REGIONAL FLOOD FREQUENCY ANALYSIS FOR ACEH PROVINCE INDONESIA

ABDUL HANAN AKHMAD

REGIONAL FLOOD FREQUENCY ANALYSIS
FOR ACEH PROVINCE
INDONESIA

By

©ABDUL HANAN AKHMAD

A Thesis Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements for

the Degree of Master of Engineering

FACULTY OF ENGINEERING AND APPLIED SCIENCE

MEMORIAL UNIVERSITY OF NEWFOUNDLAND

OCTOBER, 1993

ST JOHN'S NEWFOUNDLAND CANADA

ISBN 0-315-86679-9

Abstract

This thesis discusses the development and application of two principal methods of regional flood frequency estimation for the rivers in Aceh Province, Indonesia.

The first method is based on index-flood approaches and the second method is the regression of flood quantiles on basin characteristics approach.

For the index-flood approaches, the annual maximum floods at each site were standardized by the mean annual flood of the site. Five different index-flood methods were used to derive the regional frequency curve. These are Dalrymple's method, NERC method, the probability weighted moments method, the station year method and the L-moment method.

For the multiple regression approach, the logarithm of each selected quantile of the annual maximum floods at each site of the nine river basins was regressed on its corresponding catchment variables. The least squares method was used to develop the multiple regression equations. The derived regression equations were corrected for bias caused by the logarithmic transformation.

The flood estimates obtained using the two regional methods were compared to at-site estimates and to those obtained from a previous study by the Institute of Hydrology (IOH), Wallingford (1983) for basins in Jawa and Sumatra. The estimates obtained by the various methods were also compared to each other for basins that were not used in the study to show the variability of the results.

From a comparison of the two regional methods, the findings showed that all index flood approaches provided greater consistency and similarity in the estimation of flood quantile magnitudes. The L-moment method seemed to give the best compromise estimates, while the regression approach gave lower estimates for most stations and estimates were sometimes inconsistent. The estimates based on the IOH study gave estimates well above these obtained in this study especially for higher return period floods.

Acknowledgments

I would like to express my most sincere gratitude and appreciation to my supervisor Dr. L. Lye for his guidance and advice throughout the research and the preparation of this thesis.

I would also like to express my sincere thanks to Dr. A. Robertson and Ms. Susan Richter for many valuable discussions and useful suggestions. I also wish to thank Ms. Seana Kozar for her assistance in preparing this thesis.

Help from the technical staff at the Faculty of Engineering and Applied Science is also gratefully acknowledged. The assistance of fellow graduate students is most appreciated.

I would also like to take this opportunity to acknowledge the Indonesian Government and the Canadian International Development Agency for their financial support during the study.

Finally, I would like to express my sincere thanks and appreciation to my wife Marlinda and sons Muzie Hikmah and Zaid Ramadhan for their support and encouragement during my study.

Contents

List of Figures

List of Tables

Abbreviations

ACF	Auto Correlation Function
ADU	Aceh Design Unit
EV1, EV2, EV3	Extreme Value Type 1, 2, 3 (distribution)
GEV	General Extreme Value (distribution)
GL	Generalized Logistic (distribution)
GPA	Generalized Pareto (distribution)
IOH	Institute of Hydrology, Wallingford, U.K.
LLG	Log-Logistic (distribution)
LN2	Two parameter Lognormal (distribution)
LN3	Three parameter Lognormal (distribution)
LS	Least Squares
MAF	Mean Annual Flood
MAR	Mean Annual Runoff
MML	Method of Maximum Likelihood

MOM	Method of Moments
NERC	Natural Environment Research Council, U.K.
P3	Pearson Type 3 (distribution)
PMW	Probability Weighted Moment
SIMS	Channel Slope
TcEV	Two-component Extreme Value (distribution)
USWRC	United States Water Resources Council
WMO	World Meteorological Organization

Chapter 1

Introduction

1.1. General Description

The special district of Aceh is a province of Indonesia and one of the eight provinces on Sumatra, as shown in Figure 1.1. In 1983, the Directorate of Planning and Programming, a branch of the Directorate General of Water Resources Development, established the Aceh Design Unit (ADU) with technical assistance from the British Government (Binnie and Partners, Consultants). The prime task of the ADU was to compile an inventory of water resources and to produce a water resources development plan. The plan, presented both for short and long term development, was followed by implementation and covered not only irrigation, but also aquaculture, swamp reclamation and drainage, river improvement and flood control, municipal and industrial water use, and hydro-electric/multi-purpose projects. Under this plan 15 Water Resources Development Areas (WRDA's) were set up, each of which consists of one or more adjacent river basins, as shown in Figure 1.2.

In the interest of protecting these development areas from flood damage, two of the major rivers in Aceh with flood problems have their own special project. These are the Aceh River Urgent Flood Control Project, and the Arakundo-Jambu Aye Irrigation and Flood Control Project, both started in 1981. Flooding is also a problem for most of the other rivers in Aceh. The River whose contour lies within the 25 metre altitude and areas around the river mouth for example, are of significance in terms of flood risk and erosion because these are densely populated areas.

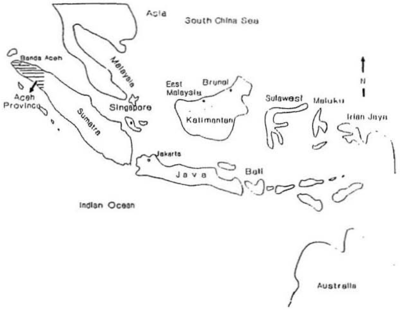

Figure 1.1 Location of Aceh Province, Indonesia.

The ADU was required to estimate flood flows in each of the WRDA's. To obtain these estimates, the ADU used a flood design manual for Java and Sumatra (published in 1983 by the Institute of Hydrology (IOH) Wallingford, UK; and the Institute of Hydraulic Engineering (DPMA), Indonesia). In general, this manual provides estimates of mean annual flood (MAF) based on different types of data and an average flood frequency growth curve based on the Gumbel distribution. The flood estimation procedure is simple to apply; the MAF is estimated, and the result is then multiplied by a growth factor related to the required return period. The resulting value is the flood quantile magnitude.

The manual however, used the Gumbel distribution for all river basins within Jawa and Sumatra, and includes only five sites within Aceh Province. Of those, only one watershed had a record length of five or more years. These five sites were Aceh River at Kampung Darang, Peusangan River at Beukah, Jambu Aye River at Rampah and at Lhok Nibong, and Susoh River at Kota Tinggi. Their record lengths were four, three, four, eight, and three years duration respectively.

Since the publication of the above mentioned manual, the data base of flood flows has increased. At present, the number of watersheds with five or more years of recorded flow has increased from one to nine. In addition to the streamflow data, more information on land use, topography, and other physiographic characteristics of the watersheds has been assembled.

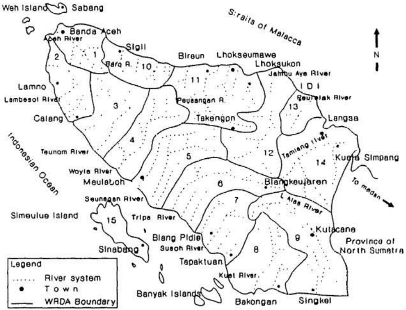

Figure 1.2 Aceh province principal rivers and WRDA's

Engineers preparing designs for the WRDA's recognize the limitations of the flood design manual, and are interested in a more accurate flood estimation procedure especially for ungauged basins. They are often required to estimate the magnitude and frequency of flood events from relatively short records often containing non-representative data. Also, there are occasionally no available records. The accuracy of the estimation procedure is important because it affects both costs and safety standards of the engineering designs.

An alternative method of estimating floods at an ungauged site or a site with a very short record is to use observed data at sites within a homogeneous hydrologic region. This method of analysis is called regional flood frequency analysis.

1.2. Objective of the Thesis

The objective of this research is to use flood flow data from the nine representative basins within the 15 WRDA's, as well as physiographic and climatic data, to develop a method of regional flood frequency analysis for the province.

Two principal regional analysis methods of flood frequency are considered in this thesis:

1) Index flood methods: these methods provide a regional flood frequency curve or table which is usually normalized by the MAF. Five estimation approaches were used: Dalrymple's method, Natural Environment Research Council (NERC, 1975) method, the Probability Weighted Moments (PWMs) method, the station year approach, and the L-moments method.

2) Multiple regional regression: this method is based on at-site flood quantiles (each Q_T) and/or the MAF at each gauged stream as the dependent variable, and catchment characteristics upstream of each corresponding gauged stream as the independent variables.

These two regional approaches of flood frequency along with at-site flood frequency analysis and the previous study (IOH, 1983) are compared.

1.3 Thesis Outline

This chapter provides an introduction to the flooding problem for the Aceh province rivers. A description of the study area is given in the next chapter. Chapter 3 summarizes the current body of literature relevant to regional flood frequency development for the present study. The preliminary procedures for at-site and regional flood frequency analyses are discussed in Chapter 4. Chapter 5 contains the development and application of the regional estimation techniques. Conclusions and recommendations are presented in Chapter 6. Appendices are provided immediately after the References.

Chapter 2

Description of the Study Area

2.1. General Description

The basin area considered in this study is located in Aceh Province, Indonesia. The region lies at the northern end of Sumatra. The climate is tropical. The total land area of 55,852 km^2 includes Weh Island in the North, and Simeulue Island and the Banyak group of islands on the south west coast. The region is mountainous, particularly the Barisan mountain range. The highest peak is about 3000 m to 3400 m above mean sea level. By contrast, the elevations of the rivers at the gauged locations lie within a range of 5 m to 430 m above mean sea level. The area and population distribution, based on 1980 data (Binnie and Partners, 1988), is shown in Table 2.1. The location of the basins within the WRDA is shown in Figure 1.2. On average, the population density of the region is about 48.3 people per square kilometre, and population growth rate is about 1.8% per annum.

Table 2.1 Area and population of WRDAs in Aceh Province

WRDA	Area (km^2)	Rural population	Urban population	Total population	population density (person/km^2)
1	2,574	161,183	142,818	304,001	123.9
2	2,208	70,713	-	70,713	32.0
3	3,321	49,169	-	49,169	14.8
4	2,447	41,235	-	41,235	16.9
5	3,846	103,718	14,847	118,565	30.8
6	4,025	40,934	-	40,934	10.2
7	2,145	152,796	4,341	157,137	73.3
8	4,060	61,695	-	61,695	15.2
9	7,557	187,111	4,334	191,445	25.3
10	2,781	361,755	5,950	367,705	132.2
11	5,465	549,842	42,358	592,200	108.4
12	5,150	132,718	-	132,718	25.8
13	2,664	138,021	-	138,021	51.8
14	5,630	226,621	19,351	245,972	43.7
15	1,979	39,216	-	39,216	19.8
Total	55,852	2,316,717	233,999	2,550,726	-

Land use in Aceh Province can be grouped into six main classes: villages (1.5%), rice fields (4.5%), estate crops (4.7%), mixed garden crops (2.5%), forest (78.1%), and open area such as bush, lakes, swamps and waste land (8.7%). Presently, indications are that the level of employment in agriculture will remain fairly stable, but may be somewhat offset by an annual decrease in numbers, as people become interested in occupations other than farming. Since the publication of the original provincial plan, the local government has viewed the development plan in terms of an industrial zone, defined as the more populous district along the north and east coasts, and an agricultural zone, including the remaining areas in the middle, western and southern regions of the province.

2.2 Location and Drainage Basins Used

The drainage basins used in the analysis lie within the WRDA plan. It should be noted that each WRDA consists of one or more river basins. The inclusion of a given basin within the WRDA was based on the availability of flood flow information and other physiographic characteristics. Nine river basins which have 5 or more years flood flow data are:

- Aceh River at Kampung Darang (KD),
- Lambeusoi River at Sango (SG),
- Seunegan River at Ujung Blang (UB),
- Kluet River at Gunung Pudung (GP),
- Lawe Alas River at Sukarimbun (SR),
- Baro River at Klibeut (KL),
- Peusangan River at Beukah (BK),
- Jambu Aye River at Lhok Nibong (LN), and
- Tamiang River at Kuala Simpang (KS).

The locations of the drainage basins and the streamflow gauges used are shown in Figure 2.1, and the basin characteristics and lengths of record are shown in Table 2.2.

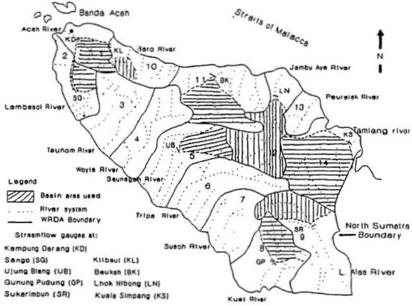

Figure 2.1 Locations of basins and streamflow gauges used

Table 2.2 The drainage basins characteristics and recorded flow duration

WRDA	River basin in WRDA / Station location	Area (km^2)	Mean stream length (km)	Main channel slope (m/m)	Site elevation (m)	Recorded flood (year)
1	Aceh at Kp.Darang	1078.0	66.04	0.0175	22	11
2	Lambeusoi at Sango	580.4	33.20	0.0614	12	12
3	Teunom at T. Kareung	2236.0	112.00	0.0154	25	2
4	Woyla at Mangi Tukut	2327.0	105.00	0.0150	21	4
5	Seunegan at U. Blang	352.0	50.80	0.0375	93	7
6	Tripa at Gunung Kong	2707.0	157.00	0.0112	26	3
7	Susoh at Kuta tinggi	193.8	24.00	0.0406	48	4
8	Kluet at G. Pudung	460.0	100.80	0.0146	22	6
9	L. Alas at Sukarimbun	1384.0	96.52	0.0160	430	8
10	Baro at Kibeut	270.0	35.60	0.0276	19	11
11	Peusangan at Beukah	2214.0	137.80	0.0169	15	13
12	Jambu aye at L.Nibong	4420.0	231.20	0.0084	5	16
13	Peurelak at G.Janeng	1050.0	91.50	0.0107	25	0
14	Tamiang at K. Simpang	4598.0	177.80	0.0140	17	7

2.3 Sources of Data

Data concerning the flood flow, topographic, demographic and physiographic characteristics were provided by the following sources:

a) Flood flow data were collected from the Aceh-Water Resources Development Services (Aceh-WRDS), in Banda Aceh. Recent data were provided by the Institute of Hydraulic Engineering in Bandung.

b) The topographic maps were obtained from the Aceh-WRDS in Banda Aceh.

Data on demographic and physiographic characteristics (e.g., population, land use, forest, basin slope, soil and isohyetal rainfall maps) were provided by the Aceh Regional Planning Board in Banda Aceh. The data were taken directly from the maps, and were compared with corresponding information available from the IOH (1983) or ADU reports.

c) Some climate information was provided by the Institute of Meteorology and Geophysics in Jakarta.

2.4 Regional Hydrology and Climate

Aceh Province is situated in the tropical zone. Characteristics of this region are significantly influenced by tropical hydrology and climate. Rainfall distribution in this region is controlled by three main factors: the monsoon systems of South-East Asia and Australia, equatorial double rainy seasons, and local topographic influences (Wild and Hall, 1982).

The east and west monsoons are characterized by distinct seasonal changes in wind direction. The west monsoon approaches during November, as the boundary between the declining and the progressing air masses moves slowly but inconsistently southwards. By December, the west monsoon is normally established and remains dominant through January and February. Moisture drawn up over the South China Sea produces heavy rainfall over Peninsular Malaysia but only moderate rainfall over Aceh Province. March to May is an inter-monsoon transition period. This is followed by the east monsoon which lasts from June to August. It carries dry air produced during the winter in high-pressure zones over Australia. Some moisture picked up from the Indian Ocean falls mainly on the south-west slopes of the Barisan mountain range and the west coastal plain. A second transition period follows until the west monsoon is reestablished by December.

The equatorial double rainy seasons are caused by regions of above average temperature which is followed by a short lag due to the annual variation of the sun's decline. In this region, the sun reaches its zenith in approximately late March and mid September. Two rainfall patterns exist with the heaviest period from October to November. The period from March to May is also wet. The driest period occurs in February (especially on the east coast) and in June and July throughout the province.

The local topography clearly influences the rainfall distribution. This is because of the interaction of the predominant monsoon and the Barisan mountain range. Annual rainfall decreases towards the north coast, reducing in some areas to as little as 1200 mm. By comparison, the west coast is considerably wetter, with typically 3500 mm of rainfall annually, rising to 4500 mm and 5000 mm in the nearby mountains (Binnie and Partners 1988).

Mean monthly temperature at Banda Aceh for example, varies by only 7.3% during the year, having a maximum of 27.4° C in June and a minimum of 25.4° C in December. Mean annual temperature varies with elevation, decreasing from about 26° C at sea level by roughly 0.52° C per 100 metre rise in elevation (Binnie and Partners, 1988).

Mean annual relative humidity falls typically in the range of 80% to 90%. There is no evidence of any correlation between relative humidity and elevation from sea level. The variation of the monthly means is nearly 5% and diurnal range is from 60% to 100%.

Sunshine duration is highly variable both spatially and seasonally. Altitude is expected to influence sunshine duration markedly, with regions of persistent cloud cover occurring in foothills of the west coast and in the valley sides in the central inter-mountain basins. Sunshine duration is traditionally measured from 0800 hour to 1600 hour local time. Mean annual sunshine duration is about 44% of maximum possible, while mean monthly sunshine varies by up to about 15% from the mean annual value.

Wind velocities are generally light through the year with little seasonal variation. In Banda Aceh, wind speed at 2 m above ground varies from 5.0 to 6.6 km/hour. Mean annual wind speed over the whole province varies from 1.8 to 5.6 km/hour.

Potential evapotranspiration for the province is approximately the same as in North Sumatra (as indicated by Wild and Hall, 1982). Potential evapotranspiration for a short green crop (an index albedo 25) varies from around 1250 mm/year on the east coastal plain to around 1430 mm/year on the west coastal plain. Annual potential evapotranspiration reduces by roughly 100 mm for each 500 m rise in elevation.

2.5 Physiography and Topography

Aceh province is mountainous, particularly in the Barisan mountain range, in which the highest peak is about 3000 to 3400 m. The region consists of upper Palaeozoic and Mesozoic sedimentary rocks with granitic intrusions (Binnie and Partners, 1988). There are three volcanoes which have shown signs of recent activity. These volcanoes lie on a fault which is particularly visible in the Aceh River valley, in the upper catchments of Teunom and Meurebo Rivers, and in the valley of Lawe Alas around Kutacane.

Coastal plains are generally wider locally at the mouths of rivers, but are narrow (< 4 km) toward the north and east coasts. Tidal swamps occur at the mouths of the principal rivers on the east coast. On the west coast, the coastal plain is generally wider (up to 20 m), and shows active aggradation of sediments from coastal and river sources. Some places on the west coast have no coastal plain.

In terms of river systems, all rivers (except the Lawe Alas River in the central rift valley) follow a normal course to the sea on either side of the central mountains of the Barisan Mountain range. Most rivers are characterized by steep boulder-strewn upper catchments with dense primary forest cover, which flatten into braided channels, then meander in their lower reaches as they emerge from the foothills onto the coastal plains.

Chapter 3

Flood Frequency Analysis

Flood frequency analyses have been widely used to obtain estimates of flood quantile magnitudes, Q_T, so as to provide a reliable decision-making tool for hydraulic works or flood alleviation programmes. Many procedures have been developed for flood frequency analysis purposes. Using these procedures, the flood quantile at a particular site in a river can be estimated from data that are specific to the site and procedure.

The method used for estimating the magnitude of flood quantiles depends on the availability of data, and on the form of the distribution and the estimation procedure used. Three methods of flood frequency analysis can be distinguished according to the amount and type of data used. The first method, based on single station analysis, uses only at-site data, and is applicable if the gauged catchments have long periods of recorded streamflow. The second method is based on a combination of at-site data and neighbouring gauged catchments and can be used in a regional context. The third method is based on a regression approach; it uses only regional data and can be applied in a regional context to develop flow characteristics which are transferred from gauged catchments to ungauged catchments. The latter two methods are applicable to hydrologically homogeneous regions. For both of these methods, there are typically five stages involved in the frequency analysis. These are:

- data collection and analysis,
- single station analysis,
- choice of a regional frequency distribution,
- delineation of homogeneous regions, and
- estimation of regional flood quantiles.

3.1. Data Collection and Analysis

Estimating the magnitude of flood quantiles is dependent upon the availability of data. The required data for a region include flood flows and catchment characteristics within the region. The flood data may be evaluated by the type of flood data and the sample properties. For a given region, the catchment characteristics are mainly divided into two groups the physiographic characteristics specific to the catchment and the climatic characteristics over the catchment. These characteristics are evaluated for each catchment.

3.1.1 Types of Flood Data

As summarized by Beable and McKerchar (1982), there are three types of flood samples: annual series, peaks over threshold series, and historical series. The most common type is the annual series, which contains the flood peak for each year of record.

This type of sampling provides data points from separate flood events. A disadvantage of this type of sample is that it may emphasize small flood peaks and neglect larger ones.

An alternative type of sample is the partial duration series, which appears to overcome the disadvantage of the annual series. This kind of sample consists of all flood peaks above a chosen base level. Sample points need to be checked more often for their serial correlation than an annual series, since large floods often contain more than one peak above the base level. According to Chow (1964, pp. 8-23), the partial duration series only has an advantage over the annual series when predictions of flood flow are required for return periods of less than 10 years.

Additional historical information on water levels during flood events is frequently available. When this information is reliable, it should be used in combination with the data sample taken from the continuous streamflow record. The inclusion of additional historical information often significantly increases the length of a sample. Therefore it is more likely that the flood frequency analysis will be improved. The information may also aid in establishing the upper threshold of the frequency curve.

In this thesis, the primary data samples were annual series, which were the maximum flood flows for each year of record for each station. It was not necessary to construct a partial duration series for this study because the main interest is in floods with return periods greater than 10 years.

3.1.2 Sample Properties

The data sample should have the following properties if the analysis is intended for generalization over the population distribution (Beable and McKerchar, 1982): sufficient length, completeness, degree of homogeneity, randomness and reliability, and representativeness.

Sample Length: Dalrymple (1960) indicated that a flood flow should have a record that is five or more years long, although it is commonly accepted that quantiles should not be determined for records shorter than 10 years. According to Beable and McKerchar (1982), 10 annual flood peak items may be used, since data samples of about 10 years in length are often the only ones available. However IOH (1983), using reliable data of 4 or more years in length was able to estimate mean annual flood and then to use that measure to derive flood frequency growth factors for Jawa and Sumatra.

Completeness: The data sample used should be taken from a continuous streamflow record. Gaps in the daily record are not important, as long as the maximum flood peak is not missing. However, the time units which include the gaps should comprise only a small portion of the total sample length.

Homogeneity: Data samples should be considered in terms of their degree of homogeneity. This means that the data samples have occurred under the same conditions. Factors affecting the homogeneity of a sample include: human activity, faulty records, and changes in the gauging control conditions.

Randomness and Reliability: When data are assumed to be independent and random, there should be no serial correlation between the consecutive flood peak items. Therefore, the samples should also be taken from reliable measuring instruments.

Representativeness: Data should also be representative of the population distribution of data points. This may be difficult to determine because the population is unknown. However, if there is a long-term streamflow record for a similar catchment nearby, the representativeness can be tested statistically (McGuinness and Brakensiek 1964).

3.1.3 Catchment Characteristics

The catchment characteristics can be divided into two groups; those describing the physical catchment, and those depicting the climate over the catchment. The physical characteristics include the size and shape of the catchment, the stream channel, and the hydraulic properties of the soil and the vegetation. Climatic information includes mean annual rainfall, and rainfall intensity of one day rainfall at a given return period.

The catchment characteristics which may influence the mean annual floods or flood quantiles for a given return period have been described by NERC (1975) and used by Beable and McKerchar (1982), and IOH (1983). The relevant characteristics are:

- catchment area,
- main channel length,
- main channel slope,
- mean catchment elevation,
- stream frequency,
- percentage of the catchment that is forested,
- mean annual rainfall over catchment, and
- the rainfall intensity of a given return period.

In terms of rainfall intensity, Beable and McKerchar (1982) used the one day rainfall with a return period of two years, while IOH (1983) used the mean annual maximum one day rainfall.

For the purposes of this thesis, the procedure for estimating the physical and climatic characteristics combined Beable and McKerchar's approach with that of IOH, with certain necessary adjustments. This procedure is summarized and presented in Chapter 4.

3.2 Single Station Analysis

Single station analysis is required to provide data for developing regional equations (suitable for ungauged sites) as well as for providing estimates for gauged rivers. The annual series of flood flows for each station are assumed to represent a random sample from a population of flood values whose distribution can be defined by a probability density function which is dependent on a few parameters. Frequency distributions having two or three parameters are usually used in the analysis. These parameters are related to the location (mean), scale (standard deviation) and shape (skewness coefficient) of the distribution. The methods for estimating the parameters are further discussed in the next section.

3.2.1 Plotting Position

Frequency analysis requires the identification of the probability or the return period of each sample point. Various formulae are available for estimating these probabilities, called plotting positions because they are used to prepare probability plots. The Weibull formula is commonly used because of its simplicity. In a comprehensive review of plotting positions, however, Cunnane (1978) noted that the Weibull formula provides biased plotting positions, which on average lead to an over estimation of flood quantiles for high return periods.

Cunnane (1978) also concluded that the selection of a plotting position formula for a sample depends on the assumed distribution to fit the sample. For example, for samples fitting the Type I extreme value (EV1) distribution, the Gringorten formula provides unbiased plotting positions. For unbiased plotting positions where a single simple formula is required for use with all distributions, Cunnane's formula provides a good compromise. This formula is given as:

$$F_i = \frac{(i-2/5)}{(n+1/5)} \tag{3.1}$$

where F_i is the probability plotting position for the i_{th} smallest of n observations.

This plotting position formula is used in the package program, Consolidated Frequency Analysis version 3 (CFA3, 1991) by the Hydrology Division of the Water Resources Branch of Environment Canada.

3.2.2 Estimation of Parameters

Many methods have been developed to obtain an estimate of distribution parameters. Kite (1988), described four parameter estimation techniques. These may be listed in ascending order of efficiency as: graphical, Least Squares (LS), Method Of

Moments (MOM) and the Method of Maximum Likelihood (MML). The MML is somewhat more difficult to apply in practice because it often requires a computer to perform iterative calculation.

WMO (1989), listed five methods of parameter estimation. These include MOM, MML, LS, Probability Weighted Moments (PWM) (Greenwood et.al., 1979), and Sextiles (Jenkinson, 1969). WMO concluded that while the MOM is easy to apply, it is not as efficient as the MML, especially in three parameter distributions. The PWM method has good statistical proper:'es for distribution that can be explicitly expressed in inverse form.

Bobee et.al., (1993) compared five methods of parameter estimations. These are: the MML, the MOM, the method of mixed moments or generalized method of moments, the PWM method, and the L-moments method. They concluded that the MML is optimally unbiased and displays minimum variance, but might result in bad estimates in small samples. The MOM is widely used because of its simplicity, as it is based on the mean, variance and skewness coefficient. The PWM method, which is based on linear combinations of order statistics, has gained wide popularity and has been used in many recent studies. The L-moment method which is a linear combination of PWMs as suggested by Hosking (1990) provides greater clarity in a statistical interpretation than the PWMs.

3.3 Choice of a Frequency Distribution

Many theoretical distributions have been developed for frequency analysis purposes. Kite (1988), described some of the frequency analysis techniques for estimating floods and drought. Kite suggested some commonly used distributions in hydrology e.g., normal, two parameter lognormal (LN2). three parameter lognormal (LN3), Type I extreme value (Gumbel), Pearson Type III (P3), log-Pearson Type III (LP3) and Type III extreme value (EV3).

WMO (1989) has suggested fourteen candidate distributions for use with annual maxima. These include: lognormal, P3, Gumbel, Type II extreme value, EV3, Gamma, LP3, General Extreme Value (GEV), Weibull, Wakeby, Boughton, Two Component Extreme Value (TCEV), Log-Logistic (LLG) and General Logistic (GL). WMO (1984, 1989), however, reports that the selection of distributions is often not chosen in any objective manner, but picking one from distributions that are widely accepted.

A study by Haktanir (1992), compared various flood frequency distributions in Anatolia, Turkey. Haktanir applied the LN2, LN3, smemax, two-step-power, Log-Boughton, Gumbel, P3, LP3, LLG and Wakeby distributions to annual series of flood flows ($\geq$ 30 observations), of 45 unregulated rivers in Anatolia. He found that the LN3, LN2, and Gumbel distributions predicted extreme right-tail events better than the other distributions.

A recent study by Vogel et.al. (1993), uses L-moment diagrams for evaluating data at 61 sites across Australia. They showed that the Generalized Pareto (GPA), LP3, LN3, GEV and Wakeby distributions are all adequate approximations to the distribution of annual flood flows in Australia.

The above discussion shows that hydrologists have reached different conclusions concerning the best method for parameter estimation and the best probability distribution to represent flood flows. The most frequently used methods for estimating parameters, however, are the MOM, the MML, and the method of PWMs. The LN2, LN3, GEV, Gumbel, P3 and LP3 are the distributions which have been most widely used in many flood frequency studies throughout the world.

For this study, the methods of parameter estimation that were used were the MOM and the MML for at-site flood frequency analysis. These were chosen because they are readily available in package programs. Five distributions were tried to model the annual maximum floods of the nine rivers in the province. These distributions were: LN2, LN3, Gumbel, GEV and LP3. To assist in the selection of the appropriate frequency distribution, the L-moment diagram developed by Hosking (1986) that compares sample estimates of the L-moment ratios with their population ratios was also used for evaluating the suitability of the various alternative distributions for modelling flood flows in the region.

3.4 Delineation of Homogeneous Regions

In regional frequency analysis, many studies have been carried out to identify homogeneous regions. There are three principal methods commonly used to delineate a study region into homogeneous regions. The first is based on geographical or administratively defined regions such as provincial, national, rivers, valley, latitude/longitude boundaries. The second is based on the similarity of physiographic and climate characteristics such as geology, land use, drainage characteristics and rainfall/runoff similarity. The third is based on flood characteristics such as the homogeneity test of Dalrymple (1960), similarity of the coefficient of variation (C_v) test within the region (Cunnane, 1987) or a heterogeneity measure based on L-moments (Hosking, 1993). In practice, there may be considerable overlap among methods, and boundaries may change as more information becomes available.

Aceh Province is a relatively small area in which to try to delineate the river basins into various homogeneous regions. Geographically, the basins within the province have similar characteristics and show similar physiographic and climatic features, with the exception of one of the basins on the west coast of the region which has a relatively high rainfall magnitude, a small catchment area and a steep channel slope. As well, identification of the region based on flood statistics may not give a clear answer due to the lack of long periods of flow records which are required to perform homogeneity tests of flood statistics. Therefore, for practical reasons, it is assumed that the region is homogeneous in terms of flood and physiographic characteristics. As additional flood flow information become available in future, perhaps this question can be more clearly answered.

3.5 Estimation of Regional Flood Quantiles

The regional method of flood quantile estimation uses a combination of at-site data and data from neighbouring gauged catchments. The concept behind a regional flood frequency analysis is that the regions are reasonably homogeneous in terms of climate, topography and physiographic characteristics, and the various catchments display flood frequency properties. The regional method usually produces a regional flood frequency curve or table. The curve is assumed to be generally applicable to catchments in the region and may be applied to both gauged and ungauged catchments. Because the regional frequency curve is based on pooling records from the region, it provides a more reliable way of estimating flood quantile than a single frequency curve fitted to a relatively short record at a site.

Many procedures have been suggested for estimating regional flood quantiles. Two principal methods are the index flood methods and the regression method. Both methods can be used for estimating flood quantiles at gauged and ungauged sites.

3.5.1 Index Flood Approaches

The earliest method of regional frequency analysis is the index flood method. This method is popular in many parts of the world because it is easy to understand and to apply and has given good results. The regional procedure estimates the distribution of a dimensionless flood variate X_T, where $X_T = Q_T / \bar{Q}$. $\bar{Q}$ is called the index flood and is usually taken as the mean annual flood at each at-site sample estimate. Assuming that the resulting variate X_T has the same form of distribution at every site, the parameters of the distribution of X_T are obtained from the combined regional data sets.

Many procedures have been developed to obtain the parameters of the X_T distribution. Six main regional index flood procedures are discussed in this section: the original method of Dalrymple (1960), regional dimensionless moments method (Nash and Shaw, 1965 or US WRC, 1976, 1977, 1981), NERC method (1975), regional PWMs method (Wallis, 1980), station year approach, and L-moments method (Hosking and Wallis, 1992 and 1993).

The first approach was initiated by the US Geology Survey (Dalrymple, 1960). The approach is based on a dimensionless regional averaging of equal length records from unregulated rivers within a given area which have been previously tested for homogeneity at the level of a ten year return period. The homogeneity test is based on the assumption that the EV1 distribution underlies the flood population, although there is no assurance that annual series of flood flows are EV1 distributed. A regional flood frequency curve is obtained by plotting the median or mean peak flow ratios at given return periods against frequencies on EV1 probability paper.

Nash and Shaw (1965) introduced regional averaging of dimensionless moments. The regional average values of the coefficient of variation and the coefficient of skewness are used. They can be used to estimate parameters of any two or three parameter distribution for $X_T = Q / \bar{Q}$. A variation of this approach is to estimate the moments and skewness of the logarithms of the at-site data and adopt a regional average value of the skewness or a weighted value of the regional and at-site estimate of the skewness (USWRC, 1976, 1977, 1981). The USWRC method assumes that all flood series in the region are distributed as an LP3 distribution and quantile estimation is obtained by moment estimation in the log-domain. However, there is also no guarantee that annual maximum floods are LP3 distributed. The use of LP3 is actually based on suggestions for institutional uniformity recommended by US Federal agencies (Benson, 1968).

The index flood method was also used by NERC (1975) for British and Irish conditions and later was implemented in New Zealand (Beable and McKerchar, 1982). The NERC method modifies Dalrymple's (1960) method by using regionally averaged standardised order statistics in a graphical procedure to estimate the X_T distribution. This averaging is intended to reduce the effect of outliers on the regional curve fitting method due to the possibility of sampling variation of each individual streamflow record for a region.

Wallis (1980) suggested an objective numerical method, based on regionally averaged standardised PWMs (Probability Weighted Moments). This method has been found to be easy to apply and is efficient for estimating flood quantiles. The PWM procedures were introduced by Greenwood et. al. (1979) and further analyzed by Hosking (1986). PWMs can be used with all distributions that can be defined in inverse form such as the uniform, exponential, EV1, Logistic, Normal, Raleigh, GPA, GEV, GL, Lognormal, Gamma, Generalized Lambda and Wakeby distributions (see also Hosking, 1986 and WMO, 1989). Thus, this method rules out data are distributed as LP3.

Cunnane (1987, 1988) included the station year method in his review of statistical flood quantile analysis. The method is based on pooling all standardised data values within a region and treating them as a single sample from a population for parameter estimation purposes. This approach is referred to as a regional pooling of data. In contrast, the other methods involve a regional averaging of data or statistics of those data.

Recent research on regional frequency analysis by Hosking and Wallis (1992, 1993) describes an index flood procedure based on L-moments. The method is based on the linear combinations of the PWMs but possesses greater clarity for statistical interpretation (Hosking, 1986). The regional analysis based on the L-moment method includes identifying unusual sites in a region, assessing a homogeneous region, and assessing a candidate distribution of adequate fit to the data. These analyses can efficiently use L-moment statistics of the at-site data (Hosking, 1990).

In addition to the index flood approach, other more complicated and less popular approaches include the Two Component Extreme Value (TCEV) and Bayesian approaches (Cunnane, 1987).

3.5.2 Logarithmic Regression for Each Q_T Approach

This approach has been used in many flood frequency studies in United States (Roskie, 1978, cited by Tasker and Moss, 1979; Tasker, 1987), and has also been used by the Newfoundland Water Resources Division (Beersing, 1990). This approach is based on estimating Q_T separately at each site in the region from a given distribution for a selection of T values such as 2, 5, 10, 20, 50 and 100 years. The logarithmic regression relation is established between Q_T for each value of T and the catchment characteristics. The flood quantiles at gauged and ungauged catchments can then be estimated using these regression equations with the relevant catchment characteristics.

3.6 Study Approach

In this thesis, two principal regional flood frequency estimation approaches were adopted, the index flood approach and the logarithmic regression for each Q_T approach. These two approaches were chosen because they are the most popular procedure used in many countries, and the short records available did not justify the use of more complicated techniques. Five index flood procedures were used in this study, including Dalrymple (1960), NERC (1975), PWMs, the station years and the L-moments, and the results were compared. In addition to the comparison study, the method of multiple regression (based on least squares parameter estimation) was used to derive the logarithmic regression equation for each Q_T on catchment characteristics. A method for bias correction due to logarithmic transformation, developed by Miller, et.al (1984) was used to provide unbiased regression equations.

Chapter 4

Preliminary At-Site and Regional Analyses

4.1 Introduction

This chapter discusses the preliminary overview of at-site and regional flood frequency analyses for the Aceh province rivers. As mentioned in the previous chapter, a prerequisite of the regional approach is the identification of regions or data sets that have homogeneous flood frequency behaviour. Aceh Province, which is a relatively small area, was assumed to be a physically homogeneous region in terms of flood flow distribution. As discussed in the previous chapter with regard to regional analysis, the following steps are required before undertaking the estimation of regional frequency distribution. These are: (1) screening of the data, (2) providing at-site flood frequency magnitudes, (3) selecting a regional frequency distribution, and (4) obtaining catchment characteristics. The years for which annual maximum streamflows were used for the 9 river basins are listed in Table 4.1.

Table 4.1: Length of record of annual maximum floods used for frequency analysis

Site	Station name	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89
1	Aceh at Kampung Darang (KD)					==	==	==		==	==	==	==	==	==	==	==		
2	Lambeusoi at Sango (SG)						==	==	==	==	==	==	==	==	==	==	==		==
3	Seunegan at Ujung Blang (UB)										==	==	==	==	==	==			==
4	Kluet at Gunung Pudung (GP)										==	==	==	==	==	==			
5	Lawe Alas at Sukarimbun (SR)									==	==	==	==	==	==	==	==		
6	Baro at Klibeut (KL)	==	==	==	==	==	==	==	==	==	==	==							
7	Peusangan at Beukah (BK)						==	==	==	==	==	==	==	==	==	==	==	==	==
8	Jambu Aye at Lhok Nibong (LN)	==	==	==	==	==	==	==	==	==	==	==	==	==	==	==		==	
9	Tamiang at Kuala Simpang (KS)				==	==	==	==	==	==	==								

4.2 Screening of the Flood Flow Data

The first stage of data screening is the inspection of the sample properties of the data. In this analysis, the records for each of the nine sites were assumed to have sufficient length, to be homogeneous, and to be random. The next stage is the identification of unusual sites in which at-site samples may exhibit different characteristics from those found at other sites. The procedure suggested by Hosking and Wallis (1993) which measures the discordancy (D_i) of the site was used in this analysis. The procedure provides an indication of unusual samples in a region as a whole based on the sample L-moments at each site. The mathematical definitions of the L-moments presented below were used in both this screening and also in the method of parameter estimation of the regional index flood by the PWM and L-moment methods.

4.2.1 PWMs and L-Moments

The L-moments defined by Hosking (1986, 1990) are linear combinations of the PWMs (Greenwood et.al., 1979). The PWMs of a random variable X with cumulative distribution function F(X) are given as:

$$\beta_r = E\,[\,X\,(F(X))^r\,] \tag{4.1}$$

where r = 0, 1, 2, 3,... The L-moments are linear combinations of the equation above and are defined by Hosking (1986, 1990) as the following quantities:

$$\begin{aligned} \lambda_1 &= \beta_0, \\ \lambda_2 &= 2\beta_1 - \beta_0, \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0, \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 + \beta_0 \end{aligned} \tag{4.2}$$

the L-moments ratios are given by:

$$\begin{aligned} \tau_r &= \lambda_r / \lambda_2, \quad r = 3, 4, \ldots \\ \tau &= \lambda_2 / \lambda_1 \end{aligned} \tag{4.3}$$

The λ_1 is a measure of location (the mean of the distribution), λ_2 is a measure of scale, λ_3 is a measure of skewness, λ_4 is a measure of kurtosis and the L-CV (τ) is the analogue of the coefficient of variation, τ_3 is the L-skewness, and τ_4 is the L-kurtosis.

In practice, Landwehr et.al (1979) showed that PWMs from a sample can be estimated using the unbiased estimator β_r according to the following equation, and the sample is arranged in order of $x_1 \leq x_2 \leq x_3 \ldots \leq x_n$.

$$b_r = n^{-1} \sum_{j=1}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_j \tag{4.4}$$

The sample unbiased estimators of the λ_r are given by:

$$\begin{aligned} l_1 &= b_0, \\ l_2 &= 2b_1 - b_0, \\ l_3 &= 6b_2 - 6b_1 + b_0, \\ l_4 &= 20b_3 - 30b_2 + 12b_1 + b_0 \end{aligned} \tag{4.5}$$

The sample values of t_r of the τ_r and t of τ are asymptotically unbiased for large n, and can be obtained using the following equations:

$$t_r = l_r / l_2, \quad r = 3, 4, \ldots$$
$$t = l_2 / l_1 \tag{4.6}$$

From the above equations, the following steps (as suggested by Hosking and Wallis, 1993) were used to perform the screening of the nine sampling sites in Aceh Province.

1) For each site, the flood flows were standardized by the mean annual flood, $\bar{Q}$. These standardized values were used to define the L-moment ratios such as L-CV (t), L-skewness (t_3) and L-kurtosis (t_4) using the equations above.

2) For each site i, $\mathbf{u}_i$ is a vector of t, t_3 and t_4 values, and is calculated using the following equation. The unweighted group average $\bar{\mathbf{u}}$ is obtained immediately after the initial definition of $\mathbf{u}_i$:

$$u_i = [t^{(i)} \quad t_3^{(i)} \quad t_4^{(i)}]^T \tag{4.7}$$

$$\bar{u} = \frac{1}{N} \sum_{i=1}^{N} u_i \tag{4.8}$$

3) The sample covariance matrix is defined using the following equation:

$$S = (N-1)^{-1} \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T \tag{4.9}$$

4) For site i, the measure of discordancy is defined using the equation:

$$D_i = \frac{1}{3} (u_i - \overline{u})^T S^{-1} (u_i - \overline{u}) \qquad (4.10)$$

(5) From the above computation, if the D_i value for site i is large, it is an indication that this site is discordant from the group as a whole and it may indicate the presence of errors in the data.

4.2.2 Results

From the flood flow data in the Aceh region, the summary statistics were calculated, and presented in Table 4.2. It shows the discordancy measure, D_i which is an indicator of unusual flood flow at each site in the region, and the various L-moment ratios.

Choosing a single value of D_i that can be used as a criterion for deciding whether a site is unusual is not easy. Hosking and Wallis (1993) suggested $D_i \geq 3$ as a criterion for declaring a site to be unusual. As well, the data for sites with the largest values should be re-examined. It can be seen from Table 4.2 that all D_i are ≤ 3, indicating that there are no unusual sites identified.

Table 4.2: Summary statistics for the Aceh province flood flow data sets

Site	River Basin	n	Mean	t	t_3	t_4	D_i
1	Krueng Aceh	11	286	0.2019	0.1242	0.3296	1.4698
2	Lambeusoi	12	549	0.2289	0.0879	0.0976	0.9028
3	Seunegan	7	183	0.0640	0.0550	0.0625	0.7385
4	Kluet	6	460	0.1772	[illegible]	0.1207	0.2667
5	Lawe Alas	8	274	0.0860	0.0787	0.0698	0.2980
6	Baro	11	79	0.1619	0.1566	0.1173	0.4618
7	Peusangan	13	460	0.1512	0.0126	-0.0338	1.3755
8	Jambu Aye	16	761	0.0900	0.1755	0.0222	1.4654
9	Tamiang	7	1184	0.0757	0.1069	0.2265	1.0215

4.3 At-Site Flood Frequency Analysis

It is necessary to determine the at-site flood frequency magnitudes at each site for four reasons. First, these magnitudes are required when applying Dalrymple's method. Second, these analyses are required to derive regression equations between these flood magnitudes at a given return period and catchment characteristics. Third, these at-site analyses are required for identifying the best flood distribution for Aceh province rivers. Finally, these at-site analyses are required for comparison with regional frequency analyses.

A general frequency model calibrated from at-site sample for the estimation of flood quantile Q_T can be expressed as (Cunnane, 1987):

$$\hat{Q}_T = \theta_1 + \theta_2 \, y_T(\theta_3) \qquad (4.11)$$

where $\hat{Q}_T$ is the event magnitude at a given return period, T. $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$ are sample estimates of location, scale and shape parameters of a selected distributional form $f(q; \theta)$. $y_T(\theta_1)$ is a standardised variate value of return period T from the f(.) distribution.

The distributions considered in this study are the Gumbel (EV1), LN2, LN3, GEV, and LP3 distributions. The method of moments was used to fit all two parameter distributions while the method of maximum likelihood was used to fit all three parameter distributions.

The computer program CFA3, 1991 developed by the Hydrology Division of Water Resources Branch of Environment Canada was used to determine the expected flood flows of various return periods at each of the nine gauging stations for the three parameter distributions. This program allows for the fitting of four distributions: the GEV, LN3, LP3 distributions, and the Wakeby (a 5 parameter distribution). For the two parameter distributions, the computer programs provided by Kite (1988) were used in this study.

Estimates of flood quantiles with 2, 5, 10, 20, 50, 100, and 200 year return periods were derived based on the five theoretical distributions and on empirical fit. Fitting for each distribution and the calculation of the so-called standard error through the least squares method was used to determine the best fitting distribution.

4.3.1 Selection of Best Fitting Distribution

The selection among the five distributions was based on the least squares method of standard error of fit for each distribution. This standard error of fit is a measure of how closely a distribution fits the actual data. The standard errors that were used in this analysis are defined by the equation (Kite, 1988).

$$SE = \left[\frac{\sum_{i=1}^{n} (Y_i - \hat{Y})^2}{n-p} \right]^{1/2} \qquad (4.12)$$

where Y_i is the actual data point, $\hat{Y}_i$ is a data point as estimated by the distribution at probabilities computed from the sorted ranks of Y_i , n is the number of points, and p is the number of parameters.

It should be noted that the computation of the standard error of fit is dependent on the plotting position. As described in Chapter 3, the unbiased plotting position suggested by Cunnane (1978), which is based on a single simple compromise formula for use with all distributions, was used in this study.

4.3.2 Results

The standard errors of fit for each distribution used for the nine rivers in the region are presented in Table 4.3, and the at-site magnitudes based on this best fit distribution are given in Table 4.4. In general, the order of best fit distributions to the nine sites in the province were the EV1, LN2, GEV, LN3 and LP3 respectively. It is not easy to choose a single frequency distribution based on at-site analysis alone, because of the different best fitting distribution derived at each site in the region. The extreme value distribution was selected for use in the regional analysis based on the above results, and on a consideration to the L-moment ratios, as discussed in the following section.

Table 4.3: Standard errors of various distributions used

Site	River basin	Standard Errors (SE) of various distributions				
		EV1	LN2	GEV	LN3	LP3
1	Krueng Aceh	22.758 *	25.040	25.203	30.967	33.093
2	Lambeusoi	143.535 *	156.436	153.783	152.090	158.459
3	Seunegan	5.729	3.646 *	5.548	9.001	6.240
4	Kluet	9.389 *	13.254	13.910	14.243	24.177
5	Lawe Alas	11.726	11.807	12.538	10.459 *	12.633
6	Baro	5.121	5.505	5.068 *	6.555	5.957
7	Peusangan	29.487 *	36.202	32.354	31.659	29.814
8	Jambu Aye	28.639	27.368 *	34.016	43.241	43.366
9	Tamiang	34.924	24.924 *	39.576	50.473	52.787
	In Region (general)	291.308 (1)	304.182 (2)	321.996 (3)	348.688 (4)	366.592 (5)

*) the best fit at each site (the minimum standard errors).

Table 4.4: At-site estimated flood quantile magnitudes in the Aceh rivers based on best fit distribution at-each site.

T years	Y Variate (EV1)	At-site estimated flood quantiles (Q_T)								
		River basins								
		Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9
2	0.367	269.8	515.4	182.5	440.8	267.0	74.9	441.9	751.7	1174.3
5	1.500	374.4	730.5	199.5	587.1	306.0	97.4	559.8	855.7	1306.1
10	2.250	443.7	872.9	209.1	683.9	333.0	112.0	637.9	915.7	1380.8
20	2.970	510.2	1009.4	217.3	776.8	360.0	126.0	712.7	968.4	1445.6
50	3.902	596.2	1186.2	227.0	897.0	397.0	144.0	809.7	1031.3	1522.3
100	4.600	660.7	1318.7	233.6	987.1	425.0	157.0	882.3	1075.5	1575.6
200	5.296	725.0	1450.7	239.9	1076.9	454.0	170.0	954.7	1117.5	1625.9

4.4 Selection of Regional Flood Frequency Distribution

This section discusses the procedure that was used in the selection of a regional frequency distribution for the Aceh region. In this context, the L-moment diagram was used to compare sample estimates of the dimensionless ratios t, t_3 and t_4 (as discussed in Section 4.2) with their population counterparts for a range of assumed distributions. Both at-site L-moment ratios and regional weighted average L-moment ratios were compared.

4.4.1 Procedure

The procedure to obtain the L-moment ratios at each site has been discussed in Section 4.2. The L-moment ratios of the at-site standardized flood flows, particularly L-kurtosis and L-skewness, are compared to the theoretical relationships between L-kurtosis and L-skewness for several distributions. Figure 4.1 compares these theoretical relationships for normal, uniform, Gumbel, GEV, LN3, gamma (P3), GPA, and the lower boundary of all distributions. These theoretical relationships were constructed using the polynomial approximations developed by Hosking (1986). It is apparent that the two parameter distributions are defined by only a single point in the L-kurtosis and L-skewness relationships, while the three parameter distributions are more flexible.

In addition, the procedure was also used in a regional context for the selection of the best fit distribution regionally based on the weighted average L-CV, L-kurtosis and L-skewness for the region. The following equations suggested by Hosking (1992) were used to derive these regional sample weighted average L-moment ratios:

$$\bar{t} = \sum_{i=1}^{N} n_i \cdot t \;/\; \sum_{i=1}^{N} n_i$$

$$\bar{t}_r = \sum_{i=1}^{N} n_i \cdot t_r \;/\; \sum_{i=1}^{N} n_i \qquad (4.13)$$

where $r = 3, 4$. $\bar{t}$ is regional weighted average L-CV, $\bar{t}_3$ and $\bar{t}_4$ are regional weighted average L-skewness and L-kurtosis, respectively, and n_i is the sample length at site i. These regional L-kurtosis and L-skewness values were compared to the theoretical relationships, and were also supported by previous at-site comparisons that were made in order to select the approximate regional distribution.

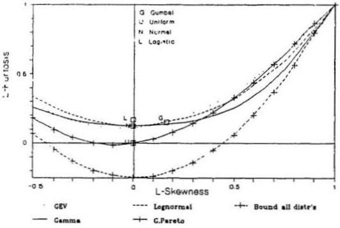

Figure 4.1: Theoretical L-moment relationships for assumed distributions (Hosking, 1986)

4.4.2 Results

Figure 4.2 shows the sample at-site estimates on the same plot as the theoretical relationships between L-kurtosis and L-skewness corresponding to the GEV, Gumbel, LN, gamma, GPA, and the lower boundary of all distributions. This figure shows that, from the models tested, the GEV, LN and gamma distributions appear to be most consistent with the sample L-skewness and L-kurtosis for the sites in the region. In Figure 4.2, nearly five of the observations are close to the GEV, LN and gamma distributions. Although there is large variability in the L-moment ratio estimates, the GEV distribution appears to provide the best overall fit, followed by the LN and gamma distributions. The overall sample L-moment ratios estimate is poorly approximated by the GPA distribution.

Figure 4.3 shows the sample regional weighted average L-moment ratios plotted on the L-moment ratio diagram together with the theoretical L-kurtosis and L-skewness relationships of assumed distributions. This figure shows that the sample regional L-kurtosis and L-skewness fall close to the theoretical GEV curve. Although the sample regional L-kurtosis and L-skewness relationships also fall close to the theoretical LN and gamma curves, from Figure 4.2 and the previous at-site analysis (Section 4.3) show that the LN and gamma distributions seem to provide a poorer representation of the flood flows distribution compared to the GEV distribution for the region considered. The GEV distribution is therefore used for the subsequent estimation of flood quantiles.

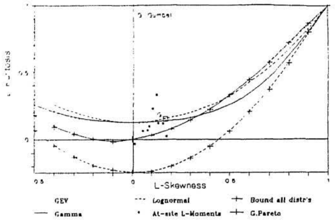

Figure 4.2: L-moment diagrams comparing the sample at-site estimates and theoretical relationships between L-kurtosis and L-skewness for nine Aceh river basins

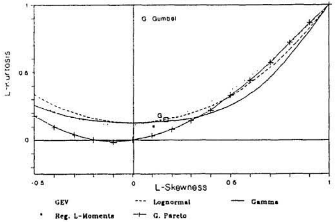

Figure 4.3: L-moment diagrams comparing the sample regional weighted average for the Aceh region with theoretical relationships for L-kurtosis and L-skewness

The GEV distribution is given by:

$$F(x) = \exp\left[-\{1 - k\,(\frac{x-m}{a})\}^{1/k}\right], \qquad k \neq 0$$

$$F(x) = \exp\left[-\exp\{-(\frac{x-m}{a})\}\right], \qquad k = 0 \tag{4.14}$$

where a and m are location and scale parameters, respectively, and k is the shape parameter which determines the class of extreme value distribution. If k = 0, the GEV distribution reverts to the Gumbel distribution. The approximate solution of k suggested in Hosking et.al.(1985) was modified for use in the regional context, is defined as:

$$\hat{k} = 7.859\,c + 2.9554\,c^2$$

$$c = \frac{2}{(3+\bar{t}_3)} - \frac{\ln(2)}{\ln(3)} \tag{4.15}$$

where $\bar{t}_3$ is the sample regional weighted average L-skewness. Testing whether $\hat{k}$ is close to zero, the test statistic Z defined in Maidment (1993) as:

$$Z = \sqrt{\frac{\bar{n}_L}{0.5633}}\cdot\hat{k} \tag{4.16}$$

where $\bar{n}_L$ is the regional average sample lengths, was used. At the 5 % significance level, for the null hypothesis that k = 0, $|Z|$ should be less than 1.96.

For the Aceh region, the computation results of regional weighted average L-moment ratios, and the regional shape factor $\hat{k}$ of the GEV distribution are presented in Table 4.5. The table shows the regional L-CV, L-skewness and L-kurtosis, and the regional shape factor $\hat{k}$ which is an indicator of the appropriate extreme value distribution in the region.

It can be seen from Table 4.5 that the obtained value of $\hat{k} = 0.102$, and $Z = 0.43 < 1.96$ is not significant at $\alpha = 5\ \%$. This indicates that the Gumbel (EV1) distribution is a feasible alternative to be used in the following regional frequency analysis.

Table 4.5: Results of regional weighted average L-moment ratios and the regional shape factor $\hat{k}$ of the GEV distribution

Site	n_i	$n_i \cdot t$	$n_i \cdot t_3$	$n_i \cdot t_4$
1	11	2.2210	1.3663	3.6250
2	12	2.7467	1.0549	1.1707
3	7	0.4480	0.3849	0.4375
4	6	1.0630	0.7975	0.7239
5	8	0.6880	0.6299	0.5581
6	11	1.7812	1.7221	1.2903
7	13	1.9660	0.1638	-0.440
8	16	1.4400	2.8074	0.3556
9	7	0.5297	0.7480	1.5853
Sum	91	12.884	9.6749	9.3065
	$\bar{n} = 10$	$\bar{t}=\bar{t}_2= 0.1416$	$\bar{t}_3 = 0.1063$	$\bar{t}_4 = 0.1023$
In regional case, $\hat{k} = 0.102$, $c = 0.0129$ and test statistic $Z = 0.43$.				

The values of t, t_3 and t_4 are presented in Table 4.2.

4.5 Catchment Characteristics

The following catchment characteristics of the Aceh river basins were obtained from available topographic and land use maps of Aceh Province.

1) *Catchment area (Area), km^2.* Watershed boundaries were drawn on topographic maps and the area measured by a planimeter. These areas are usually provided with gauging station information, and have been checked from 1:100,000 scale maps with 50 m contour intervals.

2) *Main channel length (Length), km.* The main channel length of the stream is defined as the length of the longest channel upstream of the gauging station as drawn on the topographic map, and its length was measured with a curvimeter. For gauged catchments, the *Length* for the Lambeusoi, Peusangan, and Jambu Aye rivers were checked before being taken from the IOH report, and the remaining rivers were measured. For ungauged catchments considered, the *Length* for the Tripa and Susoh rivers was also checked before being taken from that report, and the remaining rivers were measured for this study.

3) *Catchment mean elevation (Elev), m.* On the topographic map, a grid was selected and overlaid such as 15 points on the catchment boundary. The mean elevation at these points of intersection was taken as the catchment mean elevation of an area.

4) *Stream frequency (STMF), jtn/km^2.* For each catchment, the number of junctions for all stream channels was counted, and the number was divided

by the drainage area.

5) *Channel slope (SIMS), m/m.* The channel slope is defined as the difference in height between the point of interest and the highest point above the end of the main stream, divided by the main channel length (Length). The highest point is defined as the highest point on the catchment of the source of the longest tributary. The elevations were taken from 1:100,000 scale maps with 50 m contour intervals.

6) *Forest cover (Forest), %.* Land use information was sketched on the forest map and the total area of forest was determined by a planimeter. The catchment forested is expressed as a percentage of the total area. For gauged catchments, the *Forest* for the Lambeusoi, Peusangan, and Jambu Aye basins were checked before being taken from the IOH report, and the remaining basins were measured. As well, for ungauged catchments, the *Forest* for the Tripa and Susoh basins was also checked before being taken from IOH report, and the rest were measured from the above maps.

7) *Mean annual rainfall (MAR), mm.* Catchment mean annual rainfall was measured from mean annual rainfall isohyetal maps provided by the Aceh-WRDS at a scale of 1:250,000. The MAR was taken at the basin centroid for each basin.

8) *Mean annual maximum 1 day rainfall (APBAR), mm.* The APBAR was calculated by multiplying PBAR (the mean annual maximum 1 day point rainfall for the catchment) and an area reduction factor (ARF). PBAR

was taken from an isohyetal map of mean annual maximum 1 day rainfall for Sumatra island at 1:2,000,000 scale appended to the IOH (1983) report. The related ARF's with catchment areas were determined from the ARF's table used by IOH (1983).

The catchment characteristics for the nine gauged basins used in this study are presented in Table 4.6. The characteristics in Table 4.6 were used for both the index flood approaches and the multiple regression approach. They are the independent variables in the estimation of the MAF when applying the index flood approach, and in the estimation of each Q_T when applying the regression approach.

The catchment characteristics for the five ungauged basins presented in Table 4.7 were used for testing of flood quantile predictions.

Table 4.6: Catchment characteristics of gauged catchments in Aceh Province

Site	Gauged catchments								
	River basin	Area km^2	Length km	Elev m	STMF jtn/ km^2	SIMS m/m	Forest %	MAR mm	APBAR mm
1	Aceh at Kp. Darang	1078.0	66.04	557	0.020	0.0175	70	1750	93
2	Lambeusoi at Sango	580.4	33.20	734	0.019	0.0614	73	2450	123
3	Seunegan at Uj.Blang	352.0	50.80	1033	0.031	0.0375	85	3260	117
4	Kluet at Gn.Pudung	2326.0	100.80	1396	0.025	0.0146	91	3225	91
5	L.Alas at Sukarimbun	1384.0	96.52	1593	0.039	0.0160	88	3250	81
6	Baro at Klibeut	270.0	35.60	385	0.022	0.0276	36	2250	90
7	Peusangan at Beukah	2214.0	137.80	736	0.016	0.0169	73	2550	76
8	Jambu aye at L.Nibong	4420.0	231.20	476	0.010	0.0084	88	2450	77
9	Tamiang at Kl.Simpang	4598.0	177.[illegible]	1206	0.011	0.0140	75	1750	94

Table 4.7: Catchment characteristics of ungauged catchments in Aceh Province

Site	Ungauged catchments								
	River Basin	Area km^2	Length km	Elev m	STMF jtn/ km^2	SIMS m/m	Forest %	MAR mm	APBAR mm
1	Teunom at T. Kareung	2236	112.0	956	0.015	0.0154	97.0	3500	114
2	Woyla at Manu Tukut	2327	115.0	1200	0.014	0.0150	96.0	3500	105
3	Tripa at Gn. Kong	2707	157.0	1100	0.022	0.0112	86.5	3000	69
4	Susoh at Kt. Tinggi	194	24.0	625	0.072	0.0406	98.0	3200	132
5	Peureulak at Gd. Janeng	1050	91.5	425	0.011	0.0107	63.0	1750	97

4.6 Results of Preliminary Analyses: Summary

This section summarizes the preliminary at-site and regional analyses for the Aceh province rivers. Screening of the data, deriving at-site flood frequency magnitudes, selecting a regional distribution and obtaining catchment characteristics were considered in this chapter. The following items summarize the results.

1) There are no unusual sites identified in the province. This was indicated by the discordancy measure based on L-moment ratios.

2) At-site estimated flood quantile magnitudes were based on the best fit distribution at each site, regardless of which distribution provided the greatest numbers of predicted minimum standard errors in the region.

3) Using at-site and regional L-moments ratio diagrams, the GEV distribution was selected as the regional distribution. The regional shape parameter $\hat{k}$ of the GEV distribution was tested and was not significantly different from zero, based on the standard normal quantile comparison at $\alpha = 0.5$. Thus, the EV1 distribution can reasonably be used in this study.

4) The catchment characteristics of the nine gauged catchments are given in Table 4.6. As well, the characteristics of five ungauged catchments to be used for testing of flood predictions are given in Table 4.7.

Chapter 5

Application of Regional Estimation Techniques

This chapter discusses the application of the various regional flood frequency techniques to the Aceh province rivers. Two techniques of regional analysis were used in this study. The first technique was based on index flood approaches, and the second technique was based on a multiple regression approach. The development procedures and results of the techniques used are therefore discussed in two parts. The first part deals with the procedures and results of the index flood approaches, and the second part is the procedures and results of the multiple regression approach. The two regional methods were compared to each other in predicting at-sites flood quantile magnitudes for both gauged and ungauged catchments, as well as to IOH (1983) estimates, and at-site estimates for gauged catchments.

5.1 Index Flood Approaches

This section discusses the index flood method which uses at-site and regional data for developing the flood frequency curve for the Aceh province rivers. As discussed in Chapter 4, the EV1 distribution was used for the regional flood frequency analysis.

5.1.1 Theory

The assumption which underlies the index flood procedure is that the variate $X_i = Q_i / \bar{Q}$ has the same form of distribution at every site. In this case, Q_i is the data point and $\bar{Q}$ is the index flood, alternatively called the mean annual flood (MAF) at a site of interest. The parameters of the distribution are obtained from the combined regional data sets of X_i. The flood quantile Q_T is then estimated as:

$$\hat{Q}_T = \overline{Q} \, . \, X_T \qquad (5.1)$$

where $\bar{Q}$ is the observed MAF at a gauged site or the estimated MAF at an ungauged site. The MAF for an ungauged site can be estimated from catchment characteristics using regression analysis. This will be discussed in Section 5.2.

While all index flood methods are based on a similar concept, they vary in the method used to estimate parameters of the X_T distribution. For this study, estimates based on the following five approaches were compared:

a) Dalrymple (1960): Regional averaging of dimensionless at-site flood quantiles estimate $Q_T / \bar{Q}$.

b) NERC (1975) and later used in New Zealand (Beable and McKerchar, 1982):

A variation of regional averaging of dimensionless at-site order statistics $X_{(i)} = Q_{(i)} / \bar{Q}$ and fitting a distribution of these either graphically or numerically, adjusted for records of unequal length.

c) Wallis (1980): Regional analysis of dimensionless PWMs, $m_{(i)} / m_{(0)}$, $i = 0, 1, 2, 3$, and $m_{(0)}$ = regional sample mean. This method has been found to be easy to implement and is robust and efficient.

d) Station year approach: Regional pooling of all $X_i = Q_i / \bar{Q}$ values. $x_{ij} = Q_{ij} / \bar{Q}_i$, $(i = 1, 2,..., Nj;\ j = 1, 2,..., M)$. The pooled values are treated as a single sample from postulated population

e) Hosking and Wallis (1992, 1993): Regional analysis using L-moments, which are defined as linear combinations of the dimensionless PWMs. This method should give answers identical to the PWMs method.

These procedures were applied in this study and are described in the following sections.

5.1.2 Dalrymple's Method

This method, the earliest method of regional flood frequency analysis, was pioneered by the US Geology Survey (Dalrymple, 1960). The method is based on a regional dimensionless averaging of records of nearly equal length from unregulated rivers within a region.

The procedure used to develop the dimensionless regional flood frequency curve by this method consisted of the following steps:

1) From the results of the at-site analysis (Table 4.4), for each site and each selected return period, the flood quantiles were standardized by the mean annual flood.

2) For each return period of all sites, the median value of these standardized flood quantiles are determined. These are called the median ratios for a given return period.

3) Median ratios were plotted against frequencies on extreme value probability paper and a line of best fit was drawn to obtain a regional flood frequency curve.

4) The values of X_T at any given return period for all sites in the region were determined from the curve above.

5.1.3 The NERC Method

In order to take into account the unequal length of the records, the NERC (1975) method modifies Dalrymple's method by using regionally averaged standardized order statistics within a graphical procedure to estimate the X_T distribution. This procedure is intended to reduce the effect of outliers on the regional curve fitting method due to the possibility of sampling variation in the region.

The following procedure was used to develop the regional flood frequency curve by the NERC (1975) method.

1) The ratios of $Q_{obs}/\bar{Q}$ at each site of nine stations were lumped together and were plotted against their corresponding return period or reduced variate y

values.

2) To deal with the problem of different sample sizes, the y scale was divided into 0.5 class intervals, and an average Q_{obs} / $\bar{Q}$ and an average y value was determined from the data points of these ratios falling within each class. This can be expected to produced a smooth trend in the regional plot.

3) The values of X_T at any return period of interest for any site in the region were determined by using the smoothed curve obtained in step 2 above.

5.1.4 The Probability Weighted Moments (PWMs) Method

The use of PWMs has been studied for improving estimates of flood quantiles in both gauged and ungauged catchments. The method is particularly robust when the availability of record samples are either very short, highly skewed or have high kurtosis (Greenwood et.al., 1979). As discussed in Section 3.5.1, this method is applicable for distributions such as EV1 which can be expressed in inverse form.

General mathematical definition of the PWMs defined by Greenwood et.al., (1979) has been described in Section 4.2, in which the moments $M_{(i)} = \beta_r$, and $i = r = 0, 1, 2, 3$. The PWMs from a sample can be estimated using the unbiased estimator of $m_{(i)} = b_r$. Using the procedure in Section 4.2, the $m_{(i)}$ values can be determined. For operational purposes, Hosking et.al. (1985) have suggested that at-site sample values of PWMs can also be estimated from the biased PWMs estimator of $m_{(i)} = b_r$ using the following equation:

$$m_{(i)} = b_r = \frac{1}{n}\sum_{j=1}^{n} p_{j,n}^{r}\, x_j$$
$$p_{j,n} = \frac{j-0.35}{n} \qquad (5.2)$$

$p_{j,n}$ is a weighted function, and x_j represent standardized sample values ranked from smallest to largest.

The following procedure as suggested by Wallis (1980) was used for obtaining the regional frequency curve by the method of PWMs.

1) The PWMs $m_{(0)j} = b_0$ and $m_{(i)j} = b_r$ were computed using the procedure described in Section 4.2 for each site, data were standardized by the MAF at each site before using that procedure. j is the station number, and i = r = 0, 1, 2, 3.
2) All standardized moments at each site were defined as $m_{(1)j}$, $m_{(2)j}$ and $m_{(3)j}$. After standardizing, the regional $m_{(0)j} = 1$.
3) In the region, the moments of all sites were averaged to obtain a regional $\bar{m}_{(i)}$ using the following equation:

$$\bar{m}_{(i)} = \sum_{j=1}^{M} m_{(i)j}\,(n_j/L) \qquad (5.3)$$

where M is the number of stations, the regional $\bar{m}_{(0)} = 1$, and L = the total number of station years of record in the region.

4) The regional parameters of a and m were estimated using the equations provided by Greenwood et.al.(1979) for the Gumbel distribution:

$$\hat{a} = (\overline{m}_{(0)} - 2\overline{m}_{(1)}) / \ln(2)$$
$$\hat{m} = \overline{m}_{(0)} - \varepsilon\, \hat{a} \qquad (5.4)$$

where ε is Euler's number = 0.5772.

5) Regional quantiles, X_T (with T such as: 2, 5, 10, 20, 50, 100 and 200 years) were estimated using the equation (for Gumbel distribution):

$$X_T = \hat{m} - \hat{a} \ln\{-\ln(1 - \frac{1}{T})\} \qquad (5.5)$$

6) The values of X_T at any return period of interest for any site in the region were determined from the regional quantiles.

5.1.5 Station Year Approach

The station year method is based on the regional pooling of all standardized flood flows at each site in a region (Cunnane, 1987, 1988). The pooling of standardized regional data was studied by Rossi et.al (1984) using Italian data for their analysis of the TCEV (two-component extreme value) method. This method of regional pooling of data was used to compare with other methods based on regional averaging of data or the statistics of those data. The method assumes that all pooled data represent a single sample from a population, in this case, the extreme value distribution.

The following steps were used to derive the flood quantiles or frequency curve by the station year approach:

1) All standardized data were pooled together and ranked and plotted against

their corresponding return periods or y variate values using extreme value probability plot paper.

3) The plotted data were fitted by linear regression. This produced the regional flood frequency curve.

4) The values of X_T at any return period of interest for any site were determined using the regional curve.

5.1.6 L-Moment Method

L-moments defined by Hosking (1986, 1990) are linear combinations of PWMs defined by Greenwood et.al., 1979 as described in Section 4.2. Here, sample L-moments and L-moment ratios were derived using the unbiased estimator of L-moments as presented in the Section 4.2.

In this study, the procedure as defined by Hosking and Wallis (1992, 1993) was used to derive the flood frequency curve by the L-moment method.

1) The sample L-moments such as l_1, l_2, l_3 and l_4 were computed using the procedure described in Section 4.2. The at-site L-moment ratios such as t, t_3 and t_4 were also determined (Table 4.2).

2) The at-site sample L-moment estimates were combined to provide regional estimates of $\bar{t}$, $\bar{t}_3$ and $\bar{t}_4$ using the procedure outlined in Section 4.4.1, and the results in Table 4.5.

3) Regional parameters of **a** and **m** were estimated using the equations for the Gumbel and GEV distributions (Hosking, 1986):

Gumbel:

$$\hat{a} = \bar{I}_2 \, / \, \ln(2)$$
$$\hat{m} = \bar{I}_1 - \varepsilon \, . \, \hat{a} \qquad (5.6)$$

where ε is Euler's number = 0.5772, $\bar{l}_1 = \bar{t}_1 = 1.0$, and $\bar{l}_2 = \bar{t}$ = the regional averaged L-CV.

GEV:

$$\hat{a} = \frac{\bar{I}_2 \, \hat{k}}{(1-2^{-\hat{k}}) \; \Gamma(1+\hat{k})}$$
$$\hat{m} = \bar{I}_1 - \frac{\hat{a}}{\hat{k}}\{1-\Gamma(1+\hat{k})\} \qquad (5.7)$$

where $\hat{k}$ is the shape parameter given by Equation 4.15. The appropriate solution of the gamma function given by (Maidment, 1993):

$$\Gamma(1+\hat{k}) = 1+\sum_{i=1}^{5} a_i \, \hat{k}_i \qquad (5.8)$$

where a_1 = -0.5748646, a_2 = 0.9512363, a_3 = -0.6998588, a_4 = 0.425549, and a_5 = -0.1010678.

4) Regional quantiles, X_T (with T such as 2, 5, 10, 20, 50, 100 and 200 years) were estimated using Equation 5.5 for the Gumbel distribution, and using the following equation for the GEV distribution.

$$X_T = \hat{m} - \frac{\hat{a}}{\hat{k}} \{1 - [-\ln(1-\frac{1}{T})]^{\hat{k}}\} \qquad (5.9)$$

5) The values of X_T at any return period of interest for any site in the region

were determined from the regional quantiles.

5.1.7 Results

This section discusses the results of regional quantile X_T estimates for each of the five index flood approaches. Table 5.1 and Figure 5.1 show the estimated regional quantiles applying Dalrymple's approach. The flood flow ratios in Table 5.1 were based on the best fit distribution at each site. The median ratio at a given return period were obtained from these flow ratios in the region. It can be seen that all median ratios are the same flood flow ratios as site 7. The fitted curve by these median ratios is the regional quantiles.

Table 5.2 and Figure 5.2 present the estimated regional quantiles applying the NERC method. In Table 5.2, the mean y variate and the mean flow ratio were based on the points that fall within each 0.5 class interval of y values. It can be seen from Figure 5.2 that the mean ratios always represent the points which fall within 0.5 class interval of y values. The fitted curve of these mean ratios is the regional curve by the NERC method.

Table 5.3 shows the regional average PWMs, regional parameters a and m to obtain regional quantiles, X_T. The regional parameters a and m were derived based on the regional average PWMs, $\bar{m}_{(1)j}$ for the Gumbel distribution. Given return periods of T, the regional quantiles were derived using Equation 5.5.

Figure 5.3 shows the regional curve applying the station year approach. It can be seen in the regional plot that eight points at low return periods were outliers. However, it did not significantly affect the regional curve with the inclusion of these values. The fitted curve was derived by regressing these pooled flow ratios to the corresponding reduced variate.

Table 5.4 presents the regional weighted average L-moment ratios of t and $\bar{t}_r$, and regional parameters **a** and **m**. The regional parameters **a** and **m** derived based on the regional average L-moment ratios, $\bar{t}$ and $\bar{l}_1$ for the Gumbel distribution. The parameters **a** and **m** obtained were identical to the PWM method as expected because the L-moments are linear combination of the PWMs. For a comparison of estimated regional quantiles, X_T's, the regional paramaters **a** and **m** were also obtained using the GEV distribution.

The estimated regional flood quantiles X_T for each of the five index flood approaches are summarized and presented in Table 5.5 and Figure 5.4. Table 5.5 shows the estimated regional quantiles using Dalrymple, NERC, PWMs, station year, and L-moment methods based on the Gumbel distribution, and the L-moment method based on the GEV distribution. These plotted regional quantile curves presented in Figure 5.4, show that all index flood approaches gave similar estimated regional quantiles. The PWM or L-moment method seem to provide a compromise regional quantile curve. The L-moment based on the GEV distribution gave slightly lower estimated regional quantiles, particularly in high return periods.

These various X_T's based on the various index flood approaches were used to predict at-sites flood quantile magnitudes for gauged and ungauged catchments which will be discussed in Section 5.3.

Table 5.1: Flood flow ratios at certain return periods, median flood flow ratios (applying Dalrymple's method)

Site	Frequencies, in years						
	2	5	10	20	50	100	200
	Flood flow ratios (Q_T / $\bar{Q}$)						
1	0.943	1.309	1.551	1.784	2.085	2.310	2.535
2	0.939	1.331	1.590	1.839	2.161	2.402	2.642
3	0.997	1.090	1.143	1.187	1.240	1.277	1.311
4	0.958	1.276	1.487	1.689	1.950	2.146	2.341
5	0.974	1.117	1.215	1.314	1.449	1.551	1.657
6	0.948	1.233	1.418	1.595	1.823	1.987	2.152
7	0.961	1.217	1.387	1.549	1.760	1.918	2.075
8	0.988	1.124	1.203	1.273	1.355	1.413	1.468
9	0.992	1.103	1.166	1.221	1.286	1.331	1.373
Median (X_T)	0.961	1.217	1.387	1.549	1.760	1.918	2.075

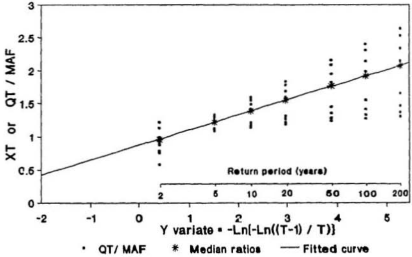

Figure 5.1: The regional plot and fitted curve for Aceh region (applying Dalrymple's method)

Table 5.2: Mean flow ratios at each 0.5 class interval of y variate (applying NERC method)

No	Y variate		Mean $Q_{obs} / \bar{Q}$
	0.5 class interval	Mean Y variate	
1	-1.193		
2	-0.693	-0.958	0.679
3	-0.193	-0.422	0.800
4	0.307	0.053	0.905
5	0.807	0.550	0.995
6	1.307	1.007	1.137
7	1.807	1.447	1.218
8	2.307	2.049	1.357
9	2.807	2.487	1.176
10	3.307	3.026	1.519

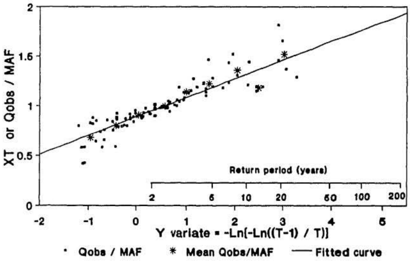

Figure 5.2: The regional plot and fitted curve for Aceh region (applying NERC method)

Table 5.3: Regional average PWMs ($\bar{m}_{(i)j}$), regional parameters **a**, **m** (applying PWMs method)

Site (j)	n_j	PWMs, $(m_{(i)j})$			n_j / L	$m_{(1)j} \cdot n_j / L$	$m_{(2)j} \cdot n_j / L$	$m_{(3)j} \cdot n_j / L$
		i = 1	i = 2	i = 3				
		$m_{(1)} = b_1$	$m_{(2)} = b_2$	$m_{(3)} = b_3$				
1	11	0.601	0.438	0.350	0.1209	0.0727	0.0529	0.0423
2	12	0.614	0.451	0.359	0.1319	0.0810	0.0595	0.0474
3	7	0.532	0.366	0.280	0.0769	0.0410	0.0282	0.0215
4	6	0.589	0.426	0.337	0.0659	0.0388	0.0281	0.0222
5	8	0.543	0.377	0.290	0.0879	0.0477	0.0331	0.0255
6	11	0.581	0.419	0.330	0.1209	0.0702	0.0506	0.0340
7	13	0.576	0.409	0.318	0.1429	0.0823	0.0584	0.0454
8	16	0.545	0.381	0.294	0.1758	0.0958	0.0670	0.0517
9	7	0.538	0.373	0.287	0.0769	0.0414	0.0285	0.0221
L = 91		Estimated regional parameters, **a** and **m**: a = 0.20440, **m** = 0.88202			$\bar{m}_{(i)j}$ =	0.5708	0.3782	0.3121

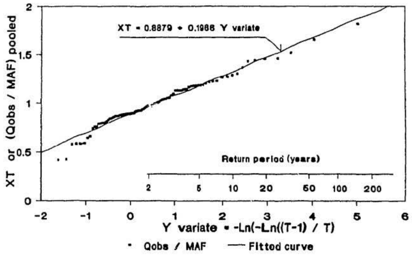

Figure 5.3: The regional plot and fitted curve for Aceh region (based on station year approach)

Table 5.4: Regional weighted average L-moment ratios (t and $\bar{t}_r$), regional parameters **a**, **m** (applying L-moment method)

Site (j)	n_j	L-Moments, (t and t_r)			$n_j \cdot t$	$n_j \cdot t_3$	$n_j \cdot t_4$
		t	t_3	t_4			
1	11	0.2019	0.1242	0.3296	2.2210	1.3663	3.6250
2	12	0.2289	0.0879	0.0976	2.7467	1.0549	1.1706
3	7	0.0640	0.0550	0.0625	0.4480	0.3849	0.4376
4	6	0.1772	0.1329	0.1207	1.0630	0.7975	0.7239
5	8	0.0860	0.0787	0.0698	0.6880	0.6299	0.5581
6	11	0.1619	0.1566	0.1173	1.7812	1.7221	1.2903
7	13	0.1512	0.0126	-0.0338	1.9660	0.1638	-0.4399
8	16	0.0900	0.1755	0.0222	1.4400	2.8074	0.3556
9	7	0.0757	0.1069	0.2265	0.5297	0.7480	1.5853
L = 91		Estimated regional parameters, a and **m**: **Gumbel:** a = 0.20430, **m** = 0.8821 **GEV:** a = 0.20110, **m** = 0.9025			Sum = 12.884 $\bar{t}$ = 0.1416	Sum = 9.6749 $\bar{t}_3$ = 0.1063	Sum = 9.3065 $\bar{t}_4$ = 0.1023

Table 5.5: Comparison of estimated X_T at various return periods based on five index flood procedures

T (yrs)	Y variate	Estimated regional quantile, X_T					
		Dalrymple Gumbel	NERC Gumbel	PWMs Gumbel	St. Year Gumbel	L-moments	
						Gumbel	GEV
2	0.367	0.961	0.962	0.9570	0.960	0.9570	0.9749
5	1.500	1.217	1.177	1.1885	1.183	1.1885	1.1822
10	2.250	1.387	1.320	1.3417	1.330	1.3417	1.3069
20	2.970	1.549	1.457	1.4888	1.472	1.4888	1.4178
50	3.902	1.760	1.634	1.6791	1.655	1.6791	1.5498
100	4.600	1.918	1.767	1.8217	1.792	1.8217	1.6409
200	5.296	2.075	1.899	1.9638	1.929	1.9638	1.7253

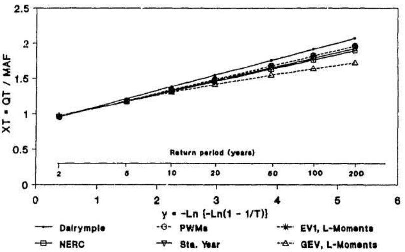

Figure 5.4: Comparison of estimated X_T at various return periods based on five index flood approaches

5.1.8 Results from the IOH (1983) Study

The previous study of flood frequency analysis by the IOH (1983) for Jawa and Sumatra is summarized here, particularly those results pertaining to the growth curve (regional frequency curve) from that study. The regional curves provided in the IOH study have been the principal source of flood estimates in the province, and an important purpose of the present study was to provide updated curves and to compare the results. The growth curve derived by the IOH was based on data from the entire regions of Jawa and Sumatra using data up to 1981. Only five stations from Aceh Province were used in that study.

The growth factors (GF) from the previous study given in terms of catchment area for various return periods are presented in Table 5.6. The GF were defined as $Q_T / \bar{Q}$.

For catchment areas not tabulated, growth factors may be obtained by interpolation. Using Table 5.6, growth factors were estimated for the basins of interest in the present study. These are presented in Tables 5.7 and 5.8.

Table 5.6: Growth factors summarized from study by IOH (1983) for Jawa and Sumatra

T	Catchment area (km^2)					
	180 or less	300	600	900	1200	1500 or more
	Growth Factor (GF)					
2	0.83	0.83	0.83	0.83	0.83	0.83
5	1.28	1.27	1.24	1.22	1.19	1.17
10	1.56	1.54	1.48	1.44	1.41	1.37
20	1.88	1.84	1.75	1.70	1.64	1.59
50	2.35	2.30	2.18	2.10	2.03	1.95
100	2.78	2.72	2.57	2.47	2.37	2.27
200	3.27	3.20	3.01	2.89	2.78	2.66
500	4.01	3.92	3.70	3.56	3.41	3.27
1000	4.08	4.58	4.32	4.16	4.01	3.85

Table 5.7: Growth factors for the nine gauged catchments in the Aceh region

T	Gauged catchments								
	River basins and Area (km^2)								
	Site 1 Aceh	Site 2 L'beusoi	Site 3 S'negan	Site 4 Kluet	Site 5 L. Alas	Site 6 Baro	Site 7 P'sangan	Site 8 J. Aye	Site 9 Tamiang
	1078	580.4	352	2326	1384	270	2214	4420	4598
	Growth Factor (GF)								
2	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
5	1.20	1.24	1.26	1.17	1.18	1.27	1.17	1.17	1.17
10	1.42	1.48	1.53	1.37	1.39	1.54	1.37	1.37	1.37
20	1.66	1.76	1.82	1.59	1.61	1.83	1.59	1.59	1.59
50	2.06	2.19	2.28	1.95	1.98	2.29	1.95	1.95	1.95
100	2.41	2.58	2.69	2.27	2.31	2.71	2.27	2.27	2.27
200	2.82	3.02	3.17	2.66	2.71	3.18	2.66	2.66	2.66

Table 5.8: Growth factors for the five ungauged catchments in the Aceh region

T	Ungauged catchments				
	River basins and Area (km^2)				
	Site 1 Teunom	Site 2 Woyla	Site 3 Tripa	Site 4 Susoh	Site 5 Peureulak
	2236	2327	2707	194	1050
	Growth Factor (GF)				
2	0.83	0.83	0.83	0.83	0.83
5	1.17	1.17	1.17	1.28	1.21
10	1.37	1.37	1.37	1.56	1.43
20	1.59	1.59	1.59	1.88	1.67
50	1.95	1.95	1.95	2.34	2.07
100	2.27	2.27	2.27	2.77	2.42
200	2.66	2.66	2.66	3.26	2.84

5.2 The Multiple Regression Approach

This section discusses the multiple regression approach for developing regional flood quantile regression equations for the Aceh province rivers. All index flood approaches as well as the IOH growth factors rely on an estimate of one flood flow, e.g., $\bar{Q}$ which is the index flood, and then obtain estimate for quantiles using multiplying factors obtained from the regional curve. The multiple regression approach, by contrast, provides a regression equation for each quantile based on catchment characteristics.

As discussed in Chapter 3, the multiple regression approach was used for two reasons. First, the approach was used in deriving regression equations which can be used to estimate the MAF, $\bar{Q}$, at ungauged catchments. From this MAF, the flood quantiles of ungauged catchments can then be estimated using one of the index flood methods. Second, this approach can be used to obtain the regression equations for each Q_T. These regression equations can be used to estimate at-site flood quantiles at gauged or ungauged catchments.

For the purposes of this study, the nine river basins which contained sufficient flood flow information and catchment characteristics were used in the multiple regression approach. The regression approach describes the statistical relationship between the catchment characteristics and the flood flows at each site in the region. The catchment characteristics are as described in Section 4.5, (Table 4.6). The at-site MAF and each Q_T for each of the gauged catchments are given in Table 4.2 and Table 4.4, respectively.

5.2.1 Theory

The function most frequently used for regressing flood flows on catchment characteristics is the nonlinear multi-variable power model (McCuen, et. al., 1990). This has the form:

$$y_i = C\, x_{1i}^{A_1} \cdot x_{2i}^{A_2} \cdots x_{pi}^{A_p} \cdot \epsilon_i \qquad (5.10)$$

where y_i = dependent variable of site i, C = constant coefficient, x_{pi} = independent variable p at site i, p = the number of independent variables, A_p = coefficients of regression equations, and ϵ_i = residual for observation i. This model is assumed to have multiplicative error term.

For the purposes of this study, Equation 5.10 is fitted to observed data by taking the logarithms of the variables, which converts Equation 5.10 into a form given by:

$$Ln y_i = Ln C + A_{1i}\, Ln x_{1i} + \ldots + A_{pi}\, Ln x_{pi} + Ln \epsilon_i \qquad (5.11)$$

The parameters in Equation 5.11 can then be estimated using linear ordinary least squares.

For the problem at hand, Equation 5.11 can be written as:

$$Ln Q_{Tx} = Ln C + \sum_{i=1}^{p} A_i\, Ln CC_i + Ln \epsilon_i \qquad (5.12)$$

where:

Q_{Tx}	:	the MAF or at-site Q_T of each the 9 sites in the region,
Ln C	:	the constant coefficient in the regression,
A_i	:	regression coefficients,

Ln CC_i : the logarithms of p catchment characteristics (such as *Area*, *Length*, *Elev*, *STMF*, *SIMS*, *Forest*, *MAR* and *APBAR*).

By using the equation of the form of Equation 5.12, it is assumed that there is a linear relationship between the logarithms of the catchment characteristics and the logarithms of the MAF or each Q_T, and that the catchment characteristics and flood flows at one basin was considered to be independent of other basins, and other assumptions of linear regression, homoscedasticity, normality of residuals, and randomness of residuals hold.

5.2.2 Regression Equations Selection Criteria

The following statistical criteria were used in the selection of the best regression equations.

1) The coefficient of determination (R^2) and the adjusted coefficient of determination (R_{adj}^2) describe the percentage of variance in the dependent variable explained by the independent variables. The adjusted R^2 considers the number of independent variables, and is defined as:

$$R_{adj}^2 = 1 - \frac{n-1}{n-p}\left[\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}\right] \qquad (5.13)$$

where p is the number of parameters in the regression equation.

2) The standard error (SE) describes the scatter of the data about the regression

line. It measures the standard deviation of the errors between observed and the predicted values. The smaller the standard error, the better the fit of the regression line.

3) The Variance Inflation Factor (VIF) measures the multicollinearity problem on the regression coefficients. The VIF is determined using the equation:

$$VIF = \frac{1}{1-R^2} \tag{5.14}$$

where R^2 is the coefficient of determination of the relationship between one predictor and the other predictors.

A VIF in excess of 10 is often taken as an indication that multicollinearity may be influencing the regression estimates (Neter, et. al, 1985).

5.2.3 Correction of Bias Due to Logarithmic Transformation

The nonlinear power model is fitted after using a logarithmic transformation. The model is thus biased and does not have the minimum expected error variance in the y-space. The method for correcting the bias suggested by Miller, et. al, 1984 was used to reduce the bias due to the logarithmic transformation.

This method adjusts the intercept coefficient using the following steps:

1) For unbiased estimates, the $\sum \epsilon_i$ in the y-space must be equal to zero. The assumed unbiased estimator of the intercept is given by:

$$f_0 = \frac{\sum y_{1obs}}{\sum x_1^{A1} x_2^{A2} \ldots x_p^{A_p}} \qquad (5.15)$$

where $f_0 = C$ = unbiased estimator of the intercept based on the assumption that bias in computed values of y can be obtained by adjusting the intercept only.

2) The final regression equations for both the estimated MAF and the estimated Q_T's after correcting of bias were then determined for use in the estimation of Q_T's at gauged and ungauged catchments.

5.2.4 Procedure

The procedure used to derive the multiple regression equations are as follows:

1) The logarithms of the nine at-site MAF values or each Q_T and the corresponding catchment characteristics were taken.

2) Stepwise regression was used to develop a reasonably well fitting regression equation for the MAF and each Q_T.

3) The following statistics were considered in choosing the regression equations:

   i) correlation matrix of the logs of the MAF or each Q_T and basin characteristics,

   ii) coefficient of determination (R^2 and R_{adj}^2),

   iii) standard errors of estimate (SE),

   iv) variance inflation factors (VIF).

4) The selected regression equations were corrected for bias.

5.2.5 Analysis of the Logarithmic Regression Equations

In estimating the regression equations for the MAF or each Q_T eight catchment characteristics (Table 4.5) were considered as independent variables initially. Table 5.9 shows the correlations between the logarithms of the observed MAF and of the catchment characteristics. In general, the highest correlations of MAF or each Q_T are with *Area* and *Length*. The highest negative correlations of the MAF or each Q_T occur with *STMF* and *SIMS*. In Table 5.9, significant correlations also occur between *Area* and *Length*, between *MAR* and *STMF* and between *APBAR* and *SIMS*. Significant negative correlations occur between *Area* and *SIMS* and between *Length* and *SIMS*, which are possible since large catchments tend to have low channel slopes and also since long channel lengths tend to also have low channel slopes.

In the selection of the regression equations, these correlations however, were analyzed based on statistical and physical criteria. First, the coefficient of correlations between flood flows and catchment characteristics should be physically meaningful. Second, the coefficient of determinations must be statistically significant at the 5 % significance level. Third, the VIF must be less than 10, and finally the regression coefficients must be statistically significant at the 5 % level.

Using stepwise regression, the alternative regression equations for estimating MAF can be summarized in Table 5.10. *The same procedure has also been used in each Q_T regression equations.* Column 7 and 8 in Table 5.10 describes the coefficient of determinations (R^2 and R_{adj}^2) between MAF with independent variable(s). It can be seen that the R^2 and R_{adj}^2 increases with more independent variables as expected.

In the Table 5.10, four possible equations for estimating the MAF were derived. Equation 1: MAF with *Area*, Equation 2: MAF with *Area* and *SIMS*, Equation 3: MAF with *Area, SIMS* and *APBAR*, and Equation 4: MAF with *Area, SIMS, APBAR* and *Elev*.

Regarding Equations 1, 2, 3 and 4, the regression coefficients (t-value in Table 5.10) for the *APBAR* parameter in Equation 3 and *Elev* in Equation 4 were not considered significant at the 5 % significant level. The results of hypothesis testing of regression coefficients for estimating the MAF are given in Table 5.11. It can be seen from Table 5.11 that the two independent variable regression equation was statistically significant at the 5 % level. Therefore, the two independent variable regression equation seem to be the best choice.

Table 5.9: Correlation matrix for logs of MAF and catchment characteristics of the Aceh river basins

	MAF	Area	Length	Elev	STMF	SIMS	Forest	MAR	APBAR
MAF	1								
Area	0.850	1							
Length	0.698	0.941	1						
Elev	0.317	0.283	0.221	1					
STMF	-0.656	-0.608	-0.563	0.434	1				
SIMS	-0.405	-0.813	-0.891	-0.055	0.420	1			
Forest	0.632	0.563	0.522	0.655	0.030	-0.303	1		
MAR	-0.250	-0.199	-0.100	0.479	0.651	0.140	0.403	1	
APBAR	-0.123	-0.588	-0.703	0.106	0.260	0.845	-0.019	0.038	1

Table 5.10: Stepwise regression for the Aceh province data for estimating MAF

Num. of Var.	Name of Variables	Constant coeff., Log C	Coeff. A_p	t-value	SE in Log	R^2	R_{adj}^2
1	Area	1.192	0.659	4.26	0.453	0.7219	0.682
2	Area SIMS	1.834	1.192 1.140	11.51 6.33	0.177	0.9637	0.952
3	Area SIMS APBAR	-4.884	1.133 0.760 1.250	13.18 3.35 2.17 *)	0.139	0.9813	0.970
4	Area SIMS APBAR Elev	-4.730	1.187 0.780 1.430 -0.187	17.72 4.73 3.34 -2.32 *)	0.101	0.992	0.984

*) Not significant (NS) at $\alpha = 5$ %.

Table 5.11: Results of hypothesis testing of regression coefficients (at $\alpha = 0.05$) for estimating MAF

Num. of Var.	Name of Variables	Constant coeff. Log **C**	Coeff. A_p	t-value	SE in Log	R^2	R_{adj}^2	VIF	P-value
1	Area	1.192	0.659	4.26	0.453	0.7219	0.682	-	0.004
2	Area	1.834	1.192	11.51	0.177	0.9637	0.952	3.0	< 0.001
	SIMS		1.140	6.33				3.0	
3	Area	-4.884	1.133	13.18	0.139	0.9813	0.970	3.3	< 0.001
	SIMS		0.760	3.35				3.9	
	APBAR		1.250	2.17 *)				7.5	
4	Area	-4.730	1.187	17.72	0.101	0.992	0.984	3.7	< 0.001
	SIMS		0.780	4.73				4.0	
	APBAR		1.430	3.34				7.6	
	Elev		-0.187	-2.32 *)				1.2	

Note : Log **C** : constant coefficient,
A_p : coefficient of independent variables,
p : the number of predictors.
SE : standard error estimate, in log,
VIF : variance inflation factors,
*) NS : not significant at $\alpha = 5$ %.
In this case, MAF is as the dependent variable.

Table 5.12 shows the logarithm of the observed MAF, catchment characteristics, estimated MAF using the logarithmic regression equation and residual error (E_{1i}).

Figure 5.5 shows the autocorrelation function (ACF) of the residuals. It can be seen that the correlation coefficients are all within the 95 % Bartlett's bands. Figure 5.6 shows the normal probability plot of the residuals. It can be seen that the plotted points are reasonably linear and was confirmed by the probability plot correlation coefficient (PPCC) test for normality at the 5 % significance level. Figure 5.7 shows the plot of the residuals versus the predicted values in log space. It can be seen that there is no evidence of heteroscedasticity. From the above tests, it appears that all regression assumptions are fulfilled.

Table 5.12: The logarithm of observed MAF, catchment characteristics, estimated MAF and the residual errors (E_{1i})

Site	Ln MAF_{obs}	Ln Area	Ln SIMS	Ln MAF_{est}	Residual error (E_{1i})
1	5.656	6.983	-4.046	5.556	-0.100
2	6.307	6.364	-2.790	6.246	-0.061
3	5.209	5.864	-3.282	5.090	-0.120
4	6.131	7.752	-4.227	6.266	0.135
5	5.613	7.233	-4.135	5.751	0.138
6	4.369	5.598	-3.590	4.424	0.054
7	6.131	7.703	-4.080	6.374	0.243
8	6.635	8.394	-4.785	6.396	-0.239
9	7.077	8.433	-4.271	7.029	-0.048
				$\sum E_{1i}$ =	0.000

```
          -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6
     +----+----+----+----+----+----+----+----+----+----
     1  -0.055          (              XX              )
     2   0.018          (               X              )
     3  -0.167          (           XXXXX              )
     4  -0.456          (    XXXXXXXXXXXX              )
     5   0.011          (               X              )
     6  -0.021          (              XX              )
     7   0.144          (               XXXXX          )
     8   0.026          (               XX             )

 The Bartlett's bands = ± 1.96 / N^1/2 = ± 0.65.
```

Figure 5.5: The ACF of the residuals for the two predictor equation of the MAF

```
Residuals

       -                                                          *
       -
       -
  0.16+
       -                                            *       *
       -
       -                                        *
       -
  0.00+
       -                                   *
       -                              *
       -                          *
       -                 *                 r  = 0.985
 -0.16+
       -
       -      *
       -
        ---------+---------+---------+---------+---------+--
               -1.20     -0.60      0.00      0.60      1.20

                                Z
```

Figure 5.6: The normal probability plot of the residuals for the two predictor equation of the MAF

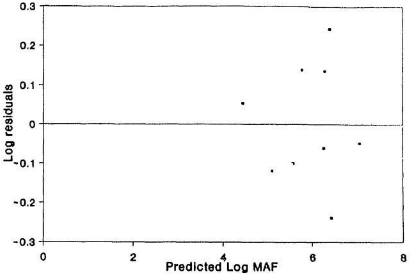

Figure 5.7: Plot of residuals versus predicted values (in log) to identify homoscedasticity of residuals of the two predictor equation for the MAF

The residuals of the regression for each Q_T were similarly checked for violations of the regression assumptions; they were all found to be approximately normally distributed, independent, and homoscedastic.

5.2.6 Results of the Multiple Regression Study

Two sets of results were found in the derivation of regression equations. The first set of results was the regression equation for the MAF, and the second set of results was the regression equations for each Q_T. The coefficients of the selected regression equations for estimating the MAF and each Q_T are given in Table 5.13.

Table 5.13: Coefficients of regression derived for estimating MAF and each Q_T for the Aceh river basins

	Coefficients of regression equations *)						
	Constant coefficient		Coeff. Area A_1	Coeff. Slope (SIMS) A_2	R^2 (%)	Adjusted R^2 (%)	SE in log (Q_{TX}) units
	log C	C					
MAF	1.834	6.259	1.192	1.138	96.37	95.16	0.177
Q_2	1.743	5.715	1.194	1.124	96.45	95.53	0.176
Q_5	2.231	9.309	1.200	1.212	95.96	94.61	0.184
Q_{10}	2.416	11.201	1.207	1.243	95.89	94.52	0.186
Q_{20}	2.607	13.558	1.208	1.272	94.85	93.13	0.208
Q_{50}	2.822	16.809	1.208	1.303	93.24	90.99	0.238
Q_{100}	2.969	19.472	1.206	1.320	91.89	89.19	0.209
Q_{200}	3.097	22.131	1.206	1.340	90.50	87.33	0.284

*) Regression equations are of the form:

$$\ln Q_{Tx} = \ln C + A_1 \ln Area + A_2 \ln SIMS \qquad (5.16)$$

or in power form:

$$Q_{Tx} = C \quad Area^{A_1} \quad SIMS^{A_2} \qquad (5.17)$$

where Q_{TX} = the estimated MAF or flood quantile at a given return period.

Bias correction of the logarithmic regression equations

Both the linear logarithmic model and nonlinear power model were used to estimate the MAF and each Q_T for the nine gauged sites which were previously used in the regression study. The estimated MAF using the logarithmic regression equation and residual error (E_{1i}) are given in Table 5.12. It apparent that the sum of the errors in log-space ($\sum E_i$) is zero when using the linear logarithmic model.

Table 5.14 shows the observed MAF, catchment characteristics, estimated MAF using the nonlinear power model Equation 5.17, residual errors (E_{2i}) and the estimated MAF and residual errors (E_{2i}') after adjusting the constant coefficients. In Table 5.14, the sum of the errors ($\sum E_{2i}$) is not zero.

Table 5.14: The observed MAF, catchment characteristics, estimated MAF and residual errors (E_{2i}) with the MAF and residual errors after adjusting the constant coefficient

Site	MAF_{obs} (m^3/s)	Area (km^2)	SIMS (m/m)	MAF_{est1} (m^3/s)	E_{2i}	$X_{i\ est}$	MAF_{est2} (m^3/s)	E_{2i}'
1	286	1078.0	0.0175	258.7	-27.27	41.3	262.4	-23.59
2	549	580.4	0.0614	515.9	-32.64	82.4	523.2	-25.31
3	183	352.0	0.0375	162.4	-20.63	25.9	164.7	-18.33
4	460	2326.0	0.0146	526.6	66.63	84.1	534.1	74.11
5	274	1384.0	0.0160	314.6	40.62	50.3	319.1	45.09
6	79	270.0	0.0276	83.4	4.40	13.3	84.6	5.59
7	460	2214.0	0.0169	586.5	126.46	93.7	594.8	134.79
8	761	4420.0	0.0084	599.5	-161.50	95.8	608.0	-152.98
9	1184	4598.0	0.0140	1128.6	-55.39	180.3	1144.6	-39.36
$\sum Y_{iobs}$ = 4236				$\sum E_{2i}$ =	-59.32	$\sum X_{iest}$ = 667.24	$\sum E_{2i}$' =	0.00
						f_0 = 6.348		

Note: $X_{i\ est}$: $Area_i^{A1}$ $SIMS_i^{A2}$, denominator variables in Equation 5.15,
MAF_{est2} : the estimated MAF after adjusting the constant coefficient,
E_{2i}' : the residual error after adjusting the constant coefficient.

The same bias correction procedure has also been used in each Q_T regression equations. Table 5.15 presents the final regression coefficients for use with Equation 5.17 for estimating the MAF and each Q_T after adjusting the constant coefficients.

Table 5.15: Final coefficients of regression derived after adjusting constant coefficient for estimating the expected MAF and each Q_T for the Aceh river basins

	Constant coefficient C	Coefficient Area A_1	Coefficient Slope (SIMS) A_2	R^2 (%)	Adjusted R^2 (%)
MAF	6.348	1.192	1.138	96.4	95.9
Q_2	5.805	1.194	1.124	96.5	95.9
Q_5	9.403	1.200	1.212	96.0	95.4
Q_{10}	11.476	1.207	1.243	96.7	96.2
Q_{20}	13.638	1.208	1.272	94.9	94.1
Q_{50}	16.943	1.208	1.303	93.3	92.3
Q_{100}	19.680	1.206	1.320	91.9	90.7
Q_{200}	22.434	1.206	1.340	90.[illegible]	89.2

5.3 Comparison of At-Site Flood Quantile Predictions

Comparison of flood quantiles estimated were divided into two parts. The first part is a comparison of flood quantiles for the nine gauged catchments that were estimated using the at-site analysis, the five index flood approaches, the multiple regression approach, and the IOH (1983) study. The second part is a comparison of flood quantiles for five ungauged catchments estimated using the same approaches, except the at-site analysis. For the L-moment approach, only the regional Gumbel distribution was considered in the comparison since the shape parameter of the GEV distribution was not significantly different from zero.

5.3.1 At-Site Flood Quantile Predictions: Gauged Catchments

A comparison of flood quantiles estimated for the gauged catchments using the methods mentioned above is given in Table 5.16 and Figure 5.8 for the site of Jambu

Aye River at Lhok Nibong as an example of estimated flood quantiles. The at-site estimated flood quantiles for the remaining eight basins are given in Appendix A. The following conclusion can be drawn from Table 5.16 and Figure 5.8 and the other estimates in Appendix A.

- The estimated quantiles by each method were quite variable, particularly for high return periods.
- The predictions of flood quantiles based on the index flood approaches were similar with little variability.
- Among the five index flood approaches of flood quantile predictions, the L-moment method gave estimates in between the other three index flood approaches (Dalrymple, NERC and station year). Dalrymple's approach predicted the highest flood quantile magnitudes, particularly in the high return periods. However, they are more consistent compared with the at-site analysis and the regression approach.
- The estimates based on the regression method provided lower predictions of flood quantiles than those given by the index flood method and the at-site analysis for five of the nine sites tested. Thus, in general, the regression approach gave lower predictions of flood quantiles.
- The estimates based on the at-site analysis were also less consistent than those generated by the index flood approaches.
- Since the regression estimates are based on at-site estimates and basin characteristics, it should be expected that the two results are similar. But,

there are no general agreement between the two methods. This is probably because of the small number of rivers used in developing the equations, and there may be much sampling error in the flood flow records or in the basin characteristics.

- The predictions of flood quantiles based on the IOH (1983) were much higher compared with the other methods that were used in this study. This could be due to the nature of data used by the IOH study, being based on flood flow records from both Jawa and Sumatra, collected in 1981 and only included five sites from Aceh Province.

Table 5.16: Comparison of at-site flood quantiles (Q_T) for the gauged site of the Jambu Aye River at Lhok Nibong

T (yrs)	Y variate	Estimated at-site Flood quantiles, Q_T						
		Dalrymple	NERC	PWMs / L-Moments	Station Year	IOH (1983)	Regression	At-site
2	0.367	731.32	732.56	728.28	730.56	631.63	612.08	751.7
5	1.500	926.14	895.70	904.45	900.26	890.37	678.22	855.7
10	2.250	1055.51	1004.52	1021.03	1012.13	1042.57	758.47	915.7
20	2.970	1178.79	1108.78	1132.98	1120.19	1209.99	790.68	968.4
50	3.902	1339.36	1243.47	1277.80	1259.46	1483.95	847.22	1031.3
100	4.600	1459.60	1344.69	1386.31	1363.71	1727.47	891.42	1075.5
200	5.296	1579.08	1445.14	1494.45	1467.97	2024.26	926.19	1117.5

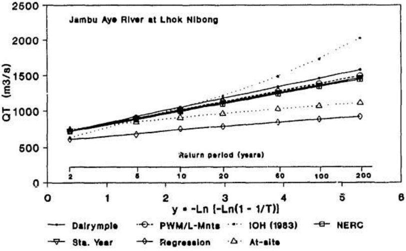

Figure 5.8: Comparison of at-site flood quantile estimates of gauged site for the Jambu Aye River at Lhok Nibong

5.3.2 At-Site Flood Quantile Predictions: Ungauged Catchments

A comparison among flood quantiles estimated for ungauged catchment using the index flood approaches, the regression approach and the IOH (1983), are presented in Table 5.17 and Figure 5.9 for the site of Tripa River at Gunung Kong, as an example of estimated flood quantiles. The at-site estimated flood quantiles for the other four basins are given in Appendix B. The following conclusions can be drawn from Table 5.17 and Figure 5.9, and other flood quantile estimates presented in Appendix B.

- The predictions of flood quantiles based on the index flood approaches were also more consistent than those generated by the regression approach, as indicated earlier for the gauged site flood predictions.
- Among the five index flood approaches of flood quantile predictions, the L-moment method remained in between the other three index flood approaches (Dalrymple, NERC and station year).
- The predictions of flood quantiles based on the IOH (1983) for the entire five ungauged sites were also higher compared with the methods that were used in this study.
- The predictions of flood quantiles based on the regression method provided lower predictions of flood quantiles in general.

Table 5.17: Comparison of at-site flood quantiles (Q_T) for ungauged site of the Tripa River at Gunung Kong

T (yrs)	Y variate	Estimated at-site Flood quantiles, Q_t					
		Dalrymple	NERC	PWMs / L-Moments	Station Year	IOH (1983)	Regression
2	0.367	454.87	455.35	452.98	454.40	391.55	490.59
5	1.500	576.05	557.11	562.56	559.95	591.47	533.72
10	2.250	656.51	624.80	635.07	629.53	710.77	600.08
20	2.970	733.19	689.65	704.70	696.75	844.23	630.53
50	3.902	833.07	773.43	794.77	783.37	1053.24	381.66
100	4.600	907.85	836.38	862.27	848.21	1243.37	721.48
200	5.296	982.17	898.86	929.53	913.06	1459.10	753.82

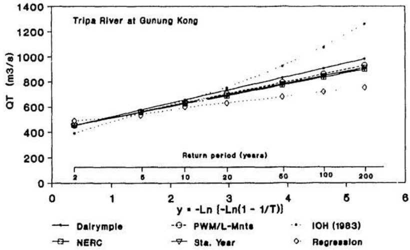

Figure 5.9: Comparison of at-site flood quantile estimates of ungauged site for the Tripa River at Gunung Kong

5.4 Summary

From the comparison of estimated flood quantiles for both gauged and ungauged sites the following points are emphasized. All five index flood approaches (Dalrymple (1960), NERC, PWMs/L-moments and station year) yielded similar flood magnitudes. The IOH (1983) flood estimations were higher compared with the current study, for all the sites that were tested. For most ungauged sites, the regression approach resulted in lower flood estimates. The flood predictions based on at-site data are inconsistent compared with the index flood approaches and regression approach for most gauged sites.

Chapter 6

Conclusions and Recommendations

Conclusions

From the findings presented in this thesis the following conclusions can be drawn.

1) From the discordancy measure (Table 4.2), the nine gauged catchments that were used in the study showed no unusual sites within the region. Therefore, the principal rivers in Aceh Province have some degree of homogeneity in terms of flood flow distribution within the region.

2) At-site flood quantile estimations (Table 4.4) were based on the best fit distribution at each site. In general, the EV1, LN2, GEV and LN3 distributions were the order of best fit distributions, respectively based on at-site analysis. Using L-moment diagrams the GEV was the selected distribution in the regional context. The shape parameter, k of the GEV distribution was found to be not significantly different from zero. Thus the EV1 distribution was the regional distribution used in the study.

3) The approximate regional quantiles X_T derived by the index flood approaches (Table 5.5) were similar for all approaches. The L-moments method (the linear combination of the PWMs method), however, was the best compromise for regional flood frequency estimations.

4) The results of the multiple regression analysis (Table 5.15) showed that the significant catchment characteristics were drainage area (*Area*) and channel slope (*SIMS*) for the prediction of flood quantile magnitudes for all return periods for the multiple regression approach and for $\bar{Q}$ for the index flood methods.

5) The flood quantile estimates at gauged sites based on at-site analysis technique resulted in inconsistent flood quantile predictions compared with the regional index flood approaches.

6) The flood estimates for both gauged and ungauged catchments based on the study by the IOH (1983) tended to over-estimate flood magnitudes, particularly in high return periods (probably due to the fact that a different data base was used for the IOH study).

7) The estimates of flood quantiles based on the regression approach gave lower estimates for both gauged and ungauged catchments compared with any of the index flood approaches.

Recommendations

1) The L-moment index flood method is recommended for regional flood frequency analysis in Aceh. The method is simple to apply for both gauged and ungauged catchments.

2) The mean annual maximum flood at ungauged sites can be estimated using multiple regression analysis with the drainage area and the channel slope as independent variables.

3) The regional index flood approaches and the regional multiple regression approach should be updated annually using new observations of flows and taking into account changing basin characteristics.

4) Development of water resources, particularly in the domains of irrigation, river improvement and flood control, municipal and industrial water use and hydroelectric projects has the potential for greater future development in this region. Improving the hydrologic data base is essential to ensure that the best development decisions are made. The principal requirements are:
   - continued high quality monitoring of data,
   - addition of streamflow measurement sites, and
   - establishment of hydrological networks.

5) With increasing number of gauged sites on the western side of Aceh Province, the Aceh region may be delineated as two hydrological regions in future studies.

References

Arnell, N. and Gabrielle, S. (1985). "Regional Flood Frequency Analysis with the Two Extreme Value Distribution: an Assessment Using Computer Simulation Experiments", *Workshop on Combined Efficiency of Direct and Indirect Estimations for Point and Regional Prediction*, Perugin, Italy.

Beable, M.E., McKerchar, A.I. (1982). *Regional Flood Estimation in New Zealand*. New Zealand National Water and Soil Conserve. Org., Water and Soil Technical Publication, No.20, 132 pages.

Beersing, A.K. (1990). *Regional Flood Frequency Analysis for the Island of Newfoundland*, Government of Newfounndland and Labrador, Water Resources Division, WRD-HM-90-1.

Benson, M.A. (1968). "Uniform Flood Frequency Estimating Methods for Federal Agencies". *Water Resources Research*, 4(5), pp.891-908.

Beran, M.A., Hosking, J.R.M. and Arnell, N. (1986). "Comment on: Two Component Extreme Value Distribution of Flood Frequency Analysis by Rossi et.al. (1984)", *Water Resources Research*, 22(2), pp.263-266.

Binnie and Partners. (1988). *Aceh De. . Unit*, Provincial Water Resources Development Plan Inventory of Water Resources Schemes: Update 1988, Report to the Director of Planning and Programming, Directorate General of Water Resources Development.

Bobee, B., Cavadias, G., Ashkar, F., Bernier, J. and Rasmussen, P. (1993). "Towards a Systematic Approach to Comparing Distributions Used in Flood Frequency Analysis", *Journal of Hydrology*, 142, pp.121-136.

Chow, V.T. (1964). *Handbook of Applied Hydrology*, McGraw-Hill, New York.

Cunnane, C. (1988). "Methods and Merits of Regional Frequency Analysis", *Journal of Hydrology*, 100, pp.269-290.

Cunnane, C. (1987). "Review of Statistical Models for Flood Frequency Estimation", in *V.P. Singh (Editor): Hydrologic Frequency Modelling*, Reidel, Dordrecht, pp.49-95.

Cunnane, C. (1978). "Unbiased Plotting Positions-A review,", *Journal of Hydrology*, 37, pp.205-222.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). "Probability Weighted Moments: Definition and Relation to Parameters of Several Distributions Expressible in Inverse Form", *Water Resources Research*, 15(5), pp.1049-1054.

Haktanir, T. (1992). "Comparison of Various Flood Frequency Distributions Using Annual Flood Peaks Data of Rivers in Anatolia", *Journal of Hydrology*, 136, pp.1-31.

Hosking J.R.M. and Wallis, J.R. (1985). "Estimation of the GEV distribution by the Method of Probability weighted moments", *Technometrics*, 27(3), pp.251-261.

Hosking J.R.M. and Wallis, J.R. (1992). "Regional Frequency Analysis Using L Moment", *Research Report RC 17843*, IBM Research, Yorktown Heights, N.Y.

Hosking, J.R.M. (1986). "The Theory of Probability Weighted Moments", *Research Report RC 12210*, IBM Research, Yorktown Heights, N.Y.

Hosking, J.R.M. (1990). "L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics", *J.R. Statist. Soc.*, Ser.B, 52(1), pp.105-124.

Hosking, J.R.M. (1993). "Some Statistics Useful in Regional Frequency Analysis", *Water Resources Research*, 29(2), pp.271-281.

Institute of Hydrology Wallingford (IOH), UK., Directorat Penyelidikan Masalah Air (DPMA), Ina. (1983). *Flood Design Manual for Java and Sumatra*, Report to Directorate General of Water Resources Development.

Jenkinson, A.F. (1969). *Statistics of Extremes*, In: Estimation of Maximum Floods, WMO No. 233, TP126, Tech.Note No:98, pp.183-228.

Kite, G.W. (1988). *Frequency and Risk Analyses in Hydrology*, Water Resources Publications, Littleton, Colorado 80161-2841, U.S.A.

Landwehr, J.M., Matalas, N.C and Wallis, J.R. (1979). "Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantities", *Water Resources Research*, 15, pp.1055-1064.

Maidment, D.R. (1993). *Handbook of Hydrology*, McGraw-Hill, New York.

McCuen, R.H., Leahy, R.B. and Johnson, P.A. (1990). "Problems with Logarithmic Transformations in Regression", *Journal of Hydraulic Engineering*, 116(3), pp.414-428.

McGuiness, J.L, Brakensiek, D.L. (1964). *Simplified Techniques for Fitting Frequency Distributions to Hydrologic Data*, US Agricultural Handbook, No.25, 42 pages.

Miller, D.M. (1984). "Reducing Transformation Bias in Curve Fitting". *The Amer. Statistician*, 38(2), pp.124-126.

Nash, J.E. and Shaw, B.L. (1966). "Flood Frequency as a Function of Catchment Characteristics", *Proc.Sympo. on River Flood Hydrology, Inst. Civ. Engrs.*, London,pp.115-136.

NERC. (1975). *Flood Studies Report Vol.1*, Hydrological Studies, Natural Environmental Research Studies (NERC), London.

Neter, J., Wasserman, W. and Kutner, M.H. (1985). *Applied Linear Statistical Models*, Irwing, Homewood, Illinois 60430.

Roskie, R.H. (1978). "Methods for Estimating Magnitude and Frequency of Floods in Arizona", *Open File Rep., U.S. Geol. Surv.*, Reston, Virginia.

Rossi, F., Firentino, M. and Versace, P. (1984). "Two Component Extreme Value Distribution for Flood Frequency Analysis", *Water Resources Research*, 20(7), pp. 847-856.

Tasker, G.D. and Moss, M.E. (1979). "Analysis of Arizona Flood Data Network for Regional Information", *Water Resources Research*, 15(6), pp.1791-1796.

Tasker, G.D. (1987). "Regional Analysis of Flood Frequencies", *in: V.P. Singh (Editor): Regional Flood Frequency Analysis*, Reidel, Dordrecht, pp.1-10.

U.S.W.R.C. (1976). *Guidelines for Determining Flood Flow Frequency*, Bulletin 17, Hydrology Committee, Water Resources Council, Washington. (Also revised versions, Bulletin 17A, 1977 and Bulletin 17B, 1981).

Vogel, R.M., McMahon, T.A and Chiew, F.H.S. (1993). "Flood Frequency Model Selection in Australia", *Journal of Hydrology*, 146, pp.421-449.

Wallis, J.R. (1980). "Risk and Uncertainties in the Evaluation of Flood Events for the Design of Hydrologic Structures", *Keynote address at Seminar on Extreme Hydrological Events-Floods and Droughts*, Erice, Italy.

Wild, J., Hall, J.K. (1982). "Aspect of Hydrology in the Province of North Sumatra-Indonesia", *Proc. Inst. Civil Engineering*, part 2, pp.85-108.

World Meteorological Organization (WMO). (1989). *Statistical Distributions for Flood Frequency Analysis*, Operational Hydrology, Report No.33, WMO-No.718, WMO, Genewa, 77 pages.

Appendix - A

<u>At-site flood quantile predictions for gauged sites</u>

The appendix gives the results of the estimated at-site flood quantiles. The results were compared to each other based on the at-site analysis, the five index flood approaches (Dalrymple, NERC, PMWs, Station year, L-moments), the multiple regression approach and the set of a study by the IOH (1983). The flood quantile magnitudes that were derived were then plotted and presented for the nine gauged sites that were used in the analysis. These nine gauged sites are the Krueng Aceh River at Kampung Darang, the Lambeusoi River at Sango, the Seunegan River at Ujung Blang, the Kluet River at Gunung Pudung, the Lawe Alas River at Sukarimbun, the Baro River at Klibeut, the Peusangan River at Beukah, Jambu Aye River at Lhok Nibong (see Section 5.3.1, Table 5.16 and Figure 5.8) and the Tamiang River at Kuala Simpang.

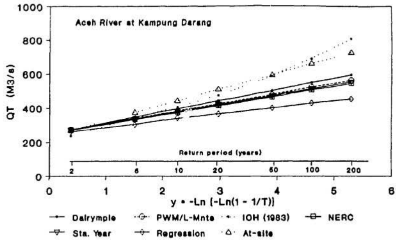

Figure A-1: Comparison of at-site flood quantile estimates of gauged site for the Aceh River at Kampung Darang

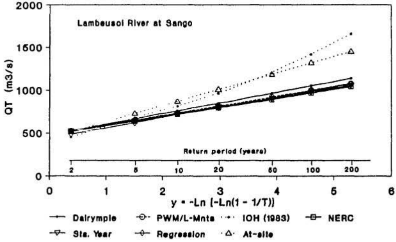

Figure A-2: Comparison of at-site flood quantile estimates of gauged site for the Lambeusoi River at Sango

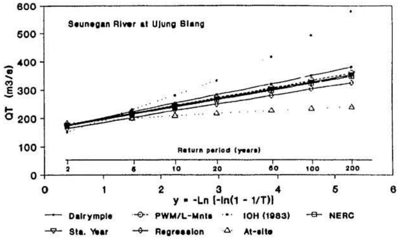

Figure A-3: Comparison of at-site flood quantile estimates of gauged site for the Seunegan River at Ujung Blang

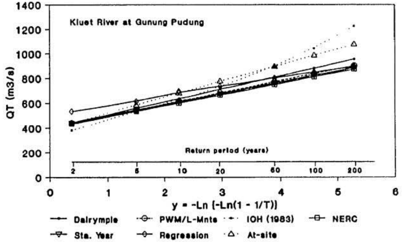

Figure A-4: Comparison of at-site flood quantile estimates of gauged site for the Kluet River at Gunung Pudung

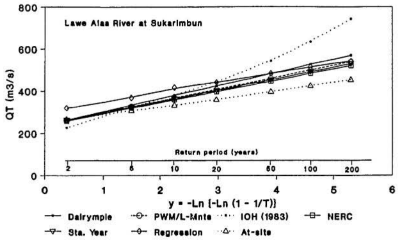

Figure A-5: Comparison of at-site flood quantile estimates of gauged site for the Lawe Alas River at Sukarimbun

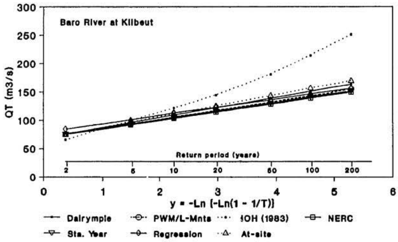

Figure A-6: Comparison of at-site flood quantile estimates of gauged site for the Baro River at Klibeut

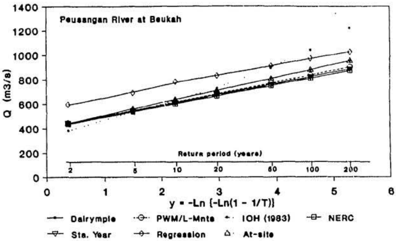

Figure A-7: Comparison of at-site flood quantile estimates of gauged site for the Peusangan River at Beukah

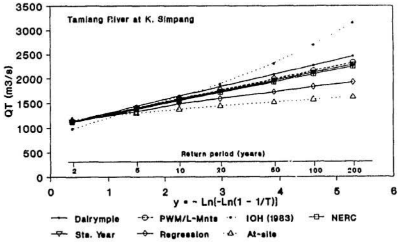

Figure A-8: Comparison of at-site flood quantile estimates of gauged site for the Tamiang River at Kuala Simpang

Appendix - B

At-site flood quantile predictions for ungauged sites.

The appendix gives the results of the at-site estimated flood quantiles. The results were compared to each other based on the five index flood approaches (Dalrymple, NERC, PMWs, Station year, L-moments), the multiple regression approach and the previous study by the IOH (1983). The flood quantile magnitudes derived were plotted and presented for the four ungauged sites that were not used in the analysis. Those five ungauged sites are the Teunom River at Tuwi Kareung, the Woyla River at Manu Tukut, the Tripa River at Gunung Kong (see Section 5.3.2, Table 5.17 and Figure 5.9), the Susoh River at Kota Tinggi and the Peurelak River at Peurelak.

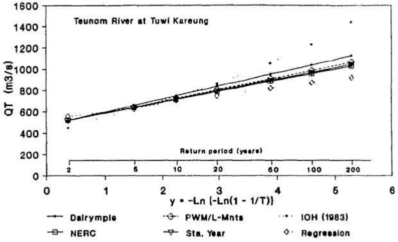

Figure B-1: Comparison of at-site flood quantile estimates of ungauged site for the Teunom River at Tuwi Kareung

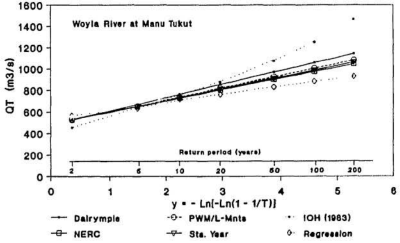

Figure B-2: Comparison of at-site flood quantile estimates of ungauged site for the Woyla River at Manu Tukut

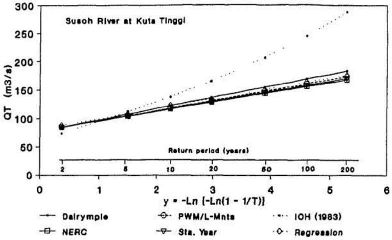

Figure B-3: Comparison of at-site flood quantile estimates of ungauged site for the Susoh River at Kuta Tinggi

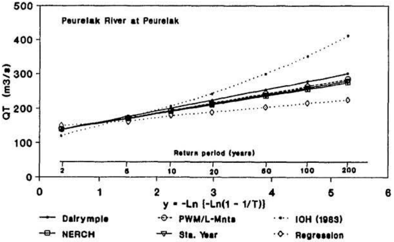

Figure B-4: Comparison of at-site flood quantile estimates of ungauged site for the Peurelak River at Peurelak

Appendix - C

Flow data used

The appendix gives the annual maximum of flood flow data of the Aceh Province region. These nine sites of flood flows along with their basin characteristics were used in the study. These flood flow records are presented in the following tables.

Table C-1: Annual maximum flood used for the nine rivers in Aceh Province

Site 1: Aceh River at Kp. Darang Elevation = 22 m		**Site 2**: Lambeusoi River at Sango Elevation = 12 m		**Site 3**: Seunegan River at Ujung Blang Elevation = 93 m	
Year	Flood flow m^3/s	Year	Flood flow m^3/s	Year	Flood flow m^3/s
1976	290	1977	230	1981	185.1
1977	352	1978	909	1982	161.0
1978	303	1979	553	1983	162.3
1980	188	1980	515	1984	194.0
1981	122	1981	323	1985	202.0
1982	226	1982	834	1986	210.0
1983	519	1983	321	1989	170.0
1984	294	1984	491		
1985	250	1985	803		
1986	247	1986	620		
1987	351	1987	438		
		1989	545		

Site 4: Kluet River at G. Pudung Elevation = 22 m		**Site 5**: Lawe Alas River at Sukarimbun Elevation = 430 m		**Site 6**: Baro River at Kp. Klibeut Elevation = 19 m	
Year	Flood flow m^3/s		Flood flow m^3/s	Year	Flood flow m^3/s
1981	545	1980	267	1972	46.0
1982	366	1981	297	1973	89.4
1983	409	1982	326	1974	113.0
1984	296	1983	326	1975	115.0
1985	481	1984	227	1976	96.3
1986	663	1985	231	1977	74.4
		1986	251	1978	70.2
		1987	269	1979	70.9
				1980	60.0
				1981	62.5
				1982	69.6

Table 5.30 Comparison of peakflows obtained using the two modelling techniques with observed peakflows.

No.	Observed peakflow (m^3/s)	Estimated peakflow based on Regression model (m^3/s)	% Error (m^3/s)	Estimated peakflow based on TH (m^3/s)	% Error (m^3/s)
1	2	3	4	5	6
1	95.6	139.5	45.9	112.02	17.1
2	193.1	202.4	4.8	226.54	17.3
3	197.2	220.3	11.7	229.01	16.1
4	261.1	411.3	57.5	285.55	9.4
5	275.5	315.9	14.7	304.91	10.7
6	284.6	503.7	76.9	401.99	41.2
7	290.7	170.1	-41.5	334.4	15
8	300.4	417.2	38.9	373.15	24.2
9	334.7	549.9	64.3	345.61	3.2
10	339.4	500.4	47.4	327.62	3.4
11	356.3	546.5	53.3	369.87	3.8
12	461.5	685.3	48.5	514.47	11.5
13	764.4	851.1	11.3	743.77	2.7
14	803.2	741.7	7.6	773.63	3.7
15	824.3	924.9	12.2	778.16	5.6
16	831.2	749.9	9.8	808.49	2.7