed like a significance test it should not be used in that way. They further stated that (Hosking and Wallis, 1997, p. 70)

> ....a significance test is of doubtful utility anyway, because even a moderately heterogeneous region can provide quantile estimates of sufficient accuracy for practical purposes. Thus a test of exact homogeneity is of little interest.

In this study, a graphical way of examining the homogeneity of a pooling group is presented which is based on L-CV, t_2. The main idea behind the approach is the comparison of the variability of t_2 from each site in the pooling group with that expected (un-weighted average pooled t_2) supposing the differences between sites to be due to sampling error. The population distribution for this purpose is selected based on the descriptive ability of the summary statistics of the region concerned.

5.3 Examination procedure

The steps of the procedure which is applied in the study are as follows

1. The gauging stations in the subject site's pooling group are identified using d_{ij} values obtained by the following equation (4.22)
   $d_{ij} = \sqrt{\left(\frac{\ln AREA_i - \ln AREA_j}{\sigma_{\ln AREA}}\right)^2 + \left(\frac{\ln SAAR_i - \ln SAAR_j}{\sigma_{\ln SAAR}}\right)^2 + \left(\frac{BFI_i - BFI_j}{\sigma_{BFI}}\right)^2}$ having a minimum of 500 station years of data in the pooling group and satisfing the 5T rule for the 100 year quantile.

2. The t_2 is obtained for each site in the pooling group and the average, without weights, of these is calculated to represent the pooled average t_2 (t_2^r).

3. Random samples are drawn from an EV1 distribution (or a GEV distribution) using the t_2^r as the population value to construct a 95% confidence interval for t_2^r. The sample size is taken as being equal to the average record length of the observed historical record at the gauging sites and the parameter values μ and α are those estimated from the value of the t_2^r. The 95% confidence interval is constructed assuming that the samples t_2 are normally distributed.

4. The number of stations in the selected pooling group whose t_2 values fall outside the confidence interval (the attribute termed here as m) is counted and reported. It is also noted whether the t_2 of the subject site is outside the confidence limits (CL).

5.4 Analysis

This study is based on AM series obtained from the 85 selected A1 and A2 Irish gauging stations. The procedure described above is applied for each of the 85 stations. Each station has its own unique pooling group. The sample values of t_2 for the stations in the group, t_2^r and the CL about t_2^r are displayed in Figure 5.6 for 22 stations. The above details for the remaining stations are displayed in Figures 5.7, 5.8 and 5.9. The EV1 is used initially as a parent distribution to construct confidence limits. The summary statistics of the procedure for each group is given in tabular form in Table 5.3 including the heterogeneity measures, H1 and H2, for each group.

The following observations and findings are obtained from the analysis.

1. At all stations, except for one, one or more of the sample t_2 of stations forming the pooling groups fall outside the 95% confidence limits.

2. Table 5.1 lists how many stations fall into the categories of one value outside the CL, 2 or 3 values outside the CL or more than 3 outside the CL. In all, 52% of stations were in the latter category.

3. In 27 cases (32% of groups) the t_2 of the subject site was outside the limits.

4. Table 5.2 summarises the results of H1 and H2 for the 85 pooling groups. 22% of groups have a H1 value lower than 4. The percentage increases to 86% when the same criterion is set for H2 and that is very similar to what was found for the U.K. pooling groups (FEH, 1999, p. 176).

5. The average range of t_2 for the 85 pooling groups was 0.11 with a minimum value of 0.06 and a maximum value of 0.18. Figure 5.1 shows a plot between H1 values and ranges of L-CV values for the 85 groups. The plot shows an upward trend, implying that a high H1 value can be expected for a high range of t_2 value in a pooling group, which is a natural tendency form the assumption of homogeneity. A similar plot is drawn for H2 in Figure 5.2, showing no obvious trend, implying that a low H2 value may be obtained for a pooling group which is in fact a heterogeneous group.

6. Figure 5.3 shows a plot between H1 and m. Different values of H1 occur for a particular m value and that is reasonable as the memberships of the groups in those cases are different even though they may have some overlap. However, the average values, marked by triangles in the plot, show an increase of H1 with m, i.e. the higher the number of t_2 values of group members outside the confidence limits, the higher the value of H1 that can be expected. If we consider a H1 value less than 4 as a good criterion for testing homogeneity, then in this approach we should not allow more than 2 values of t_2 to fall outside the confidence limits.

7. Figure 5.4 shows a plot between H1 and $d_{ij,max}$ of the pooling groups. The $d_{ij,max}$ is defined here as the distance associated with the group member which just qualified as a member of the pooling group . The plot shows an upward trend to some extent, implying that a low H1 value can be expected for a low $d_{ij,max}$ value, which is an implicit assumption of a ROI pooling scheme. However in many cases, low $d_{ij,max}$ values, even those below 1.0, can lead to a high value of H1 suggesting that the assumption may not always be true particularly for Irish conditions. A similar plot is drawn in Figure 5.5 between $d_{ij,max}$ and m. The plot leads to a similar conclusion that for Figure 5.3. While a low value of $d_{ij,max}$ is desirable, it is noted that even low values of $d_{ij,max}$ can occur where a significant number of t_2 of group members falls outside the CL.

Table 5.1: Summary of events outside the confidence limits for 85 pooling groups

No. of events outside the C.L., m	Number of ROI groups	% of ROI groups
>=1	8	9
2<=3	33	39
>3	44	52

Table 5.2: Summary of heterogeneity measure, H1 and H2 for 85 pooling groups

Heterogeneity	% of groups with het. <2	% of groups with het. <4
H1	5	22
H2	38	86

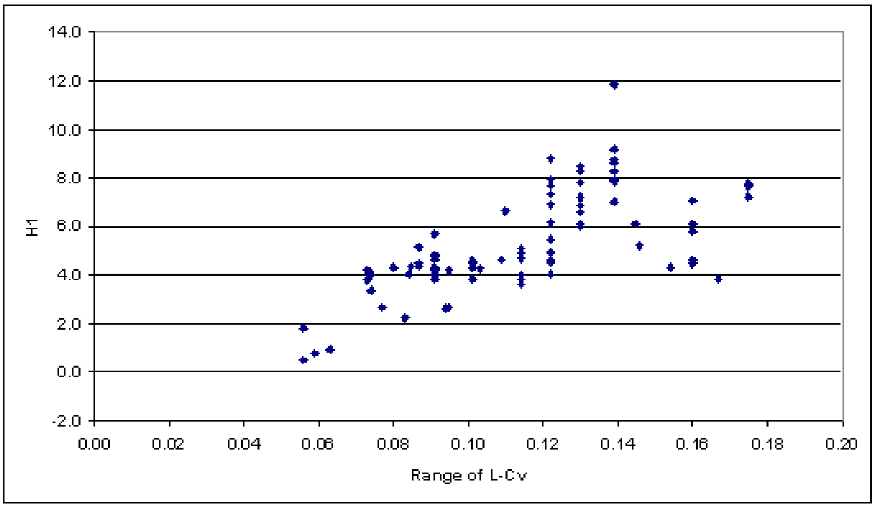


Figure 5.1: H1 plotted versus range of L-CV - Each point represents a pooling group

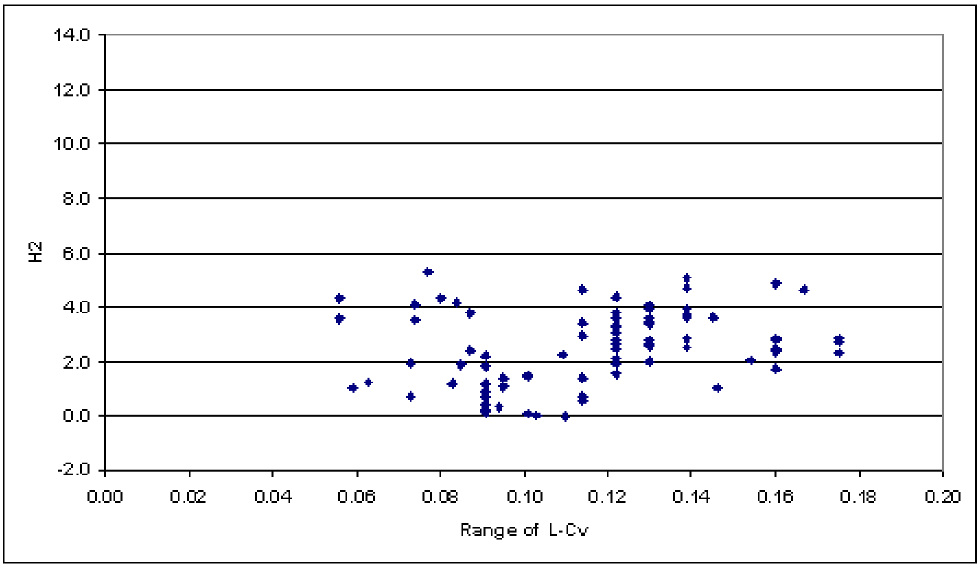


Figure 5.2: H2 plotted versus range of L-CV - Each point represents a pooling group

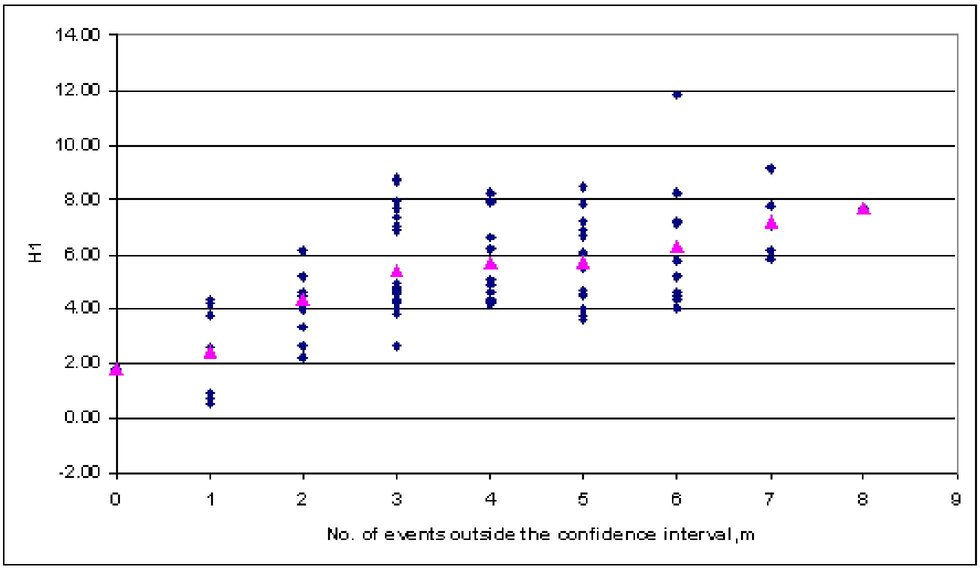


Figure 5.3: H1 plotted versus m - Each point represents a pooling group

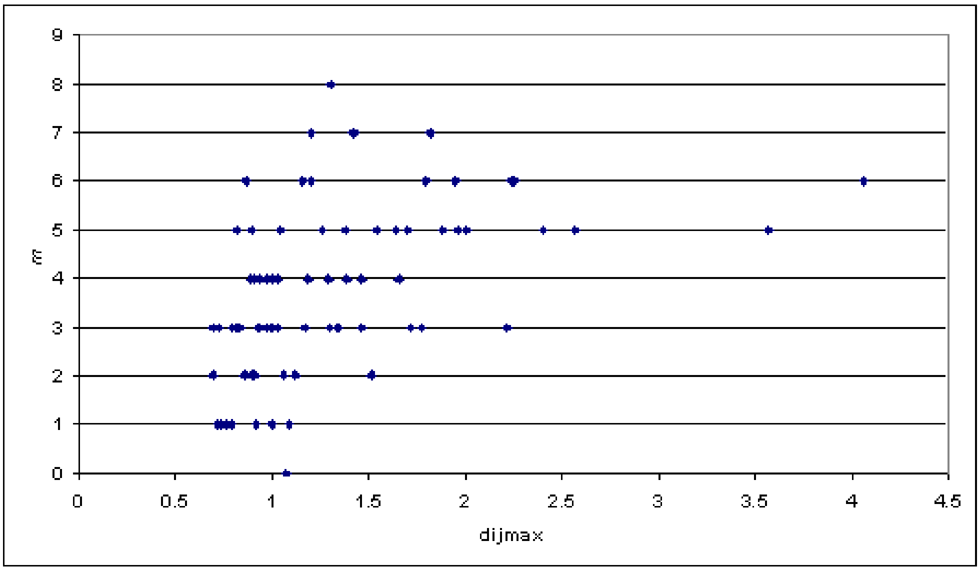


Figure 5.4: m **plotted versus** $d_{ij,max}$ - Each point represents a pooling group

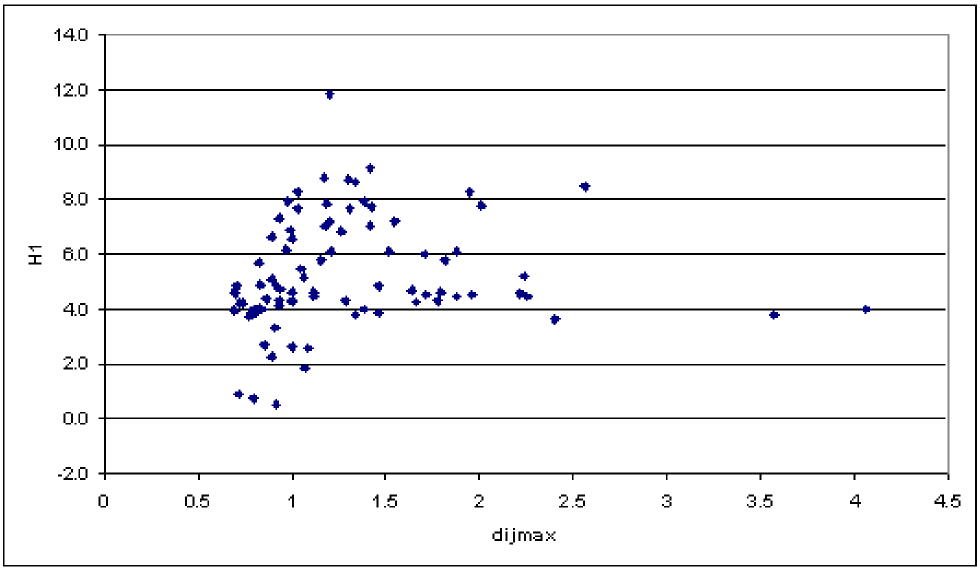


Figure 5.5: H1 plotted versus $d_{ij,max}$ - Each point represents a pooling group

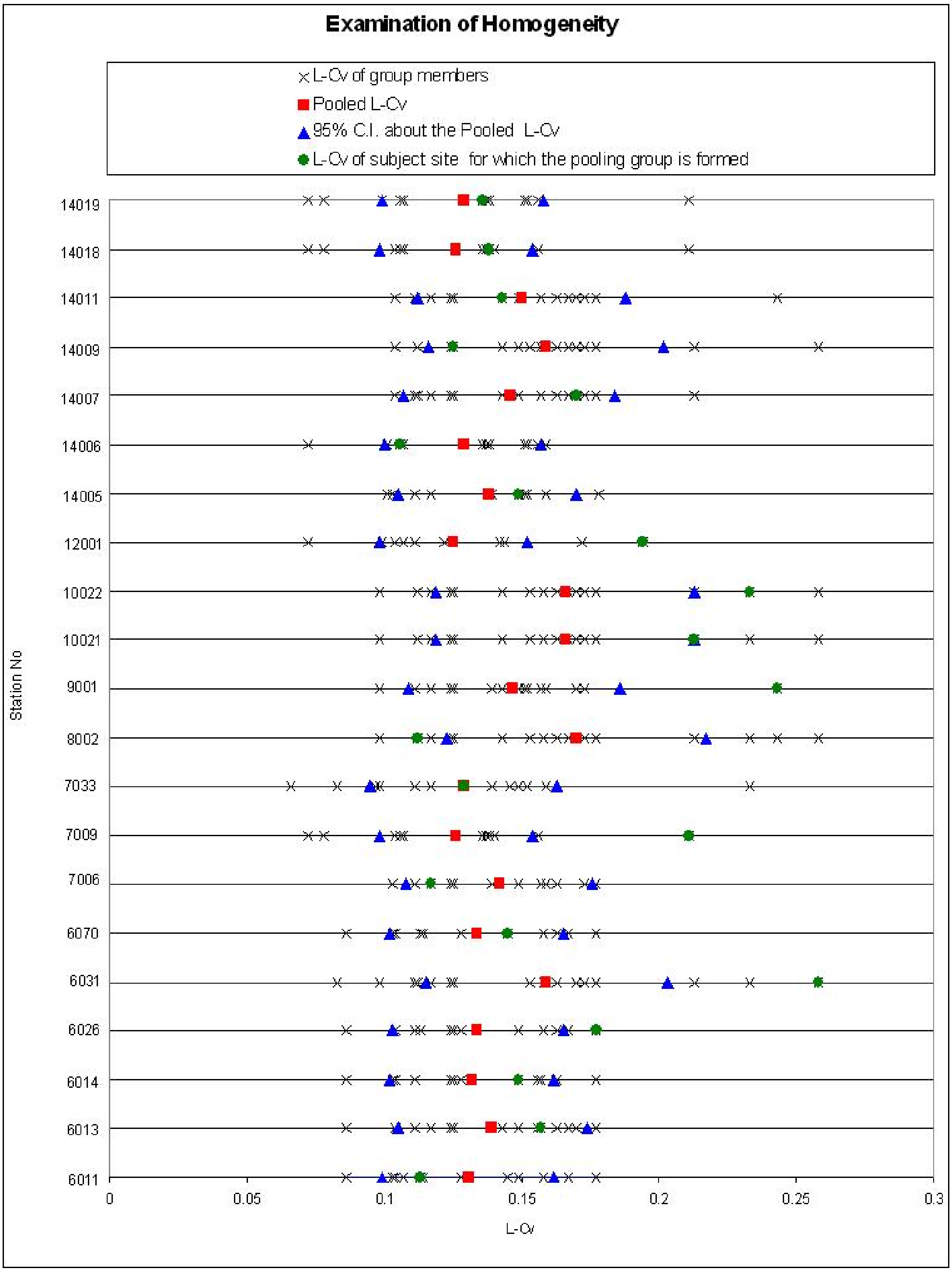


Figure 5.6: Examination of homogeneity -

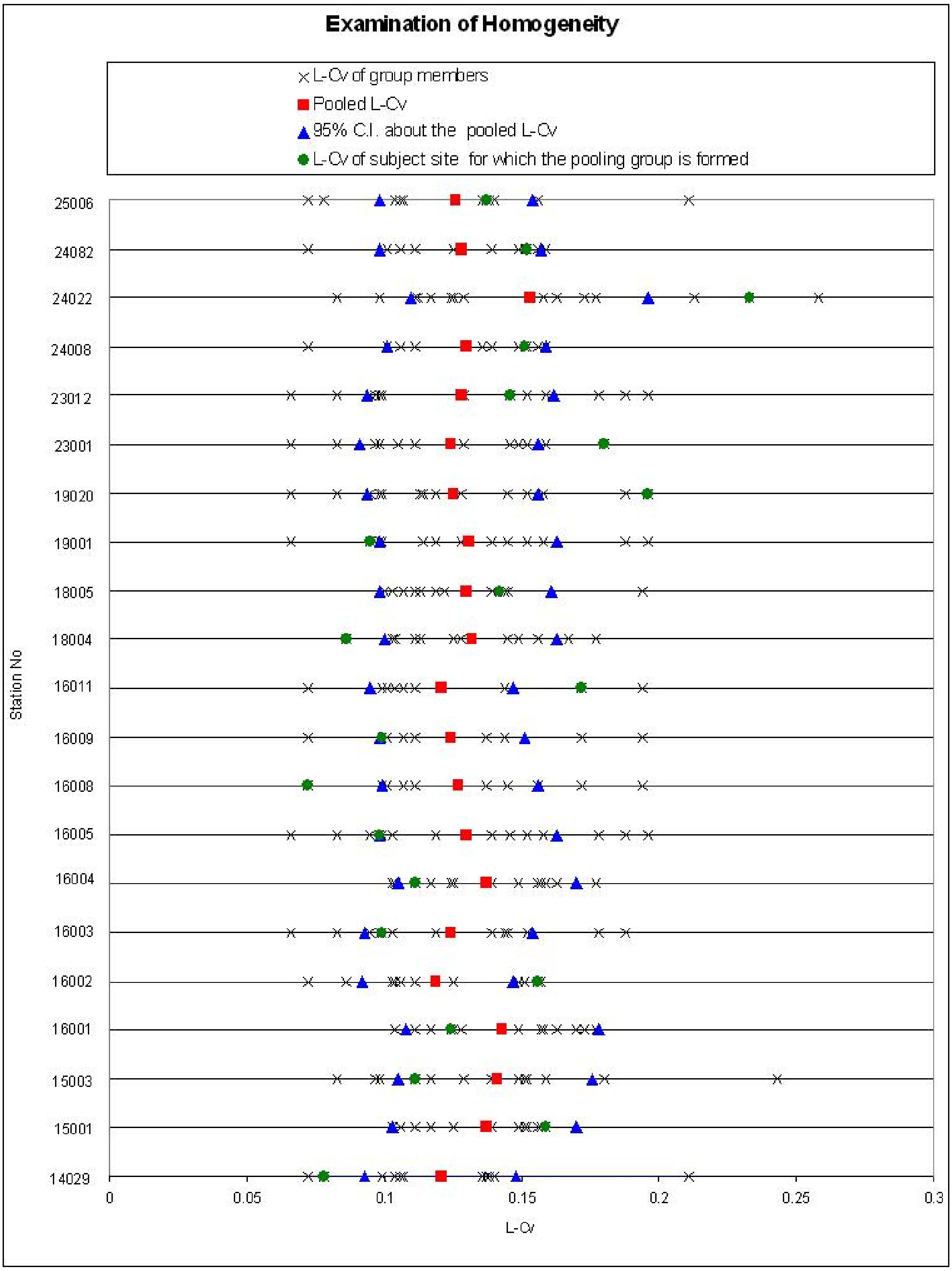


Figure 5.7: Examination of homogeneity - continued from Figure 5.6

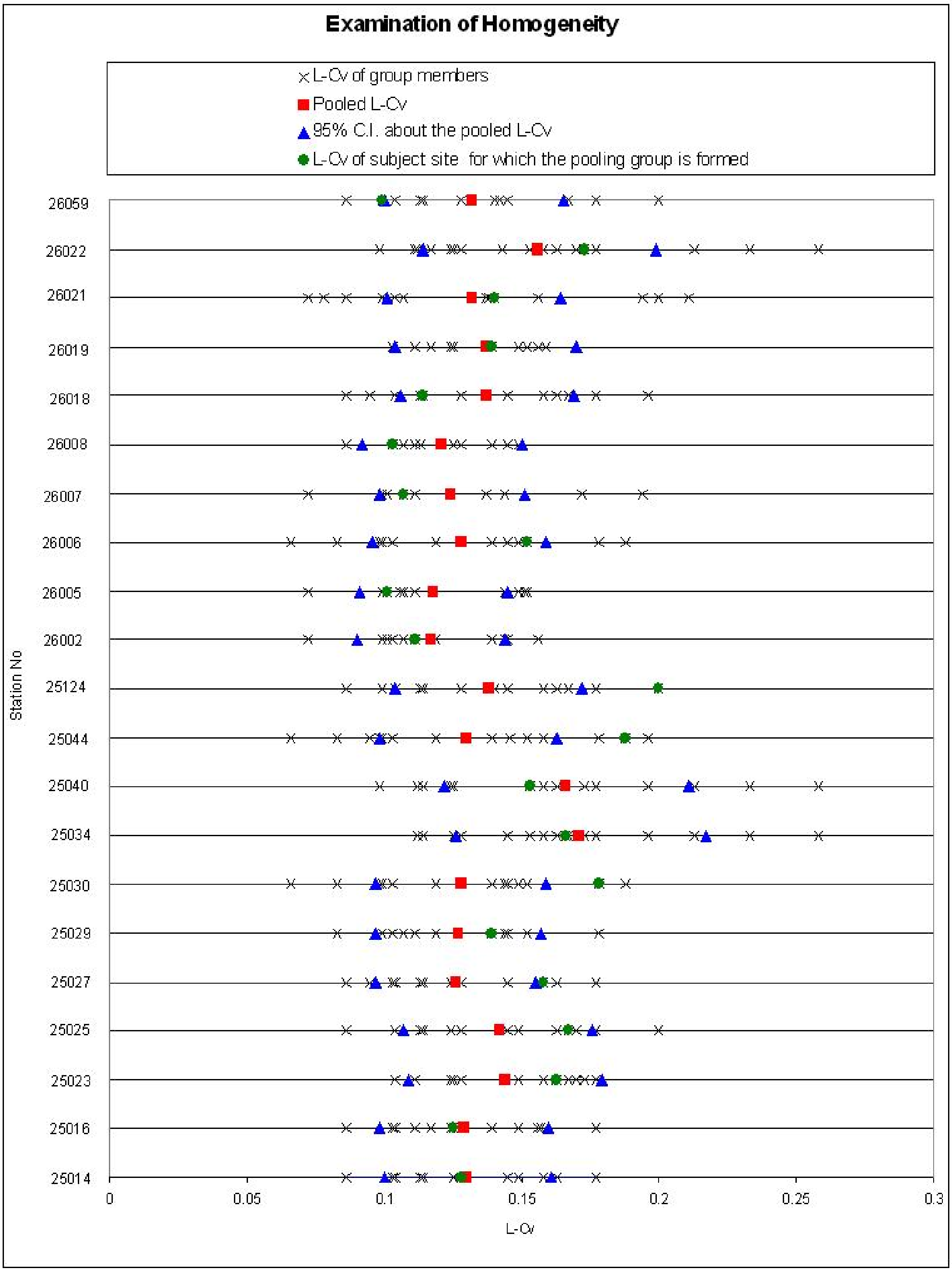


Figure 5.8: Examination of homogeneity - continued from Figure 5.7

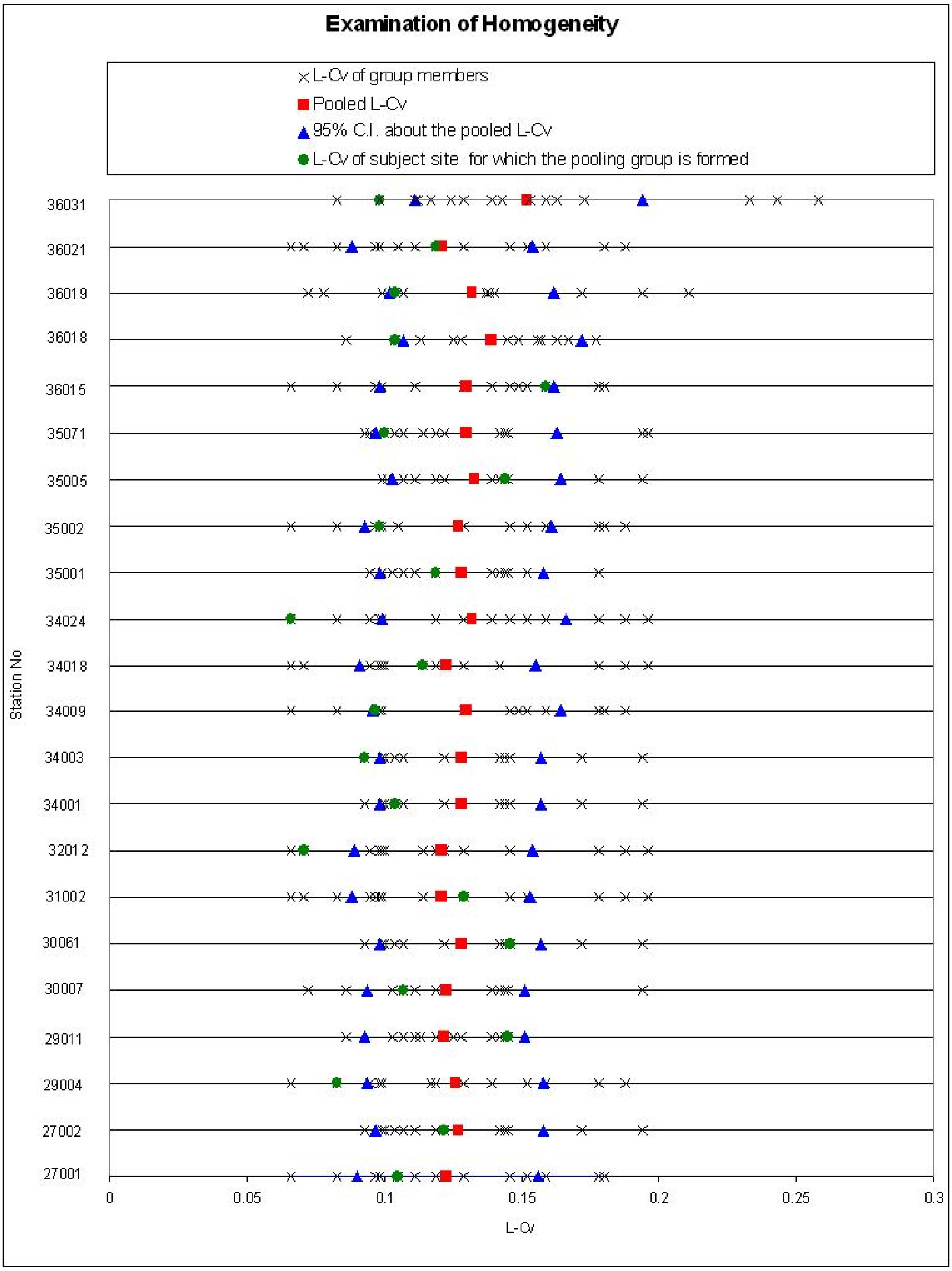


Figure 5.9: Examination of homogeneity - continued from Figure 5.8

Table 5.3: Summary results of the examination procedure for each pooling group.

St.No.	St.years	No. of Site	m	subject site ouside the C.I.		Range of L-Cv			Heterogeneity Measure		$d_{ij,max}$
				Y/N	$P-value$	Min	Max	$Diff$	$H1$	$H2$	
6011	523	13	3	N		0.09	0.18	0.09	3.99	0.68	0.836
6013	529	14	3	N		0.09	0.18	0.09	4.31	1.85	0.929
6014	531	12	3	N		0.09	0.18	0.09	4.60	1.17	0.699
6026	523	12	3	Y	$3.3E-03$	0.09	0.18	0.09	4.83	0.10	0.704
6031	506	17	8	Y	$5.2E-06$	0.08	0.26	0.17	7.66	2.82	1.306
6070	514	12	3	N		0.09	0.18	0.09	4.76	0.91	0.941
7006	523	13	2	N		0.10	0.18	0.07	3.33	4.10	0.911
7009	513	11	4	Y	$1.3E-09$	0.07	0.21	0.14	7.86	3.76	1.186
7033	512	15	3	N		0.07	0.23	0.17	3.81	4.66	1.336
8002	531	18	7	Y	$7.8E-03$	0.10	0.26	0.16	5.79	2.45	1.818
9001	515	15	2	Y	$7.0E-07$	0.10	0.24	0.14	6.13	3.65	1.517
10021	507	18	6	N		0.10	0.26	0.16	4.61	2.76	1.795
10022	507	18	6	Y	$2.6E-03$	0.10	0.26	0.16	4.45	2.85	2.254
12001	534	11	3	Y	$2.7E-07$	0.07	0.19	0.12	6.88	3.06	0.991
14005	500	12	3	N		0.10	0.18	0.08	2.66	5.29	0.999
14006	511	11	2	N		0.07	0.16	0.09	5.17	2.41	1.062
14007	534	16	2	N		0.10	0.21	0.11	4.62	2.24	1.119
14009	506	16	4	N		0.10	0.26	0.15	4.33	2.06	1.288
14011	538	15	3	N		0.10	0.24	0.14	7.03	2.49	1.174
14018	513	11	4	N		0.07	0.21	0.14	7.93	3.71	1.386
14019	546	12	3	N		0.07	0.21	0.14	8.75	3.96	1.299
14029	514	11	3	Y	$9.0E-04$	0.07	0.21	0.14	8.65	4.68	1.342
15001	519	13	0	N		0.10	0.16	0.06	1.82	3.57	1.071
15003	514	14	5	N		0.08	0.24	0.16	6.10	4.84	1.88
16001	505	13	1	N		0.10	0.18	0.07	4.20	1.95	0.742
16002	531	12	6	Y	$4.8E-03$	0.07	0.16	0.09	4.36	1.90	0.862
16003	511	13	4	N		0.07	0.19	0.12	6.17	2.79	0.972
16004	543	13	3	N		0.10	0.18	0.07	4.02	3.53	0.796
16005	523	14	6	N		0.07	0.20	0.13	7.16	3.98	1.203
16008	518	11	3	Y	$1.0E-04$	0.07	0.19	0.12	7.34	3.33	0.932
16009	500	10	3	N		0.07	0.19	0.12	7.67	3.56	1.033
16011	500	10	3	Y	$6.0E-05$	0.07	0.19	0.12	8.79	4.39	1.169
18004	541	13	3	Y	$1.8E-03$	0.09	0.18	0.09	3.83	0.94	0.81
18005	525	13	1	N		0.10	0.19	0.09	2.60	0.34	1.088
19001	503	13	4	Y	$1.4E-02$	0.07	0.20	0.13	6.59	3.47	0.997
19020	550	14	5	Y	$3.6E-06$	0.07	0.20	0.13	6.85	2.68	1.258
23001	501	14	5	Y	$3.0E-04$	0.07	0.18	0.11	3.62	4.67	2.4
23012	520	15	5	N		0.07	0.20	0.13	6.02	2.81	1.704

24008	507	11	1	N		0.07	0.16	0.09	4.34	3.78	0.997
24022	536	18	7	Y	$1.3E-04$	0.08	0.26	0.17	7.75	2.33	1.425
24082	546	12	2	N		0.07	0.16	0.09	4.49	3.81	1.116
25006	513	11	4	N		0.07	0.21	0.14	8.26	3.64	1.028
25014	533	12	3	N		0.09	0.18	0.09	4.64	0.24	0.691
25016	549	13	2	N		0.09	0.18	0.09	3.97	2.21	0.691
25023	512	13	1	N		0.10	0.18	0.07	3.75	0.71	0.77
25025	506	13	4	N		0.09	0.20	0.11	5.09	0.74	0.891
25027	542	12	5	Y	$1.5E-02$	0.09	0.18	0.09	5.67	0.25	0.823
25029	507	12	2	N		0.08	0.18	0.10	2.66	1.09	0.856
25030	511	13	4	Y	$7.9E-04$	0.07	0.19	0.12	4.62	2.47	0.999
25034	529	17	6	N		0.11	0.26	0.15	5.20	1.03	2.241
25040	537	17	7	N		0.10	0.26	0.16	7.04	1.73	1.418
25044	523	14	6	Y	$2.9E-04$	0.07	0.20	0.13	7.19	4.08	1.2
25124	508	13	4	Y	$1.8E-04$	0.09	0.20	0.11	4.86	0.55	1.465
26002	530	12	3	N		0.07	0.16	0.08	4.03	4.19	0.823
26005	503	11	4	N		0.07	0.15	0.08	4.29	4.35	1.003
26006	507	13	4	N		0.07	0.19	0.12	4.86	2.64	0.914
26007	500	10	3	N		0.07	0.19	0.12	7.92	3.78	0.977
26008	507	12	1	N		0.09	0.15	0.06	0.92	1.24	0.713
26018	519	12	5	N		0.09	0.20	0.11	6.66	-0.04	0.892
26019	527	13	1	N		0.10	0.16	0.06	0.52	4.34	0.919
26021	542	13	7	N		0.07	0.21	0.14	9.16	2.84	1.418
26022	536	17	6	N		0.10	0.26	0.16	5.77	2.34	1.155
26059	510	13	5	Y	$2.5E-02$	0.09	0.20	0.11	4.67	1.40	1.639
27001	512	15	5	N		0.07	0.18	0.11	3.78	2.96	3.568
27002	539	13	3	N		0.09	0.19	0.10	3.82	0.09	1.462
29004	513	14	5	Y	$4.2E-03$	0.07	0.19	0.12	5.48	3.82	1.043
29011	512	12	1	N		0.09	0.14	0.06	0.73	1.01	0.794
30007	522	12	3	N		0.07	0.19	0.12	4.92	2.09	0.833
30061	522	12	3	N		0.09	0.19	0.10	4.59	1.52	2.217
31002	513	15	6	N		0.07	0.20	0.13	8.26	3.58	1.949
32012	505	15	5	Y	$1.5E-03$	0.07	0.20	0.13	8.47	2.57	2.564
34001	522	12	3	N		0.09	0.19	0.10	4.53	1.48	1.718
34003	522	12	3	Y	$9.0E-03$	0.09	0.19	0.10	4.30	1.50	1.775
34009	537	15	5	N		0.07	0.19	0.12	4.50	1.95	1.877
34018	503	14	5	N		0.07	0.20	0.13	7.80	1.96	2.008
34024	540	15	7	Y	$7.1E-05$	0.07	0.20	0.13	6.12	3.36	1.204
35001	522	12	2	N		0.10	0.18	0.08	2.24	1.21	0.892
35002	519	15	5	N		0.07	0.19	0.12	4.55	1.54	1.964
35005	524	12	4	N		0.10	0.19	0.10	4.18	1.42	0.934
35071	510	14	4	N		0.09	0.20	0.10	4.28	0.03	1.658
36015	515	13	5	N		0.07	0.18	0.11	3.98	3.41	1.384

36018	504	12	3	Y	$1.9E-02$	0.09	0.18	0.09	4.20	0.37	0.724
36019	514	11	6	N		0.07	0.21	0.14	11.83	5.05	1.199
36021	521	16	6	N		0.07	0.19	0.12	4.02	3.28	4.059
36031	508	16	5	Y	$5.9E-03$	0.08	0.26	0.17	7.20	2.71	1.544

5.5 Investigation of selected heterogeneous pooling groups

The investigation has been carried out for 27 cases where the pooling groups are heterogeneous and in which the t_2 of the subject site lies outside the confidence limits. The investigation mainly focuses on identifying any inappropriateness among group members that would cause the pooling groups to be heterogeneous.

In this context, FEH (1999, 3, fig 16.9) documented a detailed review system, providing an example . It mainly considers two attributes: 1) whether the subject site has any special qualities that need to be taken into account and 2) whether any of the pooled sites has catchment descriptors that are particularly different from these of the subject site.

Sites in the pooling group can be investigated using several characteristics including at-site flood statistics and catchment descriptors. Statistics in a pooling group such as the discordancy measure (D) and distance measure the (d_{ij}) can also be used to investigate sites in the pooling group. In this part of the study, four catchment descriptors, namely, Catchment area, SAAR, BFI and FARL, and the distance measure (d_{ij}) are taken into account in the investigation process. The first three of the catchment descriptors were already used for initial selection of sites to form a pooling group.

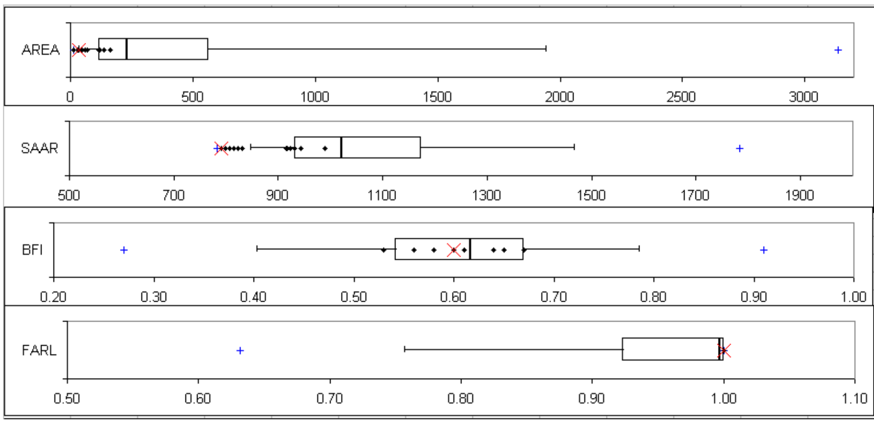
In the investigation procedure, sites are reviewed with the help of Box-plots and a summary table and, in some cases, with the help of 'examination of homogeneity' chart. Four Box-plots of catchment descriptors, such as catchment size (AREA), wetness (SAAR), soils (BFI) and lakes and reservoirs (FARL), are constructed to show the subject site in the context of the pooling group. For each of these descriptors, the distribution of values for sites in the pooling group is shown against a backdrop of the relative distribution of the 85 sites. This facilitates the identification of any particularly inappropriate sites. In the summary table, statistical properties such as t_2, t_3 and d_{ij} values of sites in a group are listed. The investigation procedure for pooling groups of station nos 6026 and 6031 are described in detail as they serve as examples. In case of the remaining 25 pooling groups, significant outcomes of the investigation procedure are listed in Table 5.4. The examination aids, i.e. Box-plots and a summary table for those cases, are shown at the end of this chapter.

.

.

Station No.6026 on the River Glyde

The pooling group for the subject site 6026 contains 12 sites. Figure 5.10 displays a summary table which lists those 12 sites including their t_2, t_3 and d_{ij} values, and four Box-plots of catchment descriptors. The investigation of Box-plots reveals the presence of a reservoir/lake on the subject site. The FARL value for the catchment is 0.92. The selected group also includes a number of other catchments with a strong reservoir/lake effect, notably station nos 36018 and 6011. The examination of homogeneity chart in Figure 5.6 showed that the t_2 of three sites, including the subject site 6026 fell outside the CL, the other two sites being 25025 and 18004. The d_{ij} values of these sites are 0.518 and 0.655 respectively which indicate the catchment descriptors for these sites are fairly similar to these of the subject site. Although the t_2 values of these sites fall outside the CL, they are not very far from it and thus it can be concluded that the group is only moderately heterogeneous. The heterogeneity measures, H1 and H2 for the group are 4.83 and 0.1 respectively. Therefore there are no strong grounds for forming a new group for the subject site by excluding sites no. 25025 and 18004 as being inappropriate.



St.No	RL	t2	t3	dij
6026	46	0.177	0.24	0
25023	52	0.163	0.142	0.249
16001	33	0.124	0.012	0.398
25014	54	0.128	0.103	0.409
36018	50	0.104	0.045	0.418
25027	43	0.158	0.046	0.509
25025	31	0.167	0.083	0.518
6014	30	0.149	0.224	0.526
25016	42	0.125	0.082	0.607
18004	46	0.086	0.043	0.655
16004	48	0.111	0.069	0.682
6011	48	0.113	0.09	0.704

Figure 5.10: Four Box-plots and a summary table for investigating a pooling group - The subject site is marked with an×. Small dots denote sites included in the pooling group. The underlying distribution of each catchment descriptor is shown in the Box-plots. Each Box-plot gives the minimum and the maximum value (+) and percentiles for the frequencies 0.05, 0.25, 0.5, 0.75, 0.95. The summary table lists record length, t_2, t_3 and d_{ij} values for a 100-year pooling group for subject station 6026.

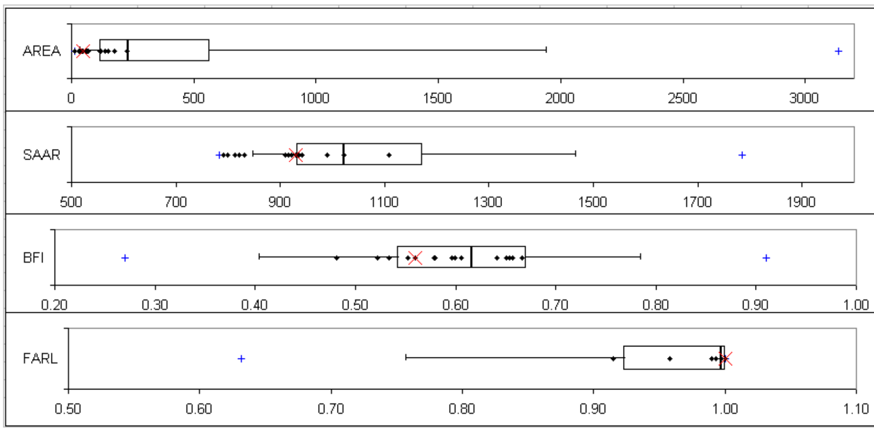
.

.

Station No.6031 on the River Flurry

There are 17 sites in the pooling group of which eight, including the subject site, have t_2 values which fall outside the CL, thus indicating a strongly heterogeneous group. The heterogeneity measures H1 and H2 for the group are 7.66 and 2.82 respectively. The examination of Box-plots in Figure 5.11 reveals the catchment area of the subject site is small (46.2 km^2) and it is very near to the 5 percentile mark on the Box-plot of AREA. The site is not positioned at the heart of the group of gauged catchments. There are 5 sites on the left of the subject site and there are as many as 11 sites on the right. The attribute certainly includes some sites that have large catchment area compared to the subject site. This may lead to d_{ij} values exceeding the value 1.0 in several cases. The d_{ij} values for the last three sites are around 1.3 and these sites are among the seven other sites that fall outside the CL. The examination of the summary table 5.11 shows that the subject site has large values of both t_2 and t_3 and that these are the largest among the group members.

Hence, the conclusion can be drawn here that the pooling group in its present structure may not be ideal for that subject site 6031. Leaving out some sites at the bottom of the table on the right side of Figure 5.11 might be considered in this context.



St.No	RL	t2	t3	dij
6031	18	0.258	0.392	0
24022	20	0.233	0.123	0.277
26022	33	0.173	0.091	0.296
36031	30	0.098	0.365	0.76
25040	20	0.153	0.203	0.809
16001	33	0.124	0.012	0.938
25023	52	0.163	0.142	0.988
8002	20	0.112	0.271	1.021
7006	19	0.117	-0.205	1.068
14009	25	0.125	0.136	1.092
25027	43	0.158	0.046	1.144
10021	24	0.213	0.186	1.146
6026	46	0.177	0.24	1.192
14007	25	0.17	0.219	1.234
10022	18	0.233	0.057	1.274
16004	48	0.111	0.069	1.276
29004	32	0.083	0.144	1.306

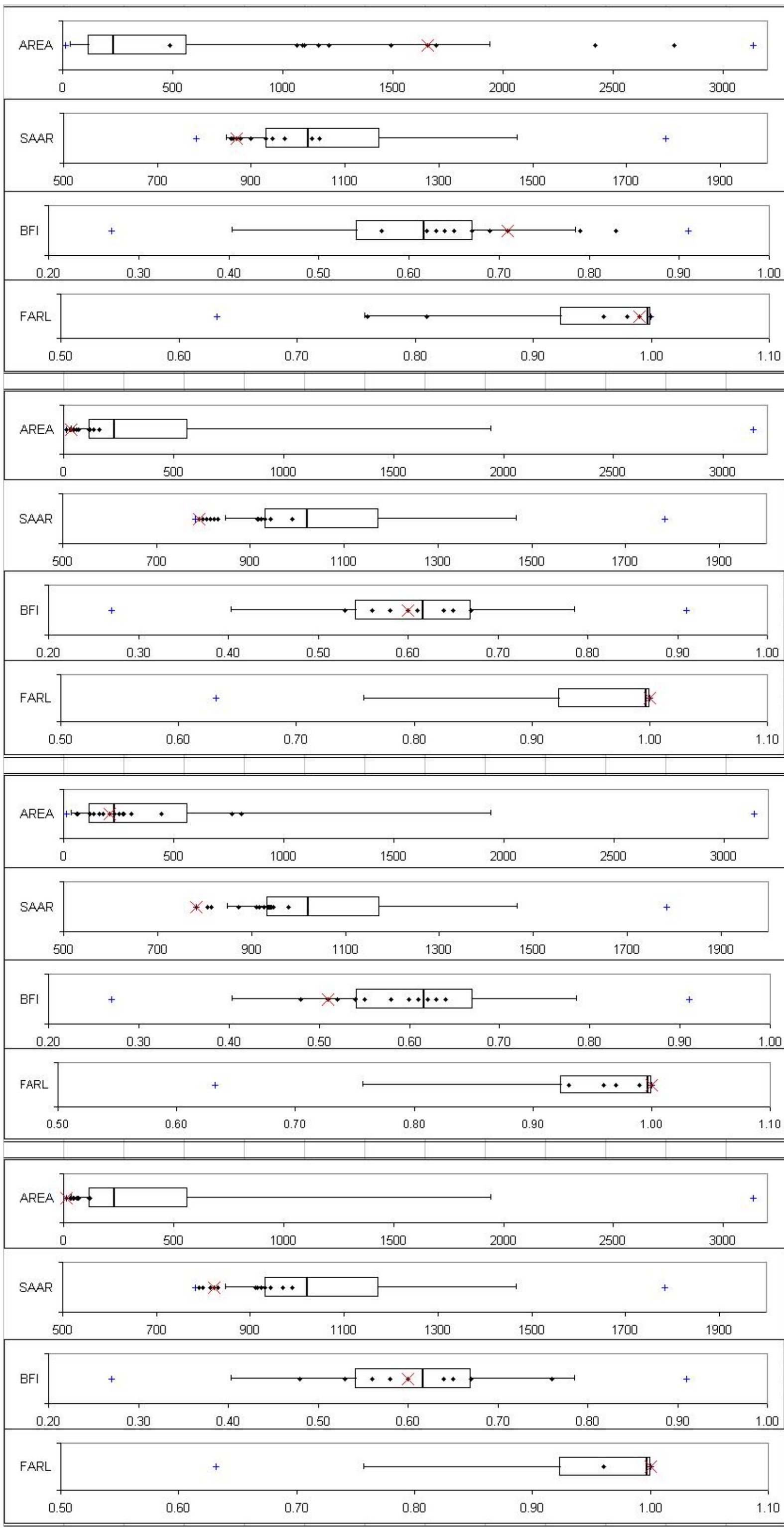
Figure 5.11: Box-plots and summary statistics for evaluating the pooling group for the subject site 6031 -

Table 5.4: Examination outcomes of the remaining 25 pooling groups out of the 27 which were considered heretogeneous and whose t_2 values fell outside CL value

Pooling group for Station No.		Special features of the pooling group
7009	subject site	Subject site possesses a unique feature with a large AREA, a small SAAR and a large BFI value
	group members	Pooling group includes a number of sites having a strong reservoir/lake influence
8002	subject site	The magnitude of AREA and SAAR are near to the 5 percentile mark
	group members	The dij values in several group memebers are above 1.5
9001	subject site	The magnitude of SAAR is the lowest among 85 catchments. The value of t2 is the largest among the group members and it falls a distance away from the upper end of the C.L.
	group members	Except subject site, only one t2 value of another site falls outside the C.L
10022	subject site	The magnitude of AREA and SAAR are below the 5 percentile mark
	group members	The dij values in several group members (6 out of 18) are above 2.0
12001	subject site	No special quality in the context of the pooling group investigation system
	group members	Several group members have strong reservoir/lake influence
14029	subject site	The magnitude of AREA is near to the maximum value
	group members	Several group members have strong reservoir/lake influence which might contribute to have negative skewness in many sites
16002	subject site	No special quality in the context of the pooling group investigation system
	group members	The magnitude of t2 of several sites that fall outside the C.L. are not very far from the C.L.

16008	subject site	The value of t3 is the lowest among the group members
	group members	Group members are fairly similar to the subject site in terms of key C.D.
16011	subject site	The magnitude of AREA is above the 95 percentile mark
	group members	Several group members have strong reservoir/lake influence
18004	subject site	No special quality in the context of the pooling group investigation system
	group members	Several sites have strong reservoir/lake influence. The magnitude of t2 of several sites that fall outside the C.L. are not very far from the C.L.
19001	subject site	No special quality in the context of the pooling group investigation system
	group members	Several sites have strong reservoir/lake influence
19020	subject site	The magnitude of AREA is near to the 5 percentile mark
	group members	Several sites have strong reservoir/lake influence
23001	subject site	The value of BFI is near to the 5 percentile mark
	group members	The dij values in many cases are above 2.0
24022	subject site	The magnitude of AREA is near to the 5 percentile mark
	group members	The value of t2 of seven sites fall outside the C.L.
25027	subject site	No special quality in the context of the pooling group investigation system
	group members	The magnitude of t2 of several sites that fall outside the C.L. are not very far from the C.L
25030	subject site	strong reservoir/lake influence
	group members	Several sites have strong reservoir/lake influence and they have FARL value less than 0.9

25044	subject site	The magnitude of AREA is near to the 5 percentile mark
	group members	Several sites have strong reservoir/lake influence
25124	subject site	Subject site have strong reservoir/lake influence and that might contribute to have negative skewness
	group members	Most of the sites have strong reservoir/lake influence
26059	subject site	The value of BFI is the maximum value among 85 stations. The value of FARL is also very small implying strong reservoir/lake influence
	group members	Most of the sites have strong reservoir/lake influence
29004	subject site	Moderate reservoir/lake influence
	group members	Several sites have negative skewness
32012	subject site	The magnitude of SAAR is the maximum value among 85 stations
	group members	The dij values in most group members are above 2.0. As many as 5 sites have t2 fall outside the C.L.
34003	subject site	A unique feature with a large AREA, large SAAR, large BFI and a small FARL value
	group members	Several sites have strong reservoir/lake influence
34024	subject site	No special quality in the context of the pooling group investigation system
	group members	Seven sites have t2 fall outside the C.L. although the dij values in the group are under 1.2.
36018	subject site	Strong reservoir/lake influence
	group members	Sites have t2 fall outside the C.L. are very close to the C.L.
36031	subject site	The magnitude of AREA is near to the 5 percentile mark
	group members	Some sites have high t2 values



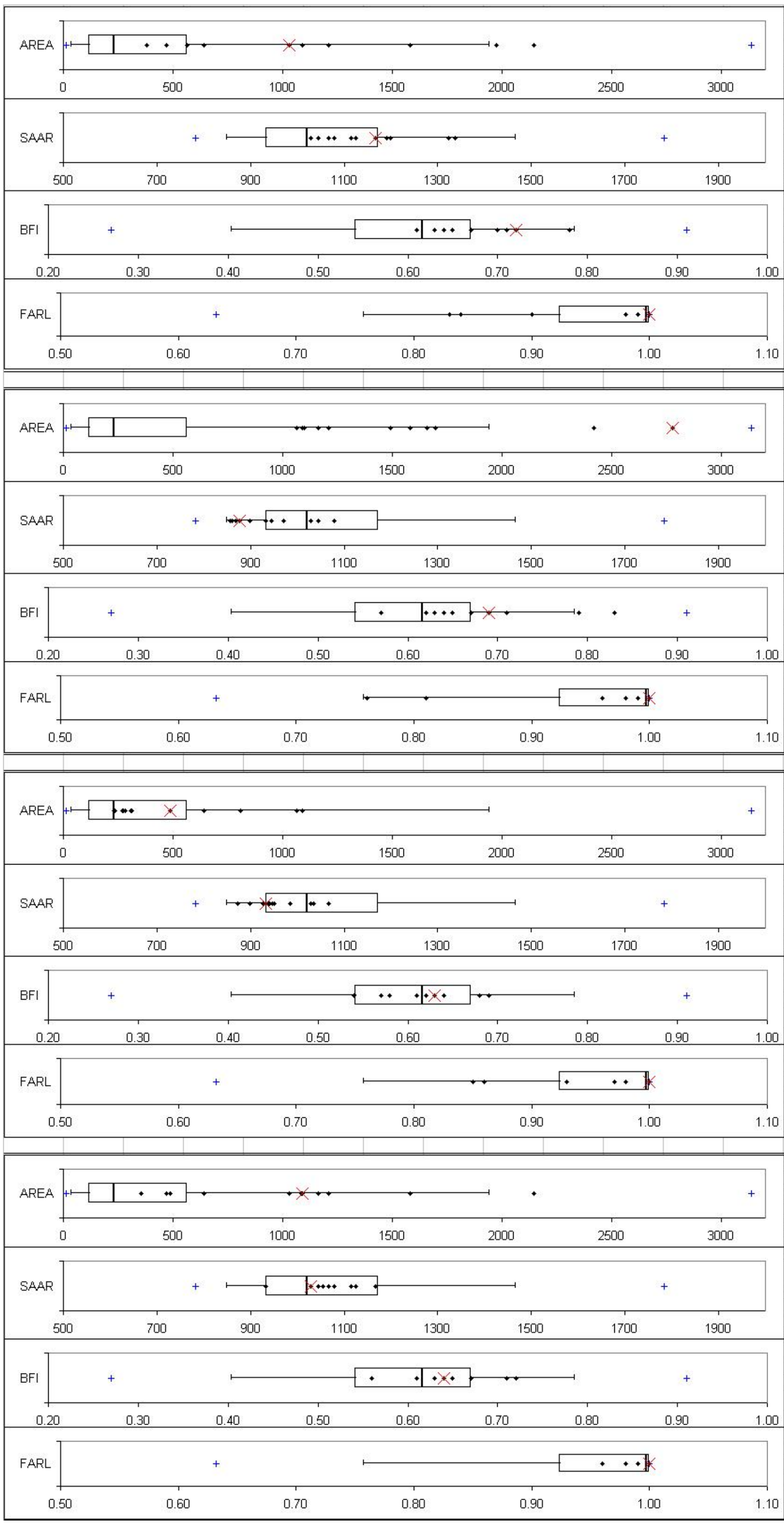
St.No	RL	t2	t3	dij
7009	29	0.211	0.206	0
14018	51	0.138	0.036	0.406
14029	47	0.078	0.058	0.431
25006	52	0.137	0.14	0.494
14019	51	0.136	0.092	0.622
36019	47	0.104	-0.029	0.812
26021	30	0.14	-0.112	0.924
14006	51	0.106	0.238	1.082
16008	51	0.072	-0.062	1.139
26007	53	0.107	0.164	1.171
16002	51	0.156	0.172	1.186

St.No	RL	t2	t3	dij
8002	20	0.112	0.271	0
10021	24	0.213	0.186	0.371
10022	18	0.233	0.057	0.782
14009	25	0.125	0.136	0.809
26022	33	0.173	0.091	0.986
6031	18	0.258	0.392	1.021
14007	25	0.17	0.219	1.056
24022	20	0.233	0.123	1.169
14011	26	0.143	0.02	1.254
25040	20	0.153	0.203	1.331
25023	52	0.163	0.142	1.362
16001	33	0.124	0.012	1.393
36031	30	0.098	0.365	1.399
6026	46	0.177	0.24	1.605
9001	48	0.243	0.191	1.631
7006	19	0.117	-0.205	1.687
25025	31	0.167	0.083	1.716
25027	43	0.158	0.046	1.818

St.No	RL	t2	t3	dij
9001	48	0.243	0.191	0
14011	26	0.143	0.02	0.789
7006	19	0.117	-0.205	1.095
6013	30	0.157	0.019	1.133
14007	25	0.17	0.219	1.153
15001	42	0.159	0.003	1.182
16004	48	0.111	0.069	1.214
16001	33	0.124	0.012	1.269
36031	30	0.098	0.365	1.309
26019	51	0.139	0.234	1.324
25016	42	0.125	0.082	1.384
6014	30	0.149	0.224	1.387
26022	33	0.173	0.091	1.445
24082	28	0.152	-0.036	1.476
24008	30	0.151	0.055	1.517

St.No	RL	t2	t3	dij
10022	18	0.233	0.057	0
8002	20	0.112	0.271	0.782
10021	24	0.213	0.186	0.831
25040	20	0.153	0.203	1.267
6031	18	0.258	0.392	1.274
24022	20	0.233	0.123	1.334
26022	33	0.173	0.091	1.394
14009	25	0.125	0.136	1.41
25034	24	0.166	-0.08	1.444
36031	30	0.098	0.365	1.725
14007	25	0.17	0.219	1.776
25023	52	0.163	0.142	1.879
16001	33	0.124	0.012	1.958
14011	26	0.143	0.02	2.001
6026	46	0.177	0.24	2.125
25027	43	0.158	0.046	2.184
7006	19	0.117	-0.205	2.238
25025	31	0.167	0.083	2.254

Figure 5.12: Box-plots and summary tables for evaluating pooling group for subject sites 7009, 8002, 9001, 10022. -



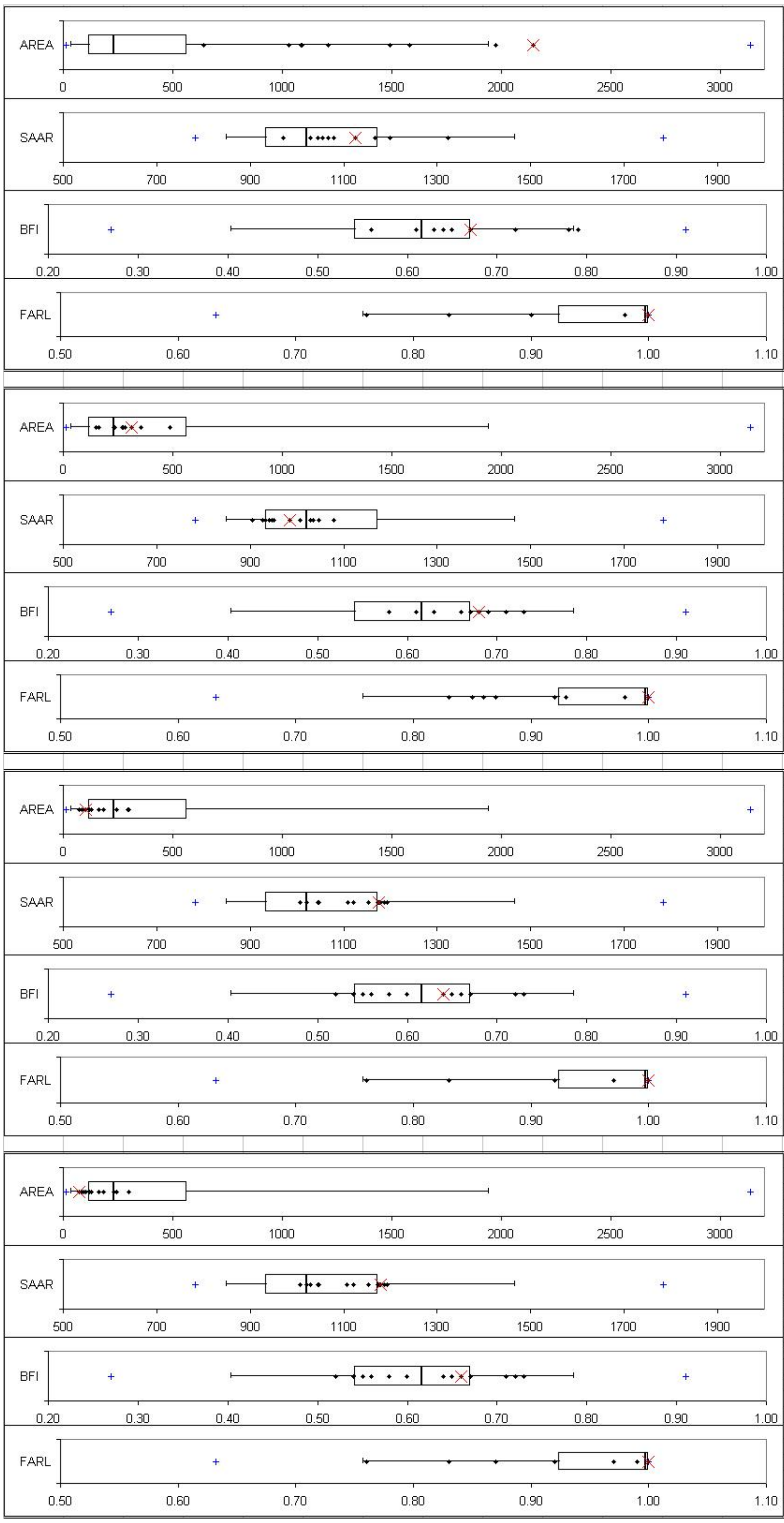
St.No	RL	t2	t3	dij
12001	50	0.194	0.176	0
16011	52	0.172	0.089	0.699
26007	53	0.107	0.164	0.798
18005	50	0.142	0.224	0.802
30007	31	0.107	0.112	0.821
16009	52	0.099	-0.096	0.834
35005	55	0.144	0.205	0.86
16008	51	0.072	-0.062	0.903
27002	51	0.122	0.181	0.923
34001	36	0.104	0.08	0.957
26002	53	0.111	0.224	0.991

St.No	RL	t2	t3	dij
14029	47	0.078	0.058	0
14018	51	0.138	0.036	0.215
7009	29	0.211	0.206	0.431
14019	51	0.136	0.092	0.633
25006	52	0.137	0.14	0.785
36019	47	0.104	-0.029	0.987
14006	51	0.106	0.238	1.166
26021	30	0.14	-0.112	1.2
16008	51	0.072	-0.062	1.236
26007	53	0.107	0.164	1.242
16009	52	0.099	-0.096	1.342

St.No	RL	t2	t3	dij
16002	51	0.156	0.172	0
6014	30	0.149	0.224	0.464
25016	42	0.125	0.082	0.481
6013	30	0.157	0.019	0.525
18004	46	0.086	0.043	0.592
16004	48	0.111	0.069	0.707
36018	50	0.104	0.045	0.719
26008	49	0.103	0.193	0.76
14006	51	0.106	0.238	0.798
24008	30	0.151	0.055	0.812
26002	53	0.111	0.224	0.823
16008	51	0.072	-0.062	0.862

St.No	RL	t2	t3	dij
16008	51	0.072	-0.062	0
26007	53	0.107	0.164	0.139
16009	52	0.099	-0.096	0.404
26002	53	0.111	0.224	0.516
26005	51	0.101	0.035	0.626
25006	52	0.137	0.14	0.747
16011	52	0.172	0.089	0.768
30007	31	0.107	0.112	0.812
16002	51	0.156	0.172	0.862
12001	50	0.194	0.176	0.903
29011	22	0.145	0.41	0.932

Figure 5.13: Box-plots and summary tables for evaluating the pooling groups for the subject sites 12001,14029,16002 and 16008. -



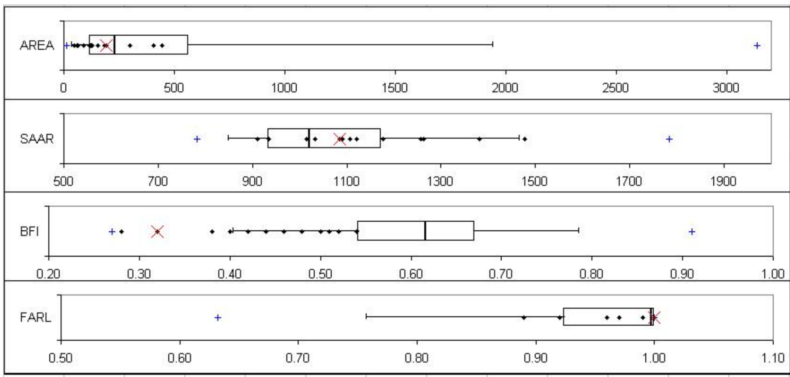
St.No	RL	t2	t3	dij
16011	52	0.172	0.089	0
16009	52	0.099	-0.096	0.442
26007	53	0.107	0.164	0.635
12001	50	0.194	0.176	0.699
16008	51	0.072	-0.062	0.768
26005	51	0.101	0.035	1.049
26002	53	0.111	0.224	1.089
35005	55	0.144	0.205	1.109
34001	36	0.104	0.08	1.166
36019	47	0.104	-0.029	1.169

St.No	RL	t2	t3	dij
18004	46	0.086	0.043	0
36018	50	0.104	0.045	0.313
6011	48	0.113	0.09	0.398
6014	30	0.149	0.224	0.506
25014	54	0.128	0.103	0.523
25016	42	0.125	0.082	0.555
26008	49	0.103	0.193	0.579
16002	51	0.156	0.172	0.592
29011	22	0.145	0.41	0.639
6026	46	0.177	0.24	0.655
6070	24	0.145	0.135	0.699
25025	31	0.167	0.083	0.785
16004	48	0.111	0.069	0.81

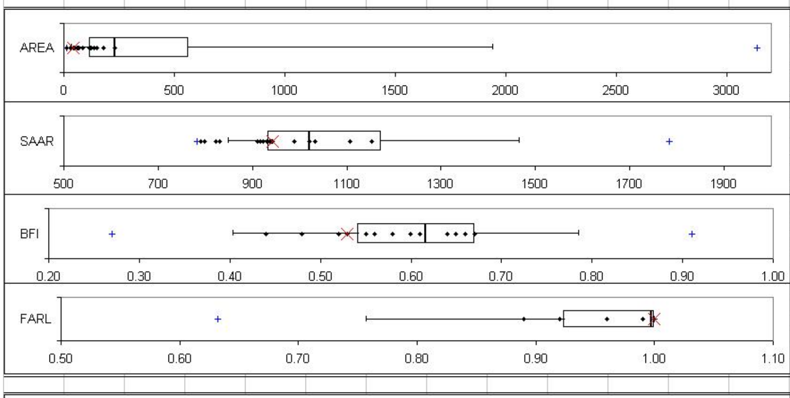
St.No	RL	t2	t3	dij
19001	48	0.095	0.095	0
19020	28	0.196	0.029	0.3
25044	33	0.188	0.248	0.462
16005	30	0.098	0.204	0.641
25027	43	0.158	0.046	0.823
26018	49	0.114	0.126	0.879
35001	29	0.119	-0.009	0.892
26006	53	0.152	0.397	0.944
34024	28	0.066	-0.051	0.963
6070	24	0.145	0.135	0.97
16003	51	0.099	0.207	0.972
25014	54	0.128	0.103	0.985
25029	33	0.139	-0.02	0.997

St.No	RL	t2	t3	dij
19020	28	0.196	0.029	0
19001	48	0.095	0.095	0.3
25044	33	0.188	0.248	0.618
16005	30	0.098	0.204	0.768
26018	49	0.114	0.126	0.892
25027	43	0.158	0.046	0.913
6070	24	0.145	0.135	1.035
25014	54	0.128	0.103	1.106
34024	28	0.066	-0.051	1.171
35001	29	0.119	-0.009	1.188
26006	53	0.152	0.397	1.205
29004	32	0.083	0.144	1.213
6011	48	0.113	0.09	1.236
16003	51	0.099	0.207	1.258

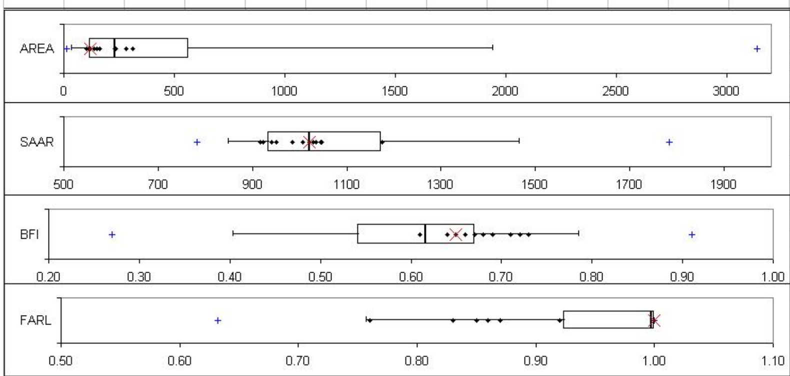
Figure 5.14: Box-plots and summary tables for evaluating the pooling groups for the subject sites 16011,18004,19001 and 19020. -



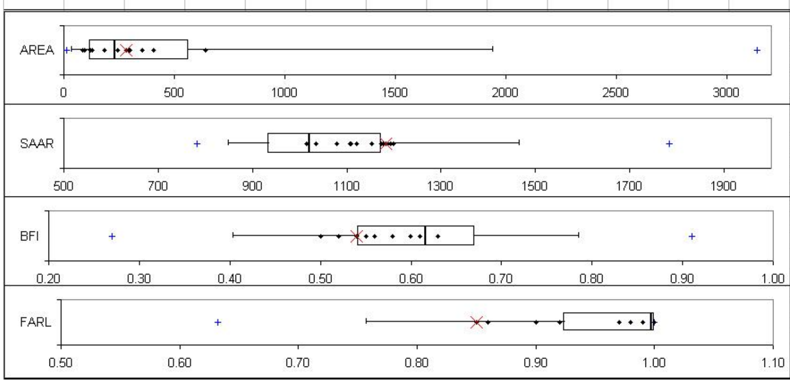
St.No	RL	t2	t3	dij
23001	45	0.18	0.128	0
15003	50	0.111	-0.146	1.216
36015	33	0.159	0.325	1.256
34009	33	0.097	0.081	1.385
7033	25	0.129	0.162	1.521
35002	34	0.098	0.026	1.964
23012	18	0.146	0.386	2.083
14005	48	0.149	0.291	2.158
27001	30	0.105	0.193	2.19
29004	32	0.083	0.144	2.252
36031	30	0.098	0.365	2.281
34024	28	0.066	-0.051	2.293
15001	42	0.159	0.003	2.389
26006	53	0.152	0.397	2.4



St.No	RL	t2	t3	dij
24022	20	0.233	0.123	0
6031	18	0.258	0.392	0.277
26022	33	0.173	0.091	0.548
36031	30	0.098	0.365	0.603
25040	20	0.153	0.203	0.957
16001	33	0.124	0.012	1.148
7006	19	0.117	-0.205	1.167
8002	20	0.112	0.271	1.169
25023	52	0.163	0.142	1.237
29004	32	0.083	0.144	1.266
16005	30	0.098	0.204	1.319
7033	25	0.129	0.162	1.33
10022	18	0.233	0.057	1.334
25027	43	0.158	0.046	1.336
10021	24	0.213	0.186	1.345
14009	25	0.125	0.136	1.353
16004	48	0.111	0.069	1.416
6026	46	0.177	0.24	1.425

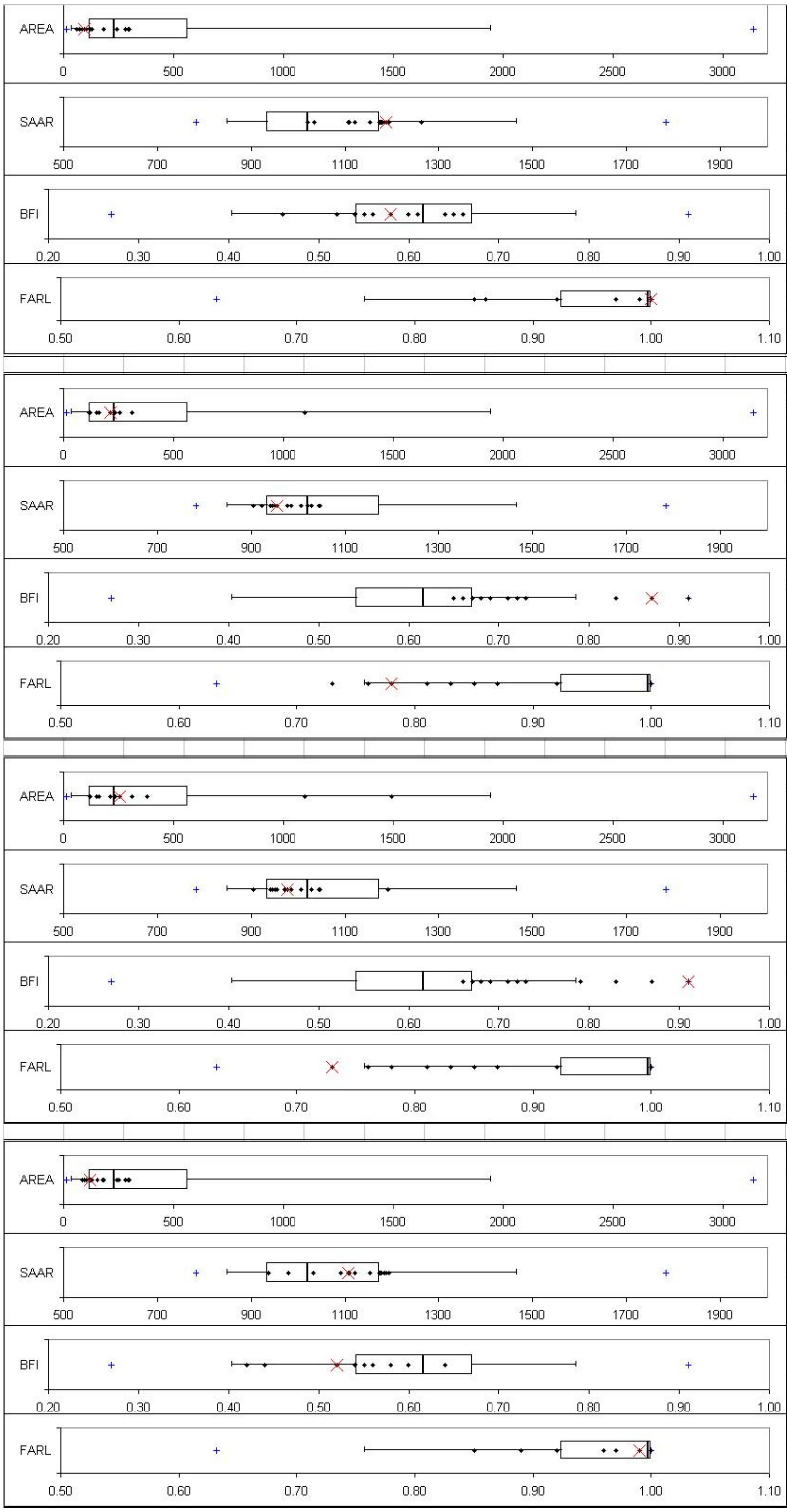


St.No	RL	t2	t3	dij
25027	43	0.158	0.046	0
25014	54	0.128	0.103	0.301
26018	49	0.114	0.126	0.485
6026	46	0.177	0.24	0.509
25023	52	0.163	0.142	0.587
6070	24	0.145	0.135	0.601
6011	48	0.113	0.09	0.659
16001	33	0.124	0.012	0.696
36018	50	0.104	0.045	0.732
26008	49	0.103	0.193	0.742
18004	46	0.086	0.043	0.812
19001	48	0.095	0.095	0.823



St.No	RL	t2	t3	dij
25030	48	0.178	0.185	0
16003	51	0.099	0.207	0.145
26006	53	0.152	0.397	0.456
35001	29	0.119	-0.009	0.488
25029	33	0.139	-0.02	0.501
34024	28	0.066	-0.051	0.648
29004	32	0.083	0.144	0.783
35005	55	0.144	0.205	0.861
29011	22	0.145	0.41	0.903
25044	33	0.188	0.248	0.934
26008	49	0.103	0.193	0.952
16005	30	0.098	0.204	0.977
14005	48	0.149	0.291	0.999

Figure 5.15: Box-plots and summary tables for evaluating the pooling group for the subject sites 23001,24022,25027 and 25030. -



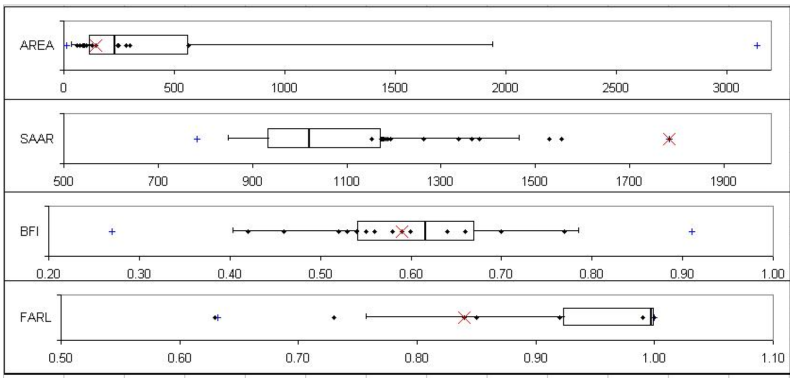
25044	33	0.188	0.248	0
16005	30	0.098	0.204	0.242
19001	48	0.095	0.095	0.462
34024	28	0.066	-0.051	0.561
19020	28	0.196	0.029	0.618
29004	32	0.083	0.144	0.674
26006	53	0.152	0.397	0.717
16003	51	0.099	0.207	0.802
25030	48	0.178	0.185	0.934
35001	29	0.119	-0.009	0.943
25029	33	0.139	-0.02	0.991
25027	43	0.158	0.046	1.031
23012	18	0.146	0.386	1.165
26008	49	0.103	0.193	1.2

St.No	RL	t2	t3	dij
25124	18	0.2	-0.069	0
26059	17	0.099	0.111	0.279
25025	31	0.167	0.083	0.891
6070	24	0.145	0.135	0.986
6011	48	0.113	0.09	1.025
36018	50	0.104	0.045	1.062
26018	49	0.114	0.126	1.109
18004	46	0.086	0.043	1.177
25014	54	0.128	0.103	1.252
6026	46	0.177	0.24	1.299
26021	30	0.14	-0.112	1.307
25023	52	0.163	0.142	1.439
25027	43	0.158	0.046	1.465

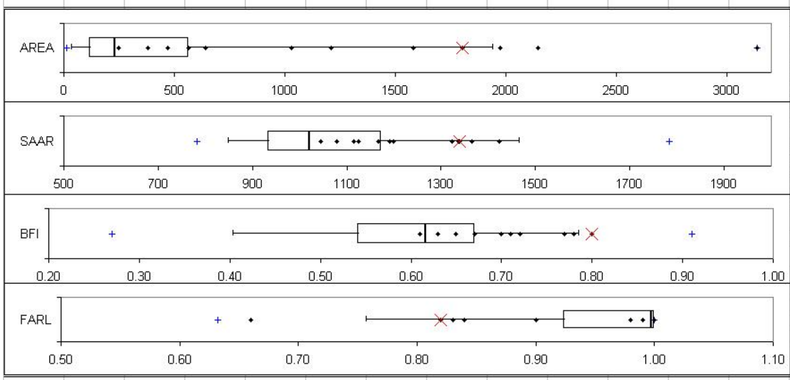
St.No	RL	t2	t3	dij
26059	17	0.099	0.111	0
25124	18	0.2	-0.069	0.279
6070	24	0.145	0.135	1.143
25025	31	0.167	0.083	1.16
6011	48	0.113	0.09	1.177
26021	30	0.14	-0.112	1.238
36018	50	0.104	0.045	1.277
26018	49	0.114	0.126	1.288
18004	46	0.086	0.043	1.341
25014	54	0.128	0.103	1.454
36019	47	0.104	-0.029	1.534
6026	46	0.177	0.24	1.546
18005	50	0.142	0.224	1.639

St.No	RL	t2	t3	dij
29004	32	0.083	0.144	0
34024	28	0.066	-0.051	0.355
26006	53	0.152	0.397	0.38
16005	30	0.098	0.204	0.505
25044	33	0.188	0.248	0.674
16003	51	0.099	0.207	0.739
25030	48	0.178	0.185	0.783
25029	33	0.139	-0.02	0.856
7033	25	0.129	0.162	0.865
26019	51	0.139	0.234	0.931
36015	33	0.159	0.325	0.997
19001	48	0.095	0.095	1.018
35001	29	0.119	-0.009	1.023
7006	19	0.117	-0.205	1.043

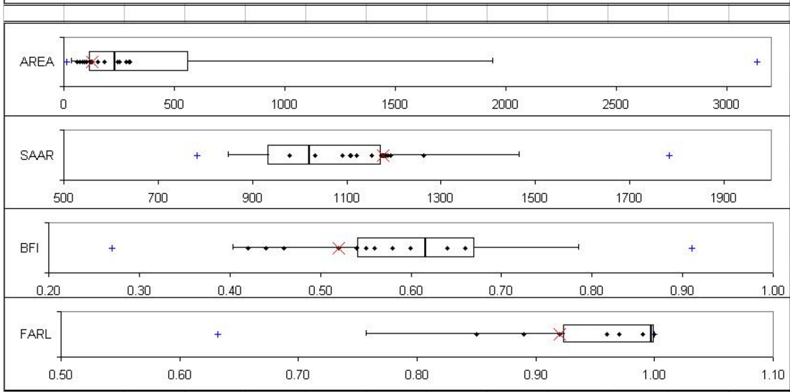
Figure 5.16: Box-plots and summary tables for evaluating the pooling group for the subject sites 25044,25124,26059 and 29004. -



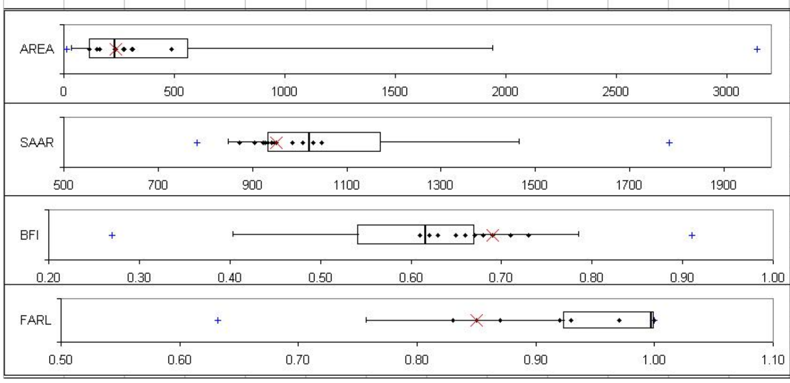
St.No	RL	t2	t3	dij
32012	24	0.071	0.001	0
34018	27	0.114	0.177	1.004
31002	26	0.129	0.308	1.16
35071	30	0.1	0.118	2.012
27002	51	0.122	0.181	2.128
35002	34	0.098	0.026	2.181
25044	33	0.188	0.248	2.38
16003	51	0.099	0.207	2.382
23012	18	0.146	0.386	2.39
19001	48	0.095	0.095	2.449
25030	48	0.178	0.185	2.455
34024	28	0.066	-0.051	2.468
35001	29	0.119	-0.009	2.486
19020	28	0.196	0.029	2.502
16005	30	0.098	0.204	2.564



St.No	RL	t2	t3	dij
34003	29	0.093	0.114	0
34001	36	0.104	0.08	0.155
30061	33	0.146	0.41	0.57
12001	50	0.194	0.176	1.028
27002	51	0.122	0.181	1.102
16011	52	0.172	0.089	1.3
18005	50	0.142	0.224	1.511
35071	30	0.1	0.118	1.583
35005	55	0.144	0.205	1.619
16009	52	0.099	-0.096	1.663
26007	53	0.107	0.164	1.744
30007	31	0.107	0.112	1.775



St.No	RL	t2	t3	dij
34024	28	0.066	-0.051	0
29004	32	0.083	0.144	0.355
26006	53	0.152	0.397	0.445
16005	30	0.098	0.204	0.486
25044	33	0.188	0.248	0.561
16003	51	0.099	0.207	0.577
25030	48	0.178	0.185	0.648
25029	33	0.139	-0.02	0.896
23012	18	0.146	0.386	0.9
35001	29	0.119	-0.009	0.941
19001	48	0.095	0.095	0.963
7033	25	0.129	0.162	1.077
36015	33	0.159	0.325	1.081
19020	28	0.196	0.029	1.171
26019	51	0.139	0.234	1.204



St.No	RL	t2	t3	dij
36018	50	0.104	0.045	0
18004	46	0.086	0.043	0.313
6026	46	0.177	0.24	0.418
6014	30	0.149	0.224	0.453
25014	54	0.128	0.103	0.46
6011	48	0.113	0.09	0.478
25025	31	0.167	0.083	0.484
25016	42	0.125	0.082	0.578
25023	52	0.163	0.142	0.655
6070	24	0.145	0.135	0.679
16002	51	0.156	0.172	0.719
6013	30	0.157	0.019	0.724

Figure 5.17: Box-plots and summary tables for evaluating the pooling group for the subject sites 32012,34003,34024 and 36018. -

5.6 Examination of homogeneity considering the GEV as the parent distribution

In the 'examination of homogeneity' procedure described in the last section, the CL about the t_2^r was constructed assuming EV1 as the population distribution. The examination is extended in this section using the GEV as the population distribution. Two GEV population shape parameters, $k = -0.05$ and $k = 0.03$, are selected in this context which correspond to L-skewness ≈ 0.21 and L-skewness ≈ 0.15 respectively. The results of the procedure for both cases are presented in Table 5.6 while a summary of the results for both cases is given in Table 5.5. The EV1 cases from Table 5.1 are also included for comparison purposes. This lists how many stations fall into the categories of one value outside the CL, 2 or 3 values outside the CL or more than 3 outside the CL. In all, 34 groups (40%) were in the latter category for the case where the negatively shaped GEV is employed while in the case of positively shaped GEV, the number of groups was 47 (55%).

From Table 5.5, it is seen on the increase in the shape parameter from $k = -.05$ to $+0.03$ leads to increase in the number of the events outside the CL.

Table 5.5: Summary of events outside the CL for the 85 pooling groups

events outside the CL (m)	GEV (k =-0.05)		EV1(k = 0)		GEV(k =+0.03)	
	No. of groups	% of groups	No. of groups	% of groups	No. of groups	% of groups
>=1	15	18	8	9	7	8
2<=3	36	42	33	39	31	36
>3	34	40	44	52	47	55

Table 5.6: Results of the examination procedure for each pooling group employing GEV as parent distribution.

			GEV(k=-0.05)			GEV(k=+0.03)		
St.No.	St. years	No.of Site	m	subject site ouside the C.I.		m	subject site ouside the C.I.	
				Y/N	$P - value$		Y/N	$P - value$

6011	523	13	3	N		3	N	
6013	529	14	1	N		3	N	
6014	531	12	2	N		3	N	
6026	523	12	2	Y	$6.6E-03$	3	Y	$2.5E-03$
6031	506	17	6	Y	$2.6E-05$	8	Y	$3.2E-06$
6070	514	12	2	N		3	N	
7006	523	13	1	N		2	N	
7009	513	11	3	Y	$3.8E-08$	4	Y	$3.4E-10$
7033	512	15	3	N		3	N	
8002	531	18	7	Y	$1.3E-02$	7	Y	$6.7E-03$
9001	515	15	2	Y	$3.7E-06$	2	Y	$3.7E-07$
10021	507	18	5	N		7	Y	$2.0E-02$
10022	507	18	5	Y	$4.3E-03$	7	Y	$1.8E-03$
12001	534	11	3	Y	$3.3E-06$	3	Y	$9.9E-08$
14005	500	12	2	N		3	N	
14006	511	11	1	N		2	N	
14007	534	16	1	N		2	N	
14009	506	16	4	N		4	N	
14011	538	15	2	N		3	N	
14018	513	11	3	N		4	N	
14019	546	12	3	N		4	N	
14029	514	11	3	Y	$2.5E-03$	3	Y	$5.9E-04$
15001	519	13	0	N		1	N	
15003	514	14	5	N		5	N	
16001	505	13	1	N		1	N	
16002	531	12	5	Y	$9.7E-03$	6	Y	$3.6E-03$
16003	511	13	4	N		4	N	
16004	543	13	1	N		3	N	
16005	523	14	5	N		7	Y	$2.5E-02$
16008	518	11	3	Y	$3.8E-04$	5	Y	$5.9E-05$
16009	500	10	3	N		3	N	
16011	500	10	3	Y	$2.8E-04$	3	Y	$3.2E-05$
18004	541	13	3	Y	$4.0E-03$	3	Y	$1.3E-03$
18005	525	13	1	N		1	N	
19001	503	13	3	Y	$2.2E-02$	5	Y	$1.1E-02$
19020	550	14	4	Y	$2.1E-05$	6	Y	$1.8E-06$
23001	501	14	3	Y	$8.6E-04$	5	Y	$2.0E-04$
23012	520	15	5	N		5	N	
24008	507	11	1	N		3	N	
24022	536	18	7	Y	$4.2E-04$	8	Y	$9.4E-05$
24082	546	12	1	N		2	N	
25006	513	11	3	N		4	N	
25014	533	12	2	N		3	N	

25016	549	13	2	N		2	N	
25023	512	13	1	N		1	N	
25025	506	13	2	N		4	N	
25027	542	12	3	Y	$2.5E-02$	5	Y	$1.3E-02$
25029	507	12	2	N		2	N	
25030	511	13	4	Y	$2.0E-03$	4	Y	$5.4E-04$
25034	529	17	5	N		6	N	
25040	537	17	6	N		7	N	
25044	523	14	5	Y	$7.9E-04$	7	Y	$1.9E-04$
25124	508	13	4	Y	$5.1E-04$	5	Y	$1.2E-04$
26002	530	12	2	N		4	N	
26005	503	11	4	N		4	N	
26006	507	13	4	N		4	N	
26007	500	10	3	N		3	N	
26008	507	12	1	N		1	N	
26018	519	12	4	N		5	N	
26019	527	13	0	N		1	N	
26021	542	13	6	N		7	N	
26022	536	17	4	N		6	N	
26059	510	13	3	Y	$3.6E-02$	5	Y	$2.2E-02$
27001	512	15	4	N		5	N	
27002	539	13	3	N		3	N	
29004	513	14	4	Y	$8.0E-03$	5	Y	$3.3E-03$
29011	512	12	1	N		1	N	
30007	522	12	3	N		3	N	
30061	522	12	3	N		3	N	
31002	513	15	6	N		6	N	
32012	505	15	5	Y	$3.2E-03$	5	Y	$1.1E-03$
34001	522	12	3	N		3	N	
34003	522	12	3	Y	$1.6E-02$	3	Y	$7.1E-03$
34009	537	15	5	N		5	N	
34018	503	14	5	N		5	N	
34024	540	15	6	Y	$2.4E-04$	8	Y	$4.4E-05$
35001	522	12	1	N		2	N	
35002	519	15	5	N		5	N	
35005	524	12	3	N		4	N	
35071	510	14	3	N		4	N	
36015	515	13	4	N		5	N	
36018	504	12	2	Y	$2.8E-02$	3	Y	$1.6E-02$
36019	514	11	6	N		6	N	
36021	521	16	6	N		6	N	
36031	508	16	5	Y	$9.3E-03$	6	Y	$4.1E-03$

5.7 Conclusion

The following conclusions are obtained from the above studies.

1. A visual approach for the identification of the homogeneity of ROI pooling groups has been presented. The results are compared with the heterogeneity measures H1 and H2, obtained for those groups. Overall the results show that even with a carefully considered selection procedure, it is not certain that perfectly homogeneous pooling groups are identified. As a compromise it is recommended that a group containing more than 2 values of L-CV outside the 95% confidence limits of that variable should not be considered homogeneous.

2. A thorough investigation on 27 heterogeneous pooling groups has been carried out. In many cases, special attributes of the subject site contributed to the degree of heterogeneity of the pooling groups. It is deemed necessary that the the subject site to be positioned near the centre of the group of gauging sites to which it is hydrologically similar; but in some cases the fulfillment of that condition does not guarantee that the pooling group is homogeneous.

6

Evaluation of the error associated with at-site and pooled estimates of flood magnitude

6.1 Introduction

The error associated with a flood frequency model is required in order to assess the predictive ability of the model. Three different indicators, namely, the bias, se and rmse of a flow quantile are generally used in such assessments. Bias and rmse may not be calculated in the case of assessments conducted on observed flood data because the true value of Q_T is unknown. They can only be calculated in the cases where assessments are conducted by simulation experiments using random samples drawn from a 'flood like' distribution such as EV1 and GEV. On the other hand, in certain cases, se can be calculated, where an se formula exists, using a scale parameter (σ or α) determined from observed data as well as by simulation methods.

This chapter is concerned with the following tasks:

1. The standard errors of Q_{med}, X_T and Q_T are evaluated in the context of Irish data

2. The performance of the pooled flood frequency method using simulation experiments is also evaluated in the Irish context

6.2 Standard errors of Q_{med}, X_T and Q_T

In flood frequency analysis, the standard error of a design flood estimate is commonly provided and plays a role as an indicator of reliability of an estimate, which can be taken into account in decision making.

The standard error of an estimate is normally based on the assumption that the data upon which the estimate is based are randomly drawn from a single population – an assumption that cannot be definitively proven. If an infinite number of similarly sized data sets were to be drawn from the same population and the value of Q_T obtained from each set by the same procedure, then the *se* is defined as

$se(Q_T) =$ Standard deviation of all the possible Q_T values.

This measure only represents the degree of scatter of the several estimates and does not refer to whether the mean of these is equal to the true value in the population. If this equality holds, then the procedure whereby Q_T is calculated is said to be unbiased – otherwise it is considered biased.

In flood frequency analysis, randomness of annual maximum floods and lack of trend with time is generally assumed. Likewise it is assumed that a single form of statistical distribution describes all the AM flood series in a region or country. Such assumptions cannot be fully proved. The presence of low or high outliers in some data sets also makes interpretation difficult. If the data are not truly from a unique homogeneous parent population then the *se* cannot strictly be defined. However the *se* concept is widely used in flood frequency analysis. Its value is determined on the assumption of a unique homogeneous parent population and the value of *se* so obtained may be considered as a lower bound on what the true value, if it could be found, actually is. However, it still provides a useful guide to the precision of the Q_T value obtained in any situation.

The aim in this section is to provide expressions or graphs which give an indication of the order of magnitude of $se(Q_T)$ in the at-site estimation at gauged sites and in pooling group based estimation at gauged sites and also at ungauged sites. All these analyses are restricted to the EV1 and GEV cases which are considered to be representative of what is appropriate in Irish conditions.

6.2.1 Standard Error of Qmed

The standard error of the median in a random sample is given by Kendall and Stuart (1977, I,p. 252) as

$$se\left(\hat{Q}_{med}\right) = \frac{1}{2\sqrt{n} \cdot f\left(Q_{med}\right)} \tag{6.1}$$

where n is the sample size and f is the pdf of the selected distribution. In the case of the normal distribution with standard deviation σ, the standard error of the median is given by Kendall and Stuart (1977, I, p. 252) as

$$se\left(\hat{Q}_{med}\right) = 1.2533\frac{\sigma}{\sqrt{n}} \tag{6.2}$$

Because Irish flood data are slightly more skewed than the zero skewness of the normal distribution and as there is evidence to suggest that Irish flood data are more likely to follow the EV1 distribution, the standard error of the median in the EV1 case is used for the Irish data. It can be expressed using eq (6.1), as (see section 6.2.3.1 for details):

$$se\left(\hat{Q}_{med}\right) = 1.442\frac{\alpha}{\sqrt{n}} \tag{6.3}$$

where α is the scale parameter of the EV1 distribution. The $se\left(\hat{Q}_{med}\right)$ can be expressed as a function of L-CV, i.e. t_2, as (see section 6.2.3.1 for details)

$$se\left(\hat{Q}_{med}\right) \Big/ Q_{med}[\%] = \frac{1.44 \times t_2}{\sqrt{n}\left(0.693 - 0.5772 \times t_2 + 0.367 \times t_2\right)} \times 100 \tag{6.4}$$

where $se\left(\hat{Q}_{med}\right) \Big/ Q_{med}[\%]$ is the percentage relative standard error.

The range of values of L-CV for most Irish flood data is between 0.1 and 0.2 with an average value of 0.15. It is observed, that in most of these cases, short record lengths are associated with high values of L-CV (see Figure 6.9). For different values of L-CV, the expressions for $se\left(\hat{Q}_{med}\right) \Big/ Q_{med}[\%]$ can be obtained as follows, e.g.

for $t_2 = 0.1$:

$$se\left(\hat{Q}_{med}\right)/Q_{med}[\%] \approx \frac{22}{\sqrt{n}}\% \tag{6.5}$$

for $t_2 = 0.15$:

$$se\left(\hat{Q}_{med}\right)/Q_{med}[\%] \approx \frac{33}{\sqrt{n}}\% \tag{6.6}$$

and for $t_2 = 0.2$:

$$se\left(\hat{Q}_{med}\right)/Q_{med}[\%] \approx \frac{44}{\sqrt{n}}\% \tag{6.7}$$

Therefore, for general purposes eq (6.6) can be adopted for Irish conditions. However, for short record lengths (< 20), eq (6.7) can be used which provides a more conservative estimate of the standard error of the median.

The se of Q_{med} using above expressions are compared with the values reported in the FEH (1999) in the following

If n = 5, $se\left(\hat{Q}_{med}\right)/Q_{med}[\%] \approx 20\%$ (but 22% in FEH, 3, Table 2.2)

If n = 10, $se\left(\hat{Q}_{med}\right)/Q_{med}[\%] \approx 14\%$ (and 14% in FEH, 3, Table 2.2)

If n = 20, $se\left(\hat{Q}_{med}\right)/Q_{med}[\%] \approx 7\%$ (but 8% in FEH, 3, Table 2.2)

If n = 50, $se\left(\hat{Q}_{med}\right)/Q_{med}[\%] \approx 5\%$ (not tabulated in FEH, 3, Table 2.2)

6.2.2 Standard error of at-site estimate of Q_T

The computation of the standard error of Q_T depends on both the form of the distribution and the method of parameter estimation. Moment based theoretical expressions of standard error of Q_T, for both EV1 and GEV cases, are reported in FSR (1975, I, p. 100:104) which also provides the development of these formulae and helpful discussions in this connection. Theoretical expressions of se for L-Moment based estimates of Q_T, in both EV1 and GEV cases, are given by Lu and Stedinger (1992) as follows.

$$EV1 \quad : se\left(\hat{Q}_T\right) = \frac{\alpha}{\sqrt{n}}\sqrt{[1.1128 + 0.4574y + 0.8046y^2]} \tag{6.8}$$

where $\hat{Q}_T$ is the estimate for the T-year flow event ; α is the EV1 scale parameter; $y = y_T = -\ln\left(-\ln\left(1 - 1/T\right)\right)$; and n is the number of observations in the sample.

$$GEV \quad : se\left(\hat{Q}_T\right) = \frac{\alpha}{\sqrt{n}}\left[\exp\left\{a_0\left(T\right) + a_1\left(T\right)\exp\left(-k\right) + a_2\left(T\right)k^2 + a_3\left(T\right)k^3\right\}\right]^{1/2} \tag{6.9}$$

where $\hat{Q}_T$ is the estimate for the T-year flow event; $a_0\left(T\right)$, $a_1\left(T\right)$, $a_2\left(T\right)$, and $a_3\left(T\right)$ are coefficients that depend on the return period, T ; α and k are scale and shape parameters respectively; and n is the number of observations in the sample. The values for the coefficients for different return periods are tabulated by Lu and Stedinger (1992).

While the expression for the EV1 is an analytical one derived using the Taylor expansion approach, the expression for the GEV is an empirical one where the coefficients for selected T values are obtained from Monte Carlo simulations. In this study, the focus is on the percentage standard error, $se(Q_T)/Q_T[\%]$ and, in the EV1 case this percentage can be expressed as a function of L-CV, as follows: (see section 6.2.3.1 for details).

$$EV1 \quad : se\left(\hat{Q}_T\right)/Q_T[\%] = \frac{t_2\sqrt{[1.1128 + 0.4574y + 0.8046y^2]}}{\sqrt{n}\,(0.693 - 0.5772 \times t_2 + t_2 \times y)} \times 100 \tag{6.10}$$

The relationship between L-CV and se is presented later in Figure 6.1(b).

In the GEV case with fixed k value, the percentage can also be expressed as a function of L-CV as follows.

$$GEV(k=-0.1) \quad : se\left(\hat{Q}_T\right)/Q_T[\%] = \frac{t_2\sqrt{[1.1376 + 0.7753y_k + 1.1816y_k^2]}}{\sqrt{n}\,(0.693 - 0.5772 \times t_2 + t_2 \times y_k)} \times 100 \tag{6.11}$$

where $y_k = y_T = \left(1 - (\ln(1 - 1/T))^k\right)/k$ and the coefficients 1.1376, 0.7753 and 1.1816 in the above equation are for fixed $k = -0.1$ (Lu and Stedinger, 1992, p. 251). The relationship between L-CV and se for fixed k (-0.1) is presented in Figure 6.2(b)

6.2.3 Standard error of the pooled estimate of X_T and of Q_T

Several researchers have presented theoretical expressions for estimating approximate standard errors of quantile estimates obtained through the index flood method. Rosbjerg and Madsen (1995) presented such expressions for the EV1 distribution with the model parameters estimated by the method of moments. Similarly, Stedinger and Lu (1995) presented expressions for the case of the GEV distribution with the parameters estimated using PWM, while the corresponding expressions presented by Kjeldsen and Jones (2006) are for the case of the GLO distribution with the parameters estimated using the method of L-moments.

Hosking and Wallis (1997) presented an L-moment based parametric simulation procedure and discussed their preference of the simulation procedure over analytical approaches such as those listed above. They pointed out that analytical approaches based on L-moments are too complicated and are based on a number of assumptions that might not necessarily be fulfilled to a satisfactory degree when considering observed

flood data. Hence, a simulation based approach is adopted in this study to estimate the se of the pooled estimate of Q_T although an analytical approach of deriving the se of Q_T is also presented for the case of the EV1 distribution, this being the preffered distribution for Irish data.

6.2.3.1 Se of the pooled estimate of X_T and of Q_T- theoretical

As noted above, Rosbjerg and Madsen (1995) presented the approximate theoretical expression for the EV1 distribution, with the parameters estimated by the method of moments, while these of Kjeldsen and Rosbjerg (2002) used the method of L-moments instead. In both of the cases, the annual mean flood is used as the index flood. In this section an approximate theoretical expression of se of Q_T for the case of the EV1 distribution is also obtained but with the median flood being used in this case instead of the mean flood as the index flood.

Using the index flood method, the at-site quantile estimator is obtained by multiplying the normalized quantile estimator, X_T, by the estimate of the site specific median flood, Q_{med}, i.e.

$$\hat{Q}_T = \hat{Q}_{med} \times \hat{X}_T \tag{6.12}$$

The variance of $\hat{Q}_T$ is approximately estimated using the first-order Taylor series expansion from the above equation as follows

$$Var(\hat{Q}_T) = Var(\hat{Q}_{med} \times \hat{X}_T) \tag{6.13}$$

$$\approx X_T^2 Var(\hat{Q}_{med}) + Q_{med}^2 Var(\hat{X}_T) + 2Q_{med} \cdot X_T \cdot cov(\hat{Q}_{med} \cdot \hat{X}_T) \tag{6.14}$$

Ignoring effects arising from covariance between the sample median and the estimated regional growth curve leads to the equation

$$Var(\hat{Q}_T) \approx X_T^2 Var(\hat{Q}_{med}) + Q_{med}^2 Var(\hat{X}_T) \tag{6.15}$$

As described in the following, $Var\left(\hat{Q}_{med}\right)$ and $Var(\hat{X}_T)$ are evaluated considering the EV1 as the population distribution.

. .

Derived expressions for $Var\left(Q_{med}\right)$ and $Var\left(X_T\right)$

The CDF and PDF of an EV1 distribution and the T-year event are given by Hosking and Wallis (1997) as follows

$$F\left(Q\right) = \exp\left[-\exp\left(-\frac{Q-\xi}{\alpha}\right)\right] \tag{6.16}$$

$$f(Q) = \frac{1}{\alpha}\exp\left[-\left(\frac{Q-\xi}{\alpha}\right) - \exp\left(-\left(\frac{Q-\xi}{\alpha}\right)\right)\right] \tag{6.17}$$

$$Q_T = \xi + \alpha y_T \tag{6.18}$$

where ξ is a location parameter, α is a scale parameter and $y_T = -\ln\left(-\ln\left(1-1/T\right)\right)$ is the EV1 reduced variate for a T-year return period.

The expressions for the median flood Q_{med} and the growth curve X_T respectively can be obtained as

$$Q_{med} = Q_2 = \xi + \alpha y_2 = \xi - \alpha\ln\left(\ln 2\right) \tag{6.19}$$

$$X_T = 1 + \beta\left(y_T - y_2\right) \tag{6.20}$$

where $\beta = \frac{\alpha}{\xi+\alpha y_2}$ and y_2 is the EV1 reduced variate for a 2-year return period. The parameters of an EV1 distribution are estimated by the first two L-moments, λ_1 and λ_2 (Hosking and Wallis, 1997) as

$$\lambda_1 = \xi + \alpha\gamma \tag{6.21}$$

$$\lambda_2 = \alpha\ln 2 \tag{6.22}$$

In terms of these L-moments, the β parameter takes the form

$$\beta = \frac{t_2}{\ln 2 - t_2\left[\gamma - y_2\right]} \tag{6.23}$$

where $t_2 = \frac{\lambda_2}{\lambda_1}$ is the L-CV and γ is Euler's constant $= 0.5772$.

Derivation of an expression for $Var\left(\hat{Q}_{med}\right)$:

The approximate variance of the sampling median is given by Kendall and Stuart (1977) as

$$Var\left(\hat{Q}_{med}\right) \approx \frac{1}{4n\left[f\left(Q_{med}\right)\right]^2} \tag{6.24}$$

The term $f\left(Q_{med}\right)$ can be obtained by substituting eq (6.19) into eq (6.17) as

$$f\left(Q_{med}\right) = \frac{0.3465}{\alpha} \tag{6.25}$$

Substituting eq (6.25) into eq (6.24) gives

$$Var\left(\hat{Q}_{med}\right) \approx \frac{2.08\alpha^2}{n} \tag{6.26}$$

Derivation of an expression for $se\left(\hat{Q}_{med}\right)/Q_{med}$ in terms of t_2:

Using $t_2 = \frac{\lambda_2}{\lambda_1}$, an expression of α can be obtained in terms of t_2 as,

$$\alpha = \frac{t_2 \cdot \lambda_1}{\ln 2} \tag{6.27}$$

Substituting the expressions of ξ from eq (6.21) and α from eq (6.27) into eq (6.18), gives

$$Q_T = \lambda_1\left(1 - \frac{t_2 \cdot \gamma}{\ln 2} + \frac{t_2 \cdot y_T}{\ln 2}\right) \tag{6.28}$$

Q_{med} can be obtained from above eq as

$$Q_{med} = Q_{T=2} = \lambda_1\left(1 - \frac{t_2 \cdot \gamma}{\ln 2} + \frac{t_2 \cdot y_2}{\ln 2}\right) \tag{6.29}$$

Taking the square root of $Var(Q_{med})$ in eq (6.26) and using eq (6.27) gives

$$se\left(\hat{Q}_{med}\right) = 1.44\frac{\alpha}{\sqrt{n}} = 1.44\frac{t_2 \cdot \lambda_1}{\sqrt{n} \cdot \ln 2} \tag{6.30}$$

Taking the ratio of eq (6.29) and eq (6.30) gives the expression for, $se\left(\hat{Q}_{med}\right)/Q_{med}$ as

$$se\left(\hat{Q}_{med}\right)\Big/Q_{med} = 1.44\frac{t_2 \cdot \lambda_1}{\sqrt{n} \cdot \ln 2}\Big/\lambda_1\left(1 - \frac{t_2 \cdot \gamma}{\ln 2} + \frac{t_2 \cdot y_2}{\ln 2}\right) = \frac{1.44 \times t_2}{\sqrt{n}\left(0.693 - 0.5772 \times t_2 + 0.367 \times t_2\right)} \tag{6.31}$$

Derivation of an expression for $Var(\hat{X}_T)$:

From eq (6.20) the variance of the growth curve can be written as:

$$Var(\hat{X}_T) = (y_T + y_2)^2 \, Var\left(\hat{\beta}\right) \tag{6.32}$$

Using eq (6.23), the variance of the $\hat{\beta}$ parameter can be obtained as

$$\hat{\beta} = \frac{\hat{t}_2^R}{\ln 2 - \hat{t}_2^R\,[\gamma - y_2]} = -\frac{1}{(\gamma - y_2)} + \frac{\ln 2}{(\gamma - y_2)\left(\ln 2 - \hat{t}_2^R\,[\gamma - y_2]\right)} \tag{6.33}$$

Hence,

$$Var\left(\hat{\beta}\right) = \frac{\ln 2}{\left(\ln 2 - [\gamma - y_2]\,\hat{t}_2^R\right)^2} Var\left(\hat{t}_2^R\right) \tag{6.34}$$

where $\hat{t}_2^R$ is the pooled L-CV estimator obtained as

$$\hat{t}_2^R = \sum_{i=1}^{M} w_i t_{2,i} \Big/ \sum_{i=1}^{M} w_i \tag{6.35}$$

The variance of $\hat{t}_2^R$ can be obtained as

$$Var(\hat{t}_2^R) = \frac{1}{\left(\sum_{i=1}^{M} w_i\right)^2} (\sum_{i=1}^{M} w_i^2 Var(\hat{t}_{2,i}) + 2\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} w_i w_j cov(\hat{t}_{2,i}, \hat{t}_{2,j}) \tag{6.36}$$

Madsen and Rosbjerg (1997) also gave the following estimate of $Var(\hat{t}_2^R)$, having considered inter-site correlation among sites in a pooling group

$$Var(\hat{t}_2^R) = \frac{1}{\left(\sum_{i=1}^{M} w_i\right)^2} \left(\sum_{i=1}^{M} w_i^2 Var(\hat{t}_{2,i}) + 2\sum_{i=1}^{M-1}\sum_{j=i+1}^{M} w_i w_j Var(\hat{t}_{2,i})^{1/2} Var(\hat{t}_{2,j})^{1/2} \rho_{\hat{t},ij}\right) \tag{6.37}$$

where $\rho_{\hat{t},ij}$ is the correlation coefficient between $\hat{t}_{2,i}$ and $\hat{t}_{2,j}$; which they assumed to be equal to ρ_{ij}^2.

If the inter-site correlation is ignored, then the above equation would reduce to

$$Var(\hat{t}_2^R) = \frac{1}{\left(\sum_{i=1}^{M} w_i\right)^2} \left(\sum_{i=1}^{M} w_i^2 Var(\hat{t}_{2,i})\right) \tag{6.38}$$

The variance of the at-site estimator of $\hat{t}_2$ is obtained from a Taylor series expansion of $\hat{t}_2 = \frac{\hat{\alpha}\ln 2}{\hat{\xi}+\gamma\hat{\alpha}}$ as follows:

$$Var\left(\hat{t}_2\right) = \left(\frac{\partial \hat{t}_2}{\partial \alpha}\right)^2 Var(\hat{\alpha}) + \left(\frac{\partial \hat{t}_2}{\partial \xi}\right)^2 Var(\hat{\xi}) + 2\left(\frac{\partial \hat{t}_2}{\partial \alpha}\right)\left(\frac{\partial \hat{t}_2}{\partial \xi}\right) cov(\hat{\alpha},\hat{\xi}) \tag{6.39}$$

where

$$\frac{\partial \hat{t}_2}{\partial \alpha} = \frac{\ln 2(\xi+\gamma\alpha) - \alpha\cdot\gamma\cdot\ln 2}{(\xi+\gamma\alpha)^2} \tag{6.40}$$

$$\frac{\partial \hat{t}_2}{\partial \xi} = -\frac{\alpha \ln 2}{(\xi+\gamma\alpha)^2} \tag{6.41}$$

The variance and covariance of the parameters as ξ and α estimated using L-Moments are given by Lu and Stedinger (1992) as

$$Var(\hat{\alpha}) = 0.8046 \times \frac{\alpha^2}{n} \tag{6.42}$$

$$Var(\hat{\xi}) = 1.1128 \times \frac{\alpha^2}{n} \tag{6.43}$$

$$Cov(\hat{\alpha},\hat{\xi}) = 0.4574 \times \frac{\alpha^2}{n} \tag{6.44}$$

The term $Var\left(\hat{t}_2\right)$ is obtained by substituting equations (6.40 to 6.44) into eq (6.39). From the result of $Var\left(\hat{t}_2\right)$ for each site in a pooling group, the term $Var(\hat{t}_2^R)$ is obtained using eq (6.38). Substituting the results of $Var(\hat{t}_2^R)$ into eq (6.34) yields an equation for $Var\left(\hat{\beta}\right)$. Finally, substituting the expression for $Var\left(\hat{\beta}\right)$ into eq (6.32) yields the variance of the growth curve, $Var(\hat{X}_T)$.

6.2.3.2 Se of the pooled estimate of X_T and of Q_T- simulation

The simulation procedure is carried out for both the EV1 and GEV distributions with parameters estimated by the method of L-moments. The median annual flood is used as the index flood and the pooling group is formed using the region of influence approach. The simulation procedure considers the implication of heterogeneity but does not consider the implication of inter site correlation among sites in a pooling group.

The steps of the procedure which is applied to obtain an estimate of the order of magnitude of the pooled estimate of X_T, $se(X_T)$, and Q_T, $se(Q_T)$ for selected values of T are described below.

1. Identify the gauging stations in the subject site's pooling group using d_{ij} values of eq (4.23) and with a minimum of 500 station years of data in the pooling group, which satisfies the 5T rule for the 100 year quantile.

2. Random samples are drawn from EV1 populations for the subject site and for each site in the pooling group. For each site the sample size is taken as equal to the length of the observed historical record at the site and the parameter values μ and α are those estimated from the observed record by L-Moments.

3. The sample Q_{med} is obtained for the subject site.

4. The L-CV value is obtained for each sample in the pooling group and the average of these is calculated, without weights, to represent the pooled average L-CV.

5. The pooled average L-CV is used to determine the pooling group's EV1 growth curve parameter β

6. The subject site's $X_T = 1 + \beta\,(\ln(\ln 2) - \ln(-\ln(1 - 1/T)))$ is calculated for T = 5, 10, 25,.........500 years

7. The subject site's $Q_T = Q_{med} \cdot X_T$ is calculated for T = 5, 10, 25,.........500 years

8. Steps 2 to 7 are repeated 10,000 times to provide 10,000 values of X_T and Q_T at the subject site and the $se(X_T)$ and $se(Q_T)$ are calculated for the subject site by the following equations:

$$se(X_T)\,[\%] = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{X}_{i,r}^{T} - \bar{\hat{X}}_{i}^{T}}{X_{i}^{T}}\right)^{2}} \times 100 \tag{6.45}$$

$$se(Q_T)\,[\%] = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{Q}_{i,r}^{T} - \bar{\hat{Q}}_{i}^{T}}{Q_{i}^{T}}\right)^{2}} \times 100 \tag{6.46}$$

where $\hat{X}_{i,r}^{T}$ and $\hat{Q}_{i,r}^{T}$ are the estimated T-year growth factors and T-year flood quantiles respectively at a site i at the r^{th} repetition; $\bar{\hat{X}}_{i}^{T}$ and $\bar{\hat{Q}}_{i}^{T}$ are the means of those estimated measures; X_{i}^{T} and Q_{i}^{T} are the true T-year growth factors and the T-year quantile at site i in the assumed parent population; M is the number of sites in the pooling group and R is the number of repetitions.

These expressions average the standard error over all the sites in the pooling group to take heterogeneity in the pooling group into account. It should be noted that the actual standard error for any individual site could be either smaller or larger than the calculated value depending on the at site value of the L-CV and L-skewness.

The simulation procedure described above was also applied with the GEV distribution used instead of EV1 in steps 2 to 6. In step 4, the L-skewness and the average L-skewness is calculated as well as the L-CV. In step 5, the GEV growth curve parameters k and β are calculated by equations (4.10) and (4.12) respectively; and in step 6 the expression for X_T given in equation (4.9) is used.

The results of this simulation method are provided later, in section 6.2.4.2.

6.2.4 Comparison of se of Q_T based on the at-site estimate and the pooled estimate

6.2.4.1 Se of Q_T based on the at-site estimate

This study is based on the AM series obtained from 85 A1 and A2 Irish gauging stations– the same data sets used in Chapter 5. Equations (6.8) and (6.18) have been applied to the data of 85 stations in which the values of ξ, α and Q_T were estimated from each site's data separately and the ratio $se(Q_T)/Q_T[\%]$ is formed in the EV1 case. While the value of $se(Q_T)$ obtained using estimated values of parameters in the above equations is not the true value , it is still considered useful for examining the range of values so obtained because part of their scatter is caused by inter-site heterogeneity, assuming that such exists, as well as by random sampling effects and unequal record lengths varying from 18 to 55 with an average of 37 years. The box plot of these 85 values for 6 different return periods is shown in Figure 6.1(a) in the case of the EV1.

It is suggested that the upper and lower values indicated by that box plot i.e. containing the middle 50% of values, ought to give a good indication of the range of values within which the true se values fall. The extreme high and low values can be considered to be unrepresentative of the whole and hence the suggestion of focusing on the boxes. What these show, under the EV1 assumption, is that the se can be assumed to be between 5% and 7% for the 10 year quantile and between 7% and 10% for the 100 year quantile.

In Figure 6.1(b), the same Box-plots are re-presented together with the theoretical EV1 percentage standard errors given by eq (6.10) for the average sample size of 37 years and for the range of L-CV values experienced among Irish AM flood data.

Corresponding results for the GEV case (using eqs 6.9 and D.5) are presented in Figure 6.2. The theoretical figures in Figure 6.2(b) are now for a single value of shape parameter $k = -0.1$ which it is felt ought to cover the most extreme underlying population case that would arise in Irish conditions.

From the box plots, using the same arguments as above, we can conclude that under the GEV assumption the se can be assumed to be between 4% and 7% for the 10 year quantile and between 8% and 16% for the 100 year quantile. However it can be seen that the most extreme values are extremely large but these are not representative of

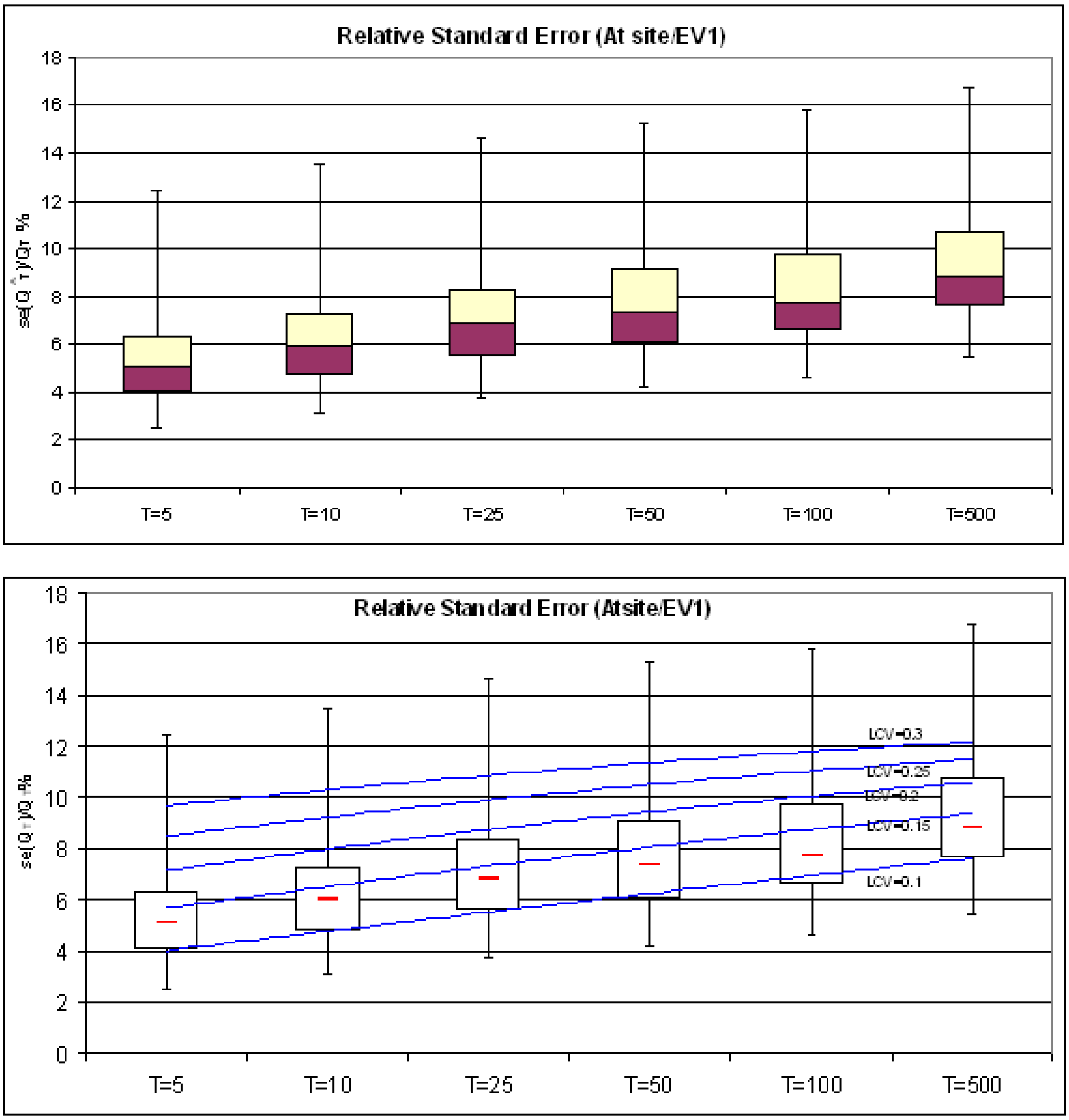
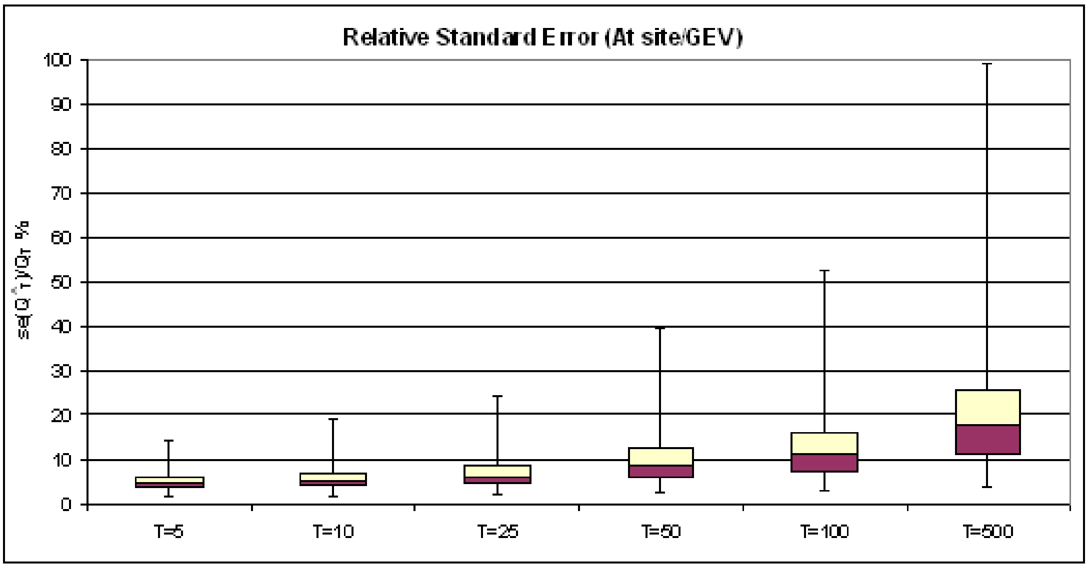


Figure 6.1: Box-plot of percentage standard error of flood quantile estimates for the EV1 distribution - (a) for 85 stations for the 5, 10, 25, 50, 100, 500-year return period(b)theoretical based:for the average sample size of 37 years and for a range of L-CV values



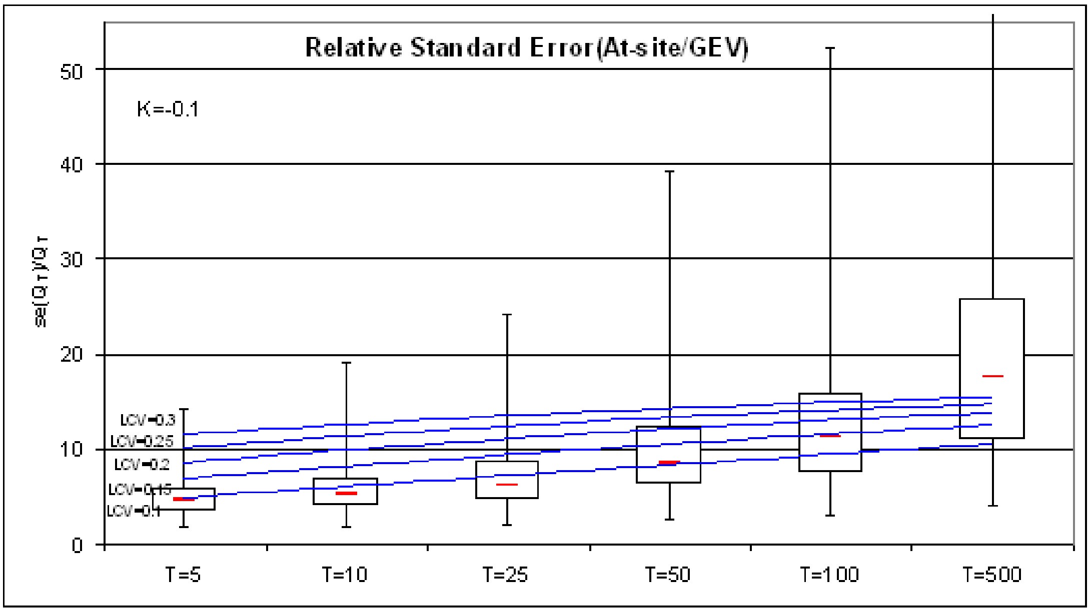


Figure 6.2: Box-plot of percentage standard error of flood quantile estimates for the GEV distribution - (a) for 85 stations for the 5, 10, 25, 50, 100, 500-year return period(b)theoretical based GEV($k = -0.1$):for the average sample size of 37 years and for a range of L-CV values

the true *se* value. The range of values of percentage *se* for all return periods and both distributions, EV1 and GEV, are summarized in tabular form in Table 6.1.

Table 6.1: Ranges of percentage *se* for quantile estimates, Q_T

T	Se(QT)%,EV1	Se(QT)%,GEV
5	4.0 to 6.3	3.6 to 5.8
10	4.8 to 7.3	4.1 to 7.0
25	5.6 to 8.3	4.8 to 8.7
50	6.1 to 9.1	6.3 to 12.4
100	6.7 to 9.7	7.5 to 15.8
500	7.6 to 10.8	11.0 to 26.0

6.2.4.2 *Se* of Q_T based on the pooled estimate

The simulation procedure described in section 6.2.3.2 is applied for each of the 85 groups, one for each station, and the percentage standard errors for X_T and Q_T are presented for the EV1 case in box plot form in Figures 6.3 and 6.4. The same remarks apply to the interpretation of these plots as for those of the at-site standard errors in the 6.2.4.1. The values of percentage *se* for X_T vary from 1.1 to 1.2 for T = 10 and to 1.8 to 2.0 for T = 100 while the corresponding percentage se values for Q_T remain relatively constant over the whole range of T value, varying between from 3.7 to 8.5 for T = 10 and from 4.0 to 9.0 for T = 100.

Corresponding results for the GEV case are presented, in box plot form, in Figures 6.5 and 6.6. The values of percentage *se* for X_T vary from 1.4 to 1.8 for T = 10 and from 3.7 to 5.0 for T = 100 while the corresponding percentage se values for Q_T show a increasing trend, varying between 3.8 to 9.0 for T = 10 and from 4.6 to 10.6 for T = 100.

These ranges for a range of return periods and both distributions, EV1 and GEV, are summarized in tabular form in Table 6.2.

Table 6.2: Ranges of percentage *se* for growth factors, X_T, and quantile estimates, Q_T

T	$se(X_T)$%,EV1	$se(Q_T)$%,EV1	$se(X_T)$%,GEV	$se(Q_T)$%,GEV
5	0.7 to 0.8	3.7 to 8.4	0.8 to 1.0	3.7 to 9.0
10	1.1 to 1.2	3.7 to 8.5	1.4 to 1.8	3.8 to 9.0
25	1.4 to 1.6	3.8 to 8.7	2.3 to 3.0	4.1 to 8.9
50	1.6 to 1.8	3.9 to 8.9	3.0 to 3.9	4.3 to 9.6
100	1.8 to 2.0	4.0 to 9.0	3.7 to 5.0	4.6 to 10.6
500	2.1 to 2.3	4.1 to 9.3	5.4 to 7.3	5.7 to 12.2

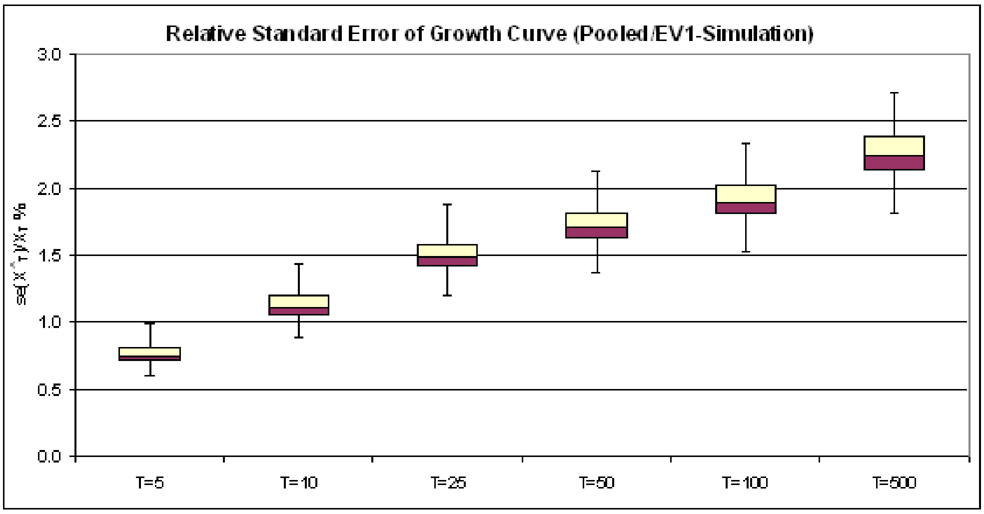


Figure 6.3: Box-plot of percentage standard error of growth factors estimates for the EV1 distribution - using simulation method for 85 stations for the 5, 10, 25, 50, 100,500-year return period

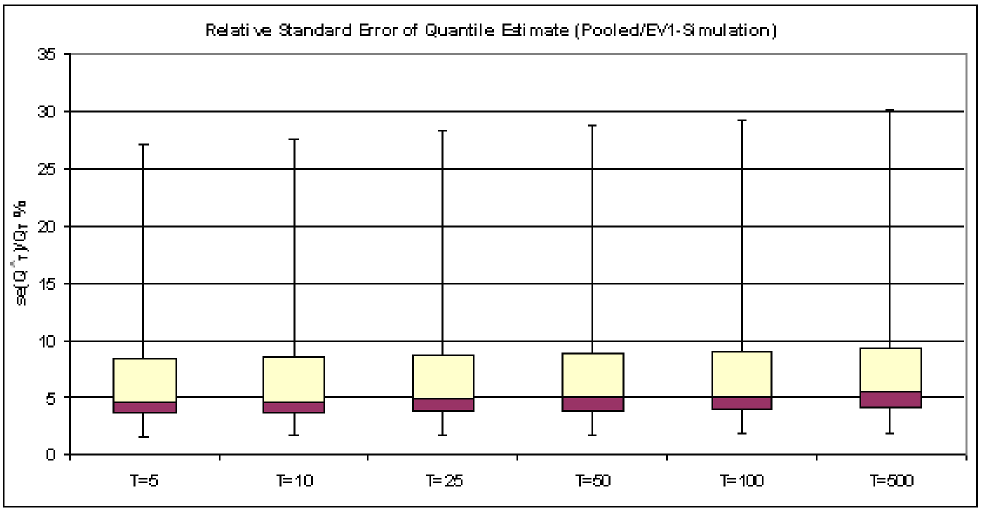


Figure 6.4: Box-plot of percentage standard error of quantile estimates estimates for the EV1 distribution - using simulation method for 85 stations for the 5, 10, 25, 50, 100,500-year return period

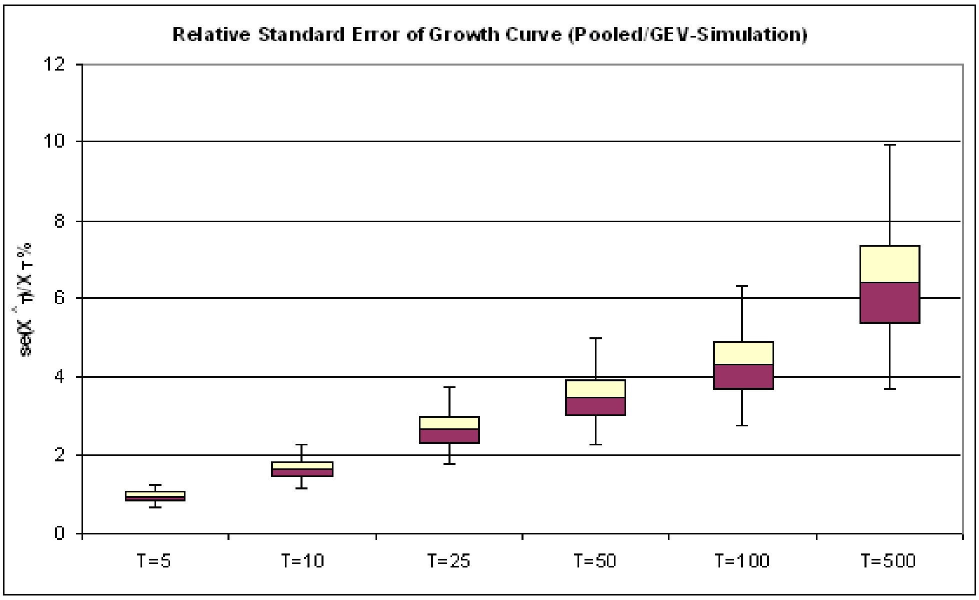


Figure 6.5: Box-plot of percentage standard error of growth factors estimates for the GEV distribution - using simulation method for 85 stations for the 5, 10, 25, 50, 100,500-year return period

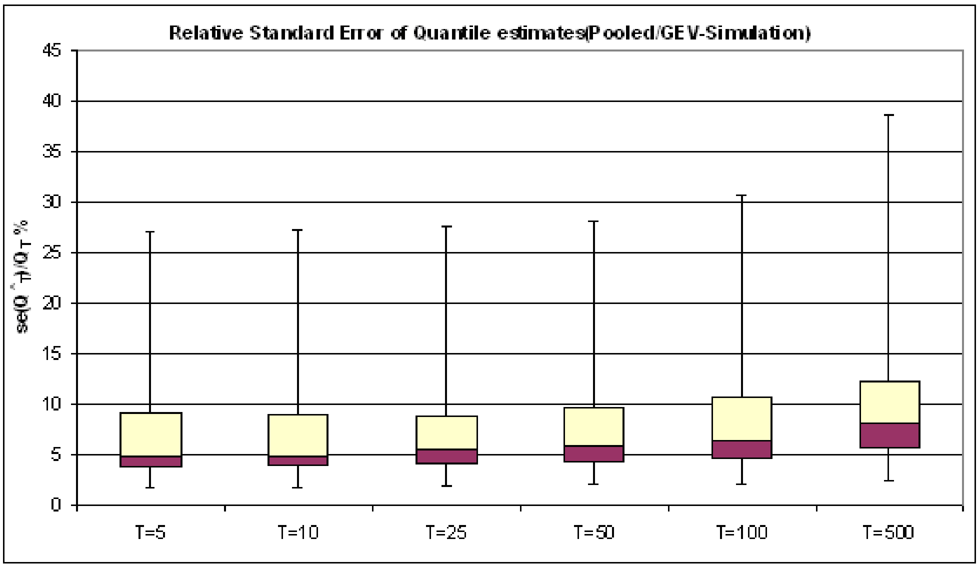


Figure 6.6: Box-plot of percentage standard error of quantile estimates estimates for the GEV distribution - using simulation method for 85 stations for the 5, 10, 25, 50, 100,500-year return period

6.2.5 Estimate of se of Q_T at ungauged site

Pooled analysis using the index-flood method allows estimates of the design flood to be obtained at ungauged sites (e.g. FEH, 1999; Hosking and Wallis, 1997). The design flood, Q_T, at an ungauged site is estimated in two steps. In the first step, the growth curve X_T is estimated from a pooling group assigned for the ungauged site. In the second step, an estimate of the index flood, Q_{med}, is obtained.

At the ungauged site, the value of Q_{med} can be obtained from the catchment descriptor data through the application of a regression model. As part of the Irish FSU, a multivariate regression equation was developed on the basis of data from 199 gauged catchments, linking the Q_{med} to a set of catchment descriptors as FSUWP2.3 (2009)

$$Qmed_{rural} = 1.083 \times 10^{-5} AREA^{0.938} BFIsoils^{-0.88} SAAR^{1.326} FARL^{2.233} DRAIND^{0.334} \times S1085^{0.188} (1 + ARTDRAIN2)^{0.051} \tag{6.47}$$

The Factorial Standard Error (FSE) of $Qmed_{rural}$ in the above equation is 1.36.

As shown in section 6.2.3.1, the variance of Q_T can be estimated approximately using the first order Taylor series expansion as:

$$Var(\hat{Q}_T) \approx x_T^2 Var(\hat{Q}_{med}) + Q_{med}^2 Var(\hat{X}_T)$$

This expression for $Var(\hat{Q}_T)$ is dominated by the $Var(\hat{Q}_{med})$ term as $Var(\hat{X}_T)$ affects only the 3rd or 4th decimal point in the value of $Var(\hat{Q}_T$ (see Table 6.2). Hence the expression effectively reduces to

$$se(Q_T) \approx X_T \cdot se(Q_{med})$$

giving

$$se(Q_T)/Q_T \approx X_T \cdot se(Q_{med})/(X_T \cdot Q_{med})$$

$$se(Q_T)/Q_T \approx se(Q_{med})/Q_{med}$$

$se(Q_T)/Q_T \approx \frac{0.44}{\sqrt{N}} \cdot \frac{Q_{med}}{Q_{med}}$(From eq 6.7)

$$se(Q_T)/Q_T \approx \frac{0.44}{\sqrt{N}} \tag{6.48}$$

Therefore, $se(Q_T)/Q_T$ is independent of T but dependent on value of N associated with the catchment descriptor based estimate of Qmed which is considered to be equivalent to either 1 year of AM data or perhaps a little more (Nash and Shaw, 1965; FSR, 1975; Hebson and Cunnane, 1987).

Hence for ungauged catchments,

with N=1, $se(Q_T)/Q_T = \frac{0.44}{\sqrt{1}}$= 0.44= 44%.

with N=1.5, $se(Q_T)/Q_T = \frac{0.44}{\sqrt{1.5}}$= 0.36= 36%.

These results apply regardless of whether the growth curve is obtained using EV1 or GEV. This may seem counter intuitive but it is due to the fact that the uncertainty in Q_T is dominated by the uncertainty in Qmed.

With N= 1.5, the 36% value of percentage standard error in this derivation is consistent with factorial standard error (FSE) =1.36 in eq (6.47) which represents corresponding to the catchment descriptor based estimate of Qmed derived in the FSU being worth 1.5 years of AM data.

6.3 Performance of a pooled flood frequency analysis in the Irish context

An implicit assumption with the index-flood procedure of pooled analysis is that the X_T-T relation is the same at all sites in a homogeneous pooling group although this assumption would generally violated to some extent in practical cases, i.e. some degree of heterogeneity exists. In fact, in only a few cases is the homogeneity criterion effectively satisfied for Irish conditions (see chapter 5). Therefore it is necessary to assess the performance of the index-flood pooled analysis procedure considering that heterogeneity may have to be taken into account.

Several studies e.g. (Lettenmaier et al., 1987; Stedinger and Lu, 1995; Hosking and Wallis, 1997) have examined the performance of index flood quantile estimators considering regions with various degrees of heterogeneity. Based on Monte Carlo simulation experiments those studies show that the performance of regional index-flood estimators degrades with the degree of departures from the homogeneity assumption but should still be useful if the amount of heterogeneity for sites in a region is moderate. In those studies the performance of index flood estimators has been assessed in most cases by hypothesizing pooling regions that resemble moderate to high skewness regimes. In this part of the present study

- the sensitivity of the index flood estimator to station years/ number of sites
- the sensitivity of the index flood estimator to record length
- the improvement in the performance of the index flood quantile estimator over the at-site quantile estimator

are explored considering pooling regions hypothesized according to Irish conditions and the results so obtained are compared with the results of the above mentioned previous studies where relevant.

6.3.1 Experimental Design

Definition of a pooling group is the first step in the experimental design procedure. The number of sites in the pooling group is then specified along with the statistical

properties, e.g. the frequency distribution and record lengths at each site. These properties are chosen to accord with Irish conditions.

6.3.1.1 Selection of pooling group size

The FEH (1999) used an error measure, the pooled uncertainty measure (PUM) as defined in eq (6.49) in its search to find a suitable pooling group size where pooling groups are formed using the ROI technique. Using UK data, a range of pooling group sizes was investigated and it was decided to adopt the 5T rule, namely, that the total number of station years of data to be included when estimating the T year flood should be at least 5T. Later, Kjeldsen et al. (2008) found that a fixed pooling-group size consisting of 500 station years performed well for a range of return periods. Hosking and Wallis (1997) used 15 sites with each having 30 years of records when performing simulation experiments .i.e. a total of 450 station years.

In this study, PUM and the heterogeneity measures, H1 and H2, are employed to select the pooling group size for Irish conditions. The magnitudes of PUM, and H1 and H2, have been computed for groups selected by the distance measure d_{ij} of eq (4.22). Figure 6.7 shows the variation in the mean PUM values with the size of the pooling group obtained using 85 ROI pooling groups . It is seen from Figure 6.7 that the PUM value does not decrease substantially for a pooling group of sizes greater than 450 station years or more which is equivalent to 12 or 13 sites in a pooling group. Figure 6.8 shows the variation in the mean H1 and H2 values with the size of the pooling group, based on the same format as used in calculating the PUM values. It is seen from Figure 6.8 that the heterogeneity, on average, increases with group size, an occurrence that is intuitively to be expected. A summary of the mean PUM values and the mean H1 and H2 values arranged according to the size of the pooling group is further tabulated in Table 6.3. Based on these results, it is decided to form a pooling group of 13 sites with record lengths on average 35 per site, i.e. a pooling-group of 455 station years. It should be noted however that the variation of PUM with pooling group size is relatively small.

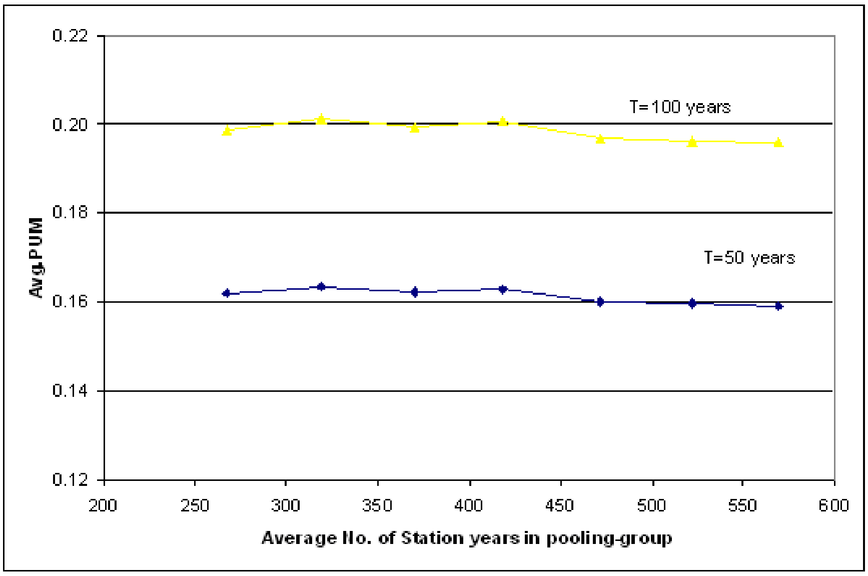


Figure 6.7: Variations in the mean PUM values - for different numbers of station years in the pooling group.

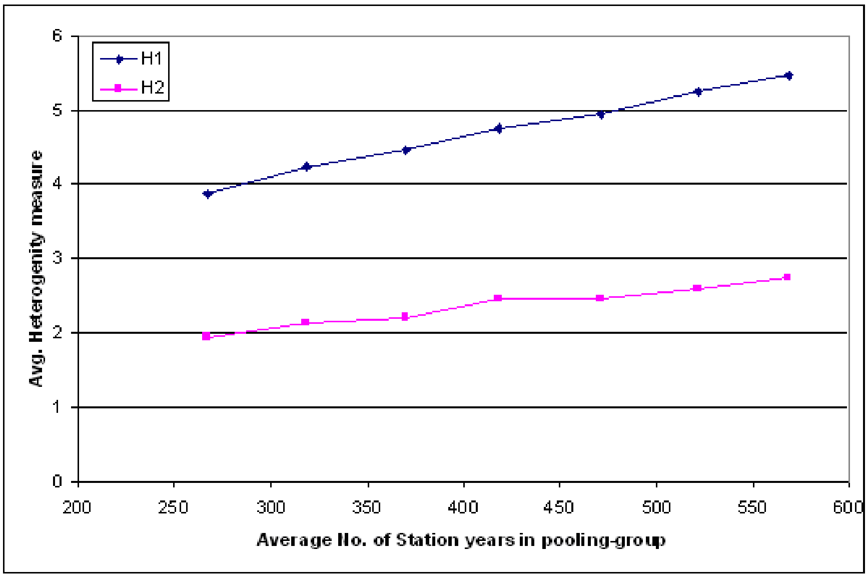


Figure 6.8: Variations in the mean H1 and H2 values - for different numbers of station years in the pooling group.

Table 6.3: Variation in the mean PUM, H1 and H2 values for different number of station years in pooling group

Avg. No. of Sites	Avg. No. of st. years	Avg.H1	Avg.H2	Avg.PUM50	Avg.PUM100
7	267	3.87	1.94	0.1266	0.1618
8	319	4.23	2.14	0.1278	0.1636
10	370	4.47	2.19	0.1267	0.1621
11	418	4.74	2.45	0.1269	0.1628
12	471	4.94	2.45	0.1249	0.1598
13	522	5.23	2.59	0.1247	0.1596
15	569	5.47	2.74	0.1244	0.1592

Pooled uncertainty measure, PUM

The PUM is a form of weighted average of the differences between each site growth factor and the pooled growth factor measured on a logarithmic scale. The pooled uncertainty measure for return period T, PUM_T is defined by FEH (1999) as

$$PUM_T = \sqrt{\frac{\sum_{i=1}^{M_{long}} n_i \left(\ln x_{T_i} - \ln x_{T_i}^{p}\right)^2}{\sum_{i=1}^{M_{long}} n_i}} \tag{6.49}$$

where M_{long} is the number of long-record sites, n_i is the record length of the i^{th} site, x_{T_i} is the T-year site growth factor for the i^{th} site, and $\ln x_{T_i}^{p}$ is the T-year pooled growth factor for the i^{th} site. In this study, $M_{long} = 13$ is used in all cases.

6.3.1.2 Choice of statistical properties

For simulating heterogeneous regions, a model is needed to describe the variation of flood distribution parameters from site to site. This study follows the type of procedure that is described in Hosking and Wallis (1997, p. 131).

This study selects the GEV distribution to describe the distribution of flood data. GEV has been widely used in the index flood procedure (FSR, 1975; Hosking and Wallis, 1997; Lettenmaier et al., 1987; Stedinger and Lu, 1995). It was also suggested by FSR (1975) as the best suited distribution for Irish flood data and the results from the goodness of fit test in chapter 4 lend further support this choice . The GEV distribution, with different combinations of t_2 and t_3, is employed to generate the simulated data series using Monte Carlo simulation.

Figure 6.9 shows the variation of t_2 with respect to record lengths in Box-plot form, based on 110 A1 and A2 category stations while Figure 6.10 shows the variation for t_3. It can be noted here that the sample estimates of t_3 have relatively larger bias than t_2 (Hosking and Wallis, 1997, p. 28). It is seen that the shortest record lengths are associated with the largest t_2 and t_3 values. Table 6.4 summarizes the average values of t_2 and t_3, for each category of range of record lengths, including the number of stations which are associated with each range. The average values of t_2 and t_3 depict an approximately linear relationship between these two statistics. Hosking and Wallis (1997) performed their experiment considering t_2 and t_3 to vary linearly for the sites in a group. While they took the average value of t_2 and t_3 for a pooling group as 0.25 in evaluating GEV/L-mom estimator, it is taken to be 0.2 in this part of the study.

Table 6.4: Average values of L-CV and L-skewness for different size of record lengths

Size of Record Length	No.of Stations	Avg L-CV	Avg L-skewness
11-20	13	0.24	0.23
21-30	34	0.14	0.12
31-40	18	0.13	0.13
41-50	25	0.13	0.1
51-60	20	0.13	0.13

6.3.1.3 Hypothesized pooling groups

Six pooling groups are hypothesized as representative examples for illustrating the performance of pooled analysis. Table 6.5 summarises some relevant properties of those groups including their heterogeneity statistic H1. The groups PG0 (corresponding to a perfectly homogeneous group) and PG1 (corresponding to a nearly a homogeneous group) represent practically homogeneous conditions while PG2 and PG3 represent moderately heterogeneous conditions and PG4 and PG5 represent highly heterogeneous conditions. This range of heterogeneity arose from the ROI approach of forming groups using Irish data as described in chapter 5.

As an example, Table 6.6 describes PG2 in detail. It serves as a prototype structure for all the six pooling groups. The quantities t_2 and t_3 both vary linearly from 0.25 at site 1 to 0.15 at site 13 while record lengths have the tendency to vary inversely proportional to the t_2 and the t_3 values from 20 at site 1 to 50 at site 13. Lettenmaier

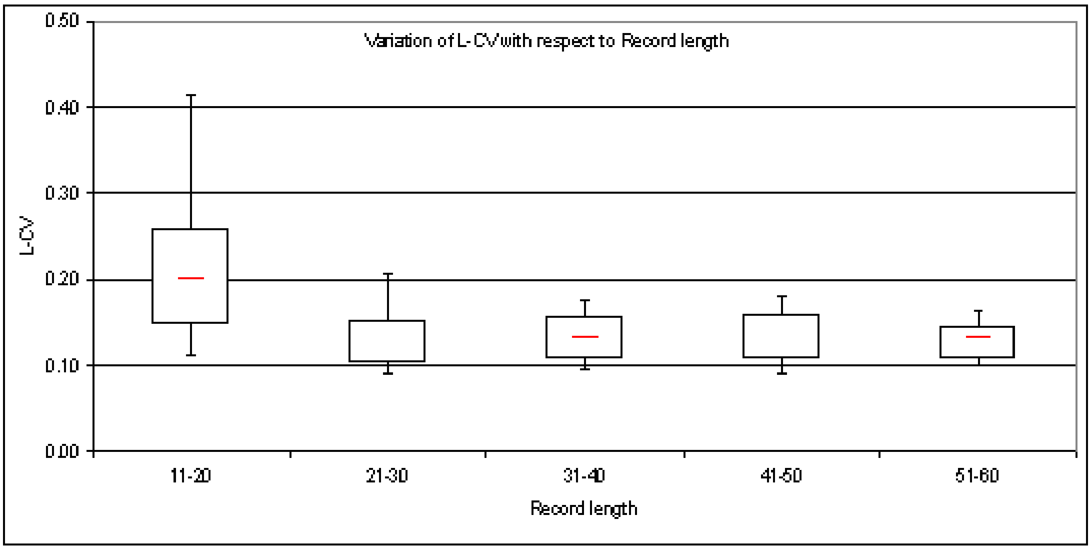


Figure 6.9: Box plots of L-CV values showing variation with respect to record lengths. - Each box plot gives percentiles for the frequencies of 0.10, 0.25, 0.50, 0.75, and 0.95.

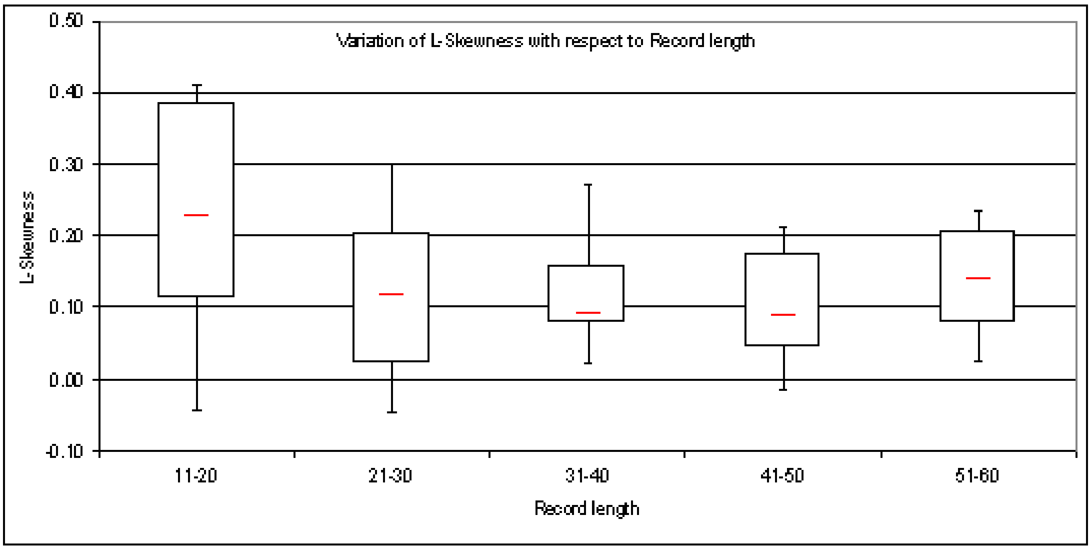


Figure 6.10: Box plots of L-skewness values showing variation with respect to record lengths. - Each box plot gives percentiles for the frequencies of 0.10, 0.25, 0.50, 0.75, and 0.95.

et al. (1987) and Stedinger and Lu (1995) included this structure of record lengths in their experiments.

Table 6.5: Characteristics of six pooling groups

Pooling Group	Avg. L-Cv	Avg.L-skewness	Range	H1
PG0	0.2	0.2	0	0
PG1	0.2	0.2	0.05	0.56
PG2	0.2	0.2	0.1	1.83
PG3	0.2	0.2	0.15	3.62
PG4	0.2	0.2	0.2	5.21
PG5	0.2	0.2	0.25	6.91

Table 6.6: Variation of the L-CV-L-skewness pairs and the record lengths for each individual site in the pooling group PG2

Site Number	L-CV	L-skewness	Record Length
1	0.25	0.25	20
2	0.242	0.242	22
3	0.234	0.234	25
4	0.226	0.226	27
5	0.219	0.219	30
6	0.211	0.211	32
7	0.2	0.2	35
8	0.192	0.192	37
9	0.184	0.184	40
10	0.176	0.176	42
11	0.169	0.169	45
12	0.161	0.161	47
13	0.15	0.15	50

6.3.2 Quantile estimators

The quantile estimation procedures used in this study are all based on the method of L-moments. Two cases are considered for estimating quantiles: the at-site and the pooled estimator. The at-site estimator is applied independently to each site. The pooled estimator, on the other hand, based on an index flood approach, uses the information from all the sites in a pooling group.

The estimators considered are as follows

1. GEV/Site – an at-site GEV/L-mom quantile estimator with all three parameters estimated using only at-site data.

2. GEV/Pool – a GEV/L-mom index-flood quantile estimator which uses the at-site sample median to scale a pooling estimator of the dimensionless quantile. It uses pooled average estimates of t_2 and t_3.

The GEV distribution is defined by Hosking and Wallis (1997) as

$$F\left(Q\right) = \exp\left\{-\left(1 - k\left(Q - \xi\right)/\alpha\right)^{1/k}\right\} \quad for \quad k \neq 0 \tag{6.50}$$

and the T-year event, which is the at-site estimator for the GEV distribution, is given as

$$Q_T = \xi + \left(\alpha/k\right)\left\{1 - \left(-\log\left(\frac{T-1}{T}\right)\right)^k\right\} \tag{6.51}$$

where $\kappa = 7.8590c + 2.9554c^2$ in which, $c = \frac{2}{3+t_3} - \frac{\ln 2}{\ln 3}$

$$\alpha = \frac{\lambda_2 k}{\left(1 - 2^{-k}\right)\Gamma\left(1 + k\right)} \tag{6.52}$$

$$\xi = \lambda_1 - \alpha\left\{1 - \Gamma\left(1 + k\right)\right\}/k \tag{6.53}$$

where ξ is a location parameter, α is a scale parameter and k is a shape parameter, λ_1 and λ_2 being the first two L-moments.

The GEV /Pooled quantile estimator at each site is $Q_T = Q_{med} \times X_T$, where Q_{med} is the sample median for that site and X_T is the T-year return period normalized quantile estimator (growth curve factor), defined in eq (4.9) and whose parameters β and k are

estimated by eqs (4.12) and (4.10) respectively. The pooled L-moment ratios (t_2 and t_3) which are needed to estimate β and k are computed by taking the arithmetic average of L-moment ratios of individual sites in the pooling group.

6.3.3 Performance Criteria

The bias, rmse and the se are estimated at each site and for each estimation method. The following performance measures, the relative bias, the relative root mean square error and the relative standard error are calculated for each method as

$$BIAS_T^i\,[\%] = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{Q}_{i,r}^T - Q_i^T}{Q_i^T}\right) \times 100 \tag{6.54}$$

$$RMSE_T^i\,[\%] = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{Q}_{i,r}^T - Q_i^T}{Q_i^T}\right)} \times 100 \tag{6.55}$$

$$SE_T^i\,[\%] = \sqrt{\frac{1}{R}\sum_{r=1}^{R}\left(\frac{\hat{Q}_{i,r}^T - \bar{\hat{Q}}_i^T}{Q_i^T}\right)} \times 100 \tag{6.56}$$

where, $\hat{Q}_{i,r}^T$ is the estimated T-year quantile at a site i at r_{th} repetition, Q_i^T is the true T-year quantile at site i, and R is the number of repetitions taken as 10,000 in this part of the current study.

A summary is given below of the accuracy of estimated quantiles over all of the sites (M) in the pooling group given by the pooling average relative bias, rmse, and se of the estimated quantile.

$$P - RMSE_T = M^{-1}\sum_{i=1}^{M} RMSE_T^i \tag{6.57}$$

$$P - ABIAS_T = M^{-1}\sum_{i=1}^{M} \left|BIAS_T^i\right| \tag{6.58}$$

$$P - SE_T = M^{-1}\sum_{i=1}^{M} SE_T^i \tag{6.59}$$

In the above case, instead of bias, absolute bias is considered. The reason for that, in a heterogeneous group, the estimated pooled growth curve tends to overestimate the

true at-site growth curve at some sites and to underestimate it at others. In such cases abs.bias indicates the magnitude of the bias at a typical site and is more useful than bias, in which the contributions of negative and positive biases may cancel out to give a misleadingly small value of the bias (see Hosking and Wallis, 1997, p. 107).

6.3.4 Results

The bias, se and rmse for the GEV/pool and GEV/site quantile estimators for each site are calculated based on the specified record lengths and the assigned t_2-t_3 pairs for individual sites using Monte Carlo simulation. Inter-site correlations are neglected in the pooling analysis; Hosking and Wallis (1997) found that inter-site dependence has no effect on the bias and has little effect on the variance of index-flood quantile estimators when inter-site dependence is moderate.

Figure 6.11 shows the estimated bias, rmse and se for each site in the PG2 category , which is that of a moderately heterogeneous pooling group, for quantiles estimated with the GEV/pool at T=10, 100, 1000. The set of three plots demonstrates the variation in the statistical performance of the same estimators across the sites in the pooling group. The amount of bias of the GEV/pool quantile estimators vary almost linearly from sites 1 to 13 as the $t_2 - t_3$ pairs for individual sites vary linearly from sites 1 to 13. This bias is negative at sites 1 to 6 where the true quantile is greater than the average for the pooling group and positive at sites 8 to 13 where the true quantile is less than the pooling group average. The explanation for their behaviour is that, in the first case, the model $(t_2 - t_3)$ < the parent $(t_2 - t_3)$ whereas in the second case, the model $(t_2 - t_3)$ > the parent $(t_2 - t_3)$.

Figure 6.12 shows the estimated P-ABIAS, P-RMSE and P-SE for six pooling groups described in Table 6.5. The quantile estimator, the GEV/pool, is reported for T=10, 100, 1000. It is observed that the groups with larger heterogeneity register a larger pooled absolute bias. PG0 gives a very small bias in comparison with the other groups. The variation of se among the pooling groups does not differ widely as was the case with bias. Thus, it can be said that the main contributing factor to rmse in a heterogeneous pooling group arises from the bias. This finding corresponds to that of (Hosking and Wallis, 1997) who showed that the main effect of heterogeneity is to introduce bias into the estimated quantiles.

Figure 6.13 shows the variation in P-ABIAS, P-RMSE and P-SE of Q_{100}, estimated using the GEV/pool, with the different sizes of pooling group. The record lengths at each site in the pooling group are equal to 35. The statistics are displayed for all the pooling groups. It is observed that the size of a pooling group is not a significant factor for estimating Q_{100} as long as the number of station years is greater than 350 which is equivalent to 10 sites in a pooling group.

Figure 6.14 shows the variation in P-ABIAS, P-RMSE and P-SE of Q_{100} estimated using the GEV/pool with various record lengths. For the particular record length being considered, the record length is the same at each site in a group. The number of station years in a pooling group is considered fixed and is 450. The statistics are displayed for all six pooling groups. For a homogeneous group, sites with long record lengths produce smaller rmse values compared with sites that have shorter record lengths. But for moderate to highly heterogeneous groups it is the other way round. Sites with long record lengths produce higher rmse values compared with sites having shorter record lengths. This finding is in accordance with that of Hosking and Wallis (1997, p. 124) and Stedinger and Lu (1995, p. 69). Thus in a highly heterogeneous group with a fixed 450 station years, it is more desirable to have many sites with short record lengths than a smaller number of sites with long record lengths.

The comparison between at-site and pooled estimation is also carried out. This is done by estimating a ratio defined as

$$rmse \quad ratio = \frac{P - RMSE_T \quad for \quad at - site \quad estimation}{P - RMSE_T \quad for \quad pooled \quad estimation} \tag{6.60}$$

It is a measure of the relative accuracy of the two estimation methods. The pooled estimation would be more reliable when the value of the ratio is greater than 1. Four plots are made based on the rmse ratio (Figure 6.15 to 6.18).

Figure 6.15 shows the rmse ratio of eq (6.60) for six different pooling groups for quantiles for T=10, 100 and 1000. It is observed that pooled estimation is preferable for groups PG0 and PG1 for quantiles at T=10, 100 and 1000. For Groups in the PG2 and PG3 categories, which are mildly heterogeneous groups, pooled estimation fared well only for quantiles at the extremes such as for T=100 and T=1000. For groups in the PG4 and PG5 categories, pooled estimation does not give reliable estimation at T=10 and T=100 but the estimation is still worthwhile for quantiles at T=1000.

Figure 6.16 plots the ratio as a function of the heterogeneity measure H1. Increased heterogeneity decreases the advantage of pooled estimation over that of at-site estimation as would be intuitively imagined. A heterogeneity measure less than 4.0 renders the pooled estimation to be preferable at T=100.

Figure 6.17 shows the variation of the rmse ratio with various record lengths, for six different pooling groups, for Q_{100}. The arrangement of the experiment is similar to that described in Figure 6.14. As the record lengths increase, the relative performance of pooled estimation over at-site estimation decreases. A similar finding was also obtained by Stedinger and Lu (1995, p. 69). The rmse ratio is again plotted in Figure 6.18 this time as a function of heterogeneity measure. As the amount of heterogeneity increases, pooled estimation fares well only for shorter record lengths. For a moderately heterogeneous pooling group, it is found to be preferable to estimate Q_{100} using at-site data only if the record lengths exceeds 50.

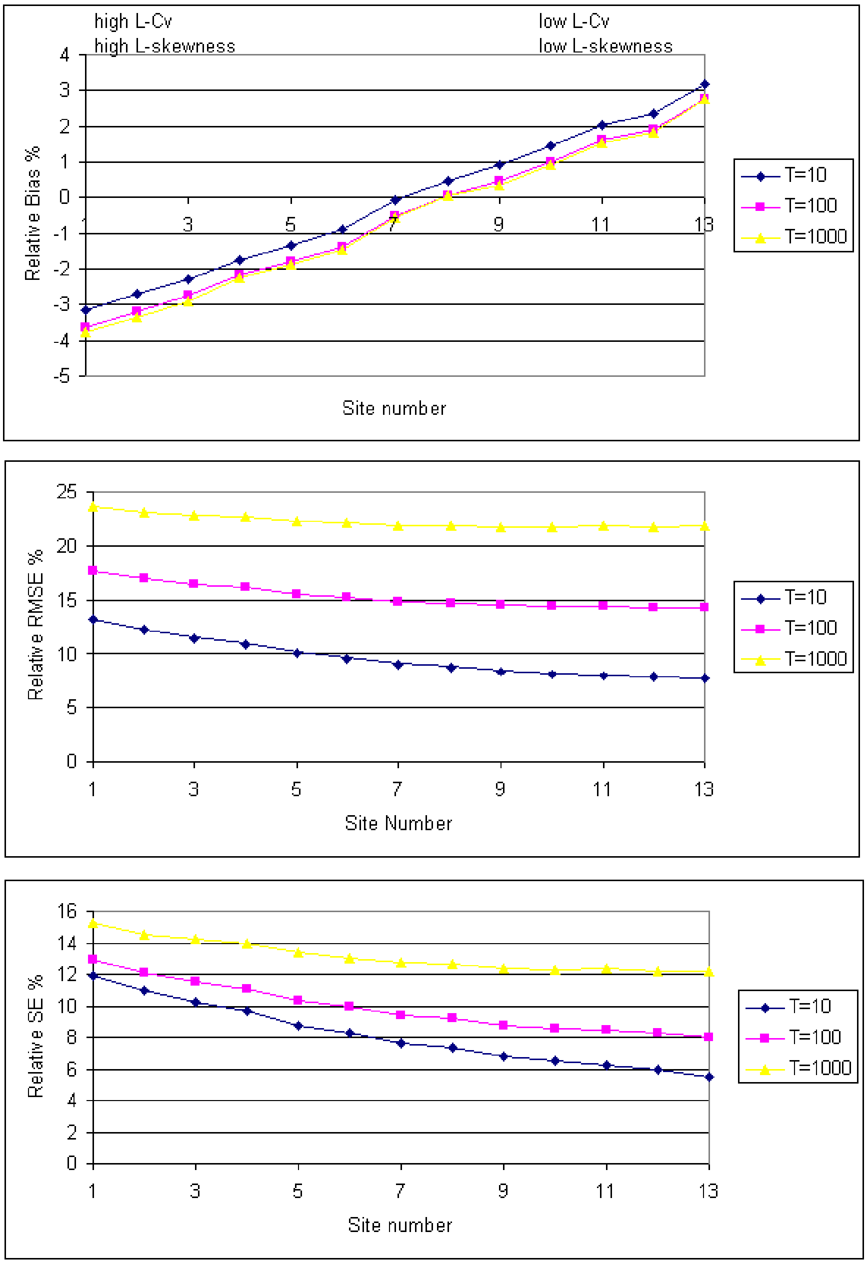


Figure 6.11: Variation of bias, rmse and se of GEV/Pool quantile estimator with respect to site number in the pooling group PG2 - , defined in Table6.6

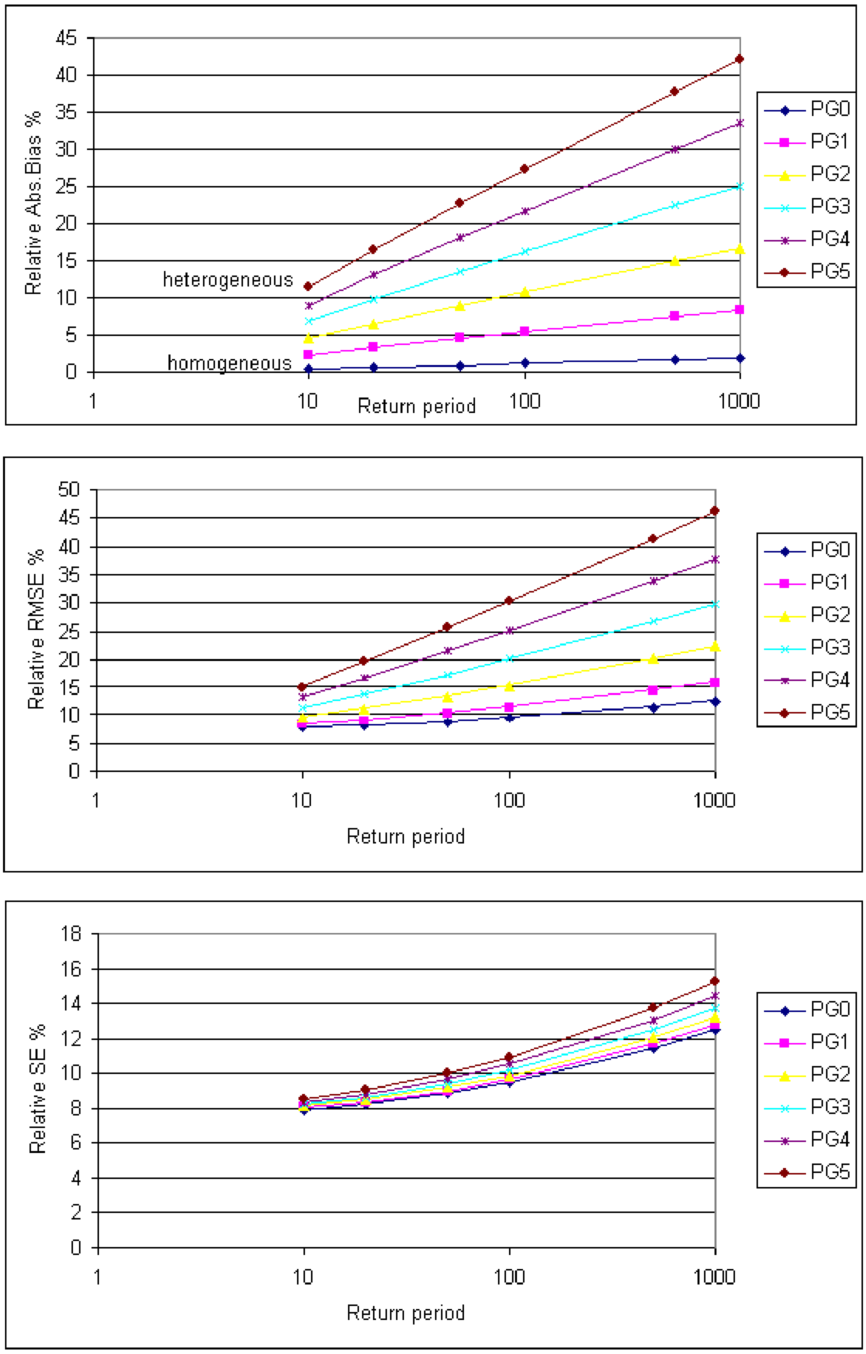


Figure 6.12: Variation of the P-ABIAS,P-RMSE and P-SE of GEV/pool quantile estimator for various return periods for six different pooling groups -

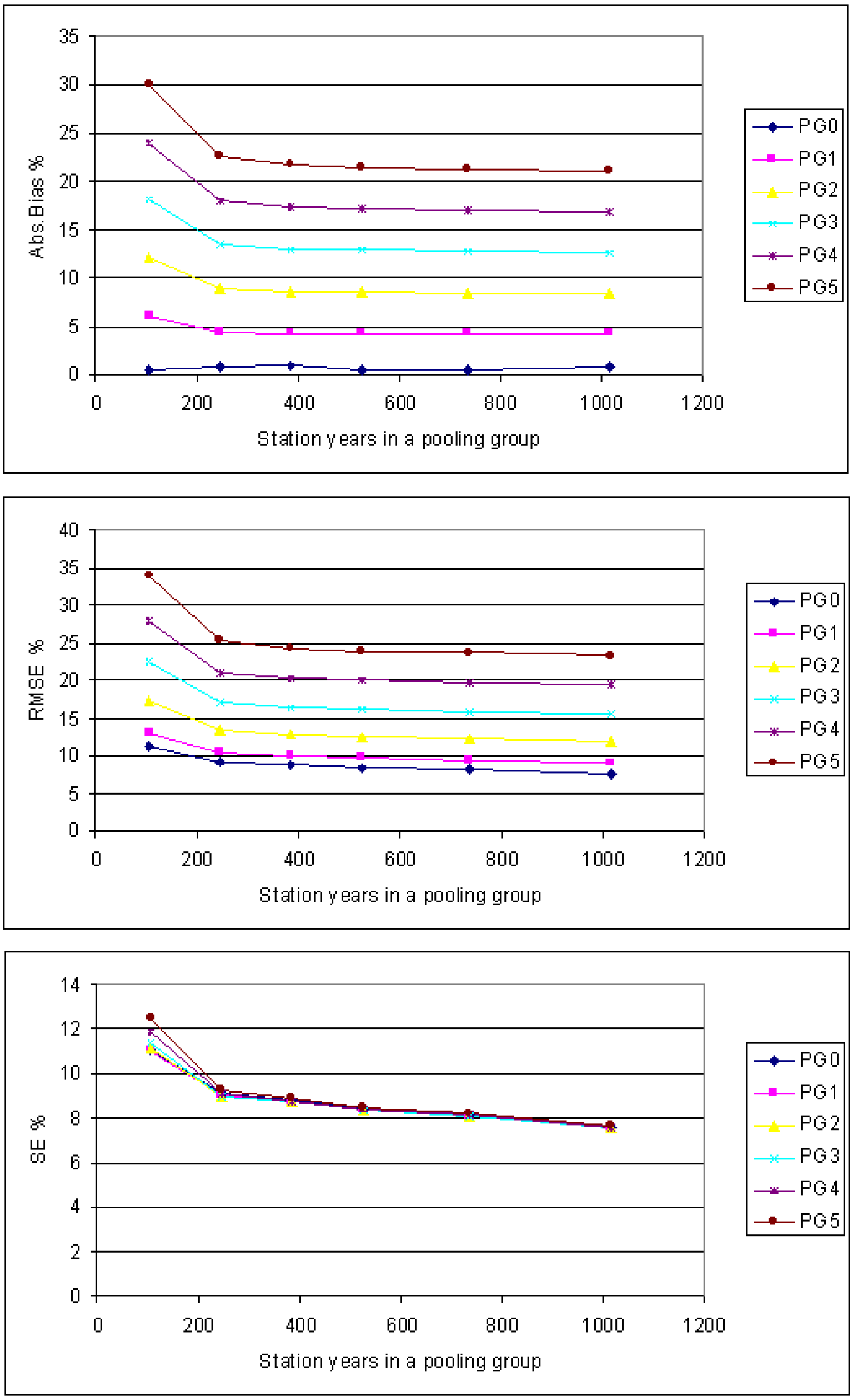


Figure 6.13: Variation of the P-ABIAS,P-RMSE and P-SE of GEV/pool quantile estimator with pooling group size - The statistics are displayed for for six pooling groups. Record lengths at each site are equal to 35

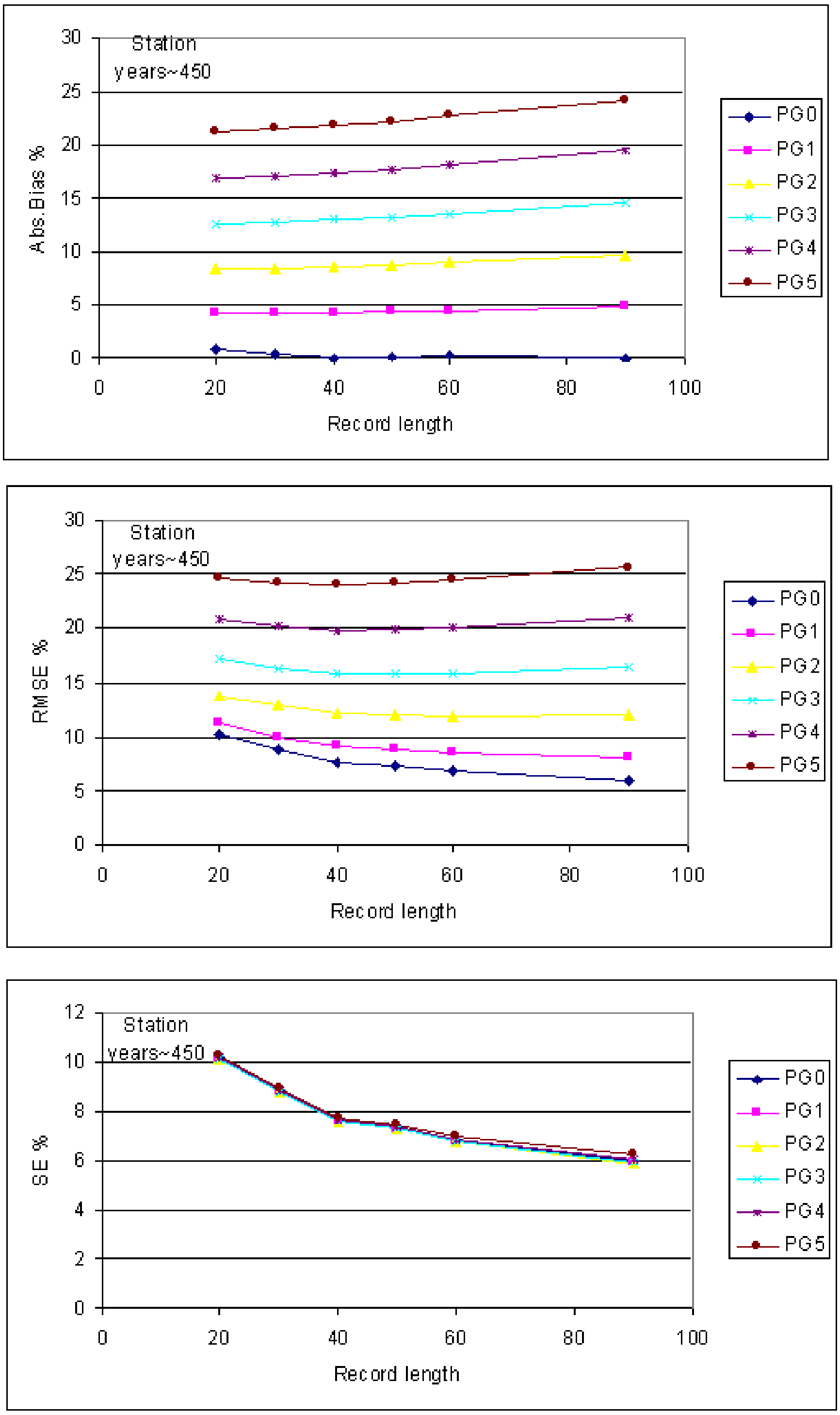


Figure 6.14: Variation of the P-ABIAS,P-RMSE and P-SE of GEV/pool quantile estimator with record lengths - The statistics are displayed for six pooling groups. Station years in a pooling group are c.450, here longer record lengths mean fewer records are included

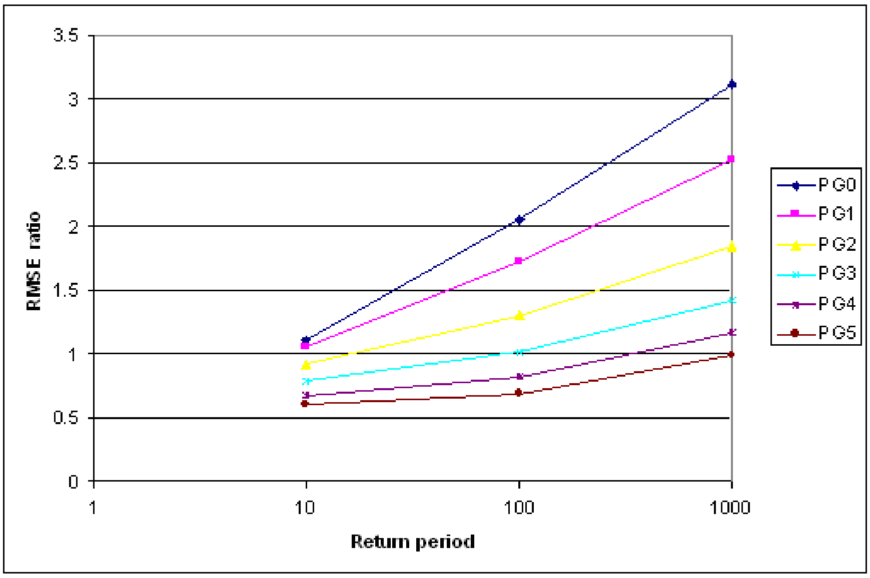


Figure 6.15: Rmse ratio for six different pooling groups for different return periods -

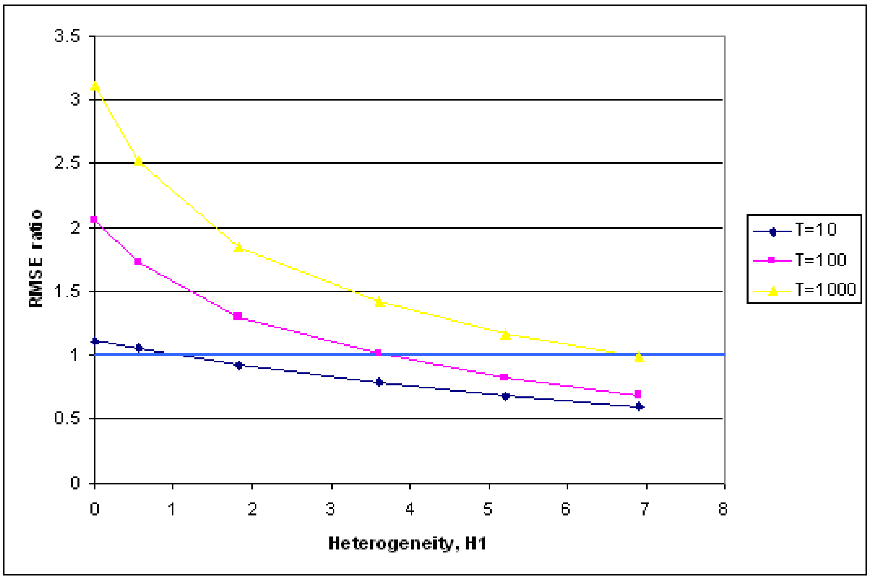


Figure 6.16: Rmse ratio as function of heterogeneity measure -

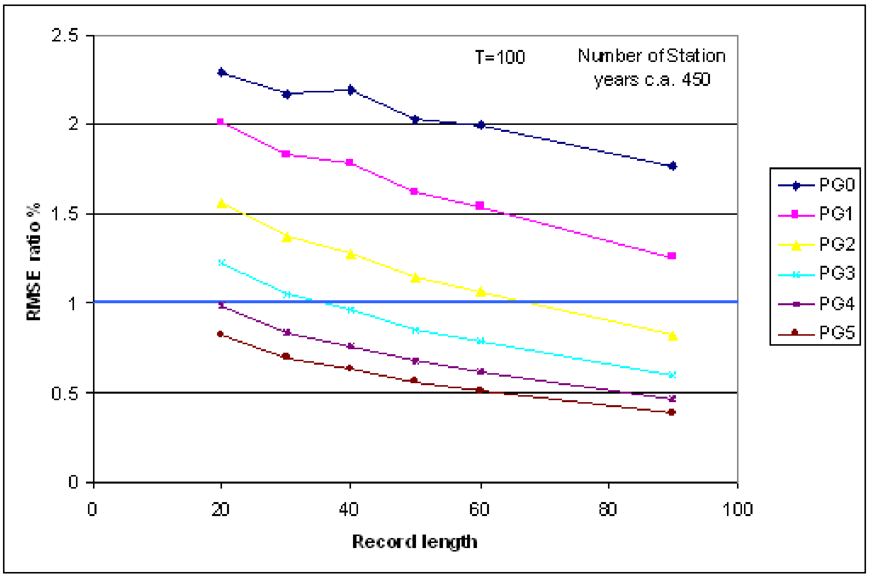


Figure 6.17: Rmse ratio for 100-year flood quantile estimators with various record lengths -

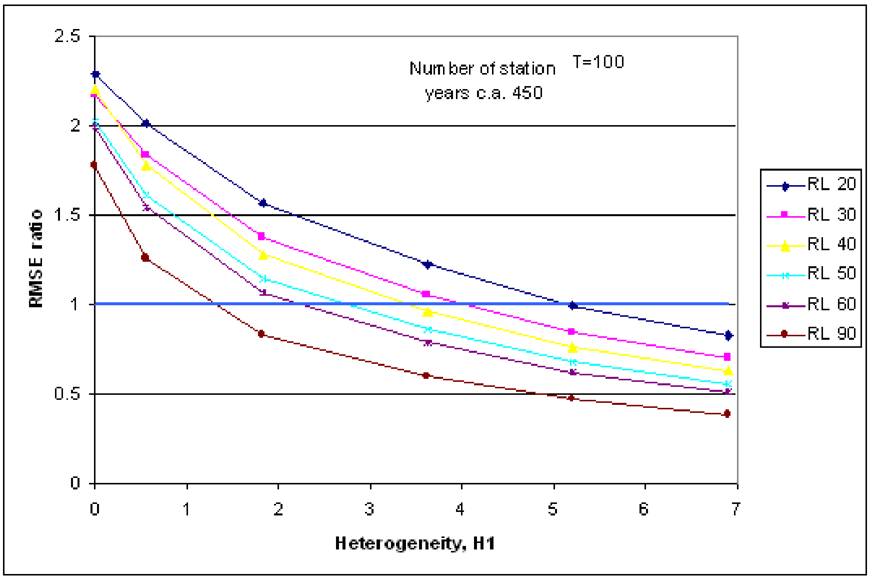


Figure 6.18: Rmse ratio for 100-year flood quantile estimators for various record lengths as function of heterogeneity -

6.4 Conclusion

The following conclusions are obtained from the above studies. Conclusion 1 below is drawn from the study described in section 6.2 while the remainder of the conclusions are drawn from the study described in section 6.3.

1. The magnitude of range of *se* that can be expected in quantile estimates for Irish conditions are summarised in Table 6.1 and Table 6.2 for gauged sites. In general, pooled estimation gives relatively smaller *se* than at-site estimation. The approximate *se* of pooling group based estimation at an ungauged site is also presented for Irish conditions. Using pooled estimation, the *se* of Q_{100}, which is often used for design purposes for many major projects in Ireland, can be considered to be approximately 10% of Q_{100} for gauged sites, while for ungauged sites this value can be considered to be approximately 36% of Q_{100}.

2. The performance of pooled estimation using Monte Carlo simulations shows that the size of a pooling group is not a significant factor for Q_{100} as long as the number of station years is more than 350 which, in the Irish context, is equivalent to 10 sites in the pooling group.

3. In a highly heterogeneous group, it is more desirable to have many sites with short record lengths than a smaller number of sites with long record lengths.

4. Increased heterogeneity decreases the advantage of pooled estimation over at-site estimation. Only a heterogeneity measure (H1) less than 4.0 can make the pooled estimation preferable to that obtained for at-site estimation for the estimation of Q_{100}.

5. It is preferable to conduct at-site analysis for the estimation of Q_{100} if the record length at the site concerned exceeds 50.

7

Guidelines for determining Q_T in the Irish context

7.1 Introduction

Q_T may be determined from at-site data if sufficient data exist and the required T value is relatively small. In some instances, although adequate at-site data may exist, interpretation may be problematical, e.g. an irregular shaped probability plot in which case it may be necessary to use a combined at-site/pooling group approach.

For large return periods, a combined at-site/pooling group approach is recommended. When no data exists at the subject site then the site's Qmed must be estimated from the catchment descriptor equation, enhanced where possible by use of suitable donor site data. The definition of donor site and its use in estimating Qmed is well documented in FEH (1999, 3,chap. 4) and FSUWP2.2 (2009, chap. 9)

While most hydrologists are greatly influenced by the pattern of at-site data on probability plots, when making choices about which procedures or estimates to adopt, it must be borne in mind that random samples from any particular statistical population show tremendous inter-sample variation when displayed on probability plots. Some such samples do not always display convincing straight line behaviour even when the parent distribution is represented by a straight line on such a plot. Figures from 3.8 to 3.13 show a number of random samples from EV1 populations with L-CV values in the range 0.1 to 0.2 (CV 0.2 to 0.4), which are relevant for Ireland. Sample sizes are in the range 20 to 50. It can be seen that, in the smaller size samples, the departures from

straight line behaviour is more pronounced than in the larger size samples. Some of the samples of 20 were given a low score in the assessment of probability plots described earlier in section 3.4. However most of the 50 size samples do not depart too markedly from straight line behaviour.

Considerations that should be borne in mind when selecting a design flood magnitude including

1. The at-site or subject station's annual maxima probability plot on ordinary and on log scales

2. The range of variation that can occur between random samples drawn from a single population – which emphasizes that, a single at-site estimate may differ from the true value.

3. The examination of the pooling group formed for the subject site

4. Preference for 2-parameter over 3-parameter distributions in at-site analysis

5. Examination of the pooling group members

6. One or more pooled growth curves

7. The overall statistical superiority of the at-site Qmed plus the pooled growth curve estimation method in homogeneous pooling groups, especially for larger return periods (c.f. Figure 6.16).

8. The robustness of 'the at-site Qmed plus pooled growth curve estimation' method to the small departures from the existence of the homogeneity of the pooling group

9. Consideration of straight line versus convex upwards relations including the issue of upper bounds on flood magnitudes

10. Comparison of estimated Q_T and corresponding water levels with "ground truth" i.e. does the proposed estimate make sense in the context of what has been previously observed or not observed in the vicinity?

11. Possible underestimation when the at-site Qmed plus the pooled growth curve estimate of Q_T is smaller than some observed at-site floods

12. The credibility of large outliers and how they might be viewed, assessed or accommodated

7.2 Determining Q_T from at-site data

The required flood quantile may be estimated from a single record of at-site data providing

- the record length N is at least 15 years. Such a record length allows $se(Q_{med})$ to be $< 10\% Q_{med}$ (see section 6.2.1)

- the required quantile return period T is less than N, or at least not appreciably larger than N and certainly not more than 2N

Ordinarily, a 2-parameter distribution should be used, either the EV1 or LN. In the case of EV1 the parameters should be estimated by the method of probability weighted moments (or the L-moment equivalent). In the case of the LN the parameters should be estimated from the mean and standard deviation of the logarithms of the data. Use of L-moments, as in the EV1 case, is statistically efficient as well as being simple to use. Use of ordinary logarithms of the data, as in the LN case, is also statistically efficient and simple to use and indeed simpler than use of L-moments in this case (for instance see Hosking and Wallis, 1997, A.8, for parameter estimation procedures using L-moments).

It is assumed that the basic statistics of the at-site data will have been calculated and that dimensionless statistics have been examined to see where they fall in the range of observed values among the gauged catchments under consideration, i.e. it is assumed that the user has carried out a thorough check on the at-site data in the context of the overall data set for the region and especially in the context of data for similar type catchments. (We may define similarity using the Euclidean distance measure d_{ij} as used in pooling group nomenclature, e.g. eq 4.22).

The at-site data should be displayed on an EV1-based probability plot, using Gringorten Plotting Positions (eq 3.10), and also on a lognormal probability plot, using Blom Plotting Positions (eq 3.11).

The EV1 parameters ξ and α are estimated using eqs (6.21) and (6.22) and the flood quantiles are estimated using eq (6.18).

7.3 Determining Q_T from pooled data

7.3.1 Data available at the subject site

7.3.1.1 Determining Qmed

If a gauging station exists at the subject site then Qmed is obtained from the AM series of floods at that site. If the record is short the Qmed so obtained is subject to larger standard error than if the record is long. However, since gauged data are so much better than information obtained from a catchment descriptor based formula it is recommended to base the Qmed value on the observed data even if the record is short.

7.3.1.2 Determining X_T

The growth factor should be determined from the data of a pooling group selected by means of the distance measure d_{ij} of eq (4.22) or (4.23) as desired. If the 5T rule is strictly applied to the 100 year flood estimation, then a minimum of 500 station years of data should be included in the pooling group. However a number of station years of data as small as 350, should still be considered adequate for the 100 year flood estimation (see Table 6.3and Figure 6.13). The examination of homogeneity based on L-CV should be undertaken at this point. In the event that the range of L-CV is greater than 0.12 (CV$\approx$ 0.25) in the group (in which case the possibility exists of having heterogeneity H1 $>$ 4.0), a thorough investigation of the pooling group members is required (e.g. Figure 4.7). A revised pooling group might be necessary in this context by excluding some sites but only on proper grounds. Consideration must also be given to the choice of a 2-parameter or a 3-parameter distribution in line with the discussion in Section 7.6 below. In the event that the at-site estimate of Q_T relation is always greater than that of the pooled estimate then consideration will have to be given to using a combination of the at-site estimate and the pooled estimate for design flood estimation. While there is a possibility that this could lead to over design, it avoids the less desirable outcome of underdesign.

7.3.2 Subject site ungauged

7.3.2.1 Determining Qmed

The value of Qmed will have to be obtained initially from the catchment descriptor data using the FSU WP2.3 equation (6.47). The Factorial Standard Error (FSE) of the estimate given by the equation is 1.36.

Use of a donor station is recommended or, in the absence of a donor station on the same river, a suitably chosen analog catchment might be used (for details see FSU WP2.2, 2009, chapter 9). However, an unsuitable analog catchment may add very little useful information. The suggested procedure to obtain an estimate of Qmed at the subject site is to use the equation

$$Q^s_{med} = Q^d_{med} \left(\frac{Q^s_{med} \quad by \quad regression}{Q^d_{med} \quad by \quad regression} \right) \tag{7.1}$$

where Q^d_{med} is the observed value at the donor site; and Q^s_{med} and $Q^d_{med} \quad by \quad regression$ are the estimate of Q_{med} using the catchment descriptors at subject and donor site .

7.3.2.2 Determing X_T

The procedure is identical to that described in the Section 7.3.1.2.

7.4 Some examples

Example 1 – Data series shows good straight line behaviour on a probability plot

A good example of such behaviour is shown by the River Clodiagh at Rahan (Stn No 25016) with 42 years of record. In Table 3.13, this station was assigned scores of 4 and L2 in the linear and non-linear pattern assessments where 4 indicates "good agreement with straight line" and L2 means "little deviation from straight line". The CV = 0.22 (L-CV = 0.13) and Hazen skewness = 0.63 (L-skewness =0.08) which is less than the theoretical value of 1.14 (0.169 using L-moments) in the EV1 distribution. It seems reasonable to suggest that the 50 year or even 100 year return period flood could be estimated by the EV1 distribution fitted to the at-site data in this case. The standard error of Q_{100} in such a case would be approximately 6% to 10%, as indicated in Figure 6.1 provided the underlying parent distribution is EV1. The at-site CV is lower than the national average value of 0.3. The pooled growth curves shown in Figure 7.1 are

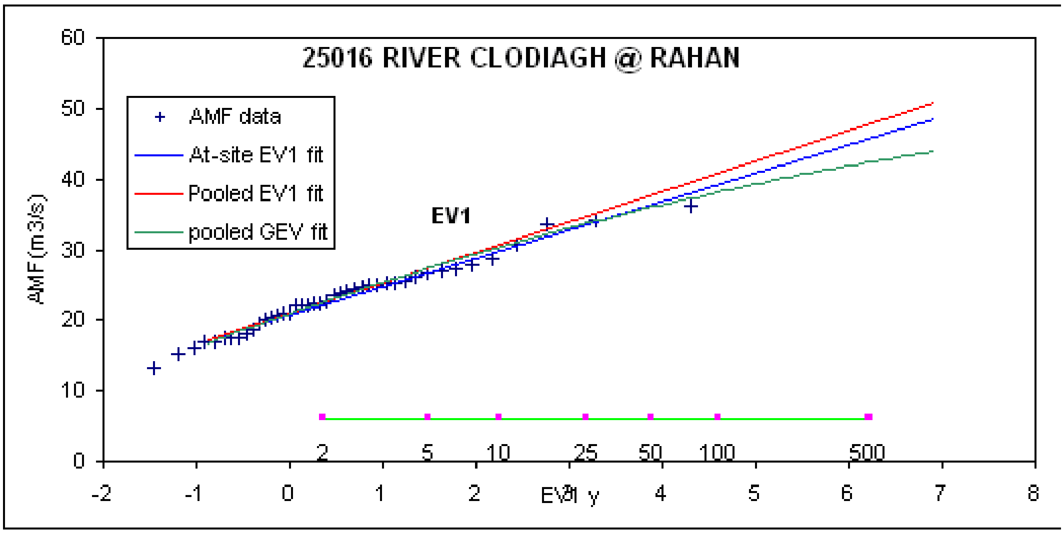


Figure 7.1: At-site and pooled quantile curves for station 25016 -

in general agreement with the at-site data and in this case the pooled estimate can be used for design flood determination.

Example 2 - Data series shows good straight line behaviour on a probability plot but lack of agreement between at-site and pooled estimates

A very good example of such behaviour is shown by the River Ryewater at Leixlip (Stn No 9001) with 48 years of record. The CV = 0.44 and Hazen skewness = 1.17 is practically equal to the theoretical value of 1.14 in the EV1 distribution. It seems reasonable to suggest that the 50 year or even 100 year return period flood could be estimated by the EV1 distribution fitted to the at-site data in this case. The standard error of Q_{100} in such a case would be approximately 4% to 8%, as indicated in Figure 6.1 provided the underlying parent distribution is EV1. The at-site CV is higher than the national average (CV = 0.3) and higher than that of other stations in the pooling group. Hence the two pooled growth curves are lower than the at-site curve shown in Figure 7.2.

However much faith is place in the superiority of the pooled estimate in general, it is a fact that the pooled estimate of Q_{50} has already been exceeded 5 times in 48 years of the record. This evidence cannot be overlooked when a final design flood is being selected even for long return periods such as 100 years. It could be that the observed sample displays a steepness on the probability plot that is at the upper end of the scale

of steepness and that the sample actually has come from a population with a less steep growth curve i.e. with a lower value of CV. However, no practical design project would ignore the at-site data in such a case.

Caveat: Random samples drawn from an EV1 distribution do not always display straight line behaviour on probability plots. The likelihood of this happening decreases as the sample size increased as seen in Figures 3.12 and 3.13. Conversely some random samples drawn from non-EV1 distributions could display straight line behaviour on probability plots. The proportion which does so decreases of course as the departure of the parent distribution from EV1 increases.

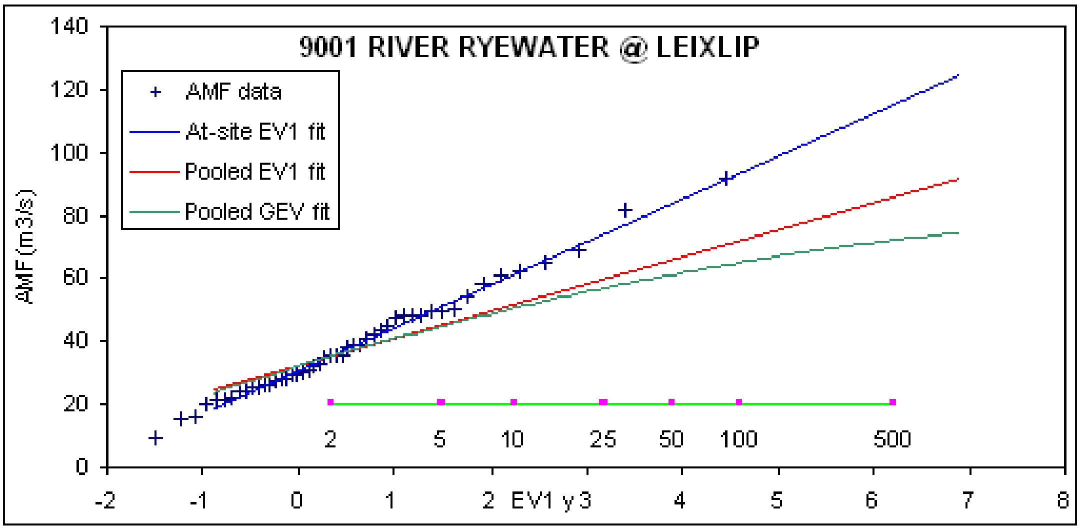


Figure 7.2: At-site and pooled quantile curves for station 9001 -

Example 3 - Data series shows concave downward behaviour on a probability plot or displays one or more high flow outliers.

A very good example of such behaviour is shown by the River Dodder at Waldron's Bridge (Stn No 9010) even though it has a relatively short period of record, 19 years. Even if the underlying parent growth curve is a straight line the short record makes it much more likely for departure from linearity to be observed on the probability plot. For this station, the CV = 0.86 and Hazen Skewness = 3.24, both of which are among the highest values recorded among the entire FSU data set. While a 3-parameter GEV distribution provides a good graphical fit (see Figure 7.3), in this case it has to be borne in mind, from a theoretical point of view, that GEV at-site estimates have high

standard error especially with short data sets. That means that other similar sized later records might not show the same pattern or be so steep. Hence, notwithstanding the good at-site fit, it would normally be recommended that a regional pooling estimate be used for flood estimation at this site unless, T $<$ 50 years. Because most other stations included in the pooling group will more than likely have smaller CV and skewness than that of the at-site data the pooling group based growth curve is lower than the at-site growth curve. However the pooled growth curve looks extremely low in comparison with the observed data as plotted in Figure 7.2.

The hydrologist then has to balance the weight of evidence provided by several publications relating to the advantages of the regional pooling method over the at-site method and his/her own beliefs about the representative nature of the observed data and whether the behaviour shown in Figure 7.2 could be repeated for another similar length of record.

For instance from Figure 7.2, the at-site GEV estimate of Q_{100} is about 350 cumec. One has to ask if this is credible. The largest flood on record (269 cumec) was caused by Hurricane Charlie in 1986 and its counterpart on the neighbouring Dargle catchment had earlier been equalled only twice in the previous 80 years. Undoubtedly therefore the 269 cumec on the Dodder could occur again but this does not necessarily support the figure of $Q_{100} = 350$ cumec. On the other hand the pooling group estimate of Q_{100} is seen on the plot to have been exceeded 4 times in 19 years which indicates that in this case the pooling based curve is definitely too low.

This sort of dilemma always presents itself when the subject site record has a CV and skewness which is considerably larger than the regional average, as has occurred in Example 2 above. In such cases, a prudent approach has to be adopted where “Ground Truth” cannot be overlooked. In such cases, it would be wise not to place too much faith in estimates of Q_T for large T.

Other very good examples of this behaviour are station 29011 on the River Dunkellin at Kilcolgan Bridge (N = 22, CV = 0.30, Hazen Skewness = 3.37) and station 36015 on the River Finn at Anlore (N = 33, CV = 0.32, Hazen Skewness = 2.62). It is noticeable that the above three examples are based on records which are considerable shorter than the longest records available in the study.

Example 4 – Data series shows convex upward behaviour on a probability plot or displays a number of high values which are almost equal and possibly some very low

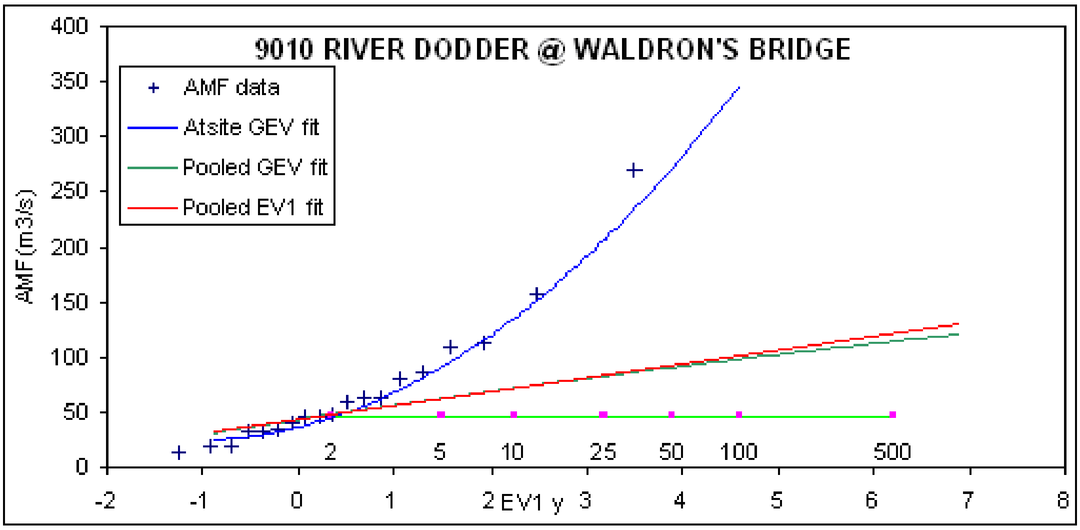


Figure 7.3: At-site and pooled quantile curves for station 9010 -

values which are among the smallest values in the series.

A very good example of such behaviour is shown by the River Newport at Barrington's Bridge (Stn No 25002) with 51 years of record. The CV = 0.16 and Hazen Skewness = -0.65, both of which are among the lowest values recorded among the entire FSU data set. While a 3-parameter GEV distribution provides a very good graphical fit (see Figure 7.4), in this case it has to be borne in mind, from a theoretical point of view, that GEV at-site estimates have high standard error especially with short data sets although 51 years may not be considered short. In addition, the k value is positive which implies an upper bound on the fitted distribution. This is not reasonable from a hydrological point of view and while it might be considered satisfactory to use the fitted distribution up to return periods of 25 years, it is recommended that a regional pooling estimate be used for flood estimation at this site. Because most other stations included in the pooling group will more than likely have larger skewnesses than that of the at-site data the pooling group based growth curve is steeper than that of the at-site growth curve. This may require that the pooling based estimate for high return periods be examined further in the context of the "Ground Truth"considerations referred to above. It is also known that this river, as well as the neighbouring Mulkear River, has been embanked since the 1920's and these embankments have an effect on flood flows as, on average, they are overtopped in one out of every 5 years. See also further

discussion below on the issue of upper bounds.

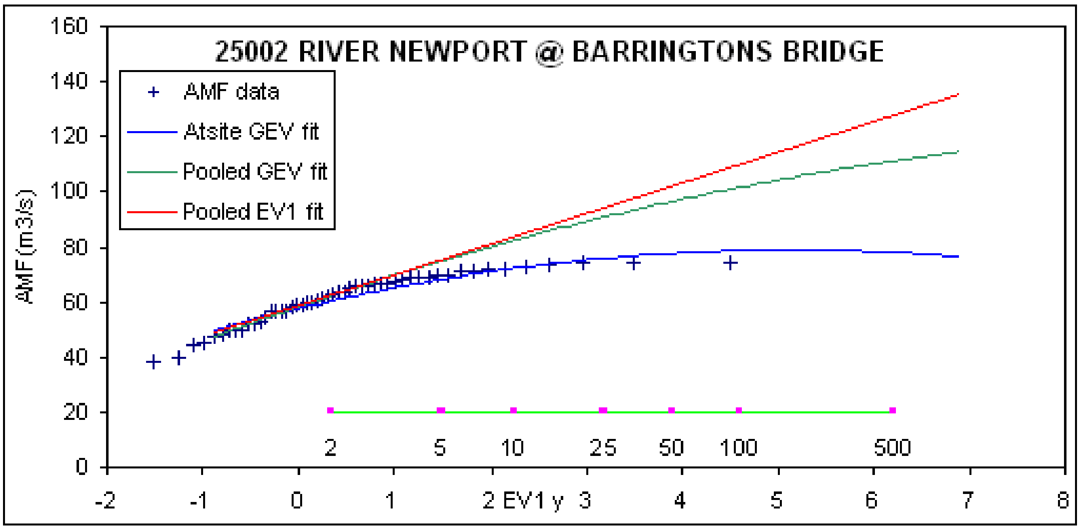


Figure 7.4: At-site and pooled quantiles for station 25002 -

Other very good examples of this behaviour are provided by station 25021 on the River Little Brosna at Croghan (N = 44, CV = 0.14, Hazen Skewness = -0.13) and station 34024 on the River Pollagh at Kiltimagh (N = 29, CV = 0.0.16, Hazen Skewness = -2.095). It is noticeable that two of these three examples are based on records which are considered to be among the longest in the study.

Example 5 - Data series contains one large outlier or two large outliers which are noticeably out of line with the other data points in a probability plot.

A very extreme example of such behaviour is provided by station 8009 on the River Ward at Balheary where 10 of 11 values available are less than 12 cumec and the largest is recorded as 53.6 cumec, which was a summer occurrence in the year 1993. The at-site and pooled quantile curves are given in Figure 7.5. The small sample size exacerbates the problem of trying to estimate a design flood for this location. This station has by far the highest estimate of the skewness, 5.58, of the whole data set. It has to be acknowledged that skewness calculated from such a short record is unreliable. The median value, which is not affected by the largest flood, is 5 cumec and even assuming a conservative growth factor of 2.5, the Q_{100} estimate would be 12.5 cumec. The probability of obtaining a largest value in 15 years which is more than 4 times the estimated Q_{100} value is virtually zero.

As a result of this statistical incompatibility, for example 5, there can be no unique recommendation made in this case. The following points should however be considered

- The meteorological conditions leading to the 1993 summer storm should be examined and the likelihood of these being repeated anywhere in the region (including the subject site) should be assessed
- Even if the rating curve or other features of the measurement are less than satisfactory the water level achieved in the locality will be known. This could be used as the basis of design of important works and floor levels of dwellings.
- The hydrologist has to balance between the belief that the 1992 flood was so large that it could never happen again or the belief that if it happened once it could happen again.

Hence for small return period events, a standard at-site Qmed and a pooled estimate of X_T could be used for the case in example 5, but for long return periods the above points would have to be taken into account.

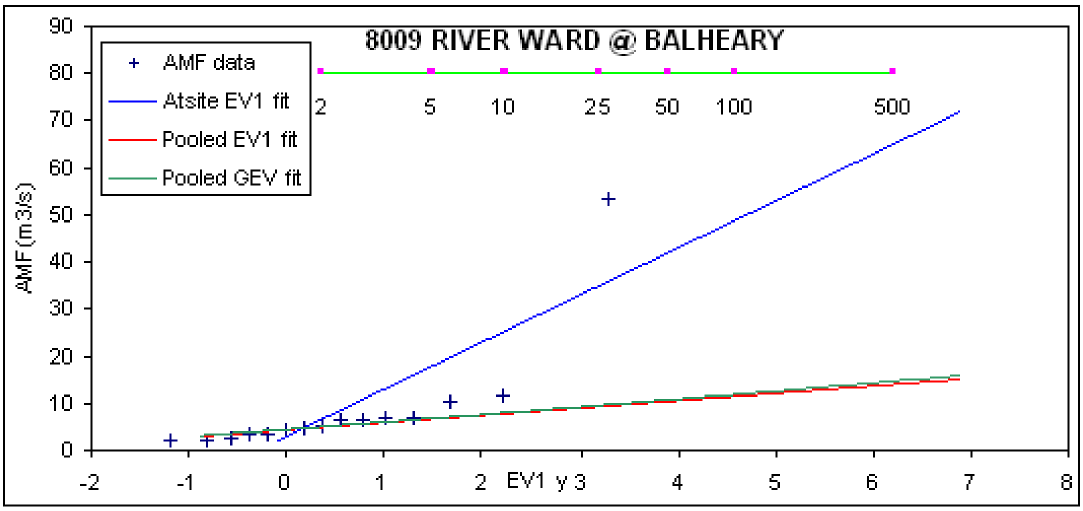


Figure 7.5: At-site and pooled quantile curves for station 8009 -

Further examples of outliers, but not of such an extreme nature, are provided by station 6021 on the River Yellow at Kiltybarden, by station 36031 on the River Cavan at Lisdarn and by station 26006 on the River Suck at Willsbrook. The same points as made above apply to these cases also.

Example 6 - Data series shows irregular behaviour of one sort or another, e.g an elongated S shape or a noticeable change of slope between lower and upper segments of the graph.

Examples are stations 15001 on the Kings River at Annamult (N = 42 years) and station 26008 on the River Rinn at Johnston's Bridge (N = 50 years). The at-site and pooled quantile curves for these stations are displayed in Figures 7.6 and 7.7 respectively.

The 15001 case is one of a probability plot with reverse curvature or elongated S shape. However the overall appearance is too far from a straight line, in the context of Figures 3.12 and 3.13 and hence application of a pooled growth factor should provide a satisfactory Q_T estimate.

The 26008 case is a more extreme version of 15001. While no single distribution could describe the at-site probability plot the upper end is not too dissimilar to some of the samples in Figures 3.12 and 3.13. Hence application of a pooled growth factor should provide a satisfactory estimate of Q_T for station 26008 also.

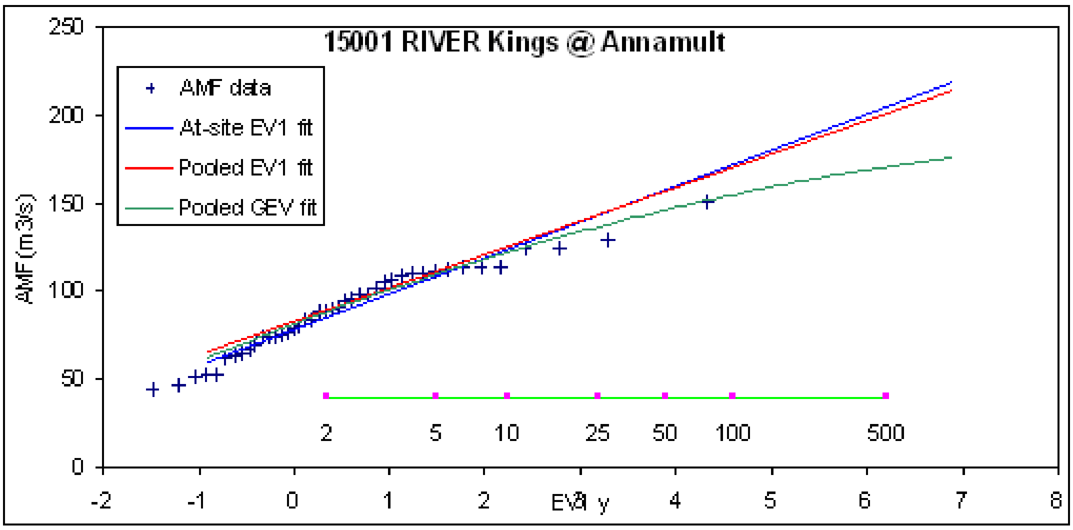


Figure 7.6: At-site and pooled quantile curves for station 15001 -

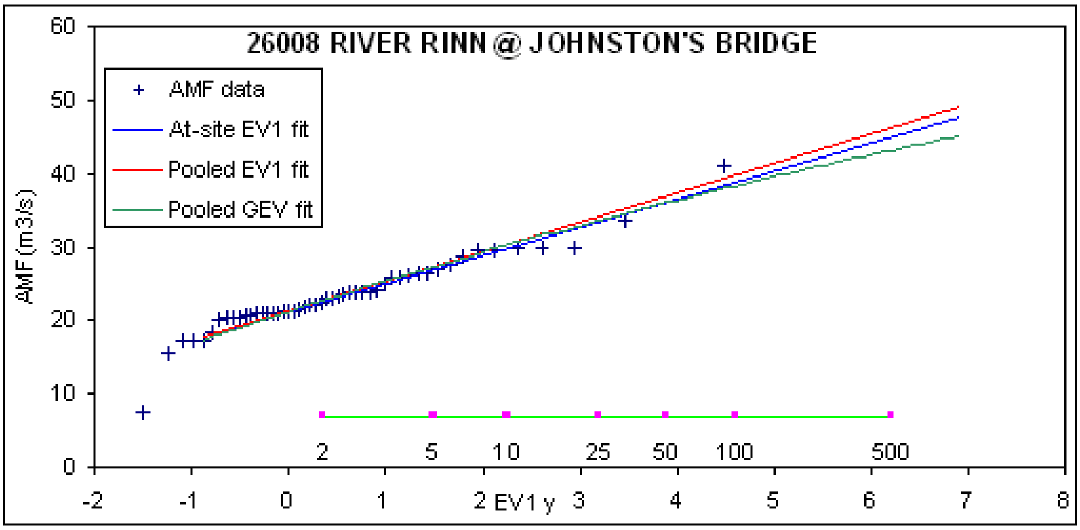


Figure 7.7: At-site and pooled quantiles for station 26008 -

7.5 Flood growth curves with an upper bound or without an upper bound?

3-parameter distributions, such as GEV and GLO, have an upper bound if the shape parameter $k > 0$. This does not seem hydrologically realistic, regardless of whether fitted to at-site or pooled data, because some unprecedented rainfall could occur in the future which would have a runoff or routing mechanism different to all previous floods and which could therefore produce a flood greatly in excess of the largest flood on record (e.g. the Nov 2009 flood). When $k > 0$, a 2-parameter distribution growth curve (i.e. a straight line growth curve) should be estimated instead. In such a case, it is nevertheless necessary to use judgement, following assessment of all relevant physical factors and water levels reached during previous floods, before a design flood can be specified. There is a possibility that a straight line 2-parameter distribution, when extrapolated to estimate very rare floods, may produce flood estimates that may seem implausibly large in the context of all relevant physical factors and the hydrologist's experience, i.e. displaying positive bias such as that in Figure 4.1(b). Hence the hydrologist will have to apply judgement to the estimation process in such a case.

7.6 Choice between at-site and pooled estimates and between 2- and 3-parameter pooled estimates

In ordinary circumstances, a 3-parameter distribution should not be used with at-site data. An exception could be made if the data series is very long, say > 50 years, and the required return period is small, say $= 25$ years. A 3-parameter distribution is more flexible and may give a better fit visually on a probability plot but the estimation of a 3^{rd} parameter has the effect of increasing the standard error of estimate of the estimated quantile, see Table 4.1. The opposite can also occur where a 3-parameter distribution, fitted to individual at-site data, can give a $Q - T$ relation that is not intuitively acceptable because of extreme upwards or downwards curvature.

In the event that the probability plot of available at-site data shows a marked departure from straight line behaviour, either by way of upward or downward curvature or because of some form of an elongated S shape, then consideration has to be given to using a pooling group estimate of the growth factor X_T for use with the at-site value of Qmed. However even this may not remove all doubts about the suitability of the method chosen as the pooled growth curve may disagree significantly from whatever general pattern is shown by the at-site data as in Examples 1 and 2 above. In such a case a decision has to be made whether to trust the pooled growth curve as a matter of good practice or principle, based on the well documented reduced standard error of estimate and robustness of the pooling method, e.g. see Figure 6.15. However when the at-site data display CV and skewness greatly in excess of the pooling group average it is not always easy to trust the pooled growth curve. Also in a case where a large number of samples is available (N > 50), an at-site estimate should be more appropriate than pooled estimate.

In the case of a very large observed flood it is also possible, under certain assumptions, to calculate the probability that such a large flood could occur. For instance if a sample of 50 floods are drawn randomly from an EV1 distribution with CV = 0.3, then the probability that the largest flood would exceed $1.25 \times Q_{100}$, where Q_{100} is the population value of the 100 year return period flood, is less than 7% and the probability that it would exceed $1.25 \times Q_{100}$ is less than 1%. In practice the true value of Q_{100} is unknown and has to be replaced by an estimate which makes these percentage probabilities less reliable. Nevertheless such calculations could be adapted to provide an

informal test of the hypothesis that the observed outlier is consistent with the assumed parent population. [The quoted probabilities are based on the fact that if Q is EV1(ξ, α) then Qmax in N years is EV1($\xi + \alpha \ln N, \alpha$).

If a data series shows negative skewness and 3 or 4 of the largest floods that differ from one another by only a very small amount then the data series will usually exhibit downwards curvature on a probability plot. Any 3-parameter distribution fitted to the at-site data will usually have an upper bound which is usually not very much larger than the largest recorded flood. However, it is unreasonable to expect that there is an upper bound of such modest size and in such a case the advice is to use the pooled growth curve rather than the at-site one. If the pooled growth curve is also convex downwards, then a 2-parameter distribution should be fitted to the pooled data so as to avoid the occurrence of an the upper bound. However, in that case care has to be taken so as not to specify design floods, for values of $T \approx N$, which are greatly in excess of the observed floods and judgement has to be used in such cases. On the other hand some unknown large rainfalls could occur in the future which would cause the catchment to respond differently to those storms which caused the recorded floods and it might be unwise not to accept the 2-parameter based estimates for higher return periods, T >> N (c.f. the 2009 floods in Ireland).

7.7 Summary guidelines

A scheme for estimating the T-year flood in the Irish context can be considered as follows.

Table 7.1: Scheme for estimating the T-year flood

Criteria	At-site information	Pooled information
1. No records or N< 3	estimate Qmed from catchment characteristics	estimate X_T using data from pooling group[1]
2. N=3 to 15	estimate Qmed from record	estimate X_T using data from at-site and pooling group combined
3. N=16 to 50	estimate Qmed from record or alternatively if $T \leq N$, estimate Q_T based on EV1 distribution	estimate X_T using data from at-site and pooling group combined
4. N>50	estimate Qmed from record or alternatively if $T \leq N/2$, estimate Q_T based on EV1 distribution	estimate X_T using data from at-site and pooling group combined
5. T>500	estimate Qmed from record	estimate X_T using data from at-site and pooling group combined

[1]pooling group should be formed using d_{ij} values provided by either eq (4.22) or eq (4.23). The GEV distribution should be used primarily in the estimation of X_T. However, in the case when the shape parameter $K > 0$, the EV1 growth curve should be estimated instead.

8

Summary, Conclusions and Recommendations

8.1 Summary and Conclusions

The work reported in this thesis was carried out to determine the most suitable method of finding the magnitude-return period relationship for Irish flood data, up to and including 2006, using the annual maximum (AM) statistical model. Data from approximately 200 gauging stations in the Republic of Ireland, from the archives of OPW, EPA and ESB were available. Of these 115 were of quality Grade A while 67 were Grade B. Data of 16 stations which had both pre and post drainage records were also available. Summary statistics and probability plots were prepared for all stations. However no further inferences were drawn from the Grade B stations data.

Flow characteristics are examined with a view to determining the regional behaviour of flood statistics, selecting appropriate statistical distributions to describe flood data and examining the seasonal aspect of flooding.

The descriptive statistics show that Irish AM data have low CV (average of 0.3) and low skewness (average Hazen skewness of 1.0) whether measured by the traditional statistics or by L-moment statistics. Furthermore, the values are similar to some extent with what is observed for Irish data in the FSR (1975). Examination of probability plots and moment ratio diagrams suggest that among 2-parameter distributions the Extreme Value Type I (EV1) and lognormal (LN) are the most appropriate, the exception being for a number of stations that display very low skewness values which in some cases

is caused by the 3 or 4 largest values in the record not being appreciably different from one another. Explanations for this latter phenomenon were sought in terms of the catchment descriptor, Flood Attenuation Index, but without success. In seeking a distribution, the classical goodness of fit tests, chi-squared and empirical distribution functions, gave different results and none of them appeared to support one distribution more than the others. Based on the Anderson Darling test the EV1 and LN are the most suitable 2-parameter distribution in about 75% of overall cases while the percentage rises to about 90% for all four 3-parameter distributions (GEV, GLO, GNO and LN3) that are considered in the study. The goodness of fit test applicable to regional pooling analysis identified the GEV as being the best fit among the three distributions (GEV, GLO and LN3). Seasonal analysis shows that two thirds of AM flows occur during the winter months of October to March and at 11 stations no AM floods occurred during the summer months. July is the least likely month in which AM floods occur while August and September are the most likely summer months to have AM floods. At 20 out of the 202 stations examined the largest flow on record occurred during summer, 8 of them in August, while 45% of stations had their largest flood on record in December.

The estimation of the T year flood quantile is considered for each of the full range of possibilities that can arise in practice, for both gauged and ungauged catchments and for flow records ranging from short to long (see 7.1). The index flood method is recommended whereby Q_T is expressed as $Q_T = Q_{med} \times X_T$, where Q_T is the flood of return period T years, Qmed is the median of the annual maximum series at the subject site and X_T is the growth factor appropriate to the subject site.

Estimation of X_T may be based on at-site data if a sufficiently long ($N \geq T$) data record exists at the subject site. Otherwise X_T is estimated from the dimensionless L-CV and L-skewness values obtained by averaging these quantities obtained over the stations of a homogeneous pooling group . The members of the pooling group are chosen with the help of a Euclidean distance or similarity measure, d_{ij}. Tests have been carried out and reported on in chapter 4.6 into the effect on the estimated value of X_T of catchment area, peat coverage, lakiness (as measured by FARL), geographical location and period of record varying from the 1950s to the 1990s. None of these were judged to be of sufficient influence to warrant special provision. Also, tests on the effectiveness of different combinations of catchment descriptors in the definition of d_{ij}

were carried out and reported on and the most effective ones were considered to be AREA, SAAR and BFI .

Examination of homogeneity is needed in the estimation of X_T because a homogeneous pooling group of sites is necessary for the minimisation of the error of estimating X_T. Tests conducted on the data of 85 gauging stations were used to assess how successful a ROI method of identifying pooling group membership is in selecting groups that actually are homogeneous. The sampling distribution of L-CV in each pooling group and the 95% confidence limits about the pooled estimation of L-CV are obtained by simulation. The L-CV values of the selected group members are compared with these confidence limits both graphically and numerically. The outcomes are also compared with the heterogeneity measures H1 and H2. The H1 values show an upward trend with the ranges of L-CV values in the pooling group whereas the H2 values do not show any such dependency. Overall the results show that even with a carefully considered selection procedure, the pooling groups identified are not perfectly homogeneous . As a compromise it is recommended that a group containing more than 2 values of L-CV outside the 95% confidence limits of that variable should not be considered homogeneous.

Some degree of heterogeneity always exists in real cases. A thorough investigation of a heterogeneous pooling group is required before it can be used to estimate X_T. Investigations on 27 heterogeneous pooling groups were carried out using exploratory diagrams of catchment descriptors and also a summary of flood statistics. The investigation results show that in many cases the special qualities of subject site lead to the pooling group being heterogeneous. The results also reveal that the subject site needs to be positioned near the centre of the gauging sites, on the respective catchment descriptor axes, to which it is hydrologically similar in order to ensure homogeneity. But in some cases it is impossible to guarantee that a group is homogeneous.

The standard errors associated with estimates of X_T and Q_T were investigated. The standard error of Q_T estimated by the index flood method is dominated by $se(\text{Qmed})$. The $se(Q_T)$, expressed as a percentage is found to vary only slightly with T. When Qmed is estimated from at-site data and X_T is estimated from a pooling group containing approximately 500 station years of data then $se(Q_T)$ is of the order of 5% to 10 % Q_T regardless of the magnitude of the return period. If Qmed is estimated from a catchment descriptor based formula alone and X_T is estimated from a pooling group

containing approximately 500 station years of data then $se(Q_T)$ is of the order of 36% Q_T.

The performance of pooled estimation has also been investigated, again in the Irish context. Experiments using Monte Carlo simulation show that the size of a pooling group is not a significant factor for estimating flood quantiles as long as the number of station years is greater than 350. An increased amount of heterogeneity decreases the advantage of pooled estimation over that of at-site estimation. For estimating extreme quantiles, a measure of heterogeneity (H1) less than 4.0 can render the pooled estimation preferable to that of at-site estimation . When the record lengths at the site concerned exceed 50, estimates based on at-site data are as effective as those based on pooled data.

Chapter 7 deals with guidelines for the estimation of Q_T, from both at-site and pooled data. Several points that need to be taken into account in practice are outlined and it is to be emphasised that the blind use of any prescribed method can sometimes lead to a Q_T estimate which is not in accordance with “ground truth”. This is especially so when the at-site data have a higher than average value of the CV and/or the skewness or in the presence of outliers. A number of examples are discussed, most of which throw up practical problems of the type met in practice where it is difficult to make the appropriate choices. Discussion is also provided on the issues of flood distributions having upper bounds and on the choice between two and three parameter distributions. In some of these cases the user may need to consider using an at-site based estimate in preference to the generally recommended regional pooling based method. What the above conclusions indicate is that flood estimation cannot be reduced to a single strict formula based procedure and that individual analysts must make choices which depend on the particular circumstances of the problem and which take into account their own knowledge and experience.

8.2 Recommendations for further research

The following recommendations are made for further development of the present study:

- Examination of behaviour of annual maximum rainfall series and investigation of any similarities in relation to the convex upwards annual maximum probability plots

- Use of rainfall runoff modelling to investigate this comparison further in the same cases.
- Investigation of intersite dependence in the pooled analysis using copula function

References

M. I. Ahmad, C. D. Sinclair, and B. D. Spurr. Assessment of flood frequency models using empirical distribution function statistics. *Water Resources Research*, 24(8): 1323–1328, 1988a. 56, 58

M. I. Ahmad, C. D. Sinclair, and A. Werritty. Log-logistic flood frequency-analysis. *Journal of Hydrology*, 98(3-4):205–224, 1988b. 12

Manuel A. Benson. Uniform flood-frequency estimating methods for federal agencies. *Water Resour. Res.*, 4(5):891–908, 1968. 0043-1397. 3, 11

K. P. Bhattarai and K. M. O'Connor. The effects over time of an arterial drainage scheme on the rainfall-runoff transformation in the brosna catchment. *Physics and Chemistry of the Earth, Parts A/B/C*, 29(11-12):787 – 794, 2004. ISSN 1474-7065. Anthropogenic impacts on catchment processes. 118

B. BOBEE and P. RASMUSSEN. Recent advances in flood frequency analysis. u.s national report to international union of geodesy and geophysics 1991-1994. *Rev Geophys*, 33(suppl):1111–1116, 1995. 89

A Brath, A Castellarin, M Franchini, and G Galeati. Estimating the index flood using indirect methods. *HYDROLOGICAL SCIENCES JOURNAL-JOURNAL DES SCIENCES HYDROLOGIQUES*, 46(3):399–418, JUN 2001. ISSN 0262-6667. 13

D. H. Burn. Evaluation of regional flood frequency-analysis with a region of influence approach. *Water Resources Research*, 26(10):2257–2265, 1990. 4, 15, 86, 88, 89, 90

Donald H. Burn. Catchment similarity for regional flood frequency analysis using seasonality measures. *Journal of Hydrology*, 202(1-4):212–230, 1997. 0022-1694 doi: DOI: 10.1016/S0022-1694(97)00068-1. 89

A. Castellarin, D. H. Burn, and A. Brath. Assessing the effectiveness of hydrological similarity measures for flood frequency analysis. *Journal of Hydrology*, 241(3-4): 270–285, 2001. 4, 12, 13, 14, 101

J. U. Chowdhury, J. R. Stedinger, and L. H. Lu. Goodness-of-fit tests for regional generalized extreme value flood distributions. *Water Resources Research*, 27(7):1765–1776, 1991. 56, 58, 129

J. M. Cunderlik and D. H. Burn. The use of flood regime information in regional flood frequency analysis. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, 47(1):77–92, 2002. 4, 14

J. M. Cunderlik and D. H. Burn. Switching the pooling similarity distances: Mahalanobis for euclidean. *Water Resources Research*, 42(3), 2006a. 88, 89

JM Cunderlik and DH Burn. Site-focused nonparametric test of regional homogeneity based on flood regime. *JOURNAL OF HYDROLOGY*, 318(1-4):301–315, MAR 1 2006b. ISSN 0022-1694. doi: {10.1016/j.jhydrol.2005.06.021}. 4

JM Cunderlik and TBMJ Ouarda. Regional flood-duration-frequency modeling in the changing environment. *JOURNAL OF HYDROLOGY*, 318(1-4):276–291, MAR 1 2006. ISSN 0022-1694. doi: {10.1016/j.jhydrol.2005.06.020}. 14

C. Cunnane. Unbiased plotting positions - a review. *Journal of Hydrology*, 37:205–222, 1978. 43

C. Cunnane. *Review of statistical models for flood frequency estimation.* Proceedings of the International Symposium on Flood Frequency and Risk Analyses. Reidel Publishing Company, Boston, 1987. 11

C. Cunnane. Methods and merits of regional flood frequency-analysis. *Journal of Hydrology*, 100(1-3):269–290, 1988. 13

C. Cunnane. *Statistical distributions for flood frequency analysis*, volume WMO Publ. No. 718. World Meteorological Organization, Geneva, operational hydrology report no. 33 edition, 1989. 3, 10, 11, 12, 43, 56, 66

R.B. D' Agostino and M.A. Stephens. *Goodness of fit techniques.* Marcel Dekker, New York, 1986. 58

T. Dalrymple. Flood frequency analysis. water supply paper 1543 a. *US Geol. Surv.*, 1960. 13, 16, 85, 129

C. De Michele and R. Rosso. Uncertainty assessment of regionalized flood frequency estimates. *Journal of Hydrologic Engineering*, 6(6):453–459, 2001. 18

FEH. *Flood Estimation Handbook*, volume 3 of *Statistical procedures for flood frequency estimation.* Institute of Hydrology, Wallingford, 1999. 4, 12, 13, 14, 15, 43, 85, 86, 88, 89, 90, 91, 95, 100, 132, 143, 162, 178, 181, 183, 199

H. D. Fill and J. R. Stedinger. Homogeneity tests based upon gumbel distribution and a critical-appraisal of dalrymple test. *Journal of Hydrology*, 166(1-2):81–105, 1995. 16

Fiorentino, M, Gabriele, S, Rossi, F, Versace, and P. *Hierarchical approach for regional flood frequency analysis.* Proceedings of the International Symposium on Flood Frequency and Risk Analyses. D. Reidel Publishing Company, Boston, 1987. 15

R. A. Fisher and L. H. C. Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 24(2):180–191, 1928. 2

H.A. Foster. Theoretical frequency curves and their application to engineering problems. *Trans. ASCE*, 87:142–173, 1924. 2

FSR. *Flood Studies Report*, volume 1. Nat. Environ. Res. Council, London, 1975. 9, 11, 13, 18, 21, 26, 27, 43, 56, 83, 85, 86, 118, 162, 179, 183, 215

FSUWP2.2. Flood studies update: Work package 2.2. 2009. 199

FSUWP2.3. Flood studies update: Work package 2.3. 2009. 178

W.E. Fuller. Flood flows. *Trans. Am. Soc. Civ. Eng*, 77:564–617, 1914. 2

L. Gaal, J. Kysely, and J. Szolgay. Region-of-influence approach to a frequency analysis of heavy precipitation in slovakia. *Hydrology and Earth System Sciences*, 12(3):825–839, 2008. 4, 13, 14, 101

S. Gabriele and N. Arnell. A hierarchical approach to regional flood frequency-analysis. *Water Resources Research*, 27(6):1281–1289, 1991. 15

J. A. Greenwood, J. M. Landwehr, N. C. Matalas, and J. R. Wallis. Probability weighted moments - definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, 15(5):1049–1054, 1979. ISI Document Delivery No.: HS865 Times Cited: 257 Cited Reference Count: 13. 11, 23

N. P. Greis and E. F. Wood. Regional flood frequency estimation and network design. *Water Resources Research*, 17(4):1167–1177, 1981. 13, 14

N. P. Greis and E. F. Wood. Correction. *Water Resources Research*, 19(2):589–590, 1983. 13, 14

PL Grover, DH Burn, and JM Cunderlik. A comparison of index flood estimation procedures for ungauged catchments. *CANADIAN JOURNAL OF CIVIL ENGINEERING*, 29(5):734–741, OCT 2002. ISSN 0315-1468. doi: {10.1139/L02-065}. 13

E. J. Gumbel. The return period of flood flows. *The Annals of Mathematical Statistics*, 12(2):163–190, 1941. 2, 3

A. Hazen. Discussion on 'flood flows' by w.e. fuller. *Trans. ASCE*, 77:626–632, 1914. 2, 43

A. Hazen. *Flood flows.* John Wiley, New York, 1932. 2, 24

C.S. Hebson and C. Cunnane. *Assessment of use of at-site and regional flood data for flood frequency estimation.* D. Reidel Publishing Company, 1987. 179

J. Holden, M. G. Evans, T. P. Burt, and M. Horton. Impact of Land Drainage on Peatland Hydrology. *J Environ Qual*, 35(5):1764–1778, 2006. doi: 10.2134/jeq2005.0477. 111

J. R. M. Hosking and J. R. Wallis. Some statistics useful in regional frequency-analysis. *Water Resources Research*, 29(2):271–281, 1993. 16, 129

J. R. M. Hosking, J. R. Wallis, and E. F. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27(3):251–261, 1985. 11, 86

J.R.M. Hosking and J. R. Wallis. *Regional Frequency Analysis*. Cambridge University Press, Cambridge, 1997. 3, 4, 11, 13, 14, 15, 16, 17, 19, 23, 59, 60, 66, 85, 86, 88, 89, 90, 91, 93, 95, 101, 106, 129, 130, 163, 165, 178, 180, 181, 183, 184, 187, 189, 190, 235, 236

John C. Houghton. Birth of a parent: The wakeby distribution for modeling flood flows. *Water Resour. Res.*, 14(6):1105–1109, 1978. 0043-1397. 12

Z. Kaczmarek. Efficiency of the estimation of the floods with a given return period. *Proceedings of the Torento Symposium, IAHS Publication No. 45*, 3:145–159, 1957. 18

A. L. Kay, D. A. Jones, S. M. Crooks, T. R. Kjeldsen, and C. F. Fung. An investigation of site-similarity approaches to generalisation of a rainfall-runoff model. *Hydrol. Earth Syst. Sci.*, 11(1):500–515, 2007. HESS 1027-5606. 93

M. Kendall and A. Stuart. *The advanced theory of statistics. Vol.1: Distribution theory.* Griffin, London, 4th edition, 1977. 161, 165

B.F. Kimball. An approximation to the sampling variances of an estimated maximum value of given frequency based on the fit of doubly exponential distribution of maximum values. *Annals of Mathematical Statistics*, 20:110–113, 1949. 18

T. R. Kjeldsen and D. A. Jones. Prediction uncertainty in a median-based index flood method using l moments. *Water Resources Research*, 42(7), 2006. 19, 163

T. R. Kjeldsen and D. Rosbjerg. Comparison of regional index flood estimation procedures based on the extreme value type i distribution. *Stochastic Environmental Research and Risk Assessment*, 16(5):358–373, 2002. 164

T. R. Kjeldsen, D. Jones, A. C. Bayliss, P. Spencer, S. Surendran, L. Stefan, P. Webster, and D. McDonald. Improving the feh statistical method. *In: Flood & Coastal*

Management Conference 2008, University of Manchester, 1-3 July 2008. Environment Agency/Defra., page 10, 2008. Available at UK Natural Environment Research Council Open Research Archive,http://nora.nerc.ac.uk/3545/. 100, 181

TR Kjeldsen, JC Smithers, and RE Schulze. Flood frequency analysis at ungauged sites in the kwazulu-natal province, south africa. *WATER SA*, 27(3):315–324, JUL 2001. ISSN 0378-4738. 14

G. Kuczera. Robust flood frequency models. *Water Resources Research*, 18(2):315–324, 1982. 12

F. Laio. Cramer-von mises and anderson-darling goodness of fit tests for extreme value distributions with unknown parameters. *Water Resources Research*, 40(9), 2004. 56

D. P. Lettenmaier, J. R. Wallis, and E. F. Wood. Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, 23(2):313–323, 1987. 13, 14, 86, 180, 183, 184

M.D. Lowery and J.E. Nash. A comparison of methods of fitting the double exponential distribution. *Journal of Hydrology*, 10(3):259–275, 1970. 18

Alfonso Gutiérrez López. Hydrological forecasting using the double index flood method. *British Hydrological Society*, 1, 2004. 13

L. H. Lu and J. R. Stedinger. Sampling variance of normalized gev pwm quantile estimators and a regional homogeneity test. *Journal of Hydrology*, 138(1-2):223–245, 1992. 16, 18, 129, 162, 163, 168

H. Madsen and D. Rosbjerg. The partial duration series method in regional index-flood modeling. *Water Resources Research*, 33(4):737–746, 1997. 167

N. C. Matalas, J. R. Slack, and J. R. Wallis. Regional skew in search of a parent. *Water Resources Research*, 11(6):815–826, 1975. 11, 18

R Merz and G Bloschl. Flood frequency regionalisation-spatial proximity vs. catchment attributes. *JOURNAL OF HYDROLOGY*, 302(1-4):283–306, FEB 1 2005. ISSN 0022-1694. doi: 10.1016/j.jhydrol.2004.07.018. 4, 14

J.E. Nash and J. Amorocho. The accuracy of the prediction of floods of high return period. *Water Resour. Res.*, 2(2):191–198, 1966. 18

J.E. Nash and B.L. Shaw. Flood frequency as a function of catchment characteristics. *Institution of Civil Engineers, symposium on River Flood Hydrology, (Published 1966)*, pages 115–136, 1965. 179

R. J. Nathan and T. A. McMahon. Identification of homogeneous regions for the purposes of regionalisation. *Journal of Hydrology*, 121(1-4):217–238, 1990. 15

B. Onoz and M. Bayazit. Best-fit distributions of largest available flood samples. *Journal of Hydrology*, 167(1-4):195–208, 1995. 0022-1694 doi: DOI: 10.1016/0022-1694(94)02633-M. 12

OPW. *Estimating Flood Magnitude/ Return Period Relationships and Effect of Catchment Drainage*. Office of Public Works, 1981. 118

TBMJ Ouarda, C Girard, GS Cavadias, and B Bobee. Regional flood frequency estimation with canonical correlation analysis. *JOURNAL OF HYDROLOGY*, 254(1-4): 157–173, DEC 10 2001. ISSN 0022-1694. 14

C. P. Pearson. *New Zealand regional flood frequency analysis using L-moments*, volume 30. New Zealand Hydrological Society, Christchurch, NOUVELLE-ZELANDE, 1991. Journal of hydrology. New Zealand. 15

MURRAY C. PEEL, Q. J. WANG, RICHARD M. VOGEL, and THOMAS A. McMAHON. The utility of l-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, 46(1), 2001. 12

Dan Rosbjerg and Henrik Madsen. Uncertainty measures of regional flood frequency estimators. *Journal of Hydrology*, 167(1-4):209 – 224, 1995. ISSN 0022-1694. doi: DOI:10.1016/0022-1694(94)02624-K. 18, 163, 164

Fabio Rossi, Mauro Fiorentino, and Pasquale Versace. Two-component extreme value distribution for flood frequency analysis. *Water Resour. Res.*, 20(7):847–856, 1984. 0043-1397. 12

C Shu and DH Burn. Homogeneous pooling group delineation for flood frequency analysis using a fuzzy expert system with genetic enhancement. *JOURNAL OF HYDROLOGY*, 291(1-2):132–149, MAY 31 2004. ISSN 0022-1694. doi: {10.1016/j.jhydrol.2003.12.011}. 4

J. R. Stedinger and L. H. Lu. Appraisal of regional and index flood quantile estimators. *Stochastic Hydrology and Hydraulics*, 9(1):49–75, 1995. 14, 18, 86, 163, 180, 183, 186, 190, 191

O. G. B. Sveinsson, J. D. Salas, and D. C. Boes. Uncertainty of quantile estimators using the population index flood method. *Water Resour. Res.*, 39(8):1206–, August 2003. ISSN 0043-1397. URL `http://dx.doi.org/10.1029/2002WR001594`. 13

OGB Sveinsson, DC Boes, and JD Salas. Population index flood method for regional frequency analysis. *WATER RESOURCES RESEARCH*, 37(11):2733–2748, NOV 2001. ISSN 0043-1397. 13

A. Viglione, F. Laio, and P. Claps. A comparison of homogeneity tests for regional frequency analysis. *WATER RESOURCES RESEARCH*, 43, 2007. 17, 130

Vogel, M. R, Fennessey, and M. N. L moment diagrams should replace product moment diagrams. *Water Resour. Res.*, 29(6):8, 1993. Water resources research. 12

R. W. Vogel and D. E. McMartin. Probability plot goodness-of-fit and skewness estimation procedures for the pearson type-3 distribution. *Water Resources Research*, 27(12):3149–3158, 1991. 56

J. R. Wallis, N. C. Matalas, and J. R. Slack. Just a moment! *Water Resour. Res.*, 10 (211-219), 1974. 24

S. E. Wiltshire. Identification of homogeneous regions for flood frequency-analysis. *Journal of Hydrology*, 84(3-4):287–302, 1986. 16

Z. Zrinji and D. H. Burn. Flood frequency-analysis for ungauged sites using a region of influence approach. *Journal of Hydrology*, 153(1-4):1–21, 1994. 4, 86

Z. Zrinji and D. H. Burn. Regional flood frequency with hierarchical region of influence. *Journal of Water Resources Planning and Management-Asce*, 122(4):245–252, 1996. 4

Appendix A

Data sets used in different parts of the study

Annual Maximum flood data from 202 gauging stations have been obtained from the OPW. Based on rating curve and water level measurement reliability, as indicated by OPW, these data fall into the following categories:

- 45 Grade A1 stations
- 70 Grade A2 stations
- 70 Grade B stations and a further
- 16 stations (comprising 8 A1, 5 A2 with the remaining 3 classified as B) that all have pre-and post-drainage records.

While data are available for 202 stations, all data were not used for every part of the study. A summary of the data sets used in different parts of the study are summarised in Table A.1.

A. DATA SETS USED IN DIFFERENT PARTS OF THE STUDY

Table A.1: Data sets used in different components of the study

Component of Study	Stations used	Reasons/Remarks
Ch.3: Descriptive Statistics	186 stations (A1, A2 and B)	
Ch.3: Regional Statistics	110 stations (A1 and A2)	Removal of 5 stations in the Mulkear Catchment reduce the number of 115 A1 and A2 stations to 110 stations
Ch.3: Probability plots	110 stations (A1 and A2)	The same data sets that are used in regional statistics
Ch.3: Diagonostic plots	110 stations (A1 and A2)	The same data sets that are used in regional statistics
Ch.3: Goodness of fit tests	74 stations (A1 and A2)	Stations which have 30 years or more of AM data
Ch.3: Flood seasonality	202 stations (all available stations)	
Ch.4: Pooled analysis	85 stations (A1 and A2)	Stations which have BFI value (The test of discordancy measure reduces the number from initially selected 88 stations to 85 stations)
Ch.4: Temporal effect on growth curve	90 stations(A1 and A2)	The study had already been completed using a total of 90 stations before the data provider indicated some of these are unsuatable for detailed statistical analysis
Ch.5: Examination of homogeneity	85 stations (A1 and A2)	The same data sets that are used in pooled analysis
Ch.6: Standard errors of X_T and Q_T	85 stations (A1 and A2)	The same data sets that are used in pooled analysis

Appendix B

Random number generation using Monte Carlo technique

It is well known that hydrologic events such as the peak flood flow vary from year-to-year in an apparently random and unpredictable fashion. It is assumed that the observed set of values, the sample, is one of the sampling realizations of the parent distribution. That is, there exists some probability density function, from which the observed sample is randomly obtained. If the nature of the parent distribution can be inferred from the properties of the sample, then the distribution provides the complete statistics of the variable under consideration. To specify the distribution, two things are needed. First, an appropriate form of the distribution must be selected from commonly used distribution functions. Second, the parameters of the candidate must be determined.

A random sample drawn from a given distribution using Monte Carlo technique can be obtained by first drawing a random sample from the uniform distribution defined over the range from 0 to 1. The F(x) is set equal to this value, where F is the cumulative distribution to be sampled. The desired value of x is then obtained by inverting the expression for F. Sampling from the uniform distribution is generally done with a random number generator returning values from 0 to 1. Most programming languages have such a built in function. In the following figure the whole procedure is illustrated.

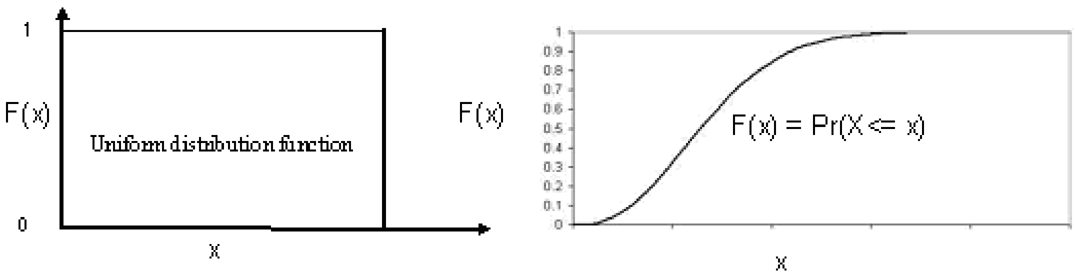


Figure B.1: Generation of random number using Monte Carlo technique. -

Appendix C

Probability plots

See in the attached CD.

Appendix D

Quantile and growth factor estimates for the EV1 and the GEV distributions

EV1

The EV1 distribution is defined by (Hosking and Wallis, 1997)

$$F(Q) = \exp\left[-\exp\left(-\frac{Q-\xi}{\alpha}\right)\right] \tag{D.1}$$

$$EV1 \quad at - site : Q_T = \xi + \alpha y_T \tag{D.2}$$

where

ξ, a location parameter and α, a scale parameter are estimated by the first two L-moments

λ_1 and λ_2 (Hosking and Wallis, 1997) . $\lambda_1 = \xi + \alpha\gamma$

$\lambda_2 = \alpha \ln 2$

$y_T = -\ln(-\ln(1 - 1/T))$ is the EV1 reduced variate for a T-year return period.

$$EV1 \quad growth : X_T = 1 + \beta(y_T - y_2) \tag{D.3}$$

where

$\beta = \frac{t_2}{\ln 2 - t_2[\gamma - y_2]}$ in which, $t_2 = \frac{\lambda_2}{\lambda_1}$ is the L-CV and γ is Euler's constant = 0.5772.

GEV

The GEV distribution is defined by (Hosking and Wallis, 1997)

$$F(Q) = \exp\left\{-(1-k(Q-\xi)/\alpha)^{1/k}\right\} \quad for \quad k \neq 0 \tag{D.4}$$

$$GEV \quad at-site: Q_T = \xi + (\alpha/k)\left\{1-\left(-\log\left(\frac{T-1}{T}\right)\right)^k\right\} \tag{D.5}$$

where

ξ is a location parameter , α is a scale parameter and k is a shape parameter.

$\alpha = \frac{\lambda_2 k}{\left(1-2^{-k}\right)\Gamma(1+k)}$

$\xi = \lambda_1 - \alpha\{1-\Gamma(1+k)\}/k$

$\kappa = 7.8590c + 2.9554c^2$ in which, $c = \frac{2}{3+t_3} - \frac{\ln 2}{\ln 3}$

$$GEV \quad growth: X_T = 1 + \frac{\beta}{k}\left((\ln 2)^k - \left(\quad \ln\frac{T}{T-1}\right)^k\right) \tag{D.6}$$

where

$k = 7.8590c + 2.9554c^2$in which, $\quad c = \frac{2}{3+t_3} - \frac{\ln 2}{\ln 3}$

$\beta = \frac{kt_2}{t_2\left(\Gamma(1+k)-(\ln 2)^k\right)+\Gamma(1+k)\left(1-2^{-k}\right)}$ in which t_2 is the sample L-CV and t_3 is the sample L-skewness.

Γ denotes the complete gamma function.