DESIGN FLOOD ESTIMATION METHODS FOR RIVERS WITH EXTENSIVE TIDAL INTERACTION ZONES

YEO HOWE LIM


CENTRE FOR NFLD. STUDIES
MEMORIAL UNIVERSITY
OF NEWFOUNDLAND

DESIGN FLOOD ESTIMATION METHODS FOR RIVERS WITH EXTENSIVE TIDAL INTERACTION ZONES

By

© Yeo Howe Lim

A thesis submitted to the School of Graduate Studies
in partial fulfilment of the requirement for the degree of
Doctor of Philosophy in Engineering

Faculty of Engineering and Applied Science
Memorial University of Newfoundland

November 2002
St. John's Newfoundland Canada

ABSTRACT

Flooding problems may never be solved as long as people live in proximity of riverbanks with high inundation risk. A rational approach well embraced by civil engineers is to derive flood estimates and adopt mitigation measures corresponding to an acceptable risk level. However similar approaches have not been well addressed in many rivers with extensive tidal interaction zones (TIZ), such as those found in Sarawak, Malaysia.

The primary objective of this research is to streamline and develop a set of design flood estimation methods applicable to the TIZ with varying degrees of sophistication. Methods in use such as the direct frequency analysis method, the regression-like rating-curve method, the joint probability method and the hydrodynamic modelling method are evaluated. Although these methods have their merits, they have many inherent limitations and constraints. Two new methods that are statistically based yet do not compromise the physical meanings of the processes involved are developed. The first is a low-pass filter method involving Fourier transforms resulting in filtering of low frequency river level series. A second new method is based on wavelet de-noising techniques that considers tidal contributions as noise in a smooth function. An algorithm ensuring a consistent separation technique is developed. Various wavelet families are investigated together with different decomposition levels and thresholding options.

Conventional flood frequency analysis techniques are subsequently applied directly to the separated flow series.

The secondary objective of this research is to establish a comprehensive flood estimation procedure for ungauged catchments in the study area based on regional flood frequency analysis. The need arises because inflow floods deriving from the major ungauged tributaries, which affect flows in the TIZ, have to be estimated in some of the methods.

After comparing all the methods, it is recommended that an appropriate number of water-level observation stations should be established at strategic locations in the TIZ on a continuous long-term basis. With these data, the effects of climatic change on sea level rise and the subsequent effects on flood estimations in the TIZ can be assessed accurately.

In conclusion, the method to be chosen for design flood estimations in the TIZ is very much dependent on data availability. If long-term water level data in the TIZ is available, the new fundamental method using wavelet de-noising gives direct and reliable results. By considering the interaction effects, it is superior to the conventional direct frequency analysis. Since wavelet methodology is still evolving, there are further potential applications for this particular method.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor Dr. Leonard M. Lye for his continual guidance, financial support and encouragement for pursuing the Ph.D. study programme at the Memorial University of Newfoundland (MUN). I wish to thank Dr. Tahir Husain, Faculty of Engineering, MUN and Ms Susan Richter, Acres Consulting Services Limited, for giving their comments and suggestions as members of the supervisory committee. Special thank goes to Dr. James J. Sharp for his teaching and general advises. Without the financial supports from MUN and the study leave package provided by the University of Malaysia, Sarawak (UNIMAS), this study could not have been possible.

My heart-felt appreciation goes to my wife, Emily, for her relentless supports and sharing my aspirations during the study period. I appreciate my beloved children Felix, Wendy, and Cindy, for being understanding and contributed joys during the period of thesis preparation.

A number of people have helped in data gathering. I wish to thank Mr. John Tan, Director of the Department of Irrigation and Drainage, Sarawak, Malaysia, for permitting the use of the hydrological records in Sarawak, and the staff at the department's Hydrology Branch, especially Mr. Lim Hiok Hwa and Mr. Chai Kian Phin, for delivering the requested data. I wish to acknowledge Mr. John Macgregor and Mr.

Thiam-Yew Wee of Shell Brunei for supplying the tide records at Lumut, Brunei. The Finders University of South Australia, Adelaide, Australia provided a copy of the report written by Easton (1977) which contains some programs for tidal analysis and prediction.

I was also fortunate to get the advice of Dr. Alwell Oyet, Department of Mathematics, MUN, and Dr. Christopher Torrence, Research Systems, Inc., Boulder, Colorado, USA, on topics related to wavelets.

Lastly, I wish to thank the academic staff, the support staff, friends and relatives, who have helped in one way or another, during the period of my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACF	Autocorrelation function
ARI	Annual recurrence interval
CD	Chart datum
cdf	Cumulative distribution function
DFT	Discrete Fourier Transform
DID	Department of Irrigation and Drainage, Sarawak, Malaysia
EV	Extreme Values
FT	Fourier Transform
FFT	Fast Fourier Transform
IFT	Inverse Fourier Transform
IFFT	Inverse Fast Fourier Transform
GEV	Generalized Extreme Values
GLO	Generalized Logistic
GPA	Generalized Pareto
JPM	Joint Probability Method
L&S	Land and Survey Department, Sarawak, Malaysia
LNO	Log-normal
MACD	metres above chart datum

MC	Monte Carlo
MOM	conventional method of moments
NOAA	National Oceanic and Atmospheric Administration, USA
PACF	Partial autocorrelation function
pdf	Probability density function
TIDEDA	Time-dependent-data
TIZ	Tidal interaction zone
WMO	World Meteorological Organization

LIST OF SYMBOLS

A	matrix of sums of squares and cross products defined in equation (5-4)
a_J	wavelet approximation coefficients
a_k	phase of harmonic component k
A_k	the amplitude of harmonic component k
AREA	catchment or basin area; transformed variable A
B_4	bias of t_4^R
Ck	coefficient of kurtosis
C_m	amplitude of the mth harmonic constituent of the function $H(t)$
Cs	coefficient of sknewness
Cv	coefficient of variation
D_i	discordancy measure for catchment i
$d_1 .. d_J$	wavelet detail coefficients
$d_{j,k}$	wavelet coefficient of expansion
$E(t)$	discharge (fluvial flows) induced water level in the Wavelet De-noising Method
f	a general function
$f(t)$	a smooth signal

F	a factor indicating the type of tides
$F(.)$	cumulative distribution function (cdf)
$f(H)$	probability density function of sea level H
$F(x)$	cumulative distribution function of annual sea level maxima x
$F_Z(\eta)$	probability of exceedance for a particular level η in the TIZ
η	an extreme water level in the TIZ
$g(Q)$	probability density function of discharge Q
$G(x)$	cumulative distribution function of annual sea level maxima x
$G_Z(\eta)$	the probability of non-exceedance of a level η in the TIZ
$G(\omega)$	transfer function of digital filter
g_m	phase lag of the mth harmonic constituent of the function $H(t)$
h	an instantaneous sea level
H	homogeneity statistic
$H(t)$	sea level function
$H(X)$	entropy of random variable X
$H(X,Y)$	total entropy of random variables X and Y
$H(Y)$	entropy of random variable Y
H_{max}	maximum level of tidal influence in a TIZ
H_0	mean sea level
$H_0(t)$	mean sea level series
H_p	annual peak water level in a TIZ
$I(t)$	interaction term due to the interaction of fluvial flows and tides

l_1	mean
$K(t)$	filtered tidal series in the TIZ derived by the Fourier Method
K_{max}	Maximum series of filtered tidal series, either annually or hourly
$LATI$	latitude; transformed variable LA
$LONG$	longitude; transformed variable LO
L_T	flood quantile (water level) for the return period of T years
L_{max}	maximum observed water level in a TIZ
m	number of year in the consideration of correction factor θ
$M_{p,r,s}$	Probability Weighted Moments
m_r	r^{th} sample moment
$m_r^{'}$	sample moment about the origin
$\mathbf{M_1}$, $\mathbf{M_2}$, ...	tidal components due to the moon
n_i	record length of site i
N	sample size; number of hourly observation in a year
$N(t)$	component depicting tidal dynamic processes in the Wavelet De-noising Method
p	probability of non-exceedance
$p_h(.)$	probability density function for sea levels (sea condition)
p_i	plotting position
$P_R(.)$	probability density function of a filtered river series (TIZ)
$p_s(.)$	probability density function for a surge regime in the sea conditions
$p_T(.)$	probability density function for a tidal regime in the sea conditions

$P_K(.)$	probability density function of a filtered tide series (TIZ)
$P_Z(.)$	joint probability density function of $P_R(.)$ and $P_K(.)$ (TIZ)
q	instantaneous fluvial discharge
$Q(t)$	fluvial discharge function
$q_i(x)$	probability that i consecutive sea levels are greater than x
Q_m	index flood, or median of the at-site annual maximum peak discharge series
Q_T	flood quantile of a site for return period T years
r	an instantaneous filtered river level; r largest maxima series
R	annual rainfall
$R(t)$	impulse function
$R(t)$	filtered river series in the TIZ derived by the Fourier Method
R_d	de-noised river series obtained by Wavelet De-noising Method
R_{max}	Maximum series of filtered river series, either annually or hourly
s	an instantaneous surge level
$s(t)$	a time signal
$S(t)$	meteorological surge component due to pressure variations in the atmosphere
$SPECQ$	specific discharge; transformed variable SQ
$\mathbf{S_1}, \mathbf{S_2}, \ldots$	tidal components due to the sun
T	return period in flood frequency analysis; period in tidal phenomenon
t	time; L-CV (a L-moment ratio)
t^R	L-CV of a region R
t^R_3	L-skewness of a region R
t^R_4	L-kurtosis a region R

$T(X,Y)$	Transinformation of random variables X and Y
T_n	tidal noise series obtained by using Wavelet De-noising Method
T_P	return period for a POT model
T_Z	return period for a level in the TIZ
T5RA	5-year return-period storm with 12-hour duration (similarly for *T10RA*, *T20RA* & *T50RA*); transformed variable *T5*, *T10*, *T20* and *T50*.
u	number of storm per year in the consideration of correction factor θ
$\mathbf{u_i}$	vector containing L-moment ratios for catchment i
$\overline{\mathbf{u}}$	unweighted regional average for $\mathbf{u_i}$
V	weighted standard deviation of at-site sample L-CVs
X	discrete random variable in entropy consideration
x	instantaneous sea level
$X(t)$	general time series
X_T	regional growth factor in index-flood equation
$Y(t)$	general water level series
y_n	a discrete observational time series
Y_n	a discrete Fourier Transform (DFT) of y_n
z	normal variate
$Z(t)$	water level series in tidal interaction zone
Z^{DIST}	goodness-of-fit measure in regional approach by L-moments
$Z_n(\omega)$	DFT of $Z(t)$

α_k	a term used to adjust for the phase at time t=0 in the equilibrium tide
α , β	special cases of Probability Weighted Moments
α , ε, k	parameters of distributions
δ	declination of perturbing body at position M
$\delta c(x)$	a function for wavelet shrinkage process
$\xi(t)$	noise of a time series
ζ	zenith angle of the moon as seen from the earth
η	an instantaneous water level in the TIZ
θ, θ_M	colatitudes of an observer on earth at position P; perturbing body at position M
θ	correction factor for hourly series
λ_1	sample mean
λ	threshold value for wavelet shrinkage
λ_i	general measures of location (L-moment)
λ_1	sample mean
λ_2	scale (L-moment)
μ_r	central moment
μ_r'	r^{th} moment about the origin
μ_v	mean of simulated Vs
σ	an estimate of the scale of noise in wavelet analysis; standard deviation
σ_4	standard deviation of t_4^R
σ_k	radial frequency of harmonic component k

σ_v	standard deviation of simulated *V*s
τ	sample L-CV
τ_3	sample L-skewness
τ_4	sample L-kurtosis
Φ, Φ_M	longitude on celestial equator plane of an observer at P and perturbing body at M
Φ'	difference between the longitude of an observer at P and perturbing body at M
$\phi(t)$	father wavelet
$\psi(t)$	mother wavelet

LIST OF APPENDICES

Chapter 1 Introduction

1.1 Importance of Design Flood Estimation

Despite the progress in science and technology, flooding of lands near rising rivers is an age-old problem that still haunts millions of people every day. The problem may never be solved, as people like to live in proximity to the riverbanks despite the risk of being flooded. A rational approach is to investigate the occurrence of historical floods and adopt appropriate mitigation measures.

The technique of associating the probabilities of occurrence with any extreme flood levels, or peak discharges for a particular locality, is generally known as *design flood estimation.* Many civil engineering projects such as roads, dikes, housing estates, industrial parks, bridges, etc. are sited along rivers. The designs of these projects require design flood levels that are usually determined based on a certain acceptable risk criterion. A design flood level or *stage* associated with a certain failure probability or probability of exceedance is normally derived from a flood frequency analysis. Engineers typically use the return period (T in year), which is the reciprocal of the probability of exceedance, $F(x) = 1/T$, as a major criterion for project planning and design for projects

related to floods. The probabilistic approach has proved attractive in view of the random nature of flood phenomenon (Caissie and El-Jabi, 1990). The technique has grown in sophistication with the advances in probabilistic and statistical methods. Nevertheless, the method is largely used in a confined domain where flood discharges are observed and measured at distinctly non-tidal zones with the water flowing in the downstream direction. Most discharge measuring stations are sited upstream of the river reaches where tides are not predominant.

Owing to the large economic and environmental impacts of floods, flood frequency analysis remains a subject of great importance and interest (Bobée and Rasmussen, 1995). It remains a very active area of investigation as evidenced by the numerous research papers in the last 20 years.

1.2 Tidal Interaction Zones (TIZ)

The lower reaches of many rivers in the world are important areas, as there lie many cities and major settlements. Many of these rivers have daily tidal excursions that create complex flow conditions along the riverine areas. In some rivers with surrounding low-lying lands, the tidal propagation can be found at some considerable distances upstream of the estuaries. The stretch along a river where the mixing occurs is commonly known as the *tidal interaction zone*, abbreviated hereafter as TIZ. Such an area is Sarawak, an eastern state of Malaysia on Borneo Island, and it has been chosen as the study area for this thesis. It has many rivers with very mild bed slopes and very extensive tidal interaction zones.

1.3 Problems of Flood Estimation in Tidal Interaction Zones

As mentioned in Section 1.1, traditional flood frequency analyses are used at gauged river stations that are free from tidal influence. The flows are unidirectional under the influence of gravity characterized by mostly stationary flows. Frequencies of discharge exceedance are accounted for and a probability plot can be developed for each gauged site. A regional flood frequency curve can subsequently be developed based on several sites with homogeneous basin characteristics. Sites without any data can rely on regression relationships of the gauged sites with some physical parameters, coupled with the regional frequency curve, to derive the desired design flood discharges. The corresponding design water levels can be ascertained by using hydraulic evaluations of the flows at the site or interpolation from the stage discharge curve specifically developed for the site.

At the mouth of rivers, where tides predominate, the rise and fall of water level essentially change in tandem with the tides of the ocean, sea or coastal waters in proximity. The behaviour of water level fluctuations at the river mouths is typically well understood with a wealth of literature available explaining the oceanography of coastal waters and tides (Dronker, 1964; Bowden, 1983). The water levels can be predicted fairly well within certain acceptable limits in the absence of storm surges caused by changes in local weather conditions.

However the situation is very much different in river reaches in the TIZ where the tides play a major influence on the river flows. The exceedance of a particular water

level can be due to many combinations of tides and floods. This makes the straightforward application of traditional flood frequency approach invalid for these locations.

Practising consultants normally approach the problem on a case by case basis. If resources are available, an elaborate hydrodynamic flow model may be established for the stretch of river concerned. The approach in computation of flows is deterministic while the approach for input of boundary conditions in design flood estimation is probabilistic. It may require months of detailed topographic survey and weeks of model calibrations. Owing to the demand of the model for data at multiple sites, not all the essential data may be available for calibration. Even then, as discussed in detail in Chapter 3, there are inherent limitations. When neither time nor resources permits the flow model to be established, very rough or conservative approximations may be adopted for deriving the design flood levels, physio-statistical approach, a combination of physical assessment and statistical interpretation, is an alternative to the two extremes as can be seen in Acres Consulting Services Limited (1987) and Polonsky (1996) to assess flooding events in estuaries.

1.4 Research Objectives and Goals

The primary objective of this research is to streamline and develop a set of efficient estimation methods that are applicable for flood frequency analysis in the TIZ. Certain methods are applicable to situations where there are limited amount of

information available while others are intended for cases where more detail data are available.

Due to the complexity of the problem, several of the latest available numerical and statistical methods, such as wavelets analysis methods and L-moments methods, are exploited. At the same time, traditional methods such as the regression methods and Fourier analysis are also elaborated.

Verification of the methods proposed are carried out whenever possible. The observed data in selected tidally affected rivers in Sarawak, Malaysia are utilised for such purpose.

The secondary objective of the proposed research is the development of regional flood estimation equations for non-tidal zones of Sarawak using the state-of-the-art L-moment approach. The results are necessary for deriving flood estimates in the downstream TIZ. From the literature review, to date regional flood frequency analysis has not been carried out for Sarawak.

1.5 Scopes and Limits of Research

1. The "study area" is limited to the basin areas that are covered by the river gauging network and not the whole land area of Sarawak.
2. The research is limited to the analysis of existing data collected by the collecting agency, namely the Department of Irrigation and Drainage, Sarawak, Malaysia (DID) and Nation Mapping and Survey, Malaysia, and Land and Survey

Department, Sarawak (L&S). Shell Berhad, Brunei is the only non-governmental agency that provided data.

3. The hydrodynamic modelling approach is one traditional approach that requires a large amount of site-specific physical data and many months of input and analysis to model a single basin. This approach is examined, discussed and compared in general in Chapter 2.
4. Any methodology proposed in general is explored thoroughly wherever possible and evaluated by comparison with results generated by other methods or local flood information.
5. Short gaps in water level data are treated as missing. Alternatively, it may be replaced by means of the existing data segments.
6. The data provided by the water level recorders are assumed to be correct. It is assumed that the operating agencies had carried out any required adjustment during the data processing stage.

1.6 Outline of Thesis

The thesis is organized into three major groups of chapters:

1. Introduction to the problem, overviews and approaches: Chapter 1, 2 and 3
2. Main methodologies: Chapter 4, 5, 6, 7 and 8
3. Summary and conclusion: Chapter 9 and 10

Chapter 2 surveys the existing literature on flood estimation methods with an objective of providing an overview of the methods that have been used and some

background theories relevant to certain specific topics such as Fourier and wavelets methods. Chapter 3 identifies the study area with an in-depth review of the existing hydrometric networks and the records available.

The main methodologies proposed are briefly introduced in Chapter 4, and then followed up by elaborations on the specific methods in Chapter 5, 6, 7 and 8. Chapter 5 discusses the regional frequency analysis in non-tidal zones, which is essential for deriving inflows for the downstream points of interest in the TIZ. The traditional Rating Curve Method and Joint Probability Method and are dealt with in Chapter 6 followed by the application of Fourier Method to this problem in Chapter 7. A whole new approach using Wavelet Method is introduced and elaborated in Chapter 8. Summary discussions can be found in Chapter 9 and the thesis is concluded in Chapter 10. MATLAB programs that are developed for various processing of the data and the subsequent analysis are found in **Appendix A1** through **Appendix A10**. Results of the analyses that are of interest but are too lengthy to be included in the main text are presented in the other appendices which are numbered with prefixes of B, C and D.

Chapter 2 Literature Review

2.1 Overview

A proposal of any solution to a problem is only justifiable after a thorough survey and understanding of the existing solution(s) to the problems in hand or problems with some similar characteristics. Conforming to this rationale, all the existing theories and methods that are relevant to the estimation of design floods in the TIZ are reviewed in this chapter. Since the research involves many interdisciplinary fields such as hydrology, hydraulics, oceanography, statistics, and signal processing, this chapter also serves as a reference document for future research in the area.

The existing approaches in design flood estimation can broadly be divided into two:

(i) Statistically based

(ii) Physically based

These approaches are critically discussed and reviewed in Sections 2.2 and 2.3 respectively. In solving many hydrologic and hydraulic problems, a thorough understanding of the boundary conditions, which govern the outcomes of potential models to be used, is essential. A brief review of the literature related to the tidal

phenomenon which governs the downstream boundary conditions of the TIZ are provided in Section 2.4. Fourier Transform Methods used in traditional signal processing and Wavelet Analysis Methods, a recent advance in non-parametric data analysis, are reviewed in Section 2.5 and 2.6 respectively.

2.2 Statistically Based Approaches

Statistically based approaches refer to the analysis of raw data collected from a site or region using state-of-the-art statistical tools in deriving probabilistic functions or frequency distributions pertaining to flood quantiles. Traditional methods that dominate the approach are flood frequency analysis and a regional version of the analysis.

2.2.1 Flood Frequency Analysis

Flood frequency analysis is a standard procedure for the planning and design of water resources projects and other civil engineering works. It provides the probabilistic assessment of the magnitude of floods associated with a certain risk tolerance level. Its application has been mainly in the determination of flood quantiles in terms of peak discharges in streams or rivers.

Existing techniques used are varied and many. The latest state of the art techniques can be found in a wealth of literature such as Vogel (1986), Cunnane (1987, 1988, 1989), Burn (1990), Stedinger et.al. (1993), Lye (1993), Bobée and Rasmussen (1995), Hosking and Wallis (1990, 1993, 1997) and, Rao and Hamed (2000).

2.2.1.1 Concept of Flood Frequency

The recurrence interval of a flood or return period T is related to the corresponding flood magnitude in a flood frequency analysis. The primary objective is to provide the basis for an inference on the flood magnitude at a particular river site given an accepted risk level that is expressed in terms of the reciprocal of the return period ($1/T$) or the exceedance probability p. An observed flood of magnitude q, normally quantified in units of flow rate, with a return period T may be exceeded once in T years. The probability of exceedance is given by

$$p = P(Q_T > q) = \frac{1}{T} \tag{2-1}$$

and the cumulative probability of non-exceedance then becomes

$$F(Q_T) = P(Q_T \leq q) = 1 - P(Q_T > q) = 1 - \frac{1}{T} \tag{2-2}$$

The return period or recurrence interval can be expressed as

$$T = \frac{1}{1 - F(Q_T)} \tag{2-3}$$

The main task of flood frequency analysis would be, on a single station basis, to find a suitable statistical distribution function that represent $F(Q_T)$.

2.2.1.2 Single Station Approaches and Flood Models

A single station refers to an individual river gauging station where river flows have been observed continuously over a certain number of years. The level of sophistication in instrumentation to observe river flows varies from station to station. The

most basic setup would be to manually observe the water level or stage at certain fixed time interval, typically twice a day. The observed water levels are then related to flows by a stage-discharge or rating curve that is established specifically for the station by measuring sufficient sets of river flows at different stages. Ideally flows at high and low stages should be measured to minimise extrapolation of the rating curves, which can be erroneous. Recently, automatic stage observing recorder is becoming a standard instrumentation at most river stations. In some small rivers, flow measurement weirs, flumes or acoustic-dopplers flow meters are installed to avoid the painstaking task of establishing rating curves.

Three different flood frequency models are identified by Cunnane(1989), namely the annual maximum series (AM) model, the partial duration series or peaks over threshold (POT) model and the time series (TS) model. The AM model considers only one peak value per year while the POT model consider all peaks above a certain base value. The later avoids the situation when the other peaks within a year are greater than the maximum peaks of other years (Kite, 1977; Chow *et al.* 1988). The return period T_P for a POT model is related to the return period T of an AM model by (Chow, 1964):

$$T_p = \left[\log\left(\frac{T}{T-1}\right)\right]^{-1} \tag{2-4}$$

Recent application of the POT model in New Brunswick, Canada is reported by El-Jabi *et al.*(1998).

The TS model considers flows at equally spaced intervals of time, typically daily interval, to signify a stochastic process in continuous time (Natural Environment

Research Council, 1975). The flow series model is then written as the sum of trend, seasonal and stochastic components.

Despite the various considerations of the other two models, the simpler AM model is still the most popular. Most recent publications still prefer to discuss the flood analysis based on the simpler model (Hosking and Wallis, 1997; Rao and Hamed, 2000). The annual peaks are assigned an initial estimate of probability of non-exceedence F called a "plotting position"(p_i) as the data are plotted on probability paper to check whether they follow a particular distribution. There are many commonly used plotting position formulas and research is still continuing (Rao and Hamed, 2000). Cunnane(1989) studied various formulas and proposed a compromised formula given by:

$$\hat{F}(x_i) = p_i = \frac{i - 0.4}{N + 0.2} \tag{2-5}$$

where i is the rank of the flood peaks in ascending order and N is the sample size. This appears to be the most commonly accepted formula, but agencies may prefer to use their own traditionally accepted formula. Rao and Hamed (2000) prefers to use a formula proposed by Hosking and Wallis (1990) which was developed specifically for use in the L-moment method of parameter estimation:

$$p_i = \frac{i - 0.35}{N} \tag{2-6}$$

2.2.1.3 Distributions Selection

After choosing a suitable flood model, the next step of a flood frequency analysis is to hypothesise the statistical functions $F(Q_p)$ and fit the underlying distribution using

one of the three main methods, namely the Method of Moments (MOM), Maximum Likelihood Estimation (MLE), and L-Moments Method (LM).

2.2.1.4 Conventional Method of Moment

Moments about the origin or about the mean are used in the MOM to characterise probability distributions. The conventional mathematics of expectations is used to obtain other measures of spread such as variance and asymmetry such as skewness. The r^{th} moment about the origin is

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x)dx \tag{2-7}$$

and

$$\mu_1' = \mu = \text{mean} \tag{2-8}$$

The central moments μ_r is given by

$$\mu_r = \int_{-\infty}^{\infty} (x - \mu_1')^r f(x)dx \,, \mu_1 = 0 \tag{2-9}$$

The relationship between μ_r' and μ_r is derived by Kendall and Stewart (1963). The sample moments m_r' and m_r are defined as

$$m_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r \tag{2-10}$$

and $m_1' = \bar{x} = \text{sample mean}$

$$m_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})_i^r \,, m_1 = 0 \tag{2-11}$$

Cunnane(1989) has proposed some correction measures to the sample moments for bias. Conventional moment ratios such as

$$\text{coefficient of variation, } C_v = \frac{\sqrt{\mu_2}}{\mu_1'} \tag{2-12}$$

$$\text{coefficient of sknewness, } C_s = \frac{\mu_3}{\mu_2^{3/2}}, \text{ and} \tag{2-13}$$

$$\text{coefficient of kurtosis, } C_k = \frac{\mu_4}{\mu_2^2} \tag{2-14}$$

are obtained by ratios of the various central moments. Sample moment ratios are also obtained similarly using m_r in place of μ_r. Extensive investigations are found on the moments of distributions and their properties such as Kirby (1974) and Landwehr *et al.* (1978).

The distributions that are found to be best suited to flood flow series are Lognormal, Log-Person Type III, Generalised Extreme Values (GEV), Generalised Pareto, Gumbel (Extreme Value Type 1 or EV1), Log-Gumbel(Extreme Value Type 2 or EV2), and Log-Logistic. Moment ratio diagrams of C_s versus C_k and C_s^2 and C_k are commonly used for preliminary check on the suitability of a particular distribution. The probability plots of the flood quantiles obtained from proposed distributions and the sample flood series should lie on a straight line for two parameter distributions. This is because the distribution function is

$$F = G\left(\frac{Q-a}{b}\right) \tag{2-15}$$

where a and b are the distribution location and scale parameters respectively. The observed values of Q is $\hat{Q}$, then the relationship between $\hat{Q}$ and $G^{-1}(F)$ is linear as given by

$$\hat{Q} = a + bG^{-1}(F) \tag{2-16}$$

The choice of distribution is often based on certain statistical test to determine the goodness-of-fit. Tests such as Least Squares Test (Kite, 1977), Chi-square Test, Kolmogorov-Simirnov Test (KS), Probability Plot Correlation Test (PCC) (Filliben 1975; Lye 1993) and Anderson-Darling Test (AD) (Lye, 2000; Lim and Lye, 2000) were used.

2.2.1.5 Maximum Likelihood Estimation Method (MLE)

This method is the most common statistical method of parameter estimation. It is based on the concept of calculating values of parameters that maximises the probability of obtaining the particular sample, in this case the annual flood peaks. The likelihood of the sample is the total probability of drawing each flood peak, which is the product of all the individual flood probabilities. The product is differentiated with respect to the parameters, and the derivatives are set to zero to enable the maximum to be evaluated.

The model parameters obtained by the maximum likelihood method are statistically efficient solutions, i.e., with minimum variance as sample sizes increases.

However, not all equations for maximisation are solvable and numerical solutions are required for those cases.

2.2.1.6 L-moment Method

The L-moment method is built on probability weighted moments (PWMs) which are defined by Greenwood et al. (1979) as

$$M_{p,r,s} = E\left\{X^p \left\{F(X)\right\}^r \left\{1-F(X)\right\}^s\right. \tag{2-17}$$

If a distribution has a quantile function $x(u)$, two special cases of PWMs are

$$M_{1,0,r} = \alpha_r = \int_0^1 x(u)(1-u)^r \, du, \tag{2-18}$$

$$M_{1,r,0} = \beta_r = \int_0^1 x(u)u^r \, du \tag{2-19}$$

These are similar to the definition of conventional moments. These PWMs were used as the basis for estimating parameters of probability distribution by Landwehr *et al.* (1979), Wallis (1981, 1982), Hosking *et al.*(1985). However, the method suffers from difficulties in interpretation. Hosking (1990) introduced L-moments, which are linear functions of PWMs and can be directly interpreted as the measures of scale and shape of probability distributions. L-moments are defined by Hosking in terms of the PWMs α and β as:

$$\lambda_{r+1} = (-1)^r \sum_{k=0}^{r} p^*_{r,k} \alpha_k = \sum_{k=0}^{r} p^*_{r,k} \beta_k \tag{2-20}$$

where $$p^*_{r,k} = (-1)^{r-k} \binom{r}{k} \binom{r+k}{k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2 (r-k)!} \tag{2-21}$$

The equations for the measures of location, such as sample mean, λ_1 and scale, λ_2, are also derived. L-moment ratios are given as:

L-CV $\tau = \lambda_2 / \lambda_1$ (2-22)

L- skewness $\tau_3 = \lambda_3 / \lambda_2$ (2-23)

L- kurtosis $\tau_4 = \lambda_4 / \lambda_2$ (2-24)

Vogel and Fennessey (1993) highlighted the advantages of L-moments compared to conventional moments. In a wide range of hydrologic applications, L-moments provide simple and efficient estimators of characteristics of hydrologic data and of a distribution's parameters (Stedinger *et al.*, 1993). Ulrych *et al.* (2000) show that probability density functions estimated from L-moments are superior estimates to those obtained from conventional moments and those based on the principle of maximum entropy. L-moments ratios have been derived for many distributions and given in Hosking and Wallis (1997) and a suite of FORTRAN routines is provided freely by the author on the Web (Hosking,1996).

2.2.2 *Regional Flood Analysis*

There are cases where the upstream inflow information is not available because the basin is ungauged. At-site flood frequency analysis is not possible. In other instances, the gauged records available at a gauged station may be very short rendering the flood frequency analysis unreliable. In those situations, a regional flood frequency analysis is recommended (Bobee and Rasmussen, 1995). Regionalisation is probably the most viable avenue for improving flood estimates and it has been the direction of research in flood frequency analysis in the 1990's (Bobee and Rasmussen, 1995). Many techniques have been used depending largely on the amount of data available for the region of interest. Physically based regression curves are typical results derived from such an exercise. The Index Flood approach originally developed by Dalrymple (1960) is often used.

The index flood method is used to determine the magnitude and frequency of flood quantiles for basins of any size, gauged or ungauged, as long as it is located within a hydrologically homogeneous region. At least two regressions are required. The first is the regression of mean annual flood with some physical parameters, e.g., basin area, basin slopes and river length, and the second relates peak flow ratio with frequency of exceedance. The peak flow ratio is the ratio of peak flow for a given frequency of exceedance to the at-site mean annual flood. A flood frequency curve can then be developed for any basin in the homogeneous region.

The index flood method has been used in earlier studies such as the UK Flood Studies by Natural Environment Research Council (1975) and Dalrymple (1960). It has become less subjective in distribution selection with the development of the L-moment

approach (Hosking, 1990). It is becoming popular as can be seen by the recent surge in usage, such as Pearson (1991a, 1991b, 1993) for New Zealand, Ben-Zvi and Azmon (1997) for Israel, Lim and Lye (2000) for Sarawak, Malaysia, Kumar *et al.* (1999) and Sankarasubramanian *et al.* (1999) for India, Glaves and Waylen (1997) in southern Ontario, Canada, Mkhandi *et al.* (2000) for southern Africa, among others. Nevertheless, the method cannot be used for areas where tidal influence is present, as in many estuarine areas of the rivers where the flow series in the river has the combined effects of fluvial flood and tides.

2.2.2.1 Regional Frequency Analysis by the L-moment Approach

Further development of the index flood method is the L-moment approach. It is an important alternative system to the Method of Moments for describing the shapes of probability distribution. Regional analysis using the L-moment approach is summarised in Hosking and Wallis (1997). It is a general approach that is best suited to environmental data and was developed primarily to analyse flood data.

L-moment ratio diagrams are analogous to the conventional moment ratio diagram and can be used to identify appropriate distributions for a particular region of interest. A plot of L-skewness vs L-kurtosis for all the at-site data in a homogeneous region can be used to identify the most appropriate distribution based on the weighted average of all the individual sites. This was strongly recommended by Vogel and Fennessey (1993).

2.2.2.1.1 Goodness-of-Fit Tests

A simulation method to test the goodness-of-fit based on the $\bar{t}_r$, the regional average of the sample L-kurtosis was proposed by Hosking and Wallis (1993, 1997). The smaller the difference between the regional average $\bar{t}_4$ and the statistic fitted for the distribution, τ_4^{DIST}, the better the goodness-of-fit.

$$Z^{DIST} = \left(\tau_4^{DIST} - t_4^R + B_4\right)/\sigma_4 \tag{2-25}$$

where B_4 and σ_4 are the bias and standard deviation of t_4^R respectively. Hosking and Wallis (1993, 1997) proposed that any three-parameter distribution could be declared as fitting adequately if $|Z^{DIST}| \leq 1.64$.

2.2.2.1.2 Homogeneity Tests

The L-moment approach provides two statistics to test for regional homogeneity, an important assumption of the regional frequency analysis. The first is a discordancy measure, intended to identify those sites that are grossly discordant with the region as a whole. The second is a heterogeneity measure, intended to estimate the degree of heterogeneity in a group of sites and to assess whether they might reasonably be treated as homogeneous. The details of the procedures are found in Hosking and Wallis (1997), and Rao and Hamed (2000).

Once the sites are found to be homogeneous based on the measures, the regional growth curve, typically consisting of a plot of the ratio of flood quantile to mean annual flood against the non-exceedance probability or reduced variates, can be established. The application of the approach can be found in Ben-Zvi and Azmon (1997), Lim and Lye

(2000). Madsen and Rosberg (1997) have extended the approach to the use of POT model.

2.3 Physically Based Approaches

Physically based approaches essentially model the actual flow conditions in river channel based on all the available physical theories and data. Two distinctive components can be delineated based on the theories used: hydrologic or hydraulic. The upstream inflows into the TIZ are often derived using hydrologic simulation models. In the TIZ and the estuaries, hydraulic theories are used to explain the hydrodynamic phenomena such as the fluctuation of water levels and the reversal of flows. Since hydrologic and hydraulic modelling have been widely used, a complete review of the vast literature is beyond the scope of this research. However some brief accounts are given in Section 2.3.1 and 2.3.2 with certain references made to the models.

2.3.1 Hydrologic Modelling

Hydrologic models can be classified as an event-based or continuous model according to the model internal structure. An event-based model typically covers durations of hours or months, while a continuous model provides simulated streamflows for as long as input data are available.

2.3.1.1 Design Flood Estimation Using Event-based Model

Flow rating has been an essential task in hydrology, which allows the flows, especially the extreme flow conditions of floods and droughts, to be accurately measured

or inferred. Hydrographs are often simulated using a hydrologic model after careful calibration. Many hydrologic models, such as HEC-1 or HEC-HMS, are used to derive some design flood flows on an event-based approach. This approach focuses on streamflow modelling of river basins considering the major factors of climatic influence and geophysical features. Since streamflow series are often available at the site of interest, a rainfall-runoff model is often used. It is also very common that design storms are used as the main input dataset resulting in the simulation of the corresponding peak discharges, which are assumed to have the same probability as the design storms. In certain circumstances where the antecedent moisture is constant, such as in equatorial rainforest of Sarawak (Lim and Lye, 2000), this assumption is reasonable. However, when variability of spatial and temporal distribution of rainfall and antecedent moisture in a large heterogeneous basin is high, the assumption of the same probability is in doubt (Bradley and Potter, 1992).

2.3.1.2 Flood Frequency of Simulated Streamflow Series

A different class of hydrologic simulation models that permit continuous streamflow simulation have now emerged, fuelled by the advent of powerful computers. Some of theses models are HSPF (U.S.EPA) used by Bradley and Potter (1992), TOPMODEL used by Cameron *et al.* (1999), and PDM (Lamb, 1999). Continuous streamflow simulation were used by Bras *et al.*(1985) and promoted by Franz *et al.*, (1989). A pre-calibrated rainfall-runoff model is driven either by a long historical rainfall series or a stochastic rainfall model. This results in a long simulated flow series. Conventional flood frequency analysis is then used to estimate flood quantiles from the

generated series. Bradley and Potter (1992) proposed a new approach for flood frequency analysis of model-simulated flows. A probability distribution is first estimated for the runoff volume. A relationship between peak discharge and runoff volume is developed using rainfall-runoff modelling of extreme flood events. The probability of peak discharge is finally derived by integrating the distribution of runoff volume with the conditional distribution of peak discharge on volume. Recent papers by Cameron *et al.* (1999), Lamb (1999) and Calver *et al.* (1999) have shown that the estimation of river flood frequencies by modelling basin runoffs on a continuous time basis is becoming more attractive given that readily available computing power has increased tremendously as compared to the 1980's. However, data requirements are not often met due to the absence of good precipitation data, both spatially and temporally over a long duration.

The estimation of flood frequency using generated streamflow series are mostly found in the non-tidal areas. Sikora and Steffen (1991) used continuous simulation at inflow points in an unsteady flow model in Illinois. The idea can be applied to the TIZ and this issue is discussed at the end of the next section.

2.3.2 Hydraulic Modelling

Most hydrometric networks do not include any flow measurement station in the tidal active zone, simply because developing a flow rating is almost not viable in a dynamic flow condition. The dynamic stage-discharge relationship in this zone has to be explained by a series of hydrodynamic model equations. The most common version used for solving a one-dimensional unsteady flow problem is the St. Venant equation. Several

hydraulic investigators derived these equations by using different procedures. These include Stoker (1957), Chow (1959), Dronkers (1964), Henderson (1966), Strelkoff (1969), Yen (1973), Liggett (1975), Cunge *et al.* (1980), Lai (1986), and Abbott and Basco (1990). These derivations essentially have to satisfy the three conservation laws governing open channel flows– mass, momentum, and energy. For flow velocity, V, flow depth, y, channel botton slope, S_o, friction slope, S_f, or the energy gradient needed to overcome friction, the dynamic or momentum equation can be written as:

$$\frac{\partial V}{\partial t} + g\frac{\partial}{\partial x}\left(\frac{V^2}{2g} + y\right) = g\left(S_o - S_f\right) \tag{2-26}$$

The deterministic approach using numerical modelling based on Equation 2-26 can provide flood stage at any specific location in the tidal interaction zones for a known flood event. These are generally known as *hydrodynamic models* to distinguish them from the steady state hydraulic models. Many software packages are readily available on commercial basis. These include MIKE11 (Danish Hydraulic Institute) and ISIS (HR Wallingford, UK). In the U.S., software such as BRANCH (USGS) is freely available.

With modern computing power, a lot of improvement has been made to these models. The models allow much more detailed data input, such as more cross sections or terrain data. In the 2D flow models, the sophistication also goes beyond the one-dimensional basis to include secondary flows across any cross-sectional areas. In coastal waters, 3D models are justified to explain vertical flows due to the presence of other dynamic factors introduced by winds and changes in atmospheric pressure.

2.3.3 Limitations of Physically Based Models

2.3.3.1 Probabilistic Aspects

Although a physically based model can estimate a water level, it cannot on itself assign a probability to the result, due to fact that there are many combinations or permutations of the tidal boundary conditions and the flood quantiles upstream. An assessment of the recurrence interval of flood levels at a particular site in the TIZ is often indirect.

2.3.3.2 Tributary Inflows

Another weakness is the need to assume that the generated streamflows from sub-basins using hydrologic models simultaneously have the same physical conditions such as the amount of precipitation and soil moisture (Bradley and Potter, 1992). The assumption is necessary to avoid further complication in assigning probability for the boundary conditions.

2.3.3.3 Limited Calibration Data

The data requirement for calibrating a hydrodynamic model is often difficult to be satisfied. Many high-resolution time series at various sites may bee required. Another drawback is that the resources required for setting up a hydrodynamic model is often beyond the scope of many small projects.

2.4 Tidal Phenomenon and Theories

The tidal phenomenon is an important aspect to be addressed in this research. Hence, a brief review of the phenomenon, features, theories and characteristics are essential.

2.4.1 Tidal Phenomenon

Tides are induced by the gravitational attraction of the moon and sun on the waters of the ocean (Dronkers, 1964). Although tidal phenomenon has caught the attention of many early scientists including Isaac Newton (1642-1727), it is still regarded as a complicated phenomenon since there are many irregularities in tides and the propagations in some coastal waters of the world remain a subject of investigation. Many modern books specifically written on tidal phenomenon include Dronkers (1964), Godin (1972), Bowden (1983), and Pugh (1987), among others. NOAA of U.S. and Canadian Hydrographic Service, at Physical Modelling Section, Division of Ocean Sciences, Fisheries and Oceans Canada have many excellent publications about tides and provide tide information and predictions at standard ports. Some observation sites are provided on a real-time basis.

The main features of commonly observed tides are:

1. The cycle of rise and fall in water occurs twice a day; each cycle lasts approximately 12 hours and 25 minutes.
2. Diurnal inequality is said to exist where two high waters on a given day is of different heights.

2.4.2 Tidal Theory

2.4.2.1 Equilibrium Tide Theory

Issac Newton first used the law of gravitation to explain tides. In this theory, the earth is assumed covered in water of the same depth everywhere and equilibrium tide is formed under the gravitational force of the moon. A mutual gravitational force between the earth and the moon governs the orbit of the moon in its elliptical path around the earth. The gravitational force of the moon caused water on the earth to be pulled towards the moon, causing water on the side facing the moon to rise and simultaneously a smaller rise on the other side of the earth. The later phenomenon is explained by two factors. The first being that there is a difference in the force of gravity from one side to the other as distance from the moon varies. The second is due to the existence of centrifugal force when the earth-moon pair system is rotating on an axis. Because the mass of moon is only about 1.2% of the earth, the centre of gravity of the rotation of the pair is located within the earth's crust. The entire earth experiences a centrifugal force as the system rotates. When the gravitational force and the centrifugal force forces exercise their influence together, two approximately equal bulges are induced. As the earth revolves once one day, the bulges are observed as two high tides every day.

2.4.2.2 Coordinate System

Figure 2-1 shows a coordinate system that is often used in explaining tidal phenomenon. Celestial equator AB is an extension to infinity of the earth equator while the polar axis coincides with the axis of rotation. P is the position of an observer on the

earth; M represents the position of a perturbing body such as the moon or the sun; and O is the centre of the celestial body or the earth. Colatitudes of P and M are the angle subtended measured from the celestial sphere summits at the north pole of the sphere N′ as θ and θ_M respectively. Longitude can be viewed on the celestial equator plane as Φ and Φ_M respectively. Φ' is the difference between the longitude of P and that of M, or $\Phi - \Phi_M$.

From these definitions, the spherical coordinates of the earth at P and the perturbing body at M are (θ, Φ) and (θ_M, Φ_M). ζ is the zenith angle of the moon as seen from the earth. The declination of M, δ, is the complement of θ_M, i.e. $\delta = 90° - \theta_M$.

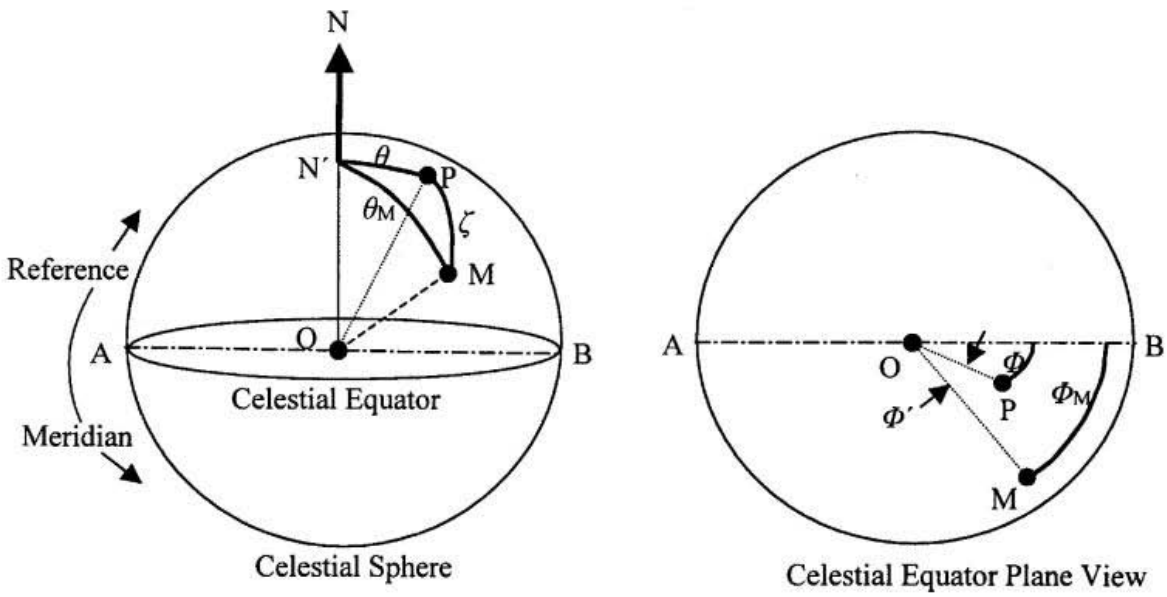


Figure 2-1 Celestial Spherical Coordinate System

2.4.2.3 Tidal Forces on Earth

A more modern approach in explaining tides formation is using the concept of tide forces (Godin, 1997). In the solar system, the earth accompanied by the moon, and the other planets are in their orbital motions that can be viewed as stable. The force of attraction that they exert on each other is exactly balanced by the centrifugal force resulting from their revolution around the centre of mass of the system. Tidal forces are forces that are not compensated by the orbital motion. These forces are minute as compared to the direct attraction of the earth. In spite of the relative minuteness of the tidal forces, they do cause tensions and displacements within the celestial bodies such as the phenomena of tidal effects on the earth.

2.4.2.4 Major Tide Forces and Periodicity

By calculating the vertical and horizontal component of the tide forces induced at P, it can be shown that these components consist of terms involving arguments θ, θ_M and Φ' (Godin, 1997). Terms involving argument Φ' are known as the diurnal term while $2\Phi'$ is for semidiurnal terms. These terms reflect the transit of the perturbing body M over the meridian in either one or two cycles per day. When the body M orbits, θ_M changes periodically over each revolution. For the moon, the period is a month and for the sun the period is a year. θ'_M signifies an anormalistic (a month) θ_M which has a period of 27.5546 solar day for the moon (27.3217 for the sidereal θ_M) and 365.2596 days for the sun (365.2422 days for the tropical θ_M).

2.4.3 *Major Tidal Components*

There are many components of the tidal forces exerted on the earth. Major components have been identified based on their arguments and certain conventional names are given. Table 2-1 shows a partial list of the major tidal components, which are as many as 60 or more.

Table 2-1 Major Components of Tidal Forces: Conventional Name and Frequency

Argument	**Due to the Moon**			**Due to the Sun**		
Semidiurnal	Name	Frequency (°/ hour)	Period (hours)	Name	Frequency (°/ hour)	Period (hours)
$2\,\Phi'$	$\mathbf{M_2}$	28.984	12.42	$\mathbf{S_2}$	30.000	12.00
$2\,\Phi' + \theta'_M$	$\mathbf{L_2}$	29.528	12.19	$\mathbf{R_2}$	30.041	11.98
$2\,\Phi' - \theta'_M$	$\mathbf{N_2}$	28.440	12.56	$\mathbf{T_2}$	29.959	12.72
$2\,\Phi' + 2\,\theta_M$	$\mathbf{K_2}$	30.082	11.97	$\mathbf{K_2}$	30.082	11.97
$2\,\Phi' - 2\,\theta_M$	$\mathbf{2N_2}$	27.886#	12.91			
Diurnal	Name	Frequency (°/ hour)	Period (hours)	Name	Frequency (°/ hour)	Period (hours)
$\Phi' + \theta_M$	$\mathbf{K_1}$	15.041	23.93	$\mathbf{K_1}$	15.041	23.93
$\Phi' - \theta_M$	$\mathbf{O_1}$	13.943		$\mathbf{P_1}$	14.959	24.07
$\Phi' + \theta'_M + \theta_M$	$\mathbf{J_1}$	15.585	23.80	$\mathbf{\Psi_1}$	15.082	23.97
$\Phi' - \theta'_M - \theta_M$	$\mathbf{Q_1}$	13.399	26.97	$\mathbf{\pi_1}$	14.918	24.13
$\Phi' \pm (\theta_M - \theta'_M)$	$\mathbf{NO_1}$	14.497	24.93	$\mathbf{S_1}$	15.000	24.00
$\Phi' - \theta'_M - \theta''$	$\mathbf{\rho_1}$	13.472	26.72			
Long Period	Name	Frequency (°/ day)	Period (days)	Name	Frequency (°/ day)	Period (days)
	Mf	1.098	13.66	**Ssa**	0.092	182.70
	Mm	0.544	27.55			

The arguments determine the periods, which will appear in a spectral analysis of the tidal forces or any induced phenomenon like sea tides. The calculation of period is obtained by:

$$T = \left[T_1^{-1} + T_2^{-1}\right]^{-1} \tag{2-27}$$

where T_1 and T_2 are the periods of Φ', θ_M and θ'_M. For example, an argument of $\Phi' - \theta_M$ has a period of 25.819 hours or a frequency of 13.943°/hour (360°/25.819 hour), given that $T_1 = T_{\phi'}$ = 24.8412 solar hours and $T_2 = T_{\theta_M}$ =655.720 solar hours. The name **O_1** has been given to this component.

2.4.4 Type of Tides

The general characteristics of tides can be represented by a few major harmonics constituents, which are described in later section. Four major types of tides are classified as shown in Table 2-2 according to a factor F defined as:

$$F = \frac{K_1 + O_1}{M_2 + S_2} \tag{2-28}$$

where K_1, O_1, M_2 and S_2 are the amplitudes of the corresponding components of **K_1, O_1, M_2 and S_2.**

Table 2-2 Types of Tides

Type of Tides	F
Semidiurnal	< 0.25
Mixed, mainly semidiurnal	0.25-1.5
Mixed, mainly diurnal	1.5-3.0
Diurnal	>3.0

2.4.5 *Harmonic Theory of Tidal Water Level*

The basis of harmonic analysis is that the tidal water level variation can be represented by a finite number of harmonic constituents. The tidal water level H at any time t can be represented by a harmonic representation:

$$H(t) = H_0 + \sum_{k=1}^{R} A_k \cos(\sigma_k t - a_k) + n(t) \tag{2-29}$$

where H_0 is the mean sea level; A_k, σ_k and a_k are the amplitude, radial frequency and phase of harmonic component k respectively; R is the number of significant harmonic components; and $n(t)$ is attributed to the contribution of physical factors other than tides. It is common to assume that $n(t)$ behaves as a white noise (Godin, 1997). Radial frequency σ_k is related to the period T_k by the relationship

$$\sigma_k = \frac{360^\circ}{T_k} = \frac{2\pi}{T_k} \tag{2-30}$$

expressed either in degrees per solar hour or in radians per solar hour. These frequencies also correspond to the frequencies of $\mathbf{M_2}$, $\mathbf{S_2}$, $\mathbf{K_1}$, $\mathbf{O_1}$, $\mathbf{P_1}$, etc. identified for the tidal forces. Full treatment of the harmonic analysis can be found in excellent texts, such as Pugh (1987) and Godin (1997).

Once the harmonic constants of A_k and a_k for the various harmonic constituents have been determined for a location, they may be used to describe the characteristics of

the tide at that location. Prediction of tides for a future time can be done through the use of a simplified version of Equation 2-30 as:

$$H(t) = H_0 + \sum_{k=1}^{R} A_k \cos(\sigma_k t + \alpha_k - a_k) \tag{2-31}$$

The noise term $n(t)$ is ignored and a term α_k is added to adjust for the phase at time t=0 in the equilibrium tide. The continuous prediction of tides can be found in numerous tide tables published by many hydrographic survey agencies in the world for various major ports and observation locations.

2.4.6 Shallow Water Tides

As ocean tides penetrate into the shallow waters of continental shelves from the deep oceans, the propagation are distorted due to various factors such as the depth, stream width, and storage width (Dronkers, 1964). Tides in coastal waters and river estuaries, referred as shallow water tides, are of interest in this research as these forms the lower boundaries of any model to be instituted for the TIZ.

2.4.6.1 Features of Tides

The main features of tides such as the type, maximum range, diurnal inequality and phase relationship between elevation and currents, are changed due to varying conditions in various locations. The tidal currents are more variable from place to place than changes of tidal levels. For South-east Asian seas and Gulf of Mexico, the diurnal inequality of tides is large, to the extent of only one high water and one low water a day, as compared to the tides found in the Atlantic coasts (Bowden, 1983).

2.4.6.2 Storm Surge Considerations

Meteorological disturbances during a storm have great effects on shallow seas. The water levels due to a coincidence of higher sea levels caused by severe storm acting on an area of shallow water and peak levels due to a spring tide often give rise to coastal floodings (Murty, 1984). Storm surge in a tidal water body is defined as the difference between the observed and the predicted water levels at the same location and time based on a deterministic tidal model in the absence of a storm (Pugh, 1987; WMO, 1988). A storm is typified by strong winds, drop in atmospheric pressure and intense rainfall. The effects of the last, mainly in the increase in fluvial discharge, are normally not included in the definition of surge (WMO, 1988). In places such as the south east coast of USA (south of Cape Hatteras), Philippines, Japan, Indo-China, Southern China, India, and Bangaladesh, tropical cyclones cause severe storm surges. In the northern Atlantic coast of USA and UK, extratropical storms are generally more prominent in causing storm surges. Tropical storms are not severe in latitudes less than 5 degrees (North or South).

2.5 Fourier Transform Method

Harmonic tidal analysis in its present form was formulated in the late 19th century for tidal predictions at coastal ports. Some researchers have attempted to analyse tidal effects in rivers, but mainly for reasons other than flood frequency analysis. The use of Fourier analysis was the natural choice as the harmonic approach has been successfully used in tidal simulation in the ocean and open seas as shown in Godin (1972). Goring (1984) used Fast Fourier Transform in deriving a river tide forecasting algorithm for a

river port in New Zealand. Sen (1991), and Lin and Lye (1994) used the harmonic analysis to investigate the cyclic features in cumulative departures of annual flow series in rivers.

2.5.1 *Fourier Analysis of Flow Series*

Harmonic analysis involves Fourier transform of a cyclic time series in terms of the distribution of its variance at different frequencies. The function can be represented by the sum of a series of sines and cosines of frequencies, which are multiples of the fundamental frequency $f = 2\pi/M\Delta t$:

$$H(t) = H_0 + \sum_{m=1}^{M/2} A_m \cos mft + \sum_{m=1}^{M/2} B_m \sin mft \qquad (2\text{-}32)$$

where the coefficients A_m and B_m may be evaluated by analysis of M values of sea level $H(t)$ sampled at constant intervals Δt, typically of hourly interval. H_0 is the mean sea level, assumed to be the average of the mean sea level over the period of observation. Equation 2-30 can be reduced to:

$$H(t) = H_0 + \sum_{m=1}^{M/2} C_m \cos(mft - g_m) \qquad (2\text{-}33)$$

where $$C_m = \left(A_m^2 + B_m^2\right)^{\frac{1}{2}} \qquad (2\text{-}34)$$

and $$g_m = \arctan\left(\frac{B_m}{A_m}\right) \qquad (2\text{-}35)$$

C_m and g_m are the amplitude and phase lag of the mth harmonic constituent of the function $H(t)$. The variance of this function about the mean sea level, H_0, is given by squaring the

terms within the summation symbol of Equation 2-33, and averaging over the period of observation. When this setup is applied and it can be found that that all the cross products terms are zero, except for the terms of the form:

$$C_m^2 \cos^2(mft - g_m) \tag{2-36}$$

The average value of this term over an integral number of cycle is $(1/2)H_m^2$. The most significant result states that the total variance of the series $H(t)$ is given by the summation of the variance at each harmonic frequency:

$$\frac{1}{2}\sum_{m=1}^{M/2} C_m^2 \tag{2-37}$$

2.5.2 Use of FFT in River Tides Analysis

A Fast Fourier Transform(FFT) algorithm basically transforms numerically the discrete observational time series, say y_n, where n = 0,1,2,....., N-1, to a series in the frequency domain. The product signified by Y_n, n = 0, 1,2,...., N-1, is termed discrete Fourier Transform (DFT). A line spectrum is a plot of the magnitude of the DFT, Y_n, at each frequency component n. Goring (1984) used the FFT method to separate river flow from the tide. For example, the DFT of a triangular shape stage hydrograph due to river flow has a maximum at the lowest frequency and decreases with increasing frequency. The DFT of the tide has just two components corresponding to the lunar and solar components and these two spikes can be spotted readily at higher frequencies than the DFT due to the river flows. The approximation in separation is to treat the frequencies lower than the DFT of the tidal spikes as belonging to the river flow. The method is

shown by Goring (1984) to be able to predict river tide in the time domain, but it lacks accuracy in the magnitude of stage.

2.6 Wavelets Method

The last decade has seen the emergence of a new approach in data analysis called wavelet analysis. Many fields, including signal processing, physics and mathematics, provided the ingredients of the theory. The use of wavelet analysis can now be found in all kinds of signal processing problems and its usage has been very successful in a wide range of fields including medicine, seismology, music, criminology, and vision. The theories on wavelets can be found in many recent introductory texts such as Antoniadis and Oppenheim (1995), Ogden (1997), Mallet (1998), and Teolis (1998), among others. A clear account of applying wavelets to solve meteorological problems can be found in Torrence and Compo (1998).

Wavelets can be used to process signals that vary with time (non-stationary) and contain discontinuities. In many cases, it has replaced Fourier analysis because of its many advantages.

Wavelet theory also provides a tool to recover functions from noise. The successful use of Fourier analysis on sea tides is largely due to the stationary nature of the process. The problem of stages in rivers with tidal interaction appears to be a good candidate for investigation by wavelet analysis because the process involved is predominantly non-stationary. The wavelet analysis involving wavelet transform can reveal structures of a non-stationary time series better than the standard Fourier and short-

term Fourier transform techniques. Indeed Jay and Flinchem (1995, 1997, 1999) have demonstrated the interaction of fluctuating river flow with a barotropic tide using the wavelet analysis. Smith *et al.* (1998) are able to characterise streamflows from different climatic regions using the wavelet transform. Gan (2001) shows the possible influence of the precipitation pattern in western Canada by large-scale climate anomalies in the Pacific Ocean using wavelet analysis.

2.6.1 Wavelet Analysis

All wavelets are based on one *mother* wavelet, $\psi(t)$, a function with some special properties. The function must oscillate and decay, preferably rapidly. The oscillation property is expressed mathematically as:

$$\int_{-\infty}^{\infty}\psi(t)dt = 0 \tag{2-38}$$

Unlike the trigonometric basis function of harmonic analysis which has only a single parameter (frequency), wavelets are essentially controlled by two parameters, dilations a and translations b of the mother wavelet. Dilation implies shrinking as well as stretching. For a continuous wavelet transform, any amount of dilation or translation can be applied and the wavelet with parameters (a, b) is defined as:

$$\psi_{(a,b)}(t) = a^{1/2}\psi(at - b) \tag{2-39}$$

for any real number a and b. For a vast majority of applications a restricted set is used: $a = 2^{j}$ and $b = k$ where j and k are integers. For a specially chosen mother wavelet the collection of $\{\psi_{j,k}(t)\}$ for all integers j, k forms an orthonormal basis for various

function spaces. This means that wavelets can represent functions analogous to what a Fourier series does. The main point is that a function f may be represented in terms of a superposition of wavelets of different dilations and translations, as

$$f(t)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}d_{j,k}\psi_{j,k}(t) \tag{2-40}$$

where $d_{j,k}$ are coefficient of expansion which can be found by

$$d_{j,k}=\int_{-\infty}^{\infty}f(t)\ \psi(t)dt \tag{2-41}$$

Wavelets can represent some functions using far fewer coefficients than can Fourier. The functions for which wavelets do better include those that have sharp jumps or discontinuities.

The simplest wavelet, as shown in Figure 2-2, is the Haar mother which can be defined on [0, 1) by

$$\psi(t)=\begin{cases}1 & \text{for } t\in[0,\ 1/2\ )\\ -1 & \text{for } t\in[1/2,\ 1)\end{cases} \tag{2-42}$$

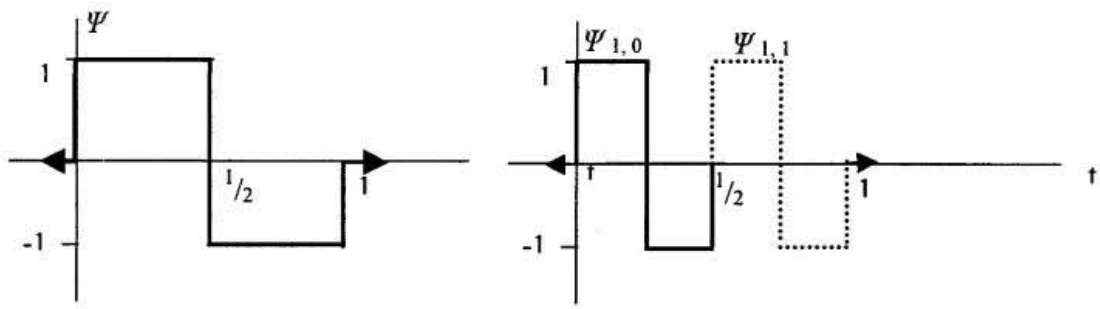


Figure 2-2 The Haar wavelet and Two Haar wavelets

There exist some instances where the mother wavelets are not enough to represent all functions. For example, the constant function on [0, 1) cannot be represented by this collection of Haar wavelets. This deficiency can be corrected by the introduction of a *father* wavelet. In the Haar case the father is defined to be:

$$\phi(t)=1 \quad \text{for } t \in [0,1\} \tag{2-43}$$

which happens to be a constant function. Any other constant function c can be represented by a multiple of the father. The introduction of the father wavelet results in the expression:

$$f(t)=\sum_{k=-\infty}^{\infty} c_k \phi_{0,k}(t)+\sum_{j=1}^{\infty}\sum_{k=-\infty}^{\infty} d_{j,k}\psi_{j,k}(t) \tag{2-44}$$

which has a natural and intuitive interpretation: a function has to satisfy both the overall "large-scale" behaviour of the function, as represented by integer translates of father wavelets, $\phi(t)$, and the "small-scale" or high-frequency or "sharp features" in terms of integer translates of mother wavelets at different scales j. The wavelet coefficients, $d_{j,k}$, allow one to "zoom-in" to small-scale detail of the function $f(t)$ just by increasing the scale factor j.

Wavelets are hence built on the concept of *multiresolution* analysis. Using wavelets, a function can be depicted as a *smooth* overall shape (*approximation*), plus detail that ranges from broad to narrow (*detail*). Different types of structure in the signal may be resolved at each scale of resolution.

There are many other different families of wavelets such as Daublets, Symmlets, Mexican hat, Morlet, Coiflets, etc. Some modern technical computing packages have

optional modules exclusively for performing wavelet analysis; MATLAB has a wavelets toolbox (The Math Works, 1999) and S-PLUS has a toolkit called S^{+}Wavelets.

2.6.2 Application of Wavelet Analysis in River Flows

Jay and Flinchem (1995, 1997, 1999) have attempted to analyse river floods using a continuous wavelet analysis. Their main objective was to study the effects of tides being affected by the river flows. The use of wavelets appears to allow inference to be made on the magnitude and frequency of the river tides, a non-stationary process. Hence a plausible approach would be to filter out the tidal process and then derive the frequency distributions of the two separate processes. A step forward in that direction is likely to yield meaningful results and this naturally leads to the investigation in Chapter 8.

Chapter 3 Aspects of Study Area

This Chapter presents five main aspects of the study area that are considered important and relevant to the elaborations on flood estimation methods in the later chapters:

1. Physical and Climatic Conditions
2. Data Used
3. Sea Tides
4. Surges
5. Hydrometric Network

3.1 Physical and Climatic Conditions

3.1.1 Physical Features

Sarawak is the largest state in the Federation of Malaysia, accounting for about 37.5% of the total land area country. Located immediately north of the Equator between latitude 0° 50' and 5° N and longitude 109° 36' and 115° 40' E, Sarawak stretches some 800 km along the north-west coast of the Island of Borneo. It borders Kalimantan province of Indonesia to the south, the Sultanate of Brunei, and Sabah, another Malaysian

state, to the northeast. The coast of Sarawak partially forms the southernmost extend of the South China Sea. Figure 3-1 shows the general location of Sarawak in South-East Asia. With about 2 million people living in a land area of 124,449 km^2, it is rich in natural resources and is known for its thick rainforest generally considered as one of the most biodiverse places on earth. The rainforest is home to more than 8,000 species of flowering plants and over 20,000 animal species, the majority of which are insects. Mangrove forests are also very common in river estuaries and coastal areas which are sustained by the nutrients brought along by the water flows during the twice daily high tides and the occasional high fluvial discharges that last between 1 day and 2 weeks in the major rivers of Sarawak.

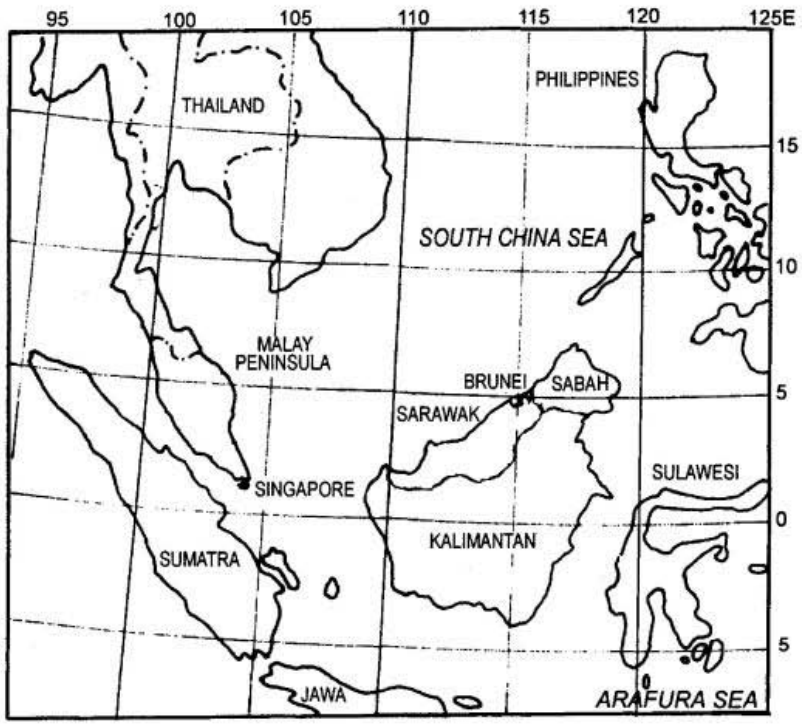


Figure 3-1 Location of Sarawak, Malaysia in South-East Asia

One-fifth of Sarawak is composed of alluvial coastal plains covered with peat swamps forming large tract of wetlands. A large portion of Sarawak is under the significant influence of tides as shown in Figure 3-2. Maximum tidal amplitudes or tidal ranges in the order of 5 to 6 m are found in many river estuaries of western Sarawak. The unusual high tidal ranges are attributed to the occurrence of co-oscillating tides in the shelf region of the South China Sea and the Arafura Sea (Zahel, 1992). In general, Sarawak has a relatively flat topography when compared to other parts of Borneo Island. The low-lying lands have many meandering rivers with very mild bed slopes. The high tidal range and mild bed slopes enable the tidal influx to reach very far inland. For example, tidal effects can be felt at a distance of over 190 km from the coast near Kapit on the Batang Rajang, the largest river in Malaysia (Lim and Lye, 2002). Hence, the tidal interaction zones are considered very extensive within most of the basins.

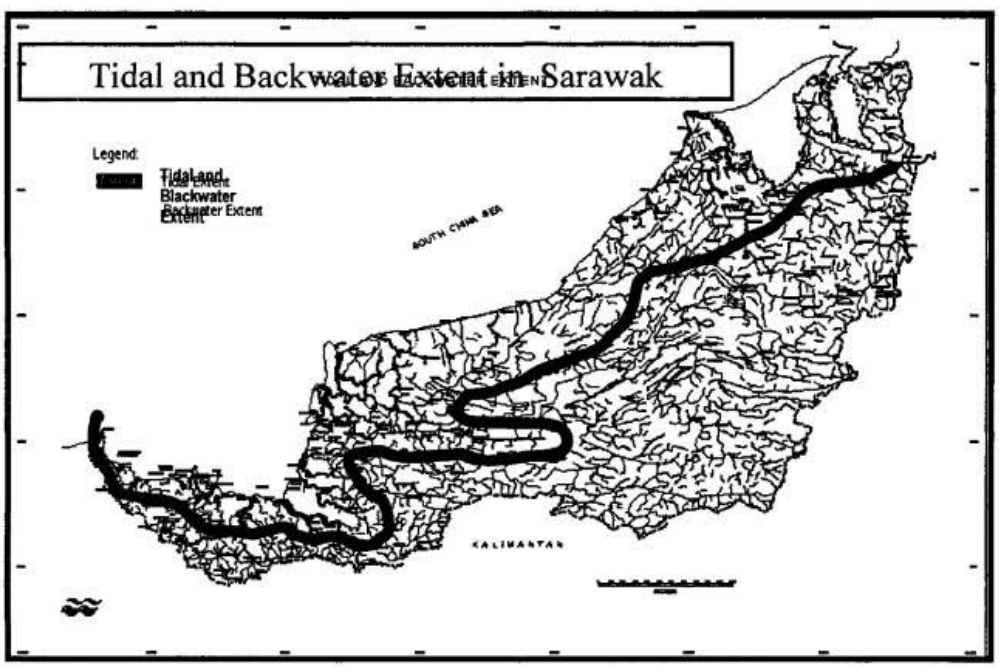


Figure 3-2 Tidal and Backwater Extent in Sarawak

3.1.2 Climatic and Runoff Factors

Bestowed with an equatorial rainforest climate, Sarawak does not have any distinct seasonal variation of temperature as found in the temperate zones. The most prominent feature is the perennial wet, hot and humid climate.

3.1.2.1 Rainfall and Evaporation

Sarawak is hot and humid throughout the year with mean daily temperature ranging from 23°C in the early morning to 32°C during the day. Rainfall in Sarawak varies from an annual average of about 5500 mm on the coastal area of western Sarawak to about 3000 mm in the interior highlands. Trade winds around the equator have significant effects on the rainfall distribution. Rain is distributed throughout the year although most falls between November and February, a period that is referred locally as the *landas* season. Due to the high volume of runoff, the state is well endowed with many perennial streams and rivers, giving a very high drainage network density. Average number of days in a year with rainfall is 280 at Kuching. Floods often occur in the *landas* seasons, frequently as a combination of spring tide and intense rainstorm.

Annual evaporation is fairly high as well, in the order of 1300 mm. There are some occasions of prolonged dry periods in the coastal zones, with durations of up to one month. Although Sarawak is very close to the typhoon active countries such as the Philippines, it has never experienced any direct onslaught of typhoons, as any cyclonic activity is weakened as it approaches the equator.

3.1.2.2 Soil Moisture

In general the soil moisture, especially in the peat swamps areas, varies from very high to full saturation point since the ground water-table often rises to the surface fed by frequent rainfalls. There are considerable extents of peat lands. About 13% of the state, mainly in the coastal region (Wong, 1991), is covered with peat soils with a depth varying from a few metres to 25m. The peat soils are often saturated and poorly drained, and waterlogged for most of the time. For regions with higher proportions of peat soil cover, the runoff of streams is characterized by a gradual depletion of the baseflows.

3.1.2.3 Vegetative Covers

The most prominent physical feature of the state is the vast tract of land covered with equatorial rainforest. More than 67% of land is under natural forest cover. The rest is secondary forests, agricultural, and urban lands. However, oil palm plantations and logging activities are fast becoming the major factors in land-use. The traditional slash-and-burn planting method of hill padi occupies about 13% of the land area. Timber extraction accounts for a significant amount of the loss of primary rainforest in the 1980's and 1990's. Generally, the runoff factor from a logged rainforest is higher from the primary rainforest.

3.2 Data Used

3.2.1 Rainfall

Most of the rainfall stations in Sarawak are maintained by the Department of Irrigation and Drainage, Sarawak (DID). Daily rainfall records (DID, 1963-1998) for up

to 9 years in western Sarawak are used in a network analysis to assess the general rainfall network design in Section 3.5.2.3. Rainfall factors are used critically in the identification of homogeneous regions as found in Chapter 5. The factors of mean annual rainfall and design storms of various return periods are obtained using statistical analysis. The length of data series used are up to the maximum record length processed by DID.

3.2.2 Discharge and Water Level

There are about 23 major river basins in Sarawak. Four categories of stations are identified based on the location and the degree of tidal influence:

1. Tidal Station (sea station that is close to the river estuary)
2. Station in the Tidal Interaction Zone (TIZ)
3. Mainstream Discharge Station
4. Basin Discharge Station (within the basin but not on the mainstream)

Table 3-1 shows the selected river systems chosen for further investigation. The main criterion is that an observed data series is available in each of the four categories of stations. A river system that possesses concurrent data series in all the categories is the best choice. Due to the non-availability of data under the category of *station in TIZ*, the Batang Samarahan and Sungai Limbang river systems are not included in the final round of data analysis. Four river systems that are finally studied in detail are Sungai Sarawak, Batang Sadong, Batang Rajang and Batang Baram. The meaning of *sungai* in the local language is river while *batang* refers to a large river.

Table 3-1 River System, Stations in Different Categories and Data Availability

River system	Tidal Station	Station in the Tidal Interaction Zone	Mainstream Discharge Station	Basin Discharge Station	Recording Period, Interval and Data Availability	
1. Sungai Sarawak	-Pulau Lakei				✓Jan 74 – Apr 89	-Hourly
		-Satok Bridge			✓Jan 90 – Mar 00	-Tideda
		-Batu Kawa			✓Jan 90 – Jan 00	-Tideda
		-Batu Kitang			✓Jan 91 – Jun 00	-Tideda
		-Siniawan			✓Jan 92 – Jun 00	-Tideda
			-Kpg Git		✓May 90 – Jun 00	-Tideda
			-Buan Bidi		✓May 90 – Jun 00	-Tideda
				-Boring	✓92 – 97	-Daily
2. Batang Samara-han	-Pulau Lakei				✓Jan 74 – Apr 89	-Hourly
		-Kpg Nanga				-N.A.
			-Kpg Maang			-Daily
			-Batu Gong		✓Jan 83 – May 00	-Tideda
				-nil		
3. Batang Sadong	-Pulau Lakei				✓Jan 74 – Apr 89	-Hourly
		-Gedong			✓Jan 93 – Nov 99	-Tideda
			Serian		✓Jan 77 – Jan 00	-Tideda
				-Krusen	✓Oct 87 – Mar 00	-Tideda
				-Meringgu	✓Jan 84 – Jun 00	-Tideda
				-Sg Bedup	✓Jan 70 – Jan 00	-Tideda
4. Batang Rajang	-Sarikei				✓Jan 95 – Dec 96	-Tideda
		-Sg Merah			✓Jan 93 – May 00	-Tideda
		-Sg Salim			✓Jan 92 – May 00	-Tideda
		-Song				-Daily
		-Machan				-N.A.
			-Kapit		✓Sep 86 – Jan 00	-Tideda
			-Mukeh		✓May 88 – Dec 99	-Tideda
			-Tj. Ayam		✓Jan 87 – Apr 97	-Tideda
				-Ng Benin	✓Aug 85 – Jul 95	-Tideda
				-Telok Buing	✓Apr 84 – Dec 99	-Tideda
				-Entawau	✓ 90 – 99	-Daily
				-Belaga	✓Aug 88 – May 99	-Tideda
				-Giam Pasang	✓Aug 77 – Jun 90	-Tideda
5. Batang Baram	-Miri Jetty				✓ 87 – 93	-Hourly
		-Bakong			✓Feb 95 – Mar 00	-Tideda
		-Marudi			✓Jan 77 – Jan 00	-Tideda
		-Benawa			✓Jan 95 – Mar 00	-Tideda
			-Pilah D		✓Jul 91 – Sep 92	-Tideda
			-Terawan		✓67 – 99	-Daily
			-Teru		✓67 – 99	-Daily
				-Lio Matu	✓Aug 77 – Jun 90	-Tideda
				-Akar Bazaar	✓Aug 77 – Jun 90	-Tideda
6. Sungai Limbang	-Limbang				✓Jan 95 – Jan 00	-Tideda
		-Batu Danau				-N.A.
		-Ukong				-N.A.
			-Insungai		✓Jan 90 – Feb 00	-Tideda
				-Saliban	✓Jan 95 – Jan 00	-Tideda
				-L. Lallang	✓Aug 89 – Feb 96	-Tideda

Note: Tideda refers to a time dependent hydrological data format in which the time step can vary to capture any abruptly changing event with higher time resolution. The finest time step is 5 minutes. ✓ = period of data available

3.2.2.1 Sea Tide Station

A tidal station at Pulau Lakei was established by the Sarawak Land and Survey Department in 1977 primarily to determine the mean sea levels in the vicinity of Kuching, Sarawak, such that a standard datum for cadastral survey purposes could be established based on the mean sea level. The Royal Malaysian Navy made tidal predictions based on this station and at other sea fronting towns of Miri and Bintulu. Later, the Malaysian National Mapping and Survey Department established observatory stations at Bintulu and Miri. In the nearby coastal water of Brunei, a station called Lumut was established by Brunei Shell Company in 1973 and discontinued in 1983. The details about these stations are summarized in Table 3-2

Data series are often discontinuous due to various reasons. For example, the observation of sea levels at Lumut, Brunei was halted mainly due to a reconstruction exercise performed on a major structure on which the recorder was attached. The whole series is plotted and shown in Figure 3-3.

Table 3-2 Sea Tide Stations

Name	Latitude	Longitude	Beginning date	Ending date	Missing or Gaps
Pulau Lakei	01°44'52"N	110°29'12"E	July, 74	April, 89	Minor
Bintulu	03°16'13"N	113°03'14"E	Nov, 94	Dec, 95	Minor
Miri	04°21'37"N	113°57'17"E	1986	1993	Minor
Lumut	04°40'39"N	114°27'32"E	July, 73	May, 83	1976-1979

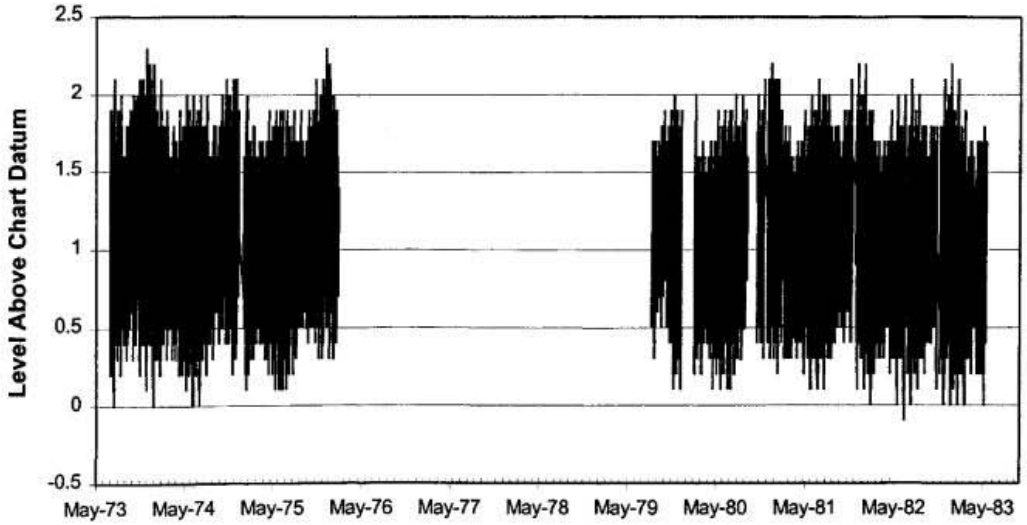


Figure 3-3 Hourly Sea Levels at Lumut (data obtained from Brunei Shell)

3.2.2.2 Typical details about recording station

Station Name: Lumut

The gauge was installed at BLNG plant at Lumut in 1973. It is placed in a single pile adjacent to the east side of the trestle. The distance from shore is about 300m and the water depth is 3.5m. The tide gauge is a Munro "float and type" type IH123 which is a combination of a Munro type IH 103 and a Fisher and Porter recorder, model 1542. A big gap of missing data between 1976 - 1979 (due to construction work on the trestle).

Reference Datum = - Chart Datum Brunei Open Waters (CD)

Relationship between CD (Brunei Open Waters) and Sarawak datum

Reference Level	
100FT. Lutong Datum	= 2.51m (8.23 ft) CD
0.00 FT Miri Datum	= 2.36m (7.73 ft) CD
CD Miri	= 0.08m (0.25ft) CD

3.2.2.3 Station in the Tidal Interaction Zone

This category of stations is rare because most agencies do not see the usefulness of acquiring seemingly sinusoidal data from these stations. However, there exist several such stations in Sarawak, under the maintenance of DID. The importance of keeping such stations must have been recognised. These stations are:

1. Satok (Sarawak)
2. Batu Kawa (Sarawak)
3. Batu Kitang (Sarawak)
4. Siniawan (Sarawak)
5. Gedong (Sadong)
6. Sungai Merah (Rajang)
7. Sungai Salim(Rajang)
8. Marudi(Baram)
9. Benawa(Baram)
10. Bakong(Baram)

The tedious task of keeping the record continuous is another step that is often difficult to achieve. For most of these stations, strip-chart recorders were used. On many occasions when the records are broken, it was mainly due to chart overruns, dried up ink or flat batteries. Modern digital recorders are being slowly introduced to replace these analogue recorders, hence minimising the problems attributed to the mechanical chart recording system.

3.2.2.4 Mainstream Discharge Station

Discharge stations are the primary source of data for developing a discharge series for the stream or river of concern. Traditionally most of the flood extremes are obtained at these stations and subsequently used for frequency analysis. All the stations selected for this study are mainstream stations located upstream of the tidal interaction zone and are outside the zones of influence of tides.

3.2.2.5 Basin Discharge Station

In order to obtain a set of more accurate flood quantile estimates, regional flood frequency analyses are performed and described in detail in Chapter 5. Those stations that are located within the basin concerned but are not on the mainstream are classified as *basin discharge station*. These stations are outside of the zones of influence of tides. The discharge data series of these stations are used in developing a representative regional flood frequency curve, such that a region-wise extrapolation of flood information can be made.

3.2.3 Data Processing and Conversion

The agencies concerned, which are custodians of those data series mentioned above, have their own standard of file formatting for archive purposes. It varies from simple columns to multi-dimensional tabulations. The most detailed data in the form of TIDEDA (*Time-Dependent-Data*) are stored at varying time intervals, depending on the rate of change in the monitored data, e.g., rapidly varying water levels would be recorded at a finer time resolution than stagnant water levels. The water level series supplied by the

data-collecting agencies were initially fitted with cubic splines before interpolation at hourly intervals could be carried out. **Appendix A1** shows the MATLAB programs specifically developed for the processing of such data.

3.3 Sea Tides

As mentioned in Chapter 1, the maximum tidal ranges found in coastal waters and river estuaries of Sarawak are large. An understanding of the characteristics of sea tides is important, as the tidal behaviours found in the TIZ are directly dependent upon its propagation. The location of Sarawak almost right on the equator is a major influencing factor; the semi-diurnal tide is at a maximum on the equator.

3.3.1 Tidal Species

The major tidal species found in the coastal waters of western Sarawak, as represented by the station of Pulau Lakei, is the semi-diurnal species. This can be seen in Figure 3-4 where two simultaneous short series plots of tidal levels observed at Pulau Lakei and Lumut are compared. Lumut is located in the Sultanate of Brunei, neighbouring northeastern Sarawak. The tidal species there consists of a mixture of semi-diurnal and diurnal tides. For a purely diurnal species, the fluctuation occurs at a frequency of one cycle per day. Figure 3-5 shows one month of simultaneous observations of tide levels at Pulau Lakei and Miri. Zahel (1992) noticed the co-oscillating semidiurnal and diurnal tidal regimes in the South China Sea and attributed that to a combination of factors. These factors include the complex topography around the

Indonesian Archipelagos and the resonant excitation of the Indian and Pacific Ocean tides. The large tidal amplitudes observed at Pulau Lakei at the western coast of Sarawak, as discussed in the next section are a manifestation of the resonant effect.

3.3.2 Maximum Amplitude

Tidal ranges in the order of 5 to 6 metres are found in the coastal waters of western Sarawak as seen in the observed Pulau Lakei series. In general the range decreases as the colatitude of the point of observation decreases. It is known that the maximum range of the tides is a maximum at the equator for semi-diurnal species of tide (Zahel, 1992). The maximum tidal amplitudes and ranges recorded for the sea-level observation stations in Sarawak and Brunei are listed in Table 3-3.

3.3.3 Sea Tidal Level Distribution

An understanding of the sea tidal level probability distribution is relevant to design flood estimation methods using the probability density functions approach. The distribution of sea tidal levels has been studied extensively in many parts of the world. However, the probability distributions of the tidal levels in the coastal waters around Sarawak are not found in the literature.

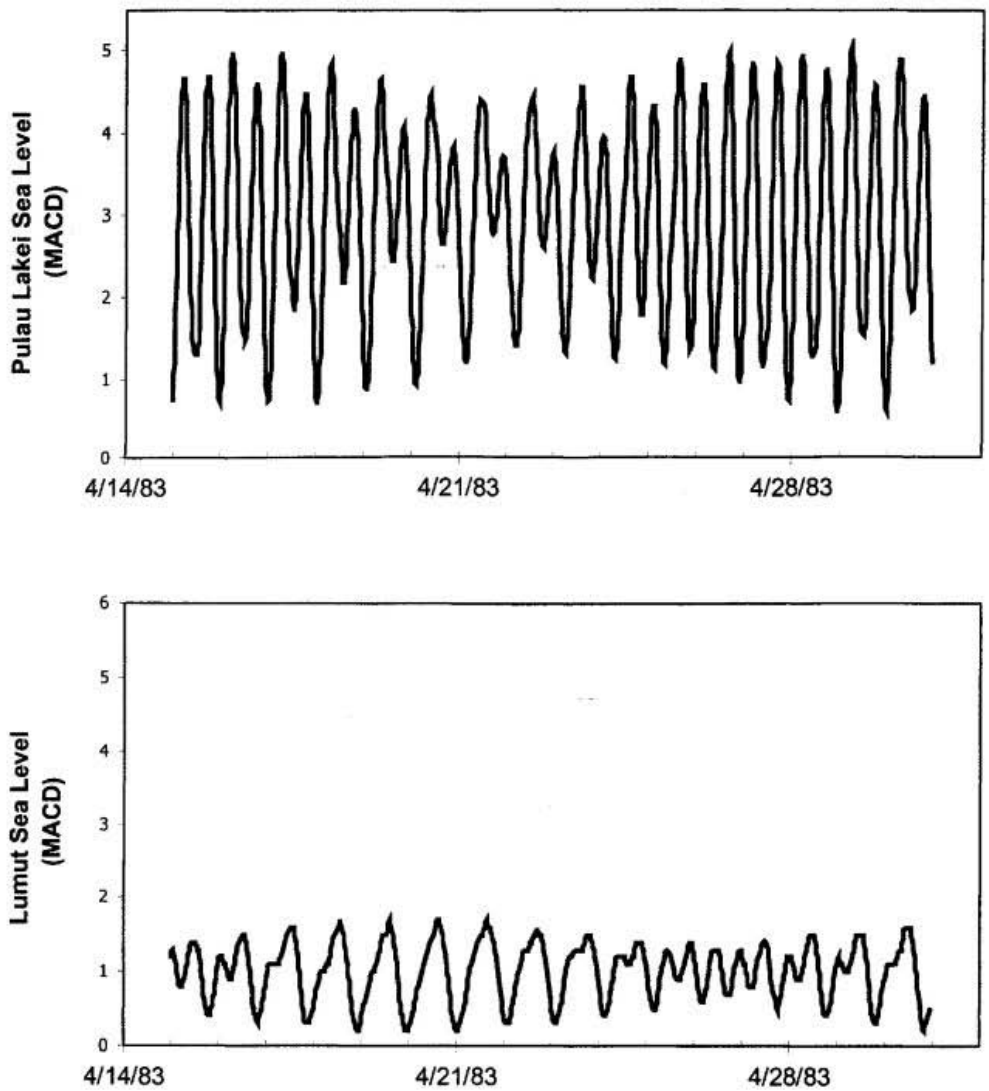


Figure 3-4 Sea Tidal Levels at Pulau Lakei (Sarawak) and Lumut (Brunei)

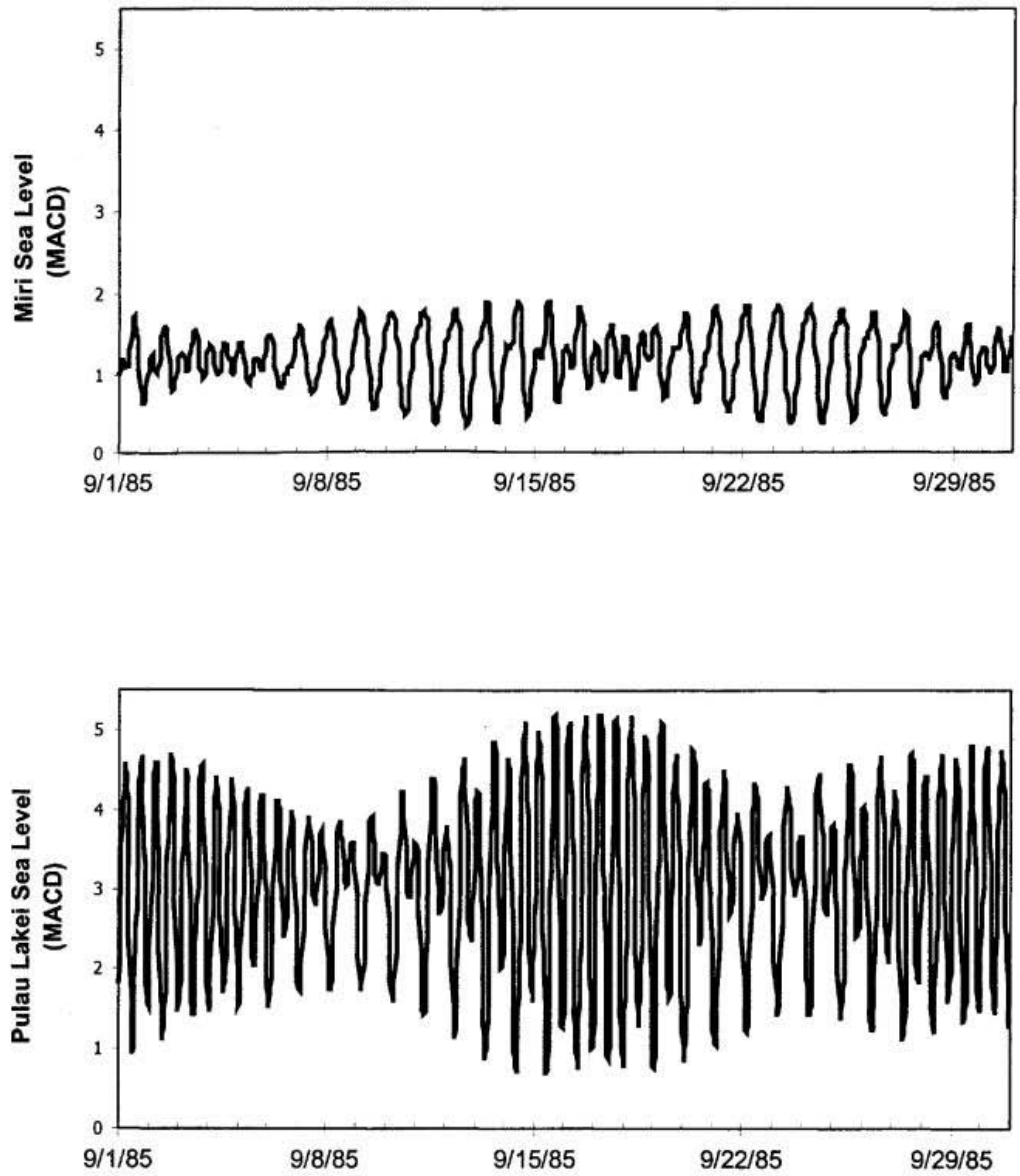


Figure 3-5 Simultaneous Observed Sea Tidal Levels in Sarawak at Pulau Lakei and Miri in September 1985.

Figure 3-6 shows four frequency distributions of sea levels at four stations in this area. A normal distribution is usually assumed in the absence of any observed data. It is clear that the normal distributions superimposed on these plots do not fit satisfactorily. More rigorous statistical tests were performed and the results are shown in Table 3-4. The p-values less than 0.0005 show that the sea levels are not normally distributed at an α value of 0.05 (for 95% confidence). It can be concluded that the sea levels observed off the coasts of Sarawak cannot be represented by normal distributions.

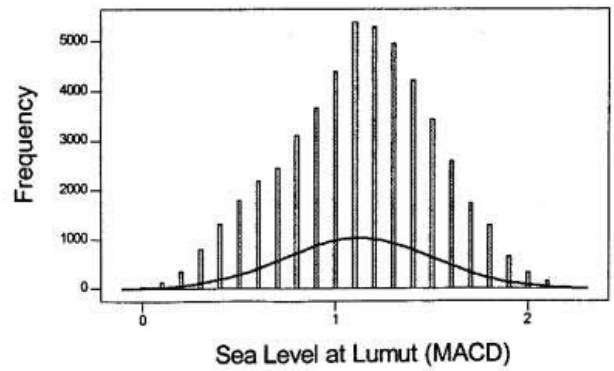
Table 3-3 Maximum Tidal Amplitude and Range Observed in South China Sea off Sarawak (all readings are in MACD)

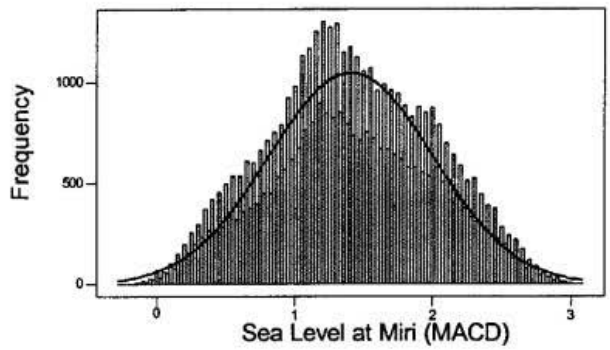
Station	Mean Sea Level (MACD)	Cycle with Maximum Amplitude			Maximum Range of Series		
		High	Low	Amplitude	Maxi.	Mini.	Range
Pulau Lakei	0.306	5.68	0.27	5.41	5.68	0.01	5.67
Bintulu	0.191	2.59	0.48	2.11	2.96	0.48	2.48
Miri	1.407	2.03	-0.22	2.25	3.07	-0.22	3.29
Lumut, Brunei	1.122	2.0	-0.1	2.1	2.3	-0.1	2.4

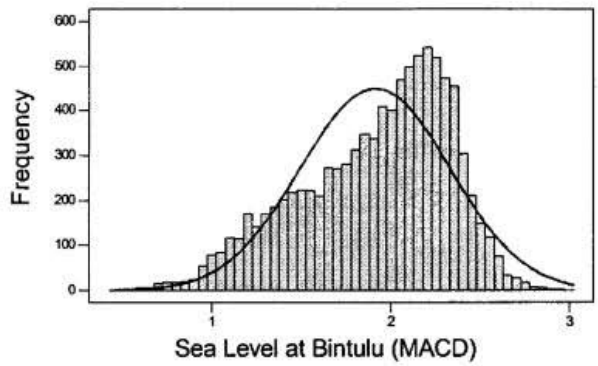
Note : MACD = Metres Above Chart Datum

Table 3-4 Test for Normality of Sea Tidal Level Distribution

Station Name	Mean	St.dev.	Number of points (hrs)	Anderson-Darling A^2	p-value
Pulau Lakei	3.0600	1.1179	123274	628.3	0.000
Bintulu	1.1912	0.4154	9360	113.1	0.000
Miri	1.4073	0.5944	62311	68.2	0.000
Lumut, Brunei	1.1221	0.3900	50370	192.0	0.000







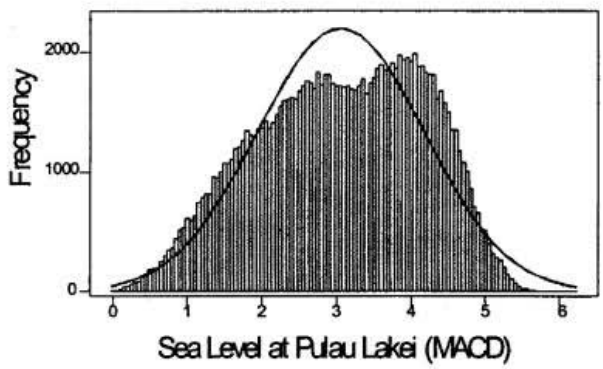


Figure 3-6 Distribution of Tides at Sea Level Observation Stations in Sarawak and Brunei, with Normal Distribution Superimposed.

Among the stations, it is clear that the distribution at Pulau Lakei has an affinity for twin peaks, or bimodal, typical of a semi-diurnal tidal species with significant diurnal inequality (Pugh, 1987),. This series is from 14 years of hourly data, the longest among the four. The rounding of observed readings to the nearest 0.1m at Lumut has resulted in distortion of the frequency distribution with 'discrete' class interval effects. The Bintulu data series is only slightly longer than one year in duration, making it the shortest series used.

3.4 Surges

A brief review of the literature regarding surges in sea conditions was given in Section 2.4.6.2. An observed sea-level series, $H(t)$, is thought to consist of three components:

$$H(t) = H_0(t) + T(t) + S(t) \qquad (3\text{-}1)$$

$T(t)$ is the tidal component of the variation and $H_0(t)$ is the mean sea-levels which changes slowly with time. $S(t)$ is normally referred to the excess sea levels generated by a severe storm and termed storm surge, or simply as surge. For the checking of surge phenomena, surge can then be taken as the difference between the observed and predicted levels in the sea condition, expressed as:

$$S(t) = H(t) - H_0(t) - T(t) \qquad (3\text{-}2)$$

For practical purposes, the mean sea level is assumed to be stationary. With all the observed data adjusted to the mean sea level datum, the surge series can be calculated by:

$$S(t) = H(t) - T(t) \qquad (3\text{-}3)$$

The differences in the series are often known as the *residual levels* in the tidal literature.

3.4.1 Surge Characteristics in Coastal Sarawak

Owing to its disturbances in the TIZ, the surge characteristics of the sea conditions around Sarawak were studied to ascertain the degree of surge phenomenon around the area using the available recorded data. Published tide tables were used to predict tidal elevations.

The station at Pulau Lakei is used in this study since it has the best sea level record in Sarawak with very few gaps. A program written in MATLAB as shown in **Appendix A2** is used to compute the surge and to derive certain statistics of the surge series. Figure 3-7 shows one year of hourly tidal observation at this station for 1988 while Figure 3-8 shows the residuals computed. The corresponding predicted tide levels were abstracted from a tide table published by The Royal Malaysian Navy (1987) for the year 1988. The extreme surge is around + 0.26 and - 0.48 m. The tidal residuals have a distribution that is slightly skewed towards the positive surge events as shown in Figure 3-9. A formal normality test was carried out with the probability plot shown in Figure 3-10. The calculated p-value is less than 0.01. With an α value of 0.05 (for 95% confidence), the null hypothesis that the residual levels are normally distributed is rejected.

An ideal tidal residuals plot should show some random scatters around zero. A casual inspection of Figure 3-8 suggest that there is a major period (around 5000 to 7500 hours or late June to Mid November) when the residuals scatter around 0 to -0.3m. A run

test on the residuals was performed to check if the order is random. The following results were obtained:

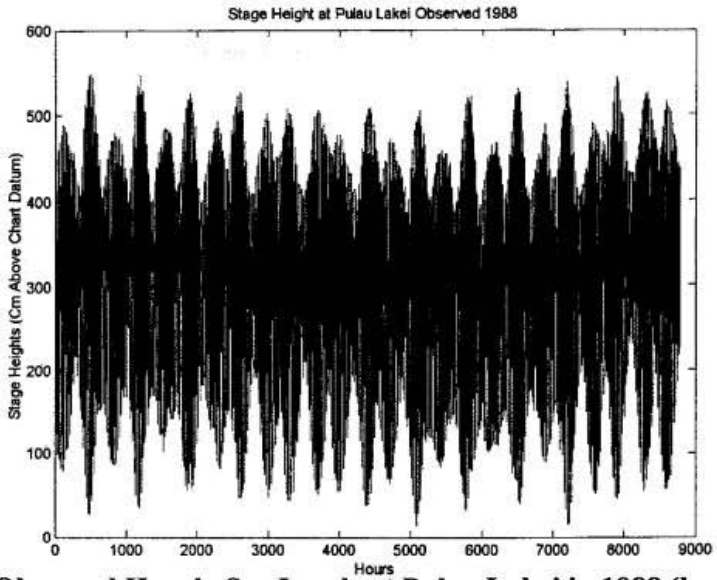


Figure 3-7 Observed Hourly Sea Levels at Pulau Lakei in 1988 (hourly from 0000 hour Jan 1, 1988)

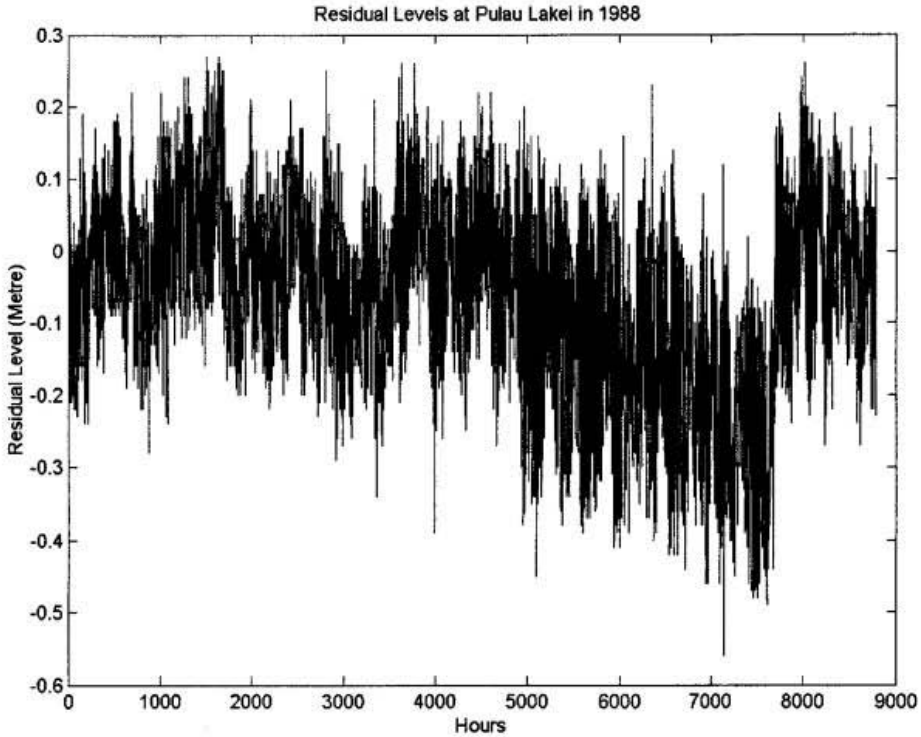


Figure 3-8 Residual Levels at Pulau Lakei in 1988

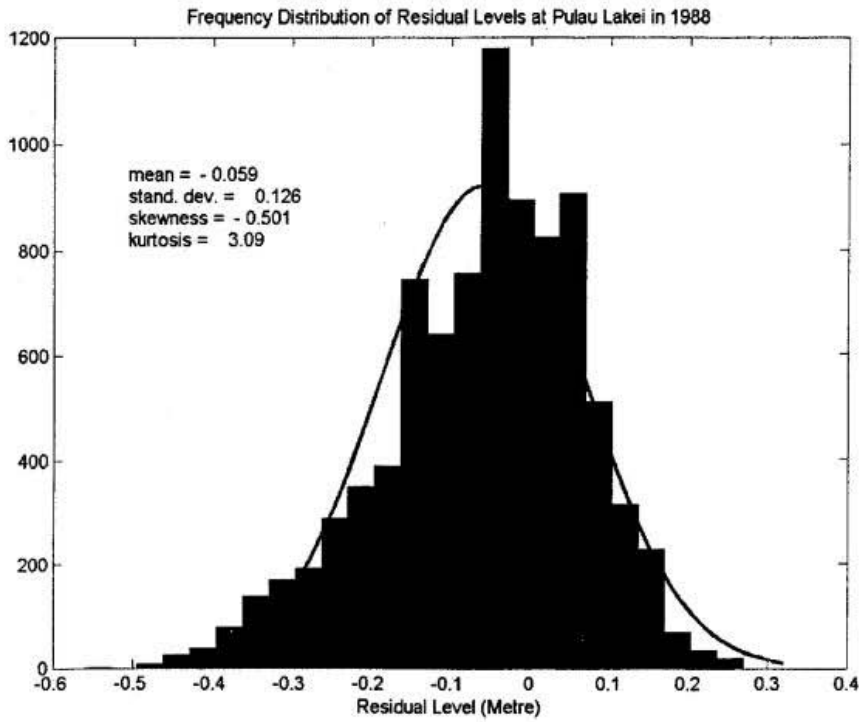


Figure 3-9 Frequency Distribution of Residual Levels at Pulau Lakei

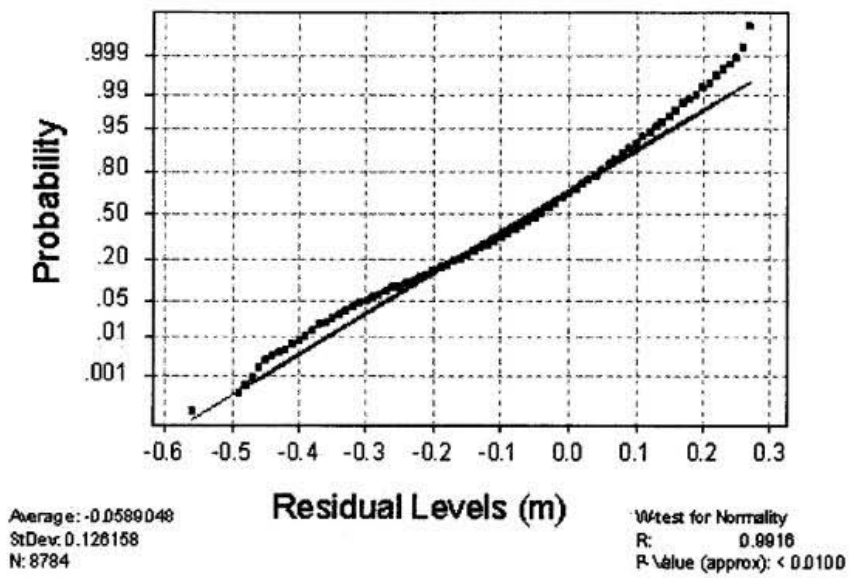


Figure 3-10 Normal Probability Plot and Normality Test of Residual Levels at Pulau Lakei, 1988

K = 0.0000

The observed number of runs =1452

The expected number of runs =3888.2

2903 Observations above K and 5881 below

The run test is significant at 0.0000

The probability of getting as extreme a number of runs or more extreme is found using a normal approximation. The probability is the attained significance level or p-value, which is less than 0.00005. This p-value indicates that for α level above 0.00005, there is sufficient evidence to conclude that the residual data are not in a random order.

Pugh (1987) illustrated a steady timing gain of 20 minutes over a week by a recorder clock can introduce non-tidal residuals of over one meter for a tidal series with maximum amplitude of 6 m. Hence the deficiency in the recorder clock at Pulau Lakei may introduce the gradual shift of the residuals as confirmed by the run test. If the gradual shift could be adjusted, the extreme surge values would have been less.

3.4.2 Summary on Surge Effects

The following remarks can be made:

1. Extreme surge values in the order of +0.26 and -0.48 m are found at Pulau Lakei, Sarawak.
2. The extreme residual levels or extreme surge levels at Pulau Lakei, Sarawak are confined to about + 2σ and - 4σ. This is small as compared with the site at Southend, UK, where + 7σ and - 6σ is the range reported in Pugh (1987). In about once in every year, British ports are reported to expect positive surge in

excess of 5 σ where as negative surge in excess of 4 standard deviation (- 4σ) occur on average less than once every 2 years.

3. The residuals are not normally distributed based on a strict normality test.

4. Run test done on the residuals shows that the residuals are not random; this is because there is a consistently higher prediction from late June to Mid November (5000 – 7500 hours).

5. The tropical cyclonic activities in Sarawak are weak since it is located near the equator. Significant storm surges have been associated with severe cyclonic activities in the Atlantic coasts of the U.S. and the North Sea in Europe.

3.4.3 Conclusion on Surge Effects

The surge component can be assumed to be negligible for the case of Sarawak as the extreme surge residual levels are small and typically confined to about + 2σ and - 4σ.

3.5 Hydrometric Networks Assessment

A critical survey of the existing hydrometric networks in Sarawak was carried out prior to using the data for the regional flood frequency analysis to be presented in Chapter 5. Hydrometric network in the context of study area refers to both the hydrological and hydrographic networks. The checking of a hydrometric network in providing sufficient information can be carried out using network assessment methods.

3.5.1 Network Status

The Department of Irrigation and Drainage, Sarawak (DID) is the main agency in Sarawak that maintains a comprehensive hydrometric network. Some other agencies also maintain a limited network of rainfall and water level stations.

3.5.1.1 DID Network

The network maintained by DID includes:

1. Rainfall stations: 235
2. Discharge measurement station: 55
3. Water level station: 19
4. Evaporation station: 15

3.5.1.2 Other Agencies and Network

The Malaysian Meteorological Service (MMS) maintains a climatic station at each of the five major airports. The Public Works Department (JKR) maintains some rainfall and water level stations primarily in the water works areas. Sea level observation stations are installed by some agencies around the coastal waters of Sarawak for various reasons. These agencies are:

1. National Survey and Mapping (for mean sea level determination)
2. Sarawak River Board (for operation of a barrage)
3. Shell Brunei (for offshore operation)

3.5.1.3 DID Rainfall Stations

There are over 235 rainfall gauges spread over the whole state of Sarawak with a coverage area of 530 km^2 per station. There are manual and standard stations. The standard station consists of a manual check gauge as well as an automatic rainfall gauge. The network is relatively dense in western Sarawak in comparison with the rest of the network.

3.5.1.4 DID Discharge Stations

Discharge measurement stations are found in most of the major river basins. There are a total of 55 discharge measurement stations in Sarawak. However, 17 of them are treated as major or principal stations while the rest are regarded as secondary stations. The principal discharge station consists of a housing that shelter an automatic water level recorder. A set of staff gauges that are related to a permanent datum are mounted on posts anchored along the riverbank to ensure a complete coverage of the expected water level range. The automatic recorder is set to run continuously while hydrological survey parties periodically take discharge measurements using current meters across designated river sections usually located near to the water level station. For the secondary stations, manual water level readings are performed twice a day by a part-time employed observer.

In Sarawak, the frequency of discharge measurement is around four per year per station. Most of the discharge stations chosen for long term gauging purposes have fairly stable beds. Ideally more measurements are necessary to ensure a good discharge-rating curve being derived for each station. The remoteness of some of these stations is the major discouraging factor for making more frequent measurements. Some of the sites are

inaccessible during high flows and/or low-flow periods where travel by riverboat or sampan is the main mode of transport, and these are the times when measurements are needed to provide better extrapolations to the rating curves.

3.5.1.5 DID Water Level Stations

Water level stations are installed at very specific locations primarily for discharge gauging purposes. However, there are 16 stations where the water level records itself are of interest. These can be found in tidal river reaches, ponds and drains. In fact those water level records in tidal reaches have provided most of the data required by the current research undertaking. Some of these stations are equipped with automatic water level recorders allowing levels to be resolved at 5 minutes interval (chart-based recorder) or preset time interval (digital recorder).

3.5.1.6 DID Evaporation Stations

The number of evaporation stations is 15, a relatively small number as compared to the rainfall stations. All the stations are equipped with a Class A Pan (US Geological Survey Standard). Whenever it rains, the day's record is discarded based on the assumption that there is no evaporation on a rain day. Owing to the very high number of rain days in a year, the interpretation of the records is always not satisfactory.

3.5.2 Network Adequacy

In an effort to meet the demand for more widespread coverage, especially in developing nations with limited resources, an excessive number of observation sites can

affect the quality of data collected. Simple measures such as network density and visual inspection using Thiessen polygon methods are usually employed to check the adequacy of network.

3.5.2.1 Broad Density Ratios

A density ratio of the total basin area covered by the whole network to the number of stations is commonly used in a World Meteorological Organisation report written by Moss (1982) as an indication of the adequacy of network. It is just an indication of an average condition and does not differentiate disparity due to differences in spatial distribution of stations. Network densities of DID Hydrologic Network are evaluated as shown in Table 3-5. If only considering reliable stations (permanent type), the network density in Sarawak has barely achieved the minimum standard set by World Meteorological Organization (WMO) for tropical areas with difficult conditions.

Nevertheless, the even spatial distribution of station may not guarantee optimum distribution of observation stations since the information content is not scrutinized. A modern approach entails using entropy concepts, as described in the next section.

Table 3-5 DID Hydrological Network Density as Compared with WMO Standard

Network Type	Density of DID Network/Station	WMO Normal Condition	WMO Difficult Condition
Rainfall station (permanent)	1:1400 km^2 (88)	1:1000 to 1:250km^2	1:1000km^2
Rainfall station (all included)	1:530 km^2 (235)		
Discharge station (permanent)	1:7300 km^2 (17)	1:300 km^2	1:5000 km^2
Discharge station (all included)	1:2500 km^2 (55)		

3.5.2.2 Entropy Approach

Singh (1997) provides a good review paper on the use of entropy in hydrology and water resources. The entropy approach has been used for the assessment of hydrometric networks by Harmancioglu (1981), Hussain (1989), Harmancioglu and Alpaslan (1992), Yang and Burn (1994), and Moss (1997). Optimization of the network in terms of station location can be doned using the approach. The aim is a more equitable network designed to achieve more effective network coverage with the same effort level or budgetary allocation.

In the entropy approach, a hydrologic network is regarded as a communication channel, which is dedicated to transmitting hydrologic information (Caselton and Husain, 1980). Shannon and Weaver (1949) defined the basic principles of information theory treating the information content of signals as "entropy" analogous to that of entropy in statistical mechanics. Information, I, is attained only when there is uncertainty about an event, which implies the presence of alternative results the event may assume. Hence information associated with certainty is zero.

Alternatives with a low probability of occurrence convey considerable information while a high probability of occurrence indicates little information being transmitted. The probability of occurrence of a certain alternative is the measure of entropy or the degree of "unexpectedness" of a signal. Signals that are sent through a communication channel must be uncertain before they are transmitted. The entropy of a random process is expressed in certain specific units depending on the statistical structure identified. The entropy function is used as an objective criterion for measuring the information content of any random process.

For a discrete random variable X, because the information associated with certainty is zero,

$$I \propto \log\left(\frac{1}{p(x_i)}\right) = -K \log p(x_i) \qquad (3\text{-}4)$$

where N is the number of elementary events with probabilities $p(x_i)$ for $i = 1, 2,\ldots N$. The entropy, $H(X)$, is defined as the expectation of I:

$$H(X) = -K\sum_{i=1}^{N} p(x_i) \log p(x_i) \qquad (3\text{-}5)$$

where K=1 if $H(X)$ is in "napiers" for logarithms to the base e. If the probability of occurrence and non occurrence of an event is equally likely, the entropy associated with the occurrence of the events will be a maximum. On the other hand, if the probability of occurrence of events is zero or one, the entropy will be zero and there will be no uncertainty associated with the occurrence of the events of variable X.

The total entropy of two independent random variables X and Y is equal to the sum of their marginal entropies:

$$H(X,Y) = H(X) + H(Y) \qquad (3\text{-}6)$$

Transinformation, T(X,Y), measures the redundant or mutual information between X and Y is described as the difference between the total entropy and the joint entropy of dependent X and Y.

$$T(X,Y) = H(X) + H(Y) - H(X,Y) \qquad (3\text{-}7)$$

or as

$$T(X,Y) = H(X) - H(X/Y) = H(Y) - H(Y/X) \qquad (3\text{-}8)$$

3.5.2.3 Rainfall Network Analysis Results

Judging from simple density evaluation, western Sarawak has a higher density than the rest of the State. Hence it is only essential to evaluate the possibility of optimizing the rainfall station network for western Sarawak. Preliminary assessment has also indicated that the network is generally not sufficiently dense according to WMO's minimum standard. Hence there is very little justification to execute very detailed assessment technique as used by Husain(1989) and Yang and Burn (1994).

The network of rainfall stations as maintained by DID Sarawak was analysed using the entropy approach, using 8 years (1990-1997) of simultaneously recorded daily rainfall series. A two-station paired approach for daily rainfall was used for this assessment. Two rainfall series obtained at two separate locations were treated as two random variables X and Y. Average daily rainfall varies between 10 to 15 mm and a daily maximum of around 100 mm was recorded. Based on this information, twenty categories were defined at a discrete interval of 5 mm. The entropies of X and Y and the joint entropies where the same rainfall categories occur at the same time were computed. Whenever there was a gap or missing data in one of the series, the corresponding data of the other series was discarded. The transinformation were subsequently computed. If the value of transinformation was low, the station-pair may have certain redundancy in the rainfall information collected. If the values were consistently high, there was less likelihood of redundancy. For manageable computation, western Sarawak was conveniently divided into three groups by geographic proximity.

Figure 3-11 shows the location of the rainfall stations used for the analysis. Table 3-6, Table 3-7 and Table 3-8 shows the computed transinformation for all the station-pairs within each group. It can be seen that the transinformation values are all greater than 3, indicating that in general the network does not have much redundancy. This result further confirms the general inadequacy of the rainfall network in western Sarawak even it has the denser network as compared to rest. It also implies that the rest of Sarawak with much lower rainfall network density certainly needs further expansion of the network.

3.5.2.4 Streamflow Network Analysis

It is often perceived that there are redundancies in measuring water levels at river stations, especially in an upstream and downstream configuration. The use of entropy approach to assess streamflow network is given in Yang and Burn (1994). For the present network in Sarawak, the absence of many long term gauging stations means that the use of an elaborate assessment technique is not warranted.

3.6 Chapter Summary

The combination of high rainfall, high evaporation, high temperature, high humidity, low sea surge range, high tidal range, and weak cyclonic activity found in the rainforest covered Sarawak is very much related to its unique geographic location almost right on the equator, surrounded largely by the South China Sea to the north and northwest and the Indonesian Borneo land masses to the south and southeast.

Data related to the study area such as rainfall, discharge, and water level data are available. However, only those river basins that have the appropriate combination of a sea level station off the coast, a water level station in the TIZ, and a discharge station on the mainstream are chosen for studies.

Investigations have shown that the tides off the coasts of Sarawak are mainly of the semi diurnal species. The maximum tidal range is around 6 m in western Sarawak and diminishes in the northern coast. The tidal level distribution cannot be represented by the normal distributions.

The surge effects can be regarded as negligible as compared to other sites around the world. The surge or residual levels at Pulau Lakei is not significant, with tidal residuals varying in the maximum range of $+2\sigma$ to $-4\ \sigma$.

Hydrometric network assessment using a broad indicator, such as the coverage area per station, shows that the network barely meets the minimum standard of the WMO specified for tropical places with accessibility problems. Although more stations are found in western Sarawak, there is an uneven spatial distribution of rainfall stations. An assessment using the entropy approach shows that the denser western Sarawak does not have any redundancy with respect to station location.

Table 3-6 Transinformation Between Stations of Group 1 (Betong Group)

Station Reference	1A	1B	1C	1D	1E	1F	1G	1H	1I	1J	1K	1L	1M	1N	1O	1P
1A. Stumbin	1.72	3.06	3.48	3.43	3.16	3.23	3.19	3.32	3.03	3.38	3.31	3.26	3.17	3.32	3.31	3.38
1B. Lingga	3.06	1.55	3.38	3.34	3.00	3.15	3.06	3.21	2.88	3.27	3.22	3.17	3.07	3.17	3.20	3.28
1C. Ng. Mujan	3.48	3.38	1.93	3.59	3.45	3.49	3.42	3.58	3.26	3.64	3.53	3.52	3.43	3.55	3.56	3.63
1D. Ng. Entalau	3.43	3.34	3.59	1.94	3.31	3.31	3.48	3.53	3.30	3.51	3.48	3.47	3.32	3.54	3.51	3.60
1E. Sebuyau	3.16	3.00	3.45	3.31	1.69	3.25	3.12	3.27	2.88	3.37	3.28	3.21	3.15	3.23	3.28	3.35
1F. Lubau	3.23	3.15	3.49	3.31	3.25	1.76	3.20	3.32	3.04	3.32	3.25	3.26	3.18	3.35	3.33	3.39
1G. Betong	3.19	3.06	3.42	3.48	3.12	3.20	1.60	3.28	2.95	3.34	3.37	3.24	3.15	3.21	3.27	3.34
1H. Debak	3.32	3.21	3.58	3.53	3.27	3.32	3.28	1.79	3.10	3.45	3.39	3.34	3.26	3.39	3.41	3.49
1I. Pusa	3.03	2.88	3.26	3.30	2.88	3.04	2.95	3.10	1.45	3.16	3.15	3.01	2.93	3.04	3.06	3.15
1J. Danau	3.38	3.27	3.64	3.51	3.37	3.32	3.34	3.45	3.16	1.85	3.39	3.41	3.32	3.44	3.47	3.53
1K. Ng. Tiga	3.31	3.22	3.53	3.48	3.28	3.25	3.37	3.39	3.15	3.39	1.88	3.32	3.23	3.46	3.38	3.46
1L. Saratok	3.26	3.17	3.52	3.47	3.21	3.26	3.24	3.34	3.01	3.41	3.32	1.75	2.94	3.30	3.23	3.37
1M. Saratok Water	3.17	3.07	3.43	3.32	3.15	3.18	3.15	3.26	2.93	3.32	3.23	2.94	1.64	3.20	3.13	3.29
1N. Sessang	3.32	3.17	3.55	3.54	3.23	3.35	3.21	3.39	3.04	3.44	3.46	3.30	3.20	1.71	3.31	3.45
1O. Roban	3.31	3.20	3.56	3.51	3.28	3.33	3.27	3.41	3.06	3.47	3.38	3.23	3.13	3.31	1.76	3.41
1P. Pakan	3.38	3.28	3.63	3.60	3.35	3.39	3.34	3.49	3.15	3.53	3.46	3.37	3.29	3.45	3.41	1.85

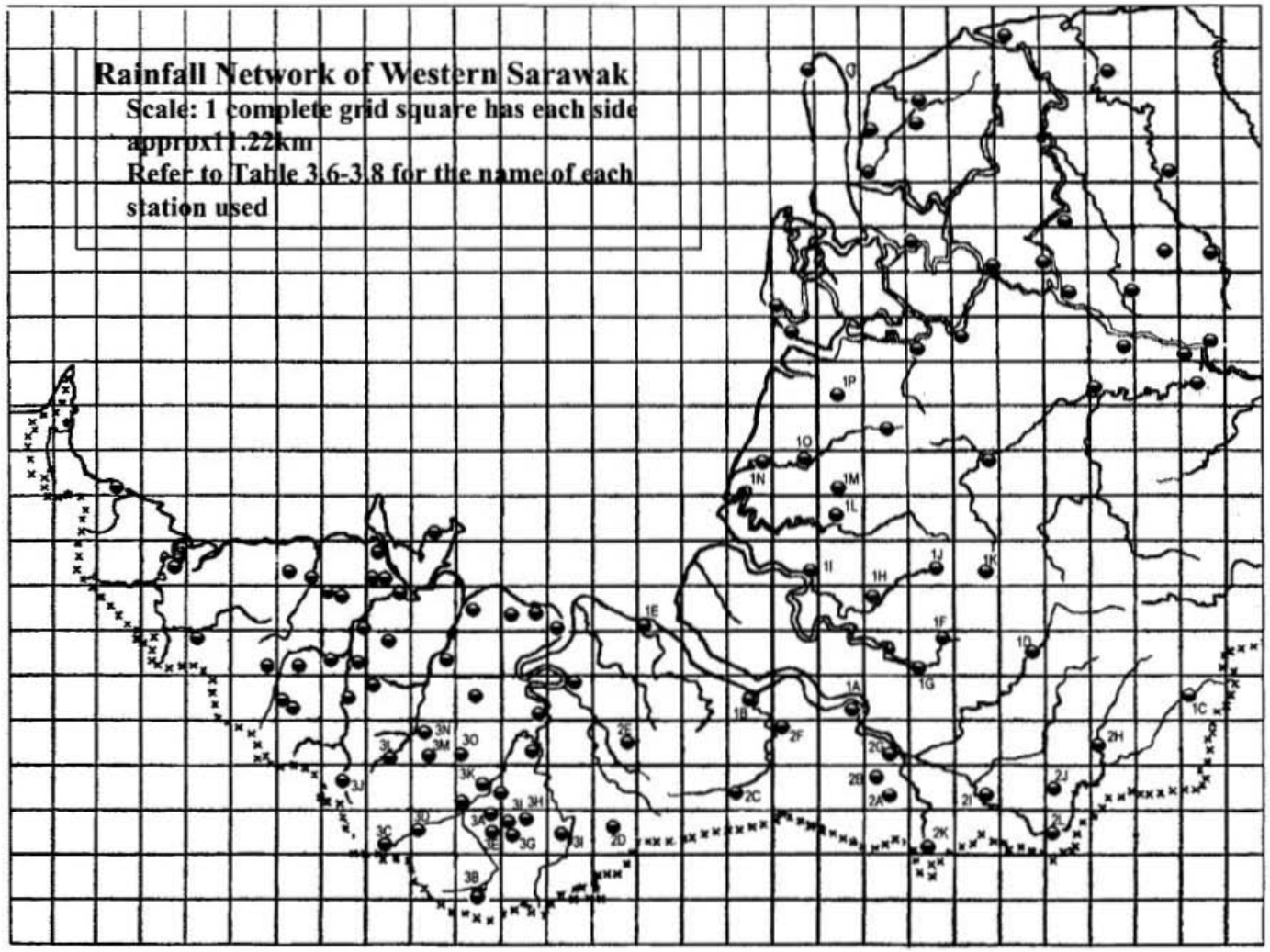
Table 3-7 Transinformation Between Stations of Group 2 (Sri Aman Group)

Station Reference	2A	2B	2C	2D	2E	2F	2G	2H	2I	2J	2K	2L
2A. Melugu	1.77	3.18	3.41	3.69	3.45	3.39	3.28	3.42	3.30	3.16	3.37	3.28
2B. Temudok	3.18	1.56	3.31	3.50	3.29	3.25	3.19	3.27	3.17	3.03	3.30	3.15
2C. Pantu	3.41	3.31	1.86	3.79	3.45	3.40	3.39	3.50	3.39	3.23	3.47	3.36
2D. Pinang	3.69	3.50	3.78	2.02	3.77	3.72	3.71	3.74	3.67	3.49	3.80	3.63
2E. Sangkalan	3.45	3.29	3.45	3.77	1.86	3.45	3.43	3.52	3.41	3.26	3.51	3.39
2F. Banting	3.38	3.25	3.40	3.72	3.45	1.79	3.34	3.47	3.35	3.20	3.46	3.33
2G. Sri Aman	3.28	3.19	3.39	3.71	3.43	3.34	1.78	3.41	3.24	3.11	3.35	3.26
2H. Tutong	3.42	3.27	3.50	3.74	3.52	3.47	3.41	1.80	3.34	3.12	3.48	3.29
2I. Engkilili	3.30	3.17	3.39	3.67	3.41	3.35	3.24	3.34	1.72	3.04	3.29	3.18
2J. Bekatan	3.16	3.03	3.23	3.49	3.26	3.20	3.11	3.12	3.04	1.55	3.22	3.01
2K. Lintang	3.37	3.30	3.47	3.80	3.51	3.46	3.35	3.48	3.29	3.22	1.86	3.27
2L. Lubok Antu	3.28	3.15	3.36	3.63	3.39	3.33	3.26	3.29	3.18	3.01	3.27	1.69

Table 3-8 Transinformation Between Stations of Group 3 (Serian Group)

Station Reference	3A	3B	3C	3D	3E	3B	3C	3D	3E	3F	3G	3H	3I	3J	3K	3L	3M	3N	3O	3P
3A. Busit	1.81	3.35	3.30	3.17	2.87	3.35	3.30	3.17	2.87	2.86	3.02	2.87	3.25	3.55	3.24	3.47	3.39	3.39	3.53	3.34
3B. Bunan	3.35	1.83	3.34	3.35	3.22	1.83	3.34	3.35	3.22	3.30	3.29	3.05	3.35	3.61	3.47	3.53	3.51	3.48	3.59	3.41
3C. Tebedu	3.30	3.34	1.80	3.26	3.19	3.34	1.80	3.26	3.19	3.27	3.27	3.06	3.36	3.51	3.42	3.47	3.43	3.38	3.57	3.41
3D. Krusen	3.17	3.35	3.26	1.82	3.12	3.35	3.26	1.82	3.12	3.19	3.22	3.01	3.36	3.55	3.33	3.47	3.38	3.40	3.56	3.39
3E. Matuh	2.87	3.22	3.19	3.12	1.70	3.22	3.19	3.12	1.70	2.79	2.63	2.74	3.10	3.47	3.19	3.37	3.30	3.27	3.43	3.24
3F. Merang	2.86	3.30	3.27	3.19	2.79	3.30	3.27	3.19	2.79	1.79	2.93	2.78	3.20	3.52	3.26	3.44	3.36	3.36	3.49	3.32
3G. Teb	3.02	3.29	3.27	3.22	2.63	3.29	3.27	3.22	2.63	2.93	1.77	2.74	3.16	3.54	3.30	3.42	3.38	3.38	3.52	3.33
3H. Bedup	2.87	3.05	3.06	3.01	2.74	3.05	3.06	3.01	2.74	2.78	2.74	1.46	2.98	3.29	3.05	3.21	3.16	3.08	3.25	3.03
3I. Ringin	3.25	3.35	3.36	3.36	3.10	3.35	3.36	3.36	3.10	3.20	3.16	2.98	1.82	3.59	3.38	3.51	3.48	3.45	3.58	3.37
3J. Padawan	3.55	3.61	3.51	3.55	3.47	3.61	3.51	3.55	3.47	3.52	3.54	3.29	3.59	1.97	3.64	3.50	3.58	3.53	3.76	3.58
3K. Serian	3.24	3.47	3.42	3.33	3.19	3.47	3.42	3.33	3.19	3.26	3.30	3.05	3.38	3.64	1.89	3.52	3.40	3.42	3.57	3.39
3L. Gayu	3.47	3.53	3.47	3.47	3.37	3.53	3.47	3.47	3.37	3.44	3.42	3.21	3.51	3.50	3.52	1.91	3.40	3.38	3.65	3.47
3M. Nyabet	3.39	3.51	3.43	3.38	3.30	3.51	3.43	3.38	3.30	3.36	3.38	3.16	3.48	3.58	3.40	3.40	1.91	3.27	3.60	3.44
3N. Dragon	3.39	3.48	3.38	3.40	3.27	3.48	3.38	3.40	3.27	3.36	3.38	3.08	3.45	3.53	3.42	3.38	3.27	1.85	3.54	3.40
3O. Tarat	3.53	3.59	3.57	3.56	3.43	3.59	3.57	3.56	3.43	3.49	3.52	3.25	3.58	3.76	3.57	3.65	3.60	3.54	1.92	3.54
3P. Gedong	3.34	3.41	3.41	3.39	3.24	3.41	3.41	3.39	3.24	3.32	3.33	3.03	3.37	3.58	3.39	3.47	3.44	3.40	3.54	1.73

Figure 3-11 Map of Western Sarawak Showing three Groups of Rainfall Stations Used in Network Assessment



Chapter 4 Overview of Methods for Estimating Design Floods in the TIZ

The literature on the methods available for estimating design floods in the tidal interaction zones is scarce. A comprehensive investigation and streamlining of the methods has not been carried out. This Chapter provides an overview of all the possible methods, including those currently in use, those revised in Chapter 6, and those that are newly proposed and developed in Chapter 7 and 8.

4.1 Direct Frequency Method

Coastal engineers normally perform direct frequency analysis of the sea elevations observed at coastal sites, typically at seaports and oil drilling platforms, for design purposes. The Gumbel distribution has been the *de facto* distribution assumed to fit a probability density function for the extreme sea elevations with a record length of $N\Delta T$ where N is an integer and ΔT is a sub-record length containing many independent extreme events. Owing to the name of the distribution in use, this method is often referred in the coastal engineering literature as the *Gumbel Method*. For most cases, ΔT is taken to be 1

year and N is preferred to be greater than 25. The method is simple and straightforward, considering annual extremes as a result of all the causative factors of tides and surges. It provides a quick indication on the return periods of extreme levels. Evaluations of the return periods of sea levels can be found in Middleton and Thompson (1986) and Bardsley *et al.* (1990).

The method can be applied to the TIZ assuming that the annual extremes are the resultants of all the prevailing physical forces in the TIZ. However, the fluvial flood component is significantly large as compared to the surge and tidal components.

One limitation of the direct frequency method is the inefficient usage of data as only one point is taken per year and hence a long record is normally required. Middleton and Thompson (1986) insisted that in a sea condition with strong nodal modulations, a better choice for ΔT would be 18.6 years, a maximum tidal period. The usage of many short to medium long-term records for analysis would be curtailed if this recommendation were strictly adhered to.

In any circumstance when this method is to be used, it is proposed that the distribution used should not be confined to the Gumbel distribution alone. Many other probability distributions should be considered, together with the application of the L-moment approach for parameter estimation that has been briefly described in Chapter 2. Further descriptions and applications of the approach can be found in Chapter 5.

4.2 Rating Curve Method

The idea behind the Rating Curve Method is similar to the traditional approach of rating discharge at a river section with the objective of deriving a relationship between

observed discharges and water levels. In this method applicable to the TIZ, three synchronous time series at three locations are required: a water level station located in the TIZ, an upstream discharge observation station, and a sea-level station in the open sea. The main feature of this method is that the observations of water levels and discharges are made at different locations.

The upstream discharge records are analyzed using flood frequency analysis as described in Chapter 5 so that flood quantiles are available. It is common to find that no correlation exists between the sea levels and the river discharges since the downstream tidal sequence is deterministic while stochastic processes govern the upstream discharges. However, a particular level Z observed in a TIZ can be associated with a certain combination of the river discharge Q and the sea level H. A regression line for each convenient category of Z can be fitted for the associated group of Q-H data. A plot of the concept is shown in Figure 4-1. The plot looks similar to stage-discharge rating curves and hence the method is known as the Rating Curve Method. Van der Made(1969) used this method to derive the return period floods in the Rhine deltas in Netherlands. The user specifies both the river flood quantiles at the upstream and the extreme levels in the sea, and the corresponding level in the TIZ is derived from the curves.

However, the Rating Curve Method suffers from the arbitrary choices that have to be made, such as the selection of lag-time between tides in the sea and at the TIZ. Since the distance between a typical discharge station and a station in the TIZ can be in the order of 20 to 100 km, the amount of unaccounted lateral inflows from sizeable tributaries can be substantial. The scatter in the regression can be very large due to the over-simplification of the physical processes.

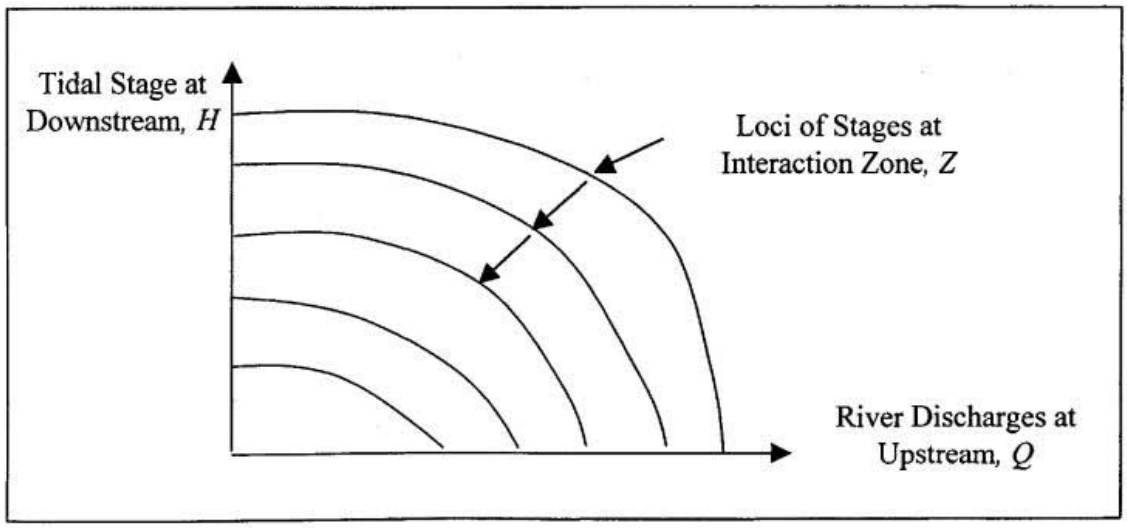


Figure 4-1'Rating Curves' at A Tidal Interaction Zone

Van der Made(1970) further developed an approach in accounting for the total frequency of exceedance based on three major cases of exceedance: (a) low discharges, high tides (b) medium discharges, medium high tides and (c) high discharges, low high tides. In the situation where the records are short, the derived *H-Q* curves can be very inaccurate. Any further elaboration on the curves can be dubious. Further descriptions and applications of the rating curve method can be found in Chapter 6.

4.3 Bivariate Distribution Method

The Bivariate Distribution Method is an extension of the Rating Curve Method. It also requires synchronous measurements to be carried out at three locations. The flood exceedance probability function at an upstream discharge station can be derived from conventional flood frequency analysis while the sea level probability distribution can be derived from observed sea levels, preferably from the records of a long-term sea-level

station. In the absence of long records, tidal harmonic models are fitted to shorter records. The models are then used to simulate tidal levels and to derive the corresponding distribution. The sea-level H and the upstream discharge Q are assumed to have probability functions say $f(H)$ and $g(Q)$ respectively. The distribution of the water level at a point of interest in the interaction zone, Z is, as in the rating curve method, given by

$$Z = h(H, Q) \tag{4-1}$$

and the probability of H and Q is a bivariate distribution

$$p(H,Q) = f(H) \cdot g(Q) \tag{4-2}$$

Figure 4-2 shows a sketch on the concept of combining the probabilities of H and Q to form a bivariate distribution. It is assumed that the fluctuation in sea levels and the variation in river discharges are independent events. A variant of the method, known as the Joint Probability Method (JPM), has been used for estimating the probability distribution of extreme sea levels. It was introduced by Pugh and Vassie (1980) and later improved by Tawn and Vassie (1989), and Tawn (1992). Tidal regime and surge phenomenon are the main concerns in the sea conditions. D'Onofrio (1999) used this method for evaluating the return periods of extreme water levels in Buenos Aires, Argentina, which is located on a tidal estuary. However, little information on the application of the method in a tidal river, where tides and fluvial floods are the main concerns, has been published. Theoretically the method can work for the TIZ assuming that the tides and river flows are generated by two independent processes. A brief introduction to a conceptual method can be found in WMO Report No. 704 (WMO, 1988). The method is further investigated in Chapter 6.

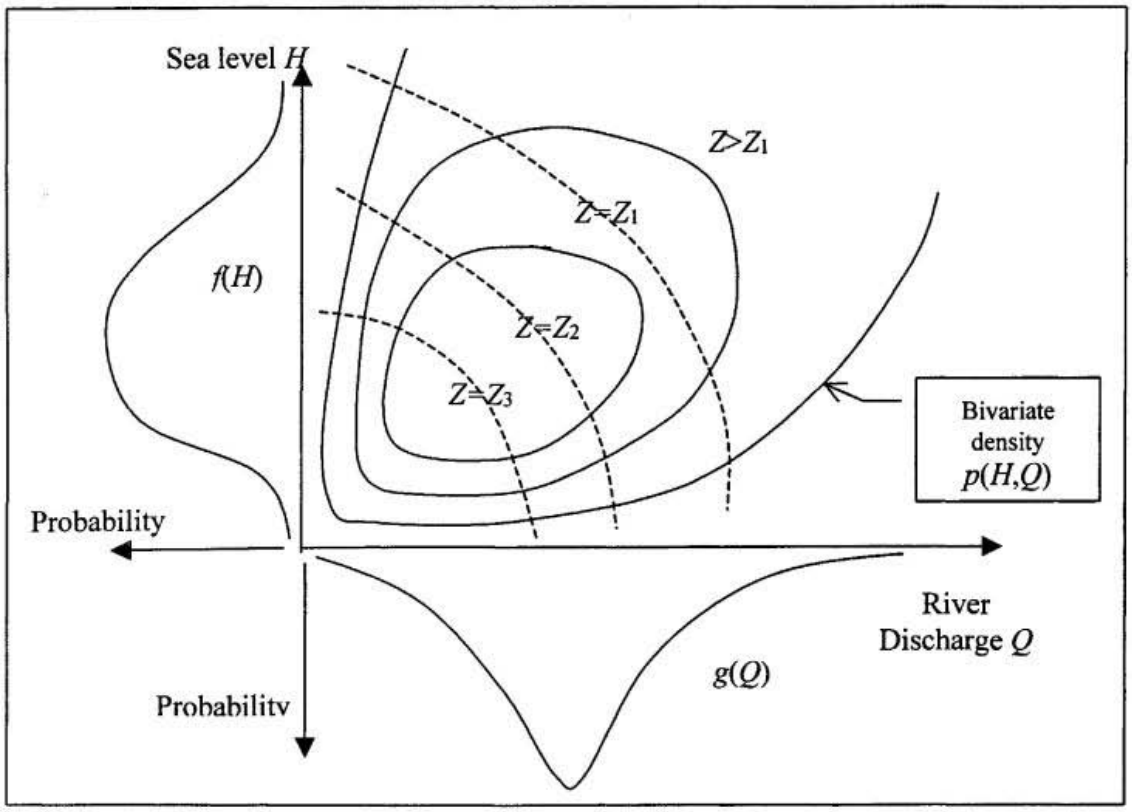


Figure 4-2 Combining Probabilities in Bivariate Distribution Method

4.4 Hydrodynamic Modelling Method

A deterministic approach using hydrodynamic modelling may provide flood information including water levels and discharges at any specific location in the tidal interaction zone for a known flood event. Figure 4-3 shows a schematic of the hydrodynamic modelling method. A rainfall-runoff model typically uses the output of a rainfall model and a hydrologic basin model to generate runoff hydrographs from sub-basins which respond to certain design storms. The runoff hydrographs could also be developed by other methods. A full St. Venant equation-solving hydrodynamic model is then coupled to generate flow conditions in the dynamic model areas based on the outputs

from a hydrologic model upstream and a tidal model downstream. A surge effect is considered in the tidal model if it is of significance.

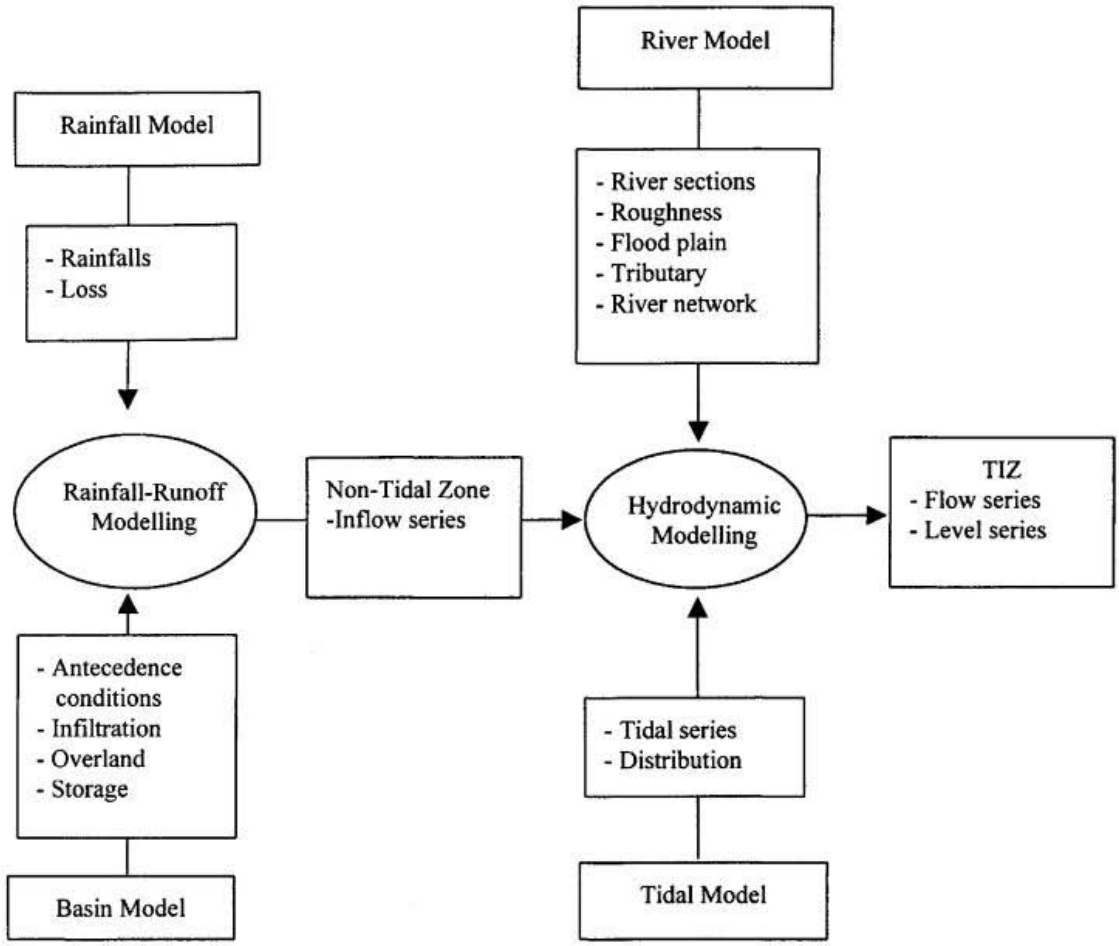


Figure 4-3 A Schematic of Hydrodynamic Method

The hydrodynamic models provide no information on the probability of the outcomes. Users often assume that the occurrence of a certain recurrence interval flood within a non-tidal zone upstream would generate a flood level of the same recurrence interval in a downstream position in the tidal interaction zone. There are also complexities involved such as the assignments of various tidal boundary conditions downstream, the

combination of different tributary inflows with different flow distribution and characteristics, and the dependence on the calibrated model parameters, amongst others.

In order to derive continuous flow simulations in the TIZ, continuous simulations of the boundary variables such as rainfall and tides are increasingly used. The simulated series can be used to establish a pdf of the water levels in the TIZ giving rise to the estimates of flood quantiles . However, many considerations have to be addressed to use the hydrodynamic method. These include the demand for many synchronous observation records at different locations, and difficulties in model calibration and verification, among others. A hydrodynamic simulation typically uses a time step not greater than 5 minutes. This implies that a lot of data would be generated for the continuous simulation cases.

From a non-technical perspective, the time and resources required for setting up a hydrodynamic model is often beyond the scope of many small projects. Apart from the gauged records, there are numerous topographic and hydrographic data that have to be collected to define the river model as closely as possible. For a river with very extensive TIZ, the area to be considered can be a very large portion of the river basin itself.

4.5 Fourier Method

Fourier methods have been employed in many time series analysis, mainly for spectral analysis and filtering purposes. In the version of this method that is applicable to the TIZ, a water level series Z(t) is subjected to a filtering process. The objective is to obtain two separate series T(t) and D(t) that correspond to the tidal regime and river flow regime respectively from the water level series Z(t). In places where surge component is significant, D(t) would be a combined effects of the river flows and surges. Surge

phenomenon is small in Sarawak as found in Chapter 3. Since the component due to the tidal regime is deterministic, as described in Chapter 2 and 3, the separation by filtering in the Fourier domain can be achieved. The frequencies of the known harmonic components are at least greater than 12 deg./hour (Easton, 1977). Utilizing this fact, a low pass filter is designed to allow the low-frequency components attributed to the slow changing river series to be separated from the high-frequency components attributed to the tides. Inverse Fourier transforms are then performed allowing the two series to be separated. The traditional methods of frequency analysis can subsequently be carried out to derive the probability of exceedance for the extreme levels using the joint probability approach. The flood level LT with a return period T can be determined from the joint probability. Figure 4-4 shows a schematic of the Fourier Method.

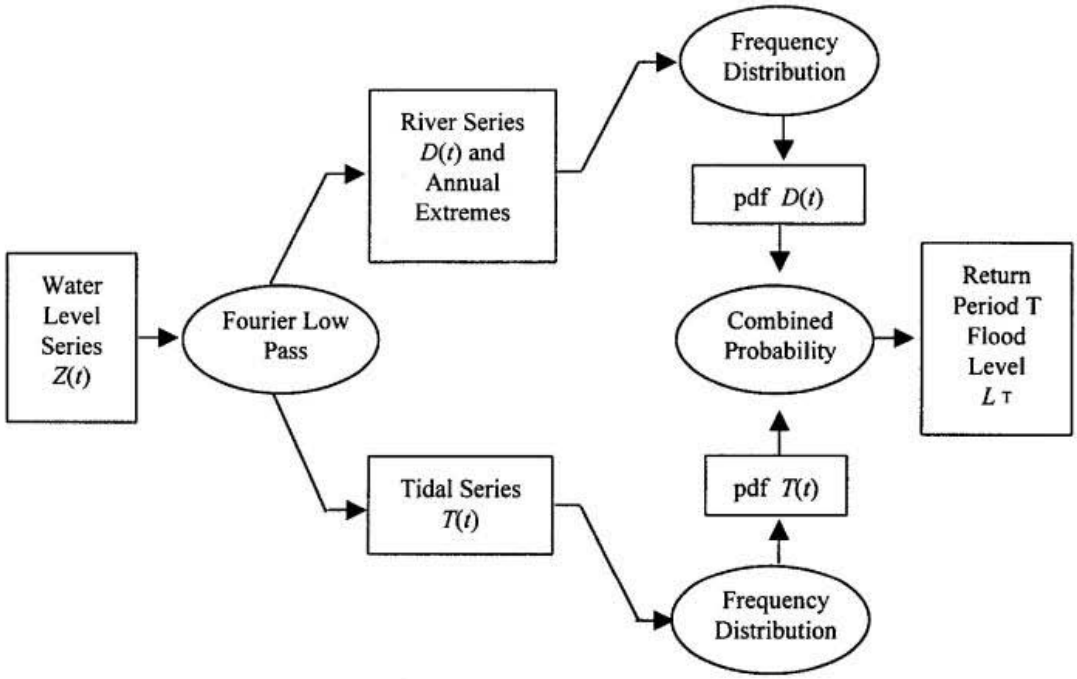


Figure 4-4 A Schematic of Fourier Method

Goring(1984) used the Fourier method to separate tides and river flow regimes in New Zealand rivers. However, the exercise was not tied to any flood frequency analysis, rather it was purely used for predicting the arrival of tides at certain river ports. Chapter 7 describes the details of the method and its application to the data obtained from Sarawak.

4.6 Wavelet De-noising Method

Wavelet analysis is a new approach applicable to solve many problems where Fourier methods have been used. Among the many wavelet methods, the wavelet de-noising method is found to be suitable for analysing the water level signals found in the TIZ. A schematic of the method is shown in Figure 4-5.

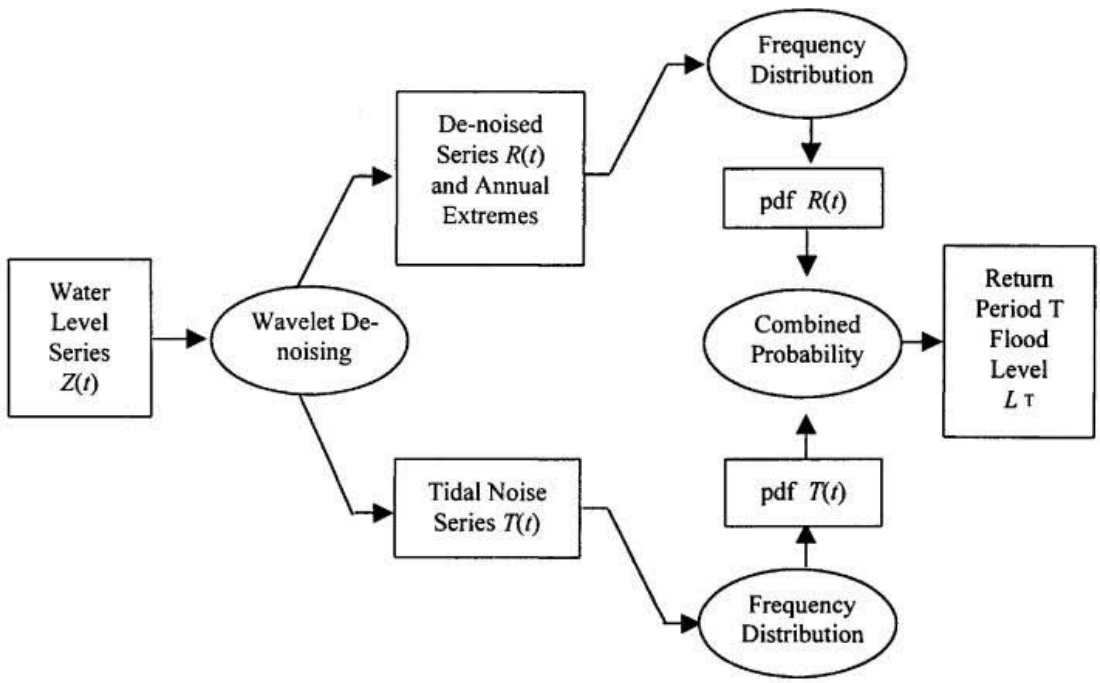


Figure 4-5 A Schematic of Wavelet De-noising Method

The application of the wavelet de-noising method to the water level series $Z(t)$ in the TIZ is based on an assumption that the tide fluctuates as noise while the low frequency streamflow component behaves as an approximation function. A de-noised series $R(t)$ that depicts a streamflow series allow the derivation of a probability distribution of extreme water levels attributed to river flows. The analysis of the series is similar to the traditional flood frequency analysis in the non-tidal zone. Since the surge phenomenon is small in Sarawak, the de-noised series $R(t)$ would be attributed to the streamflows alone rather than as a combined effect of the surges and the streamflows.

The residuals obtained from the differences between the observed series and the wavelet de-noised series form a tidal noise series $T(t)$. From the noise series an empirical probability density function can be derived. The tidal sequence is considered as deterministic based on known excitation theories of celestial bodies. Ideally an interaction factor of momentum transfer between the two flow regimes has to be accounted for. A simplistic approach is adopted by assuming the two series as independent. The joint probability of extreme levels can be derived by using a Monte Carlo simulation technique. The flood level L_T with a return period T can be determined from the joint probability. Chapter 8 describes the method in more detail and the data observed in the TIZ's of Sarawak are used for checking and verification. Owing to the unique properties of wavelets, such as the ability to handle non-stationary signals, localized spikes and discontinuities, the wavelet de-noising method is a promising technique for solving this type of problem.

4.7 Choice of Method

Design flood estimations can have many purposes. In certain situations, flood levels are required just for a very specific site, such as an industrial development, or a river port. In other cases, it may be a large area, e.g., a major agricultural development on a large parcel of land. In some circumstances, only water levels are sufficient, e.g. a minimum platform level for a power station. Yet, in other circumstances, both discharges and water levels are required, e.g., a river barrage construction. Hence, the purpose must be clear before adopting a particular method. Table 4.1 summarizes the major input and output characteristics of the methods. Data availability is often a major determining factor in the selection of appropriate model.

Table 4-1 Characteristics of Methods

Methods	Input	Output
Direct Frequency Analysis	• Water levels series in the TIZ	• Flood quantiles (water levels) in the TIZ
Rating Curve Method	• Water levels series in the TIZ • Upstream discharge series • Tidal series downstream	• Flood quantiles (water levels) in the TIZ
Bivariate Distribution Method	• Water levels series in the TIZ • Upstream discharge series • Tidal series downstream	• Flood quantiles (water levels) in the TIZ
Hydrodynamic Methods	• Water levels series in the TIZ • Upstream discharge series • Tidal series downstream • Cross-sectional Data • Rainfalls or simulated • Simulated inflows • Simulated tides • Channel physical roughness	• Flood quantiles (water levels and discharges) in the TIZ
Fourier Method	• Water levels series in the TIZ	• Flood quantiles (water levels) in TIZ
Wavelet De-noising Method	• Water levels series in the TIZ	• Flood quantiles (water levels) in TIZ

Chapter 5 Regional Flood Frequency Analysis

Apart from the propagation of tides, the major factor that influences the flow conditions in a TIZ is the inflow from sub-basins located upstream. Most of the methods identified in Chapter 4 require some prior estimation on the design flood quantiles at those strategic points. Thus the derivation of flood quantiles in non-tidal zones, particularly at a sparcely gauged geographic state such as Sarawak, is an important facet of estimating design floods in the TIZ's. This chapter presents a comprehensive treatment of flood frequency analysis for the study area using a regional approach that best utilizes the available flow information.

At present, there is no specific regional flood estimation procedure developed for Sarawak. Many practicing engineers rely on rainfall intensities to generate flood peaks using a simple method called the rational method for design flood estimations. Despite the fact that the rational method is intended for very small basins, its application in some large basins without any modification is common practice. There are many circumstances where flood estimation procedures applicable to Peninsular Malaysia, published by DID,

Malaysia, are adopted for use in Sarawak with the assumption that the basins characteristics are similar. The approach becomes irrational when a particular region in Peninsular Malaysia has to be chosen arbitrarily as similar to Sarawak. On the other hand, some rainfall-runoff simulation programs are also being used. It is observed that there are attempts to apply simulation programs using default values of internal parameters without actual comprehensive calibration of regional basins. There are also circumstances where calibrations are done on relatively ordinary flood events and then extrapolated to make estimates for extreme floods.

5.1 Objectives and Rationale

The purpose of design flood estimation is to make predictions on the magnitude of flood discharges at a particular section of a river of interest corresponding to a risk level that is acceptable to the design standards of structures. The risk level is normally taken to be a probability of non-exceedance expressed as a certain return period (T-year) or annual recurrence interval (ARI). With sufficient length of flow observations at a particular site of interest, one can make statistical inferences on the flood discharges corresponding to various acceptable risk levels. Typically a highway bridge is designed for a 50-year flood while a small dam may be designed for a 100-year flood.

5.1.1 Objectives of Regional Flood Frequency Analysis

The objectives of regional flood frequency analysis as presented in this chapter are:

(i) to derive regional flood frequency equations for the study area based on the flood data series obtained from the gauged stations located upstream of the TIZ, or outside the influence of the tides, such that design flood discharges at any location can be estimated by the regional equations; and

(ii) to provide the T-year flood quantile values at selected upstream stations that are to be associated with the floods in the tidal interaction zones.

5.1.2 Rationale for Regionalization

It has been well accepted that using a regional approach in flood frequency analysis is effective in extending the flood information at a site to sites within an homogeneous region (Riggs, 1968). The extension enables flood quantile estimates for any site in a region to be expressed or inferred in terms of flood data recorded at all gauging sites in that region (Cunnane, 1988). Estimating design floods using a regional approach can often be carried out using methods such as the index-flood method and the direct regression on quantiles method. The index-flood method provides an appropriate procedure for statistical flood estimation of extreme events (Dalrymple, 1960), and it has been applied in many countries for flood frequency estimation. Hydrological Procedure No 4 (1st edition) published by DID Malaysia (Heiler,1974) and in later versions also adopted the method.

When there is a sufficient number of reliable gauging stations, regionalization can be very helpful in pooling flood data such that design flood estimates can be made at ungauged basins. There is no current regional flood frequency curve established for the

Sarawak. With those limitations and circumstances in mind, an entirely new procedure in estimating design floods based on regional flood frequency analysis are developed in this chapter using an approach that is able to minimize the bias due to outliers and short record length.

5.2 Index-Flood Method

The index-flood method essentially assumes that within a homogeneous region, the exceedance probability distribution of annual peak discharge is invariant except for a site-specific scaling factor called the index-flood. Typically the index-flood, Q_m, is taken as the mean of the at-site annual maximum peak discharge series. Alternatively, Robson and Reed (1999) recommended using the median instead of the mean. A relation can be established between the flood quantile Q_T of a site and Q_m with the introduction of a regional growth factor X_T that defines the dimensionless frequency distribution common to all sites within a homogeneous region. The relationship is:

$$Q_T = X_T Q_m \tag{5-1}$$

A regression of basin characteristic(s) on the index-flood can be established based on available information gathered from the gauged sites. Regional growth curves showing the relationship between X_T and return period T can be derived once an appropriate probabilistic distribution can be found fitting to all the gauged flood series Q_{ij} within the region with N sites. $Q_{ij} : i = 1,2,3,\ldots, N; j = 1,2, 3\ldots, L_i$; where L_i is the record length at site i. The standardized flood peaks

$$X_{ij} = Q_{ij} / Q_{im} \tag{5-2}$$

is used in the estimation of X_T, where Q_{im} is the observed mean or median annual flood at site i. Recent development in statistical methods (Hosking, 1990; Hosking and Wallis, 1997; Robson and Reed, 1999) has reaffirmed the usage of the index-flood method, and the concept has even been extended to the analysis of site-specific environmental data.

The direct regression based method is the other significant method still in use (Pandey and Nguyen, 1999). With the exception of basin area, the information needed for defining the explanatory variables are often not readily available, especially in developing countries. In many cases, it is the unsatisfactory fits that prevent it from being used widely. The index flood procedure is still the preferred method as can be seen in many recent published papers such as those by Fill and Stedinger (1998), El-Jabi *et al.* (1998), Kumar *et al.* (1999), Burn and Goel (2000), Mkhandi *et al.* (2000), Kachroo *et al.* (2000), De Michele *et al.* (2001), Heo *et al.* (2001), Brath *et al.* (2001). The latest edition of the UK Flood Estimation Handbook (FEH) also continues to use the method. The details are found in Volume 3 of the manual (Robson and Reed, 1999).

5.3 Identification of Regions

The first step in regionalization is to delineate regions that are deemed to be homogeneous. The technique of clustering applied to environmental grouping of sites can be found in Fovell and Fovell (1993), and Kalkstein (1987). The application of cluster analysis to delineate cluster of basins for regional flood frequency analysis has been used, among others, in New Zealand (Mosley, 1981), Scotland (Acreman and Sinclair, 1986), Canada (Burn, 1989), and U.S.A. (Bhaskar *et al.*, 1989). Hosking and Wallis (1997)

illustrated the use of cluster analysis for regionalization of U.S. annual precipitation total. Burn and Goel (2000) have further improved the technique to include overlapping groups.

By identifying homogeneous regions, we can take advantage of the fact that the more homogeneous a region is, the greater is the gain in using regional instead of at-site estimation. Studies by Dubreuil (1986) on tropical basins show that physical factors such as basin area and slope are the most significant physical explanatory factors for several regression-based flood quantile estimation equations. Various method of clustering procedures were studied by Kalkstein *et al.* (1987) suggesting that the average linkage method is superior to the centroid and Ward's technique. Hosking and Wallis (1997) have also used average linkage method in the regionalization of annual precipitation total observed in the United States. The average linkage approach used by Hosking and Wallis (1997) and Burn and Goel (2000) are adopted in this study.

Variables for consideration in clustering are derived by transformation of the site characteristics that are measured in different scales. Appropriate transformation by scaling, is necessary to ensure that these factors fall between zero and unity. Initially, all sites are treated as separate clusters. Any pair of sites that are closest in terms of Euclidean distance are joined. In the next step, either a third site joins the first two, or two other sites join together into a different cluster. This process will continue until all clusters are joined into one. A dendrogram can effectively summarize the results of the clustering procedure. The similarity level at any step is the percent of the minimum distance at that step relative to the maximum inter-observation distance in the data. The decision on how many groups or regions, which essentially determines the cut-off level for similarity, is largely heuristic. However, the pattern of how similarity or distance

values change from step to step can assist in choosing the final grouping. The step where the values change abruptly may indicate a good point for cutting the dendrogram. Other methods of delineating homogeneous regions are extensively discussed in Robson and Reed (1999), which includes a software package called WINFAP-FEH.

5.4 L-moments Approach

In the past, the classification of each gauged basin as belonging to a particular homogeneous hydrological region has often based on similarity of basin characteristics. The final decision on what constitutes a region has been largely heuristic. However, very significant developments made in the last decade, particularly the development of a discordancy measure and a homogeneity test based on L-moment ratios (Hosking and Wallis, 1993) has made the decision less subjective.

5.4.1 L-moments

Probability weighted moments (PWMs) were introduced by Greenwood *et al.* (1979) as an alternative to conventional moments to minimised the squaring and cubing of observations that may give undue weighting to large observations. However, the method suffers from difficulties in interpretation. Hosking (1990) introduced L-moments, which are linear combinations of PWMs and can be directly interpreted as the measures of scale and shape of probability distributions.

Vogel and Fennessey (1993) highlighted the advantages of L-moments compared with conventional moments. In a wide range of hydrologic applications, L-moments provide simple and reasonable efficient estimators of characteristics of

hydrologic data and of a distribution's parameters (Stedinger *et al*, 1992). Ulrych *et al.* (2000) show that probability density functions estimated from L-moments are superior estimates to those obtained from conventional moments and those based on the principle of maximum entropy.

Hosking and Wallis(1993, 1997) refined the index-flood method using the L-moment algorithm and the approach presents an elegant method for flood frequency analysis. It is a general approach that is best suited to environmental data and was developed primarily to analyse regional flood data. Distribution selection has become less subjective using the L-moment approach. The usage of the approach in regional flood frequency analysis has gathered momentum over the last decade. Excellent discussion on the use of L-moment in regional flood frequency can be found in Robson and Reed (1999) in addition to Hosking and Wallis (1997).

5.4.2 Discordance

According to Hosking(1993,1997), if a proposed region has N basins, the measure of discordancy, D_i., for basin i is defined as:

$$D_i = \frac{1}{3} N(\mathbf{u}_i - \bar{\mathbf{u}})^T \mathbf{A}^{-1}(\mathbf{u}_i - \bar{\mathbf{u}}) \qquad (5\text{-}3)$$

where $\mathbf{u}_i$ is a vector containing the L-moment ratios for basin i, namely the L-CV(t), L-skewness(t_2), and L-kurtosis(t_3), $\bar{\mathbf{u}}$ is the unweighted regional average for $\mathbf{u}_i$ and $\mathbf{A}$ is the matrix of sums of squares and cross products defined as:

$$\mathbf{A} = \sum_{i=1}^{N} (\mathbf{u}_i - \bar{\mathbf{u}})(\mathbf{u}_i - \bar{\mathbf{u}})^T \qquad (5\text{-}4)$$

5.4.3 *Homogeneity Checks*

Homogeneity checks using the L-moment approach as proposed by Hosking (1993), which is also adopted by Burn and Goel (2000), is based on Monte Carlo simulation. The L-moment ratios t^R, t^R_3 and t^R_4 of the proposed region are calculated as the sample means weighted proportionally to the record length l of i sites. It follows that V, the weighted standard deviation of the at-site sample L-CVs (t_i), is given by:

$$V = \left[\sum_{i=1}^{N} l_i (t_i - t^R)^2 \Big/ \sum_{i=1}^{N} l_i \right]^{1/2}. \qquad (5\text{-}5)$$

A homogeneity statistic, H, is a measure of the departure of V from similar statistic obtained from simulation of some large number of realizations of a region, with μ_v and σ_v as the mean and standard deviation of simulated Vs:

$$H = \frac{(V - \mu_v)}{\sigma_v} \qquad (5\text{-}6)$$

Hosking and Wallis (1997) suggested that a region is considered to be "acceptably homogeneous" if $H<1$, "possibly heterogeneous" if $1 \leq H \leq 2$, and 'definitely heterogeneous" if $H \geq 2$. Robson and Reed (1999) relaxed somewhat the criteria by suggesting that $2<H\leq 4$ as heterogeneous and $H > 4$ as strongly heterogeneous.

5.4.4 *L-moment Ratios and Selection of Distribution*

The need to fit a suitable distribution cannot be over-emphasized as the flood quantile estimates rely solely on it for extrapolation purposes. Once the appropriate

regions have been delineated, the selection of a suitable probability distribution to fit the flood series can be carried out. The L-moment ratios diagram (Vogel and Fennessey, 1993) provides a simple and quick method in selecting the statistical distribution that fit the data best. The point defined by the weighted regional means of τ_4 and τ_3 is plotted on a graph of τ_4 vs τ_3 and the theoretical distribution lying closest to the plotted point is selected as an appropriate regional distribution. Peel *et al.* (2001) shows by simulations that there are some caveats such as not using the weighted regional means while using the L-moment ratio diagram. The use of heterogeneity tests in conjunction with the L-moment ratio diagrams is highlighted.

5.4.5 Goodness-of-Fit Test

Statistical tests are required to confirm the appropriateness of the distribution chosen earlier and to give a certain degree of confidence in the selected distribution. A test based on Monte Carlo simulation by Hosking and Wallis (1993, 1997) is used herein.

For each of the proposed region, a Kappa distribution with its parameters derived from the fitting of the distribution to the regional average L-moment ratios is used to simulate some large number (N_{sim}) of the same region. For each of the m^{th}-simulated region, the regional average L-kurtosis t_4^m is calculated. Typical three-parameter distributions are fitted to the sample regional L-moment ratios. For each of the fitted distribution, the corresponding L-kurtosis, τ_4^{DIST}, is found. The goodness-of-fit measure for each distribution is given by

$$Z^{DIST} = \left(\tau_4^{DIST} - t_4^R + B_4\right)/\sigma_4, \tag{5-7}$$

where the bias of t_4^R is

$$B_4 = \frac{\sum_{m=1}^{N_{Sim}} \left(t_4^m - t_4^R\right)}{N_{Sim}}, \tag{5-8}$$

and the standard deviation of t_4^R is given by

$$\sigma_4 = \left[\frac{\sum_{m=1}^{N_{Sim}} \left(t_4^m - t_4^R\right)^2 - N_{Sim} B_4^2}{N_{Sim} - 1} \right]^{1/2}. \tag{5-9}$$

Any of the distributions could be declared as fitting satisfactorily if $|z^{DIST}| \leq 1.64$ (Hosking and Wallis, 1993). The final decision on which distribution to use, however, cannot rely solely on the results of the goodness-of-fit test. Robson and Reed (1999) specifically mention that any distribution with an upper bound, such as the Generalized Pareto distribution, is not suitable for the annual peak series. This distribution is only appropriate for fitting the peak-over-threshold series.

5.5 Regional Flood Frequency Analysis for Sarawak

In this thesis the index flood method is used for regional flood frequency analysis of flood series, mainly annual peak discharges, of the gauged stations found in the non-tidal zones of Sarawak. The L-moments approach proposed by Hosking and Wallis (1993, 1997) is used, including the steps of screening data, checking discordancy measures, checking heterogeneity measures, and performing goodness-of-fit tests. Homogeneous hydrological regions are identified, apart from the initial works that have

been accomplished as reported in Lim and Lye (2000) on the western part of the study area. Suitable statistical distributions applicable for the regions are identified and tested. The expected results in the form of some regional growth curves for the generation of flood quantiles for each of the region are identified.

5.5.1 Gauging Stations and Data Screening

All the flow-gauging stations within the study area are maintained by the Department of Irrigation and Drainage, Sarawak. Some of these stations were established in the 1960's. The basin size ranges between 34 km^2 and 34503 km^2. Record lengths are between 10 and 39 years. Figure 5-1 shows the location of these gauging stations within the study area. Each basin is assigned an ID number and a full listing can be found in Table 5-3. Some of the gauged data in the study area failed to pass the initial data screening process because of having such factors as poor discharge rating curve, discontinuity of records, short record length, or influence by tidal backwater. Gauging stations with disturbed flow regime such as Lubok Antu is limited to the use of records prior to the construction of a dam upstream at the site.

The data screening process also involved checking with the discordancy statistics D_i of Equation 5-3. None of the sites that have passed the initial screening process were found to be grossly discordant, judging from the computed discordant statistics D_i as shown inTable 5-1. L-CV as a function of L-skewness and L-skewness as a function of L-kurtosis are plotted as shown in Figure 5-2. No particular site appears to be an outlier.

Table 5-1 L-moment ratios and discordant statistics

ID	Basin	n	Area	l_1	τ	τ_3	τ_4	D_i
1	Telok Buing, Rejang	18	9522	1399	0.0418	0.0343	0.0967	0.513
2	Mukeh, Rejang	17	2273	479	0.1032	0.2101	0.1205	0.722
3	Kapit, Rejang	32	34053	7188	0.0747	0.1442	0.2476	0.757
4	Benin, Rejang	19	21273	5842	0.0882	0.0808	0.1795	0.288
5	Belaga, Rejang	29	18261	5121	0.0780	0.1543	0.2484	0.763
6	Insungei, Limbang	19	2413	1215	0.1058	0.0723	0.1243	0.102
7	Long Jegan, Baram	14	2491	452	0.0675	-0.0769	-0.1266	1.516
8	Lio Matu, Baram	34	2690	2991	0.3143	0.2755	0.2163	1.166
9	Lubok Lalang, Limbang	9	180	350	0.4201	0.4023	0.2186	2.930
10	Sibiu, Kemena	16	103	30	0.0824	0.0178	0.3612	2.237
11	Terawan, Baram	17	3210	594	0.0850	-0.1645	0.1327	2.194
12	Bedup, Sadong	34	45	35	0.2233	0.1538	0.0420	0.353
13	Kpg Git, Sarawak	27	440	439	0.1892	0.0520	0.0054	0.483
14	Batu Gong, Samarahan	22	53	14	0.1745	0.2555	-0.0381	1.631
15	Sebatan, Krian	15	34	12	0.0434	0.1229	0.0675	0.978
16	Boring, Sarawak	28	125	282	0.1645	-0.0023	0.0550	0.548
17	Maang, Samarahan	22	138	185	0.1986	0.1008	0.1586	0.309
18	Meringgu, Sadong	17	338	82	0.1495	0.3719	0.3324	2.147
19	Lubau, Layar	15	321	799	0.2719	0.3153	0.2390	0.878
20	Krusen, Sadong	22	456	407	0.1851	0.0569	0.0983	0.318
21	Buan Bidi, Sarawak	28	225	354	0.2314	0.1332	-0.0226	0.742
22	Serian, Sadong	36	951	313	0.1322	0.0079	0.0240	0.339
23	Lubok Antu	10	1305	624	0.1167	0.1010	-0.0789	1.086

Note: l_1 – median peak discharge, τ – sample L-CV, τ_3 – sample L-skewness, τ_4 – sample L-kurtosis

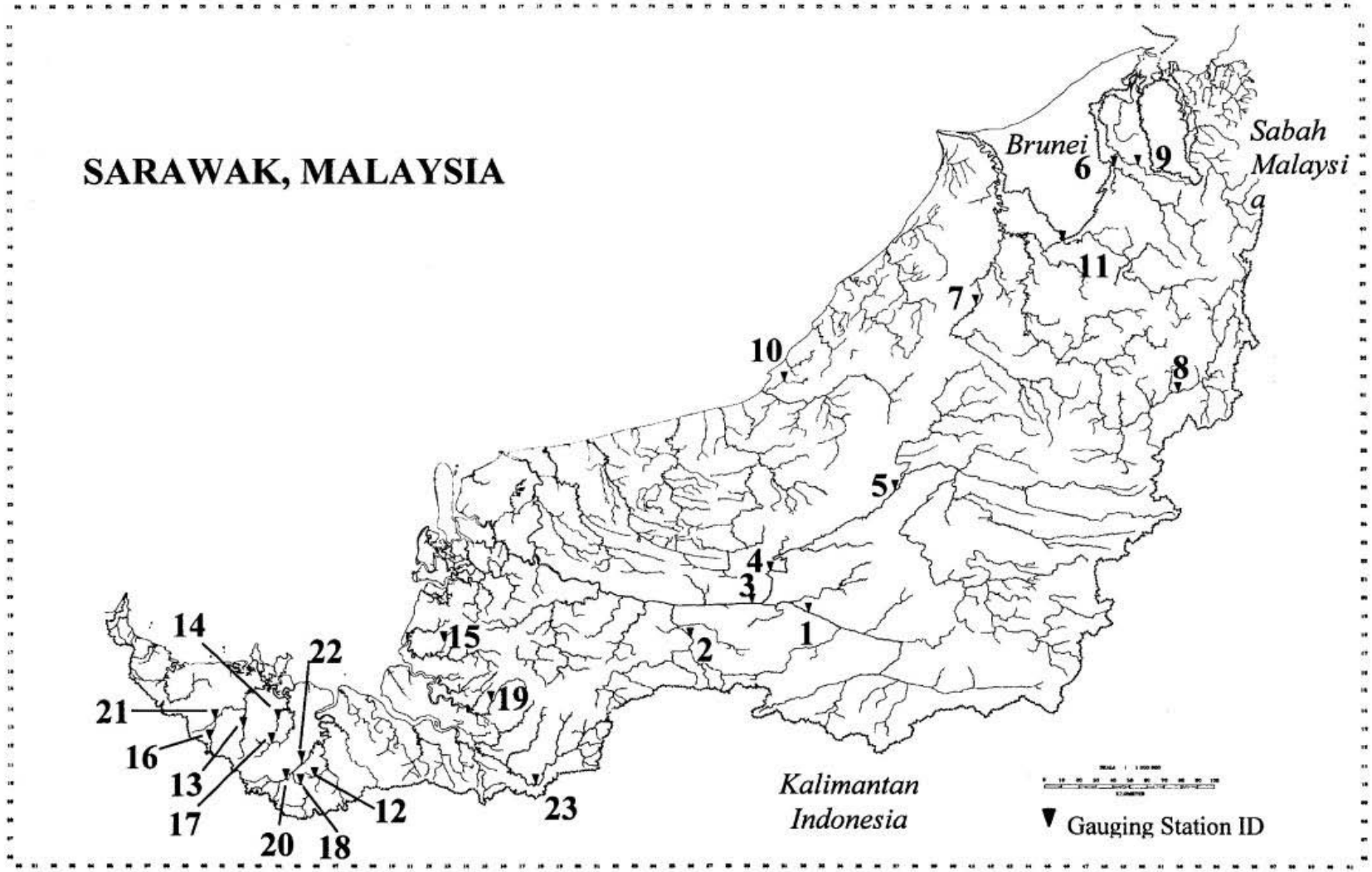


Figure 5-1 Map of Study Area: Gauging Stations

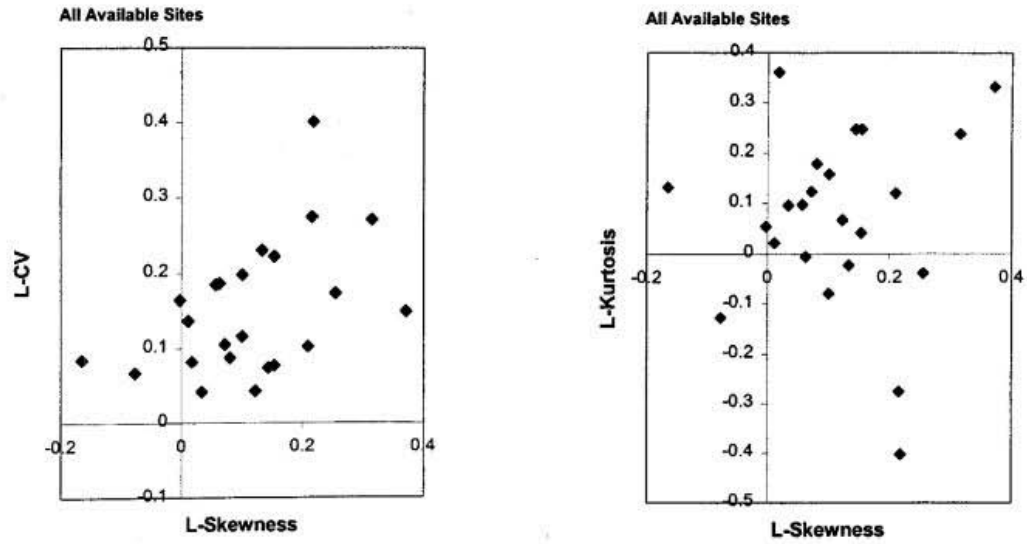


Figure 5-2 Plots of L-moment ratios for all the available sites

5.5.2 Identification of Regions by Cluster Analysis

Cluster analysis is employed as a first attempt to delineate Sarawak into regions. Variables for consideration in clustering are derived by transformation of the site characteristics that are measured in different scales. Appropriate transformation by scaling, as shown in Table 5-2, is necessary to ensure that these factors fall between zero and unity. A clustering algorithm using the average linkage method, which is available in MINTAB, is used. For this study, variables that are considered are basin area (*A*), mean annual rainfall (*R*), return-period storms with duration of 12 hours (*T5*, *T10*, *T20*, *T50*), longitude (*LO*), latitude (*LA*) and specific discharge (*SQ*). These are shown in Table 5-3. Initially, all sites are treated as separate clusters. Any pair of sites that are closest in terms of Euclidean distance are joined. In the next step, either a third site joins the first two, or two other sites join together into a different cluster. This process will continue until all clusters are joined into one.

Table 5-2 Transformation of Site Characteristics

Site Characteristics	Cluster Variable
Basin Area, *AREA*	*A* = *AREA*/35000
Annual Rainfall, *RAIN*	*R* = *RAIN*/5000
T=5-Yr 12Hr Duration Storm, *T5RA*	*T5* = *T5RA*/210
T=10-Yr 12Hr Duration Storm, *T10RA*	*T10* = *T10RA*/260
T=20-Yr 12Hr Duration Storm, *T20RA*	*T20* = *T20RA*/310
T=50-Yr 12Hr Duration Storm, *T50RA*	*T50* = *T50RA*/360
Latitude, *LATI*	*LA* = *LATI*/45
Longitude, *LONG*	*LO* = *LONG*/55
Specific Discharge, *SPECQ*	*SQ* = *SPECQ*/100

Table 5-3 Site characteristics

Site	ID	AREA (km2)	RAIN (mm)	T5RA	T10RA	T20RA	T50RA	LATI	LONG	SPECQ
1	18	9522.0	4343	147	167	186	211	17	36	78.14
2	17	2273.0	4197	147	167	186	211	16	24	71.71
3	32	34053.0	3983	133	152	170	193	21	38	73.00
4	19	21273.0	3883	119	136	153	174	24	39	78.30
5	29	18261.0	4038	119	136	153	174	25	43	81.45
6	19	2413.0	2480	150	166	182	202	41	51	66.93
7	14	2491.0	4512	146	175	202	238	32	43	86.71
8	34	2690.0	3048	110	129	148	172	34	54	60.81
9	9	180.0	3346	150	166	182	202	43	51	43.42
10	16	103.0	3784	187	214	240	274	32	31	47.18
11	17	3370.0	4194	146	175	202	238	38	50	83.38
12	34	45.0	3551	148	174	198	230	9	6	55.36
13	27	440.1	3678	188	225	261	306	12	2	74.60
14	22	52.5	4107	188	225	261	306	13	4	65.19
15	15	34.1	3495	138	155	171	192	19	13	58.54
16	28	124.5	4055	208	255	300	358	13	1	75.98
17	22	138.3	3903	159	181	202	230	12	4	59.90
18	17	338.2	3468	141	161	180	205	9	5	65.75
19	15	320.8	3873	137	156	177	198	15	16	86.38
20	22	456.1	3192	171	208	243	289	10	3	64.14
21	28	225.0	4055	208	255	300	358	13	1	70.47
22	36	950.5	3395	156	185	212	247	10	4	57.97

n record length; D_i Discordant statistics; For other factors see Table 5-2 .

Figure 5-3 shows a dendrogram that effectively summarizes the results of the application of clustering procedure on the set of data. The similarity level at any step is the percent of the minimum distance at that step relative to the maximum inter-observation distance in the data. A heuristic decision has to be made on the final number of clusters (or regions). However, the pattern of how similarity or distance values change from step to step as can be seen in Table 5-4. The step where the values change abruptly may identify a good point for cutting the dendrogram. In this study, there is a significant change of similarity level from step 20 to 21, dropping 10.66% from 56.96% to 46.3%. The drops in previous steps are within the order of 2 to 5%. Hence, 3 clusters can be associated with the cut off level at 56.96%. Similarly, 2 main clusters can be identified for the next abrupt change going from step 21 to 22 with a cut-off level of 46.3%. From this clustering exercise, there are two major cases (2 and 3 clusters) and two minor cases (1 and 4 cluster(s)). Further investigation on the homogeneity of these four cases are carried out using simulations and statistical tests decribed earlier.

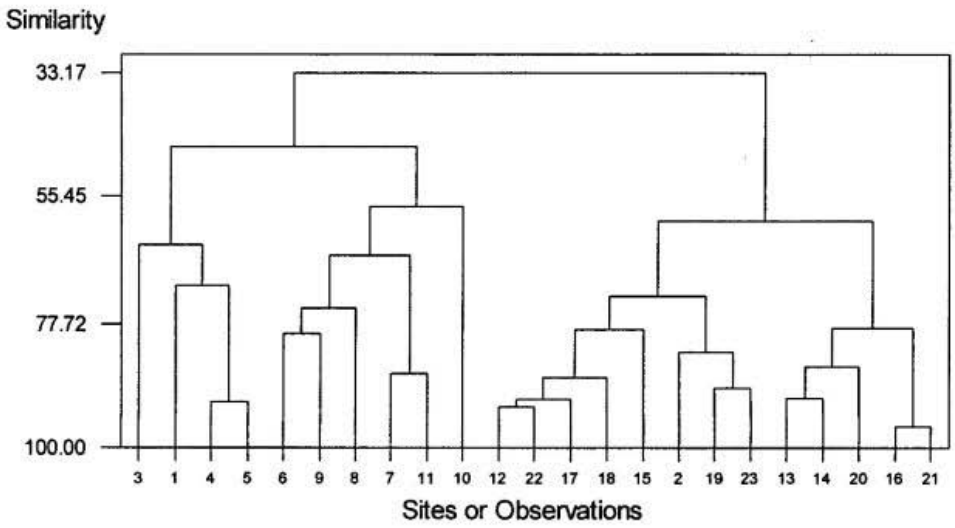


Figure 5-3 Dendrogram: A, R, T5, T10, T20, T50, LA, LO, SQ

Table 5-4 Cluster Amalgamation Steps

Step	Number of clusters	Similarity level	Drop in Similarity	Distance level	New cluster	Number of sites in new cluster
1	22	96.29	3.71	0.055	16	2
2	21	92.74	3.55	0.108	12	2
3	20	91.72	1.02	0.123	4	2
4	19	91.32	0.40	0.129	12	3
5	18	90.94	0.38	0.135	13	2
6	17	89.13	1.81	0.162	19	2
7	16	87.39	1.74	0.188	12	4
8	15	86.59	0.80	0.199	7	2
9	14	85.47	1.12	0.216	13	3
10	13	82.81	2.66	0.256	2	3
11	12	79.68	3.13	0.302	6	2
12	11	78.82	0.86	0.315	12	5
13	10	78.65	0.17	0.317	13	5
14	9	75.04	3.61	0.371	6	3
15	8	72.97	2.07	0.402	2	8
16	7	71.06	1.91	0.43	1	3
17	6	65.76	5.30	0.509	6	5
18	5	63.97	1.79	0.536	1	4
19	4	59.68	4.29	0.599	2	13
20	3	56.96	2.72	0.64	6	6
21	2	46.3	10.66	0.798	1	10
22	1	33.17	13.13	0.994	1	23

5.5.3 L-moments Approach

5.5.3.1 Discordant Statistics

None of the sets derived for Sarawak are found to be clearly discordant judging from the computed discordance measure. Table 5-1 shows D_i when treating the whole of Sarawak as a region. It can be seen that most of the values, except for site 9, are less than

the critical value of 2.632. Similar analyses for two- , three-, and four-region options are also carried out. None of the sites are found to have particularly high D_i with respect to the regions proposed indicating that sites are not clearly discordant.

Table 5-5 MC Simulation Results for Four-Cluster Option

Cluster	Parameters of Fitted Kappa Distribution				H
	ξ	α	k	h	
A1 (1, 3, 4, 5)	0.9820	0.0695	-0.1537	-1.000	0.82
A2 (6, 7, 8, 9, 10, 11)	0.9352	0.2112	-0.0401	-0.5868	8.59
A3 (2, 12, 15, 17, 18, 19, 22, 23)	0.7805	0.3418	0.1574	0.4162	7.12
A4 (13, 14, 16, 20, 21)	0.4296	0.9447	0.7233	1.0894	1.31

Note: ( ) site numbers in each group

5.5.3.2 Homogeneity Measures

A fitted Kappa distribution for each of the proposed regions in question, as shown in Table 5-5, is used in generating series of similar record lengths for a similar number of sites in a region. A very large number of realizations, say 10,000, are simulated for each region. The results show that two of the proposed clusters from the four clusters obtained from the initial clustering exercise have very high H values. By removing sites 8 and 9 from Cluster 2, the H value decreases significantly. Site 8 and 9 are known to have very steep slopes, with basin areas covering the highest peaks in Sarawak. The slope characteristic unfortunately is not available for all the sites making it unavailable as a formal factor in cluster analysis. However the plots of the moment ratios under the sub-heading "Region A" in Figure 5-5 show the outlying nature of these two

sites. The next step would be to consider combining clusters to form larger clusters. Clusters 1 and 2 are combined without sites 8 and 9. An acceptable H value of 1.56 is still achievable even including site 2 and 15 from Cluster 3. The parameters used for the MC simulations are shown in Table 5-6.

Table 5-6 MC Simulation Results for Two-Region Option

Region	Parameters of Fitted Kappa Distribution				H
	ξ	α	k	h	
A (1, 2, 3, 4, 5, 6, 7, 10, 11, 15)	0.9907	0.0798	0.0707	-1.000	1.56
B (12, 13, 14, 16, 17, 18, 20, 21, 23)	0.5391	0.7477	0.5584	0.9145	2.02

Note: ( ) site numbers in each group

Cluster 3 has very high H value of 7.12. Site 19, which is a very steep hill basin, is removed from Region B which is formed by the combination of Cluster 3 and 4. Figure 5-4 shows the plot of L-moment ratios for the site 19 in the context of Region B. Site 22 is situated on a wide flood plain, which has many large storage areas reducing the peak discharges. Its series is consistently lower than at the Site 20 located upstream. Its removal from the group is desirable as the H value is significantly reduced. Region B is then re-simulated which give a better H of around 2. The final delineation is shown in Table 5-8 and Figure 5-7 .

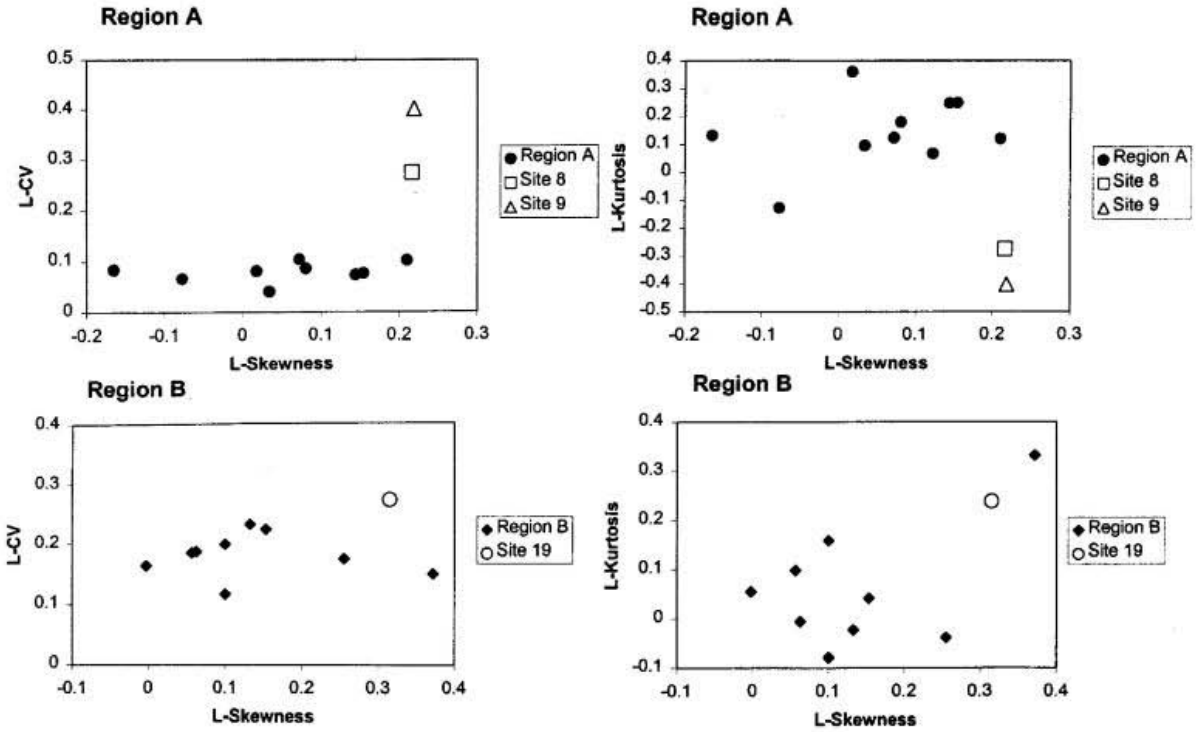


Figure 5-4 Plots of L-Moment Ratios for the Two Regions and Sites Removed

Table 5-7 Homogeneity statistics at various steps of delineating regions

1st Step		2nd Step	
Cluster 1 (1, 3, 4, 5)	0.82	Region A (1, 2, 3, 4, 5, 6, 7, 10, 11, 15)	1.56
Cluster 2 (6, 7, 8, 9, 10, 11)	8.59		
Cluster 3 (2, 12, 15, 17, 18, 19, 22, 23)	7.12	Region B (12, 13, 14, 16, 17, 18, 20, 21, 23)	2.02
Cluster 4 (13, 14, 16, 20, 21)	1.31		
		Possible Mountainous Region C (8, 9, 19) Site 22 removed	

Note: ( ) site numbers in each group

Table 5-8 Final delineation of regions

ID	**Region A Basins**	n	l_1	τ	τ_3	τ_4	D_i	V
1	Telok Buing, Rejang	18	1399	0.0418	0.0343	0.0967	0.99	
2	Mukeh, Rejang	17	479	0.1032	0.2101	0.1205	1.42	
3	Kapit, Rejang	32	7188	0.0747	0.1442	0.2476	0.38	
4	Benin, Rejang	19	5842	0.0882	0.0808	0.1795	0.11	
5	Belaga, Rejang	29	5121	0.0780	0.1543	0.2484	0.39	
6	Insungei, Limbang	19	1215	0.1058	0.0723	0.1243	0.78	
7	Long Jegan, Baram	14	452	0.0675	-0.0769	-0.1266	1.74	
10	Sibiu, Kemena	16	30	0.0824	0.0178	0.3612	1.37	
11	Terawan, Baram	17	594	0.0850	-0.1645	0.1327	1.66	
15	Sebatan, Sebatan	15	12	0.0434	0.1229	0.0675	1.17	
	Weighted Average		2873	0.0776	0.0737	0.1631		
Total /Value		196						0.01917

ID	**Region B Basins**	n	l_1	τ	τ_3	τ_4	D_i	V
12	Bedup, Sadong	34	35	0.2233	0.1538	0.0420	0.64	
13	Kpg Git, Sarawak	27	439	0.1868	0.0633	-0.0053	0.20	
14	Batu Gong, Samarahan	22	14	0.1745	0.2555	-0.0381	1.42	
16	Boring, Sarawak	28	282	0.1645	-0.0023	0.0550	0.59	
17	Maang, Samarahan	22	185	0.1986	0.1008	0.1586	0.61	
18	Meringgu, Sadong	17	82	0.1495	0.3719	0.3324	2.70	
20	Krusen, Sadong	22	407	0.1851	0.0569	0.0983	0.34	
21	Buan Bidi, Sarawak	28	354	0.2314	0.1332	-0.0226	1.04	
23	Lubok Antu	10	624	0.1167	0.1010	-0.0789	1.58	
	Weighted Average		247	0.1891	0.1287	0.0565		
Total /Value		210						0.03033

It can be seen from the exercise that in order to achieve homogeneity, a certain trade-off has to be made on the group size. Detail discussions on this aspect can be found in Robson and Reed (1999). It is also noted that while an attempt is made to achieve statistical homogeneity, the physical aspects are taken into considerations, at times heuristically due to the lack of local information such as the slope factor. In fact, it be reasonable to group the steep basins as another region. However, the number of sites is

limited to only three, well below the accepted minimum of seven sites to form a homogeneous region (Burn and Goal, 2000).

5.5.4 *Selection of Distribution*

Once the appropriate regions have been delineated, the selection of a suitable probability distribution to fit the classical with the assistance of the L-moment ratios diagrams. For the two-region case (see Figure 5-5), it is found that Region A, which include the north-eastern part of Sarawak, tends toward the Generalized Logistic distribution while Region B, covering the western part of Sarawak, inclines toward the Generalized Pareto distribution. The findings related to Region B are consistent with the results as reported earlier (Lim and Lye, 2000).

5.5.5 *Goodness-of-Fit Test for Sarawak*

The results of the goodness-of-fit test by simulation for the two regions of Sarawak are shown in Table 5-9. The Generalized Logistic distribution is found to fit well for Region A based on a $|z^{DIST}|$ of 0.259, well below the critical value of 1.64. The Lognormal, Pearson Type III, and Generalised Extreme-Value (GEV) distribution are also found to pass the test with $|z^{DIST}|$ of 1.28, 1.38 and 1.52 respectively.

For Region B, it can be seen that only the Generalized Pareto distribution fitted to the regional L-moment ratios has $|z^{DIST}| \leq 1.64$, at 0.57. The next closest fits are the GEV, Pearson Type III and Lognormal distributions, which have $|z^{DIST}|$ of 3.64, 3.65 and

3.67 respectively. It is rare to find the Generalized Pareto distribution to fit well to a regional flood frequency curve. However, the extremely wet equatorial rainforest condition (e.g. two days in three are rain days), and the flat terrain of Region B may justify the distribution which has an upper bound. The area has high flows in rivers very frequently in a year and the specific discharges are high. Another factor that may explain the absence of very extreme floods is that being on the equator, it is off the path of all the typhoons which often bring about extreme storm rainfalls in the Philippines and elsewhere in South-east Asia. The findings related to Region B are consistent with the results as reported earlier (Lim and Lye, 2000). Nevertheless, Robson and Reed (1999) prefer to avoid the bounded distribution. In this situation, the next closest distribution that best fits the regional data is used. It is the GEV distribution, which tends to give more conservative results.

Table 5-9 Results of the Goodness-of-Fit test for the L-moment Approach

	Generalised Extreme-Value	Generalised Pareto	Generalised Logistic	Lognormal	Pearson Type III
Region A					
τ_4^{DIST}	0.1201	0.0199	0.1712	0.1269	0.1242
$\lvert Z^{DIST} \rvert$	1.52‡	5.00	0.259‡	1.281‡	1.375‡
Region B					
τ_4^{DIST}	0.1355	0.0404	0.1805	0.1356	0.1277
$\lvert Z^{DIST} \rvert$	3.64	0.574‡	5.64	3.67	3.65

‡ These fits are acceptable as $\lvert Z^{DIST} \rvert \leq 1.64$

† Border line case

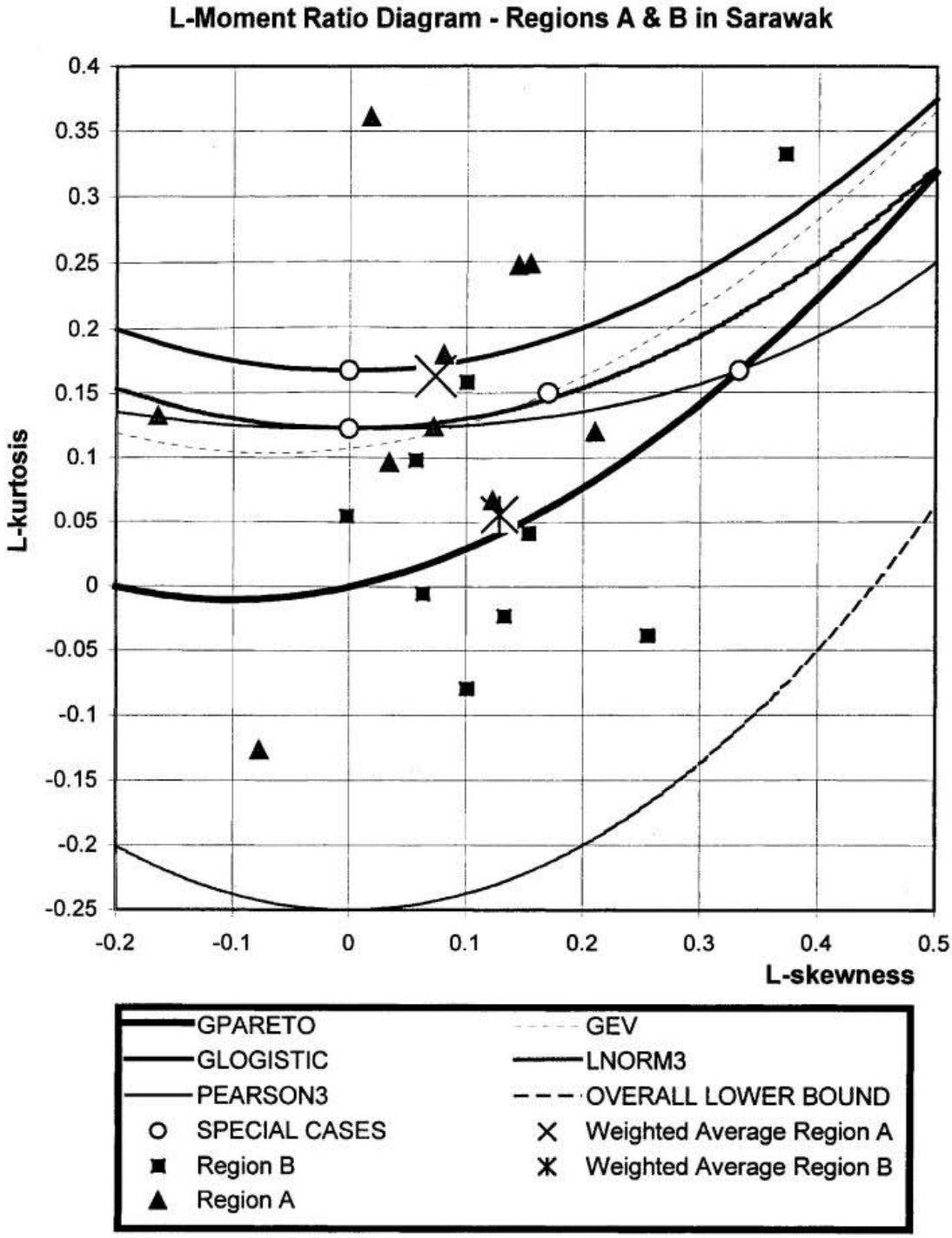


Figure 5-5 L-moment Ratio Diagram for Two-region Case

5.6 Regional Growth Curve

Once the delineated regions are shown to be homogeneous and suitable distributions are identified for the respective regions, regional growth curves can be developed based on the distributions. For this study, region A has been associated with the Generalized Logistic distribution while region B the Generalized Extreme Value distribution. Figure 5-6 shows the proposed regional growth curves fitted for the two regions in Sarawak:

$$\text{Region A}: \; Q_T / Q_m = 0.9990 - 1.0436 \{1 - (T\text{-}1)^{0.07367}\} \qquad (5\text{-}10)$$

$$\text{Region B}: \; Q_T / Q_m = 0.8961 + 4.6446 \{1 - [-\ln(1\text{-}1/T)]^{0.06558}\} \qquad (5\text{-}11)$$

where Q_T is the flood quantile, Q_m is the at-site median maximum discharge, and T is the return period. For each site, the plotting position formula $(i - 0.35)/n$ is used where i is the rank of each standardized annual maximum discharge (divided by median) recorded at the site and n is the sample size of the site. The standardized annual maximum discharge data for the sites are plotted against the set of plotting positions derived for the site. The plotting positions are related to the Gumbel variate via the return period T.

The regressions of median maximum discharges on basin area are plotted for each respective region as shown in Figure 5-8 and Figure 5-9. Ideally these regression equations should be used when estimating the median annual maximum discharge for any ungauged basin. However, the regression fit for region B is not desirable due to the poor R^2. An alternative approach is to use the regression based on all the available basins in the two regions of Sarawak as shown in Figure 5-10 , with a regression equation of:

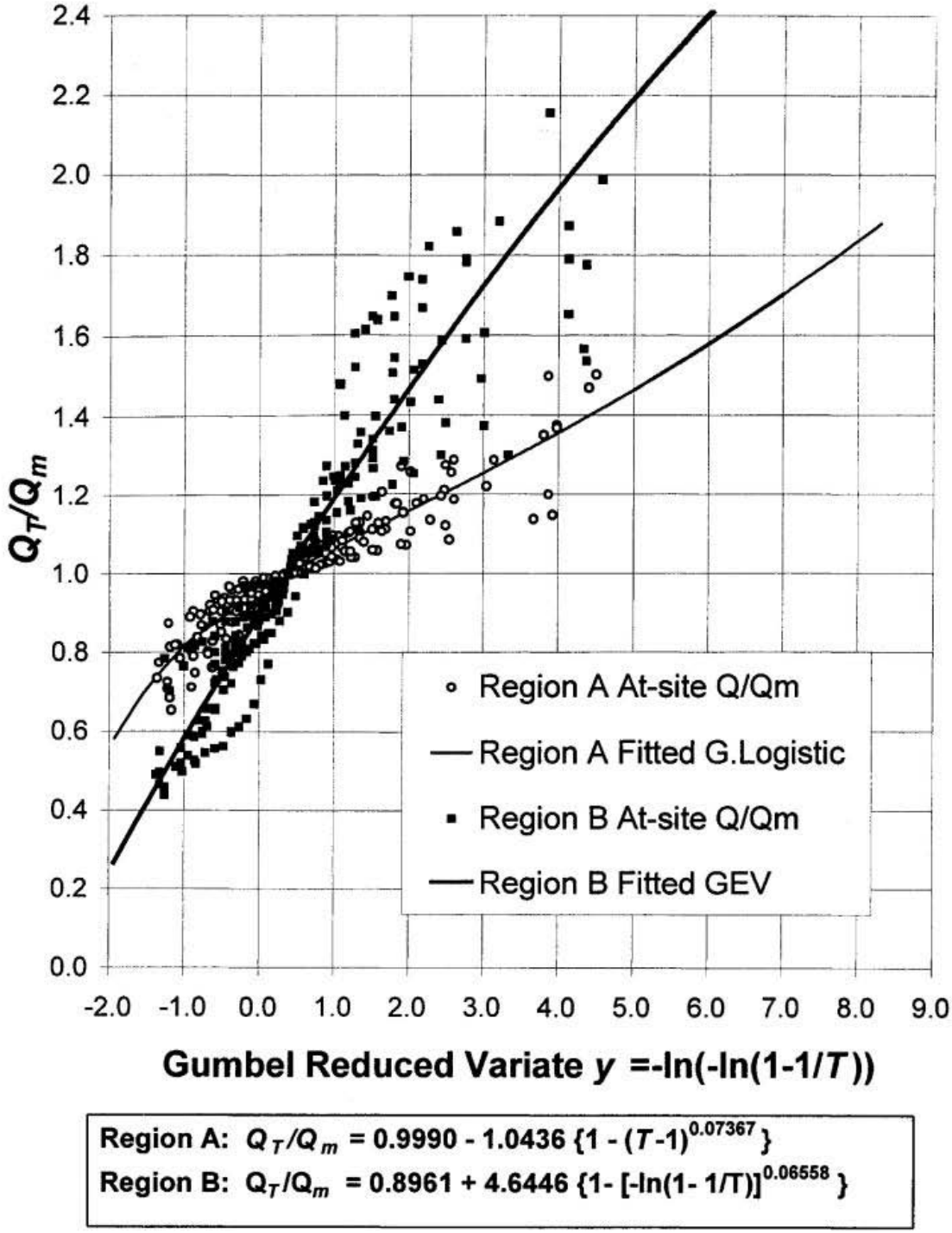


Figure 5-6 Regional Growth Curves for Sarawak

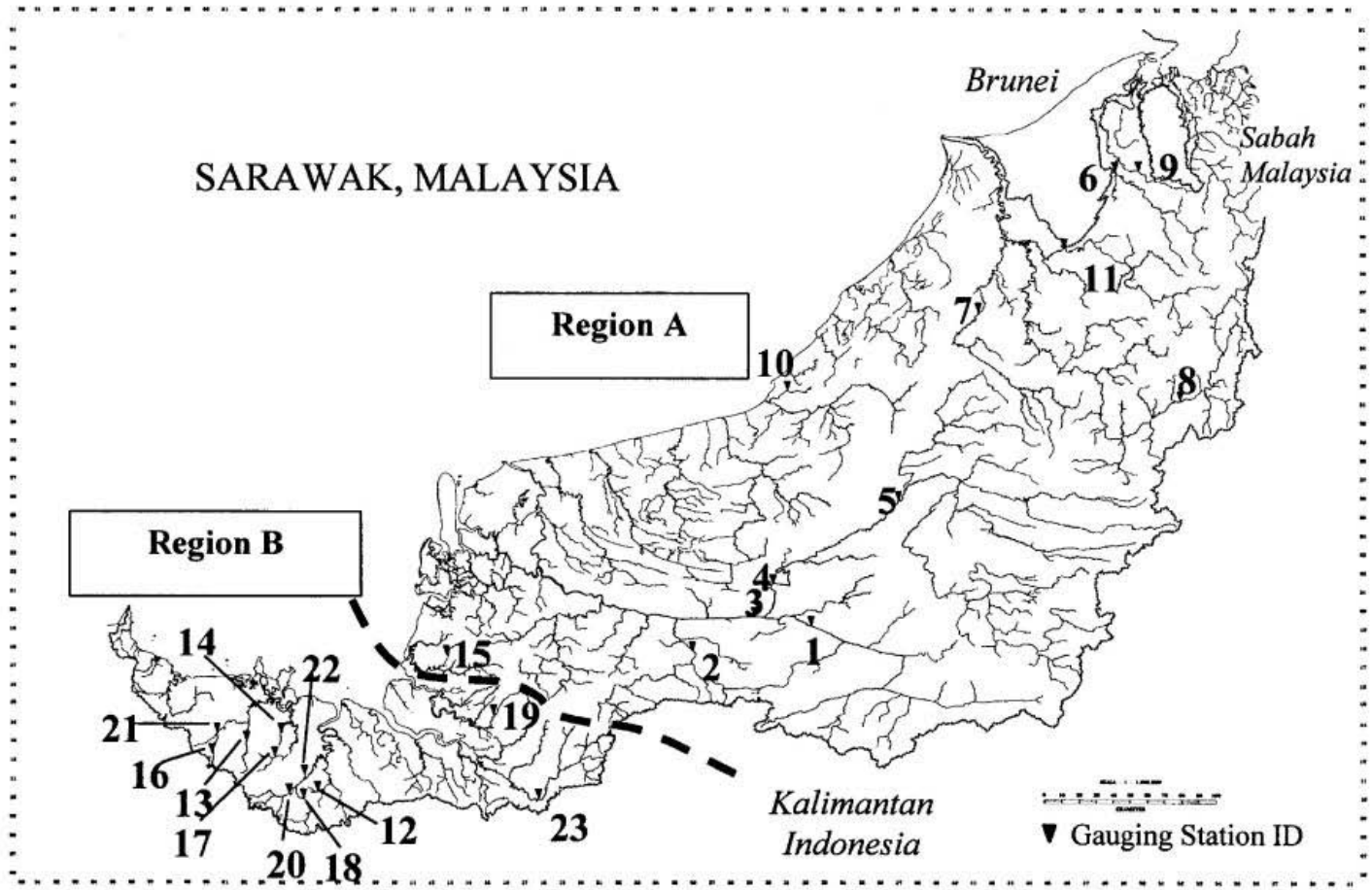


Figure 5-7 Gauged basins in Sarawak and delineated homogeneous regions

$$Q_m = 2.312\ Area^{0.775} \quad (R^2 = 0.80) \tag{5-12}$$

Beside basin area, basin slope is the other important physical variable (Dubreuil, 1986) for tropical regions. The regression is expected to be better if the data on slopes are available.

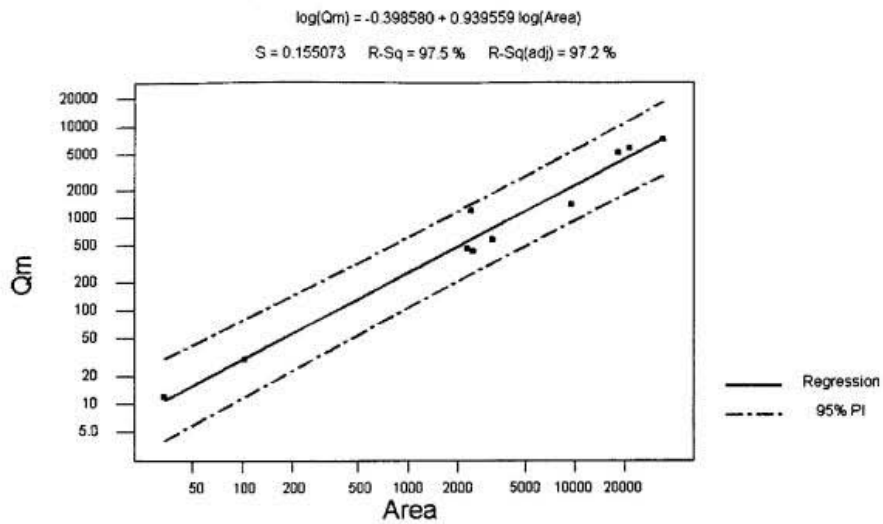


Figure 5-8 Regression of Q_m with Basin Area for Region A

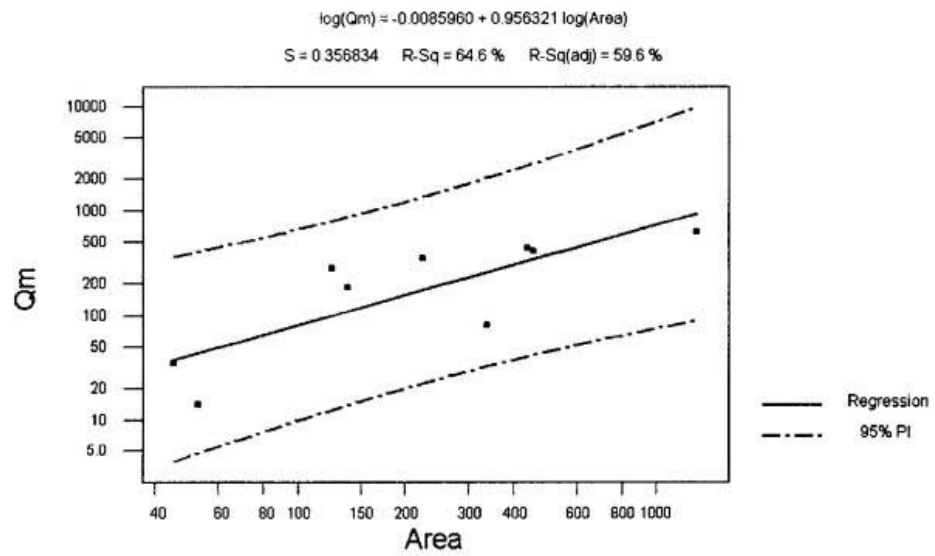


Figure 5-9 Regression of Q_m with Basin Area for Region B

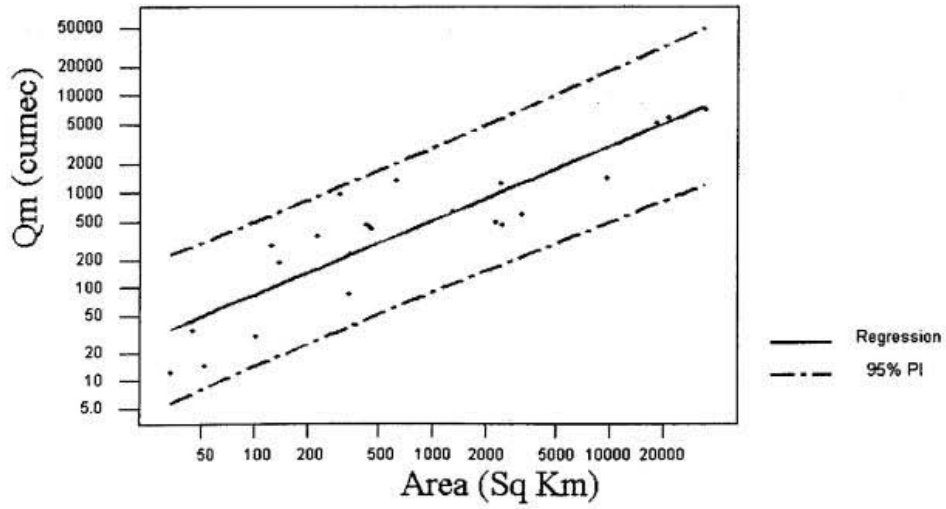


Figure 5-10 Regression of Q_m with Basin Area for All Basins in Sarawak, $Q_m = 2.312$ Area $^{0.775}$ ($R^2 = 0.80$)

5.7 Limitations of Regional Flood Frequency Analysis

In some major basins of Sarawak, representative gauging stations are absent. The application of the regional growth curves to those areas is not recommended. For a state occupying almost half of Malaysia's land mass, the number of gauging stations in Sarawak is still below the standard density in stream gauging network as recommended by WMO(1981). Hence certain large areas, such as the Batang Kemena areas, do not have any credible representative gauged station. The uncertainty in flood quantiles would be high in those out-lying areas.

The regional flood frequency analysis used assumes that the study area has sufficient number of long-term river gauging stations, with the stations being optimally

spread out in each of the hydrological regions. However, as noted in Lim and Lye (2000) the western part of the study area has better coverage and network density as compared with the rest of the area.

5.8 Chapter Summary

This chapter has presented a practical regional flood estimation procedure for Sarawak, Malaysia employing the latest techniques in regional flood frequency analysis. Although it is limited, both in terms of the spatial coverage of the gauged basins and the record length, the identification of the two homogeneous regions is viable after testing for homogeneity in a statistical sense. The selection of the appropriate frequency distribution is also conducted based on a statistical test using MC. With due respect to the limitations as discussed, the regional flood frequency results can be applied to non-tidally influenced ungauged basins in Sarawak lying within or in proximity of an identified homogeneous region. This effectively has taken care of estimating the inflow conditions of upstream sub-basins when evaluating floods in the TIZ using methods to be discussed in the next chapter.

Chapter 6 Direct Frequency, Rating Curve and Bivariate Distribution Methods

Three of the six methods of flood estimation in the TIZ identified and briefly described in Chapter 4 are elaborated in this chapter. These are the Direct Frequency Method, the Rating Curve Method and the Bivariate Distribution Method. Data series obtained from Sarawak are analysed using these methods.

6.1 Direct Frequency Method

6.1.1 Basic Method

The Direct Frequency Method is the simplest of the six methods. This method rests on the assumption that the annual extreme water levels are the resultants of all the prevailing physical forces in the TIZ. Any water level series observed in the TIZ, $Z(t)$, is then considered a sequence generated by random physical processes.

The probability of occurrence of extreme water levels can be obtained subsequently by traditional frequency analysis of extreme values. Even though there are objections to this assumption, the method is still in use, primarily due to its advantage of producing quick indications of the magnitude of design flood levels. Indeed, the method

is found to be acceptable at TIZ locations that are in proximity to the non-tidal areas, where the interaction effects of tides have diminished.

Hourly water level series in the TIZ's of rivers in Sarawak were scanned for the maximum level thus forming the annual maxima series. The latter series were then subjected to a host of frequency analysis and fitting routines, which were developed and used in Chapter 5. Beside the conventional Method of Moment (MOM) approach, the L-moment (LM) or the Probability Weighted Moments (PWM) approaches were also used in parameter estimations for distributional fitting. A macro-enabled spreadsheet written in Microsoft Excel gives the results of the goodness-of-fit (standard errors) for most common types of distributions. By ranking the standard errors computed for each distribution, the best distributions acceptable for various series were obtained, and summarized in Table 6-1. The detailed results on the quantile estimations and the plotting of the top three ranked distributions are shown in **Appendix B1** and **Appendix B2** respectively. In general, the Generalized Pareto is the best fitting distribution followed by the GEV. Satok, Batu Kawa, Batu Kitang and Siniawan are four stations in the Sungai Sarawak Basin located in order from downstream to upstream. However, since the Generalized Pareto Distribution has an upper bound as discussed in Chapter 5, its usage is avoided in the flood frequency analysis of the annual extreme series. It can be seen that the standard errors of fits increases while moving from the downstream area to the upstream area. The Satok station being closest to the estuaries has the least standard errors of fit. It is also obvious that the three-parameter variants of distributions are the distributions that fit well to all the series. There is no trend in the data series used. If there is, it should be removed by a suitable de-trending method.

Table 6-1 Best Fitting Distributions for Water Level Series in the TIZ of Sarawak

(Standard errors are shown in brackets, water levels are in mm above chart datum)

Station	Rank 1	Rank 2	Rank 3
Satok Bridge	Generalized Pareto MOM (58.7)	Generalized Pareto LM (63.6)	GEV MOM (70.0)
Batu Kawa	Log Gumbel PWM (94.7)	GEV LM (95.9)	Pearson III LM (98.5)
Batu Kitang	Generalized Logistic LM (259.2)	GEV LM (261.1)	Log Gumbel PWM (261.1)
Siniawan	Generalized Pareto MOM (296.2)	Generalized Pareto LM (326.2)	Log Gumbel PWM (330.7)
Sungai Merah	Generalized Logistic LM (38.3)	GEV LM (39.5)	Log Pearson III LM (49.6)
Sungai Salim	Generalized Pareto MOM (76.0)	Generalized Pareto LM (85.2)	GEV LM (94.5)
Marudi	Generalized Pareto MOM (135.8)	GEV MOM (164.5)	Log Normal (2) MOM (165.9)

Table 6-2 Return Period Levels in the TIZ Determined by Direct Frequency Method

(water levels are in m above chart datum)

Station	Mean	T=2	T=10	T=50	T=100	T=200
Satok	5.767	5.764	6.038	6.189	6.236	6.275
Batu Kawa	6.182	6.209	6.706	6.964	7.048	7.123
Batu Kitang	6.593	6.686	7.712	8.354	8.588	8.804
Siniawan	8.331	8.492	10.205	10.843	10.993	11.100
Sungai Merah	4.111	4.040	4.389	4.920	5.252	5.674
Sungai Salim	5.042	5.128	5.434	5.471	5.478	5.482
Marudi	4.590	4.537	5.296	5.872	6.102	6.333

6.1.2 *Limitations of the Direct Frequency Method*

It is generally recommended to use long records for this method and Pugh (1987) suggested at least 25 years for application in the sea conditions. However, as mentioned in Chapter 3, it is very rare to find observed data in the TIZ. If this guideline is strictly adhered to, virtually no analysis using the method can be carried out.

6.2 Rating Curve Method

The main objective of the Rating Curve Method is to associate a particular level Z at a TIZ with a certain combination of the river discharge Q and the sea level H. The method essentially derives a regression line for each convenient category of Z for the associated group of Q-H data. To perform this method, three synchronous time series for three locations are required: a water level station located in the TIZ, an upstream discharge observation station, and a sea-level station in the open sea.

6.2.1 *Application of the Rating Curve Method*

There are two sets of data series in Sarawak that satisfy the basic requirement of synchronous observations. These synchronous series are the Buan Bidi – Siniawan – Pulau Lakei series and the Serian – Gedong – Pulau Lakei series.

6.2.1.1 Lag time

The distance between a sea station in proximity of a river estuary and a corresponding station in the TIZ are in the order of 30 to 150 kilometers. Hence a lag time has to be estimated based on an estimated propagation speed up the river. Alternatively, the time series are lagged by a certain number of hours such that correlation coefficient

for the lagged series is a maximum. For example, by virtue of the correlation coefficients as shown in Table 6-3, a 6 hour lag time was chosen to allow for the peak discharge at Buan Bidi to be transcended in the TIZ at Siniawan. Using the same approach, the Pulau Lakei series at the sea condition was lagged by 4 hours from the Siniawan series to allow the effects of tidal propagation, especially the crest levels, to reach that locality.

Table 6-3 Correlation Coefficients of Lagged Series for Pulau Lakei and Siniawan

Lagged Time (hours)	**Correlation Coefficient for Buan Bidi - Siniawan**	**Correlation Coefficient for Pulau Lakei - Siniawan**
1.0	0.4220	0.0983
2.0	0.4354	0.3468
3.0	0.4477	0.5198
4.0	0.4579	**0.5749**
5.0	0.4649	0.4983
6.0	**0.4680**	
7.0	0.4667	
8.0	0.4612	

6.2.1.2 Pulau Lakei-Siniawan-Buan Bidi Series

At each hourly time step t, the water level at a TIZ station Siniawan, $Z(t)$, is associated with a pair of values $Q(t\text{-}6)$ and $H(t\text{-}4)$. By defining a suitable category of Z, say at 4, 5, 6, 7 and >8 metres, the datasets of the corresponding Q and H associated with each Z category are correlated and plotted (superimposed) in Figure 6-1. For the convenience of the designers, a simplified version is shown in Figure 6-2. The regression lines are not flat indicating that the stages at Siniawan are influenced almost equally by the river flows and the tidal regimes.

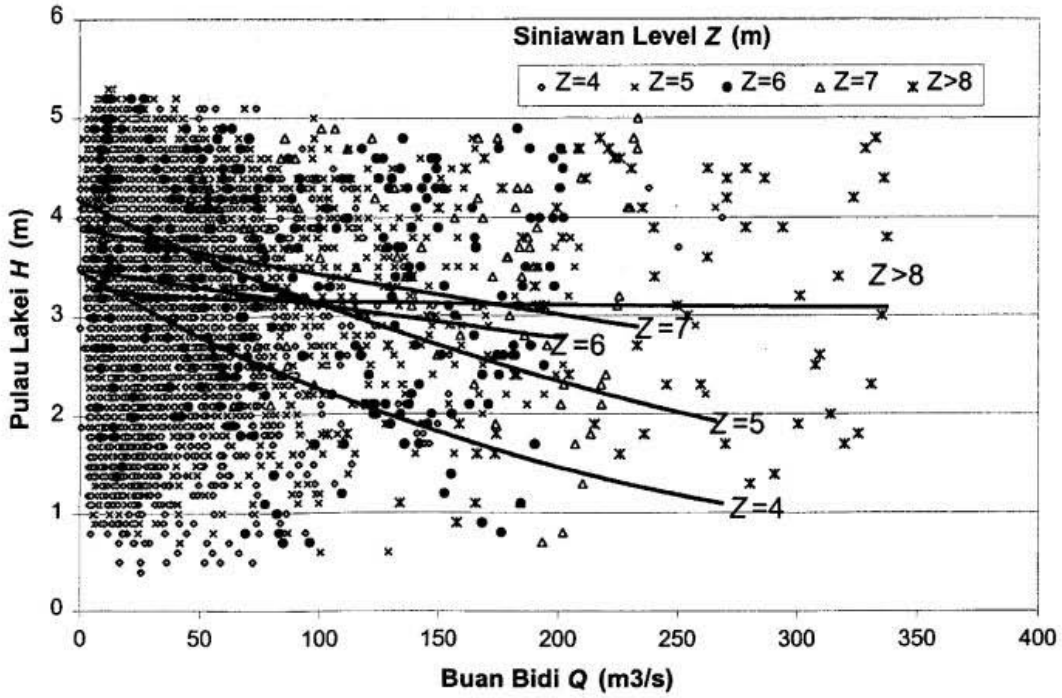


Figure 6-1 *H-Q* Curve for Siniawan

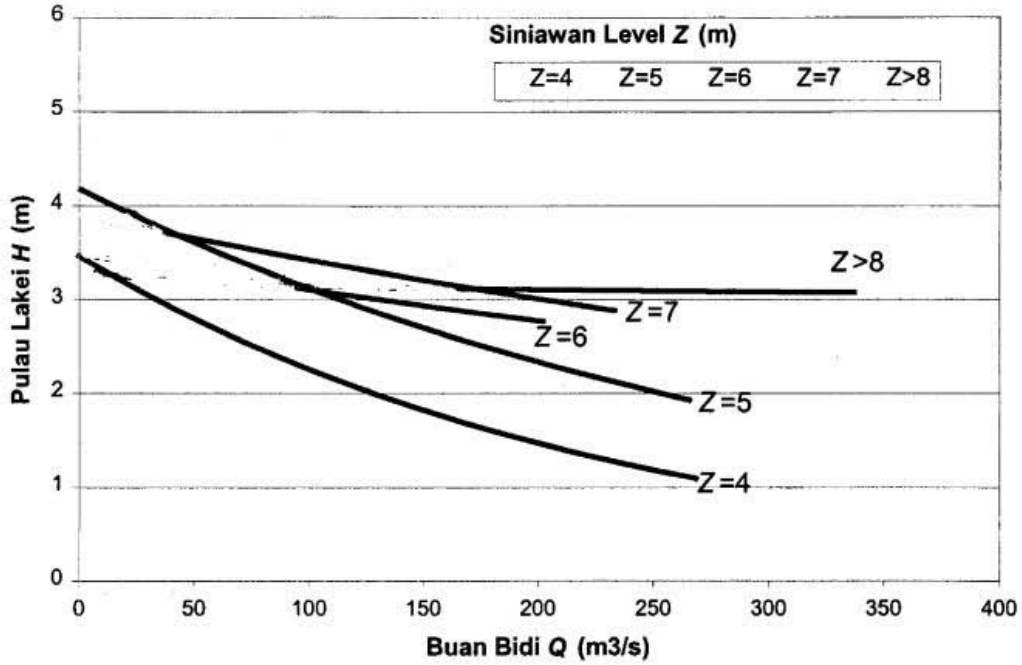


Figure 6-2 *H-Q* Curve Fitted for Siniawan, Presented for Design Purposes

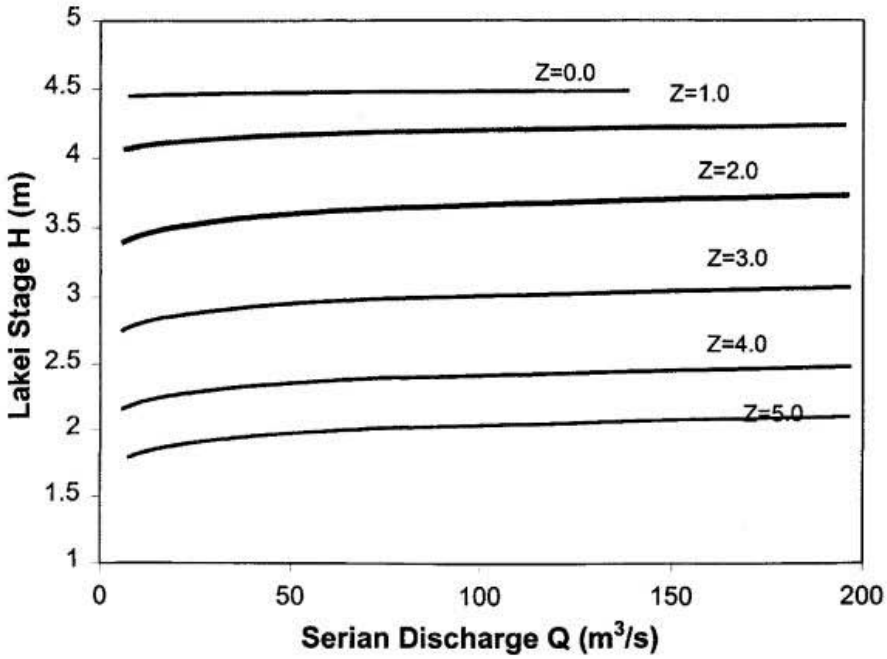


Figure 6-3 *H-Q* Curve at Gedong

6.2.1.3 Pulau Lakei-Gedong-Serian Series

Another three-station group in western Sarawak is comprised of Gedong, a water level observation station in the TIZ; Serian, an upstream discharge station, and Pulau Lakei, a sea station. One year of simultaneous hourly data at the three sites were used. Due to the presence of a significant gap in the original record, the Pulau Lakei water level series was derived from published tide tables. By lagging the series at Gedong by three hours, the correlation between Lakei and Gedong was found to be the highest at 0.90. This high correlation indicates that the station Gedong is in proximity of the tidal active zone rather than the river-regime zone. The data are plotted as shown in Figure 6-3. The curves are relatively flat indicating that the influence of the river discharges is very limited at this location.

6.2.2 *Limitations of the Rating Curve Method*

Residuals of the fits are large and R^2 of the fits are small (see **Appendix B3**). The Rating Curve Method suffers from some arbitrary choices: the lag-time between tides in the sea and at the TIZ, the selection of grouping of loci of Z, and the variant of regression curves or trend lines used for fitting the loci of Z. Since the distance between a typical discharge station and a station in the TIZ can be in the order of 30 to 150km, the amount of unaccounted lateral inflows from some sizeable tributaries can be very substantial. When the records available for analysis are short, the derived H-Q curves are not representative. Any further elaboration on the curves can be dubious.

6.3 Bivariate Distribution Method

The fluctuations of water levels in the TIZ of Sarawak are due to two major processes: the fluvial flows from storm runoff process and the tidal propagation from the open seas. It is assumed that the fluctuation in the sea levels and the variation in the river discharges are independent events. The interaction effect as felt in the TIZ is then considered a joint probability effect that can be described by a bivariate distribution. In places where sea surge is significant, the phenomenon has to be modeled either separately as a process by itself or as a combined process with the tides.

6.3.1 *Joint probability*

As defined in Chapter 4, the sea-level H and the upstream discharge Q each has a probability function $f(H)$ and $g(Q)$ respectively. The distribution of Z in the TIZ is given by

$$Z = h(H, Q) \tag{6-1}$$

and the probability of H and Q is a bivariate distribution

$$p(H, Q) = f(H) \cdot g(Q) \tag{6-2}$$

Using the relationship expressed in Equation 6-2 alone is not sufficient to calculate the distribution of Z. However, from the Rating Curve Method, or from hydraulic modeling, the loci of Z can be derived and plotted on the H-Q plot. This satisfies the need to define the function expressed in Equation 6-1.

6.3.2 Probability Density Functions

The first step of using the Bivariate Distribution Method is to derive the probability density functions at the open sea condition and at a river discharge station located outside the influence of tides.

6.3.2.1 Sea Level

Sea-level probability density function is derived preferably from sea level observations over a long period. The ideal record length is at least 18.6 years to cover all the possible tidal fluctuations due to various tidal constituents including the longest Nodal cycle of the moon (period of 18.6 years). The hourly data series observed at a downstream position is subjected to frequency tabulation based on a certain bin width selected and a suitable distribution that emulates well these bin data. The details about the selection of bin width are treated in detail in Simonoff (1996).

6.3.2.2 Discharge

The flood exceedance probability function derived from conventional flood frequency analysis at the upstream discharge station can be used. Alternatively, when the record is short, an hourly-based discharge distribution based on a long record can be developed. The derivation using the latter approach is applied to a case in the TIZ in Sarawak.

6.3.3 Simulation Models

When sea level data is limited with only short records available, a harmonic analysis is performed which produces a tidal model which can be fitted by a least squares method. The fitted model is then used to simulate at least 18.6 years of hourly tidal sequence. Frequency analysis can subsequently be performed to derive the tidal density function.

6.3.4 Joint probability Method (JPM)

Pugh and Vassie(1980) introduced the JPM for extreme sea level computations in the open seas. The situation is different from the TIZ with surge being a prominent component. Surge and tides are considered as independent except in very shallow seas. The probability density functions for sea levels (H), surge (S) and tidal (H-S) are respectively represented by $p_H(.)$, $p_S(.)$ and $p_T(.)$, with the relationship:

$$p_H(h) = \int_{-\infty}^{\infty} p_T(h-s) \cdot p_S(s) ds \qquad (6\text{-}3)$$

The distribution function of hourly instantaneous sea levels (less than x) is then given by:

$$F_H(x) = \int_{-\infty}^{\infty} p_H(h)dh \qquad (6\text{-}4)$$

Assuming hourly values are independent, the distribution function of annual maxima $G_H(.)$ is related to $F_H(.)$ by

$$G_H(x) = {F_H}^N(x) \qquad (6\text{-}5)$$

where N is the number of hourly observation in a year, taken to be 8766 on average. A relationship to obtain the return period $T(x)$ of a level x was derived as:

$$T(x) = \frac{1}{1-G_H(x)} = \frac{1}{1-{F_H}^N(x)} \approx \frac{1}{N(1-F_H(x))} \qquad (6\text{-}6)$$

A correction factor 1- $q_1(x)/q_2(x)$ was introduced to adjust for the dependent hourly data:

$$T(x) \approx \frac{1}{N(1-q_2(x)/q_2(x))(1-F_H(x))} \qquad (6\text{-}7)$$

where $q_i(x)$ are the probability i consecutive levels are greater than x.

Deficiencies are found in the original proposed method. Tawn and Vassie(1989) proposed changes mainly concerned with a suitable correction factor.

6.3.5 Adaptation of the JPM

The JPM method is essentially a variant of the Bivariate Distribution Method. However, it is specifically developed for sea levels in open sea conditions, and not directly applicable to the TIZ where meteorological forcing is different. Storm surge is not significant in the TIZ as compared to the effects of storm runoffs. However, the method can be adapted for the TIZ with some minor modifications. In the case of a tidal regime, the distribution is the same but a lag time has to be incorporated for the tide to propagate up the river channels. For the river flow regime, the distribution has to reflect

the upstream inflow conditions, for which there is often data than for the surge phenomenon in the open seas. Lag time is also important if the inflows, as well as the water level regime in the TIZ, vary rapidly.

6.3.6 An Example of the Bivariate Distribution Method

The data of Pulau Lakei–Siniawan–Buan Bidi case as mentioned in Section 6.2.1.2 is used. Pulau Lakei sea level series are studied in Chapter 2 producing the distribution of hourly sea levels. By frequency analysis of the 14 years of hourly sea level series, an empirical probability density distribution is derived, as shown in Figure 6-4. A class interval of 0.1m is used in deriving the distribution.

Buan Bidi discharge station record is 10 years in duration. Hourly discharge is converted from hourly water levels recorded using an established rating curve for the station which was supplied by DID Sarawak. Similar frequency analysis is carried out which produces another empirical density function as shown in Figure 6-5. The details of the frequency and probability derivations are shown in Table 6-5.

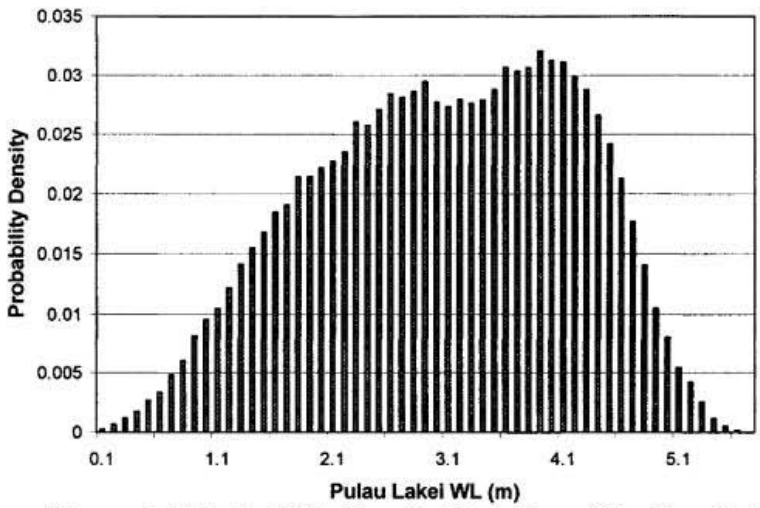


Figure 6-4 Probability Density Function of Sea Level at Pulau Lakei

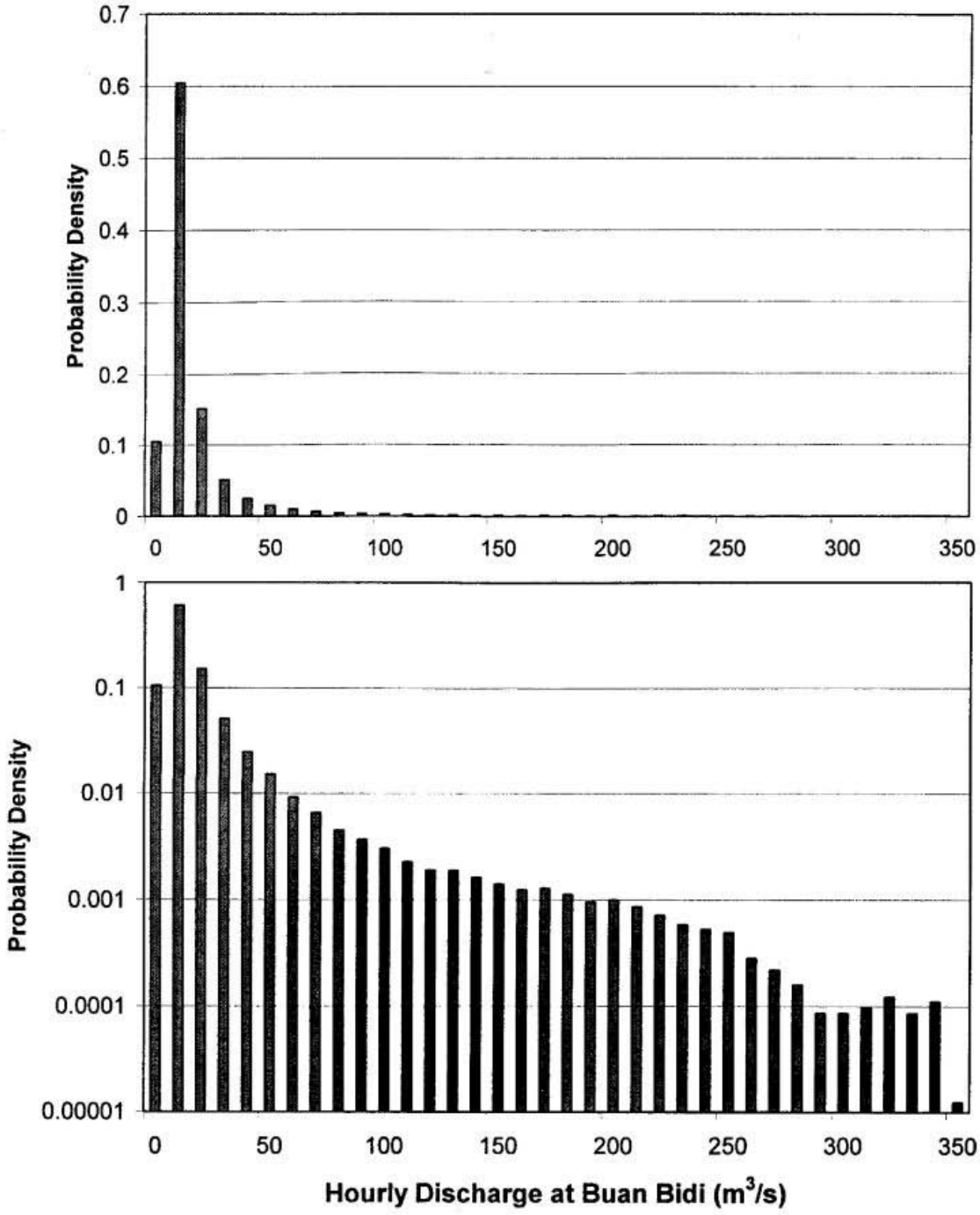


Figure 6-5 Probability density of Hourly Discharge at Buan Bidi

(Based on 10 years of hourly record; class interval 10 m^3/s; lower plot with probabilities in logarithmic scale)

Table 6-4 Buan Bidi Hourly Discharge and Pulau Lakei Hourly Water Level Distributions

Discharge at Buan Bidi (m^3/s)	Frequency	Probability
0	8507	0.105208
10	48929	0.605115
20	12300	0.152117
30	4129	0.051064
40	1999	0.024722
50	1236	0.015286
60	754	0.009325
70	537	0.006641
80	369	0.004563
90	300	0.00371
100	246	0.003042
110	184	0.002276
120	154	0.001905
130	153	0.001892
140	131	0.00162
150	114	0.00141
160	101	0.001249
170	104	0.001286
180	91	0.001125
190	78	0.000965
200	81	0.001002
210	70	0.000866
220	58	0.000717
230	48	0.000594
240	43	0.000532
250	40	0.000495
260	23	0.000284
270	18	0.000223
280	13	0.000161
290	7	8.66E-05
300	7	8.66E-05
310	8	9.89E-05
320	10	0.000124
330	7	8.66E-05
340	9	0.000111
350	1	1.24E-05
Total	**80859**	**1**

Water Level at Pulau Lakei (m)	Frequency	Probability
0.1	36	0.000292
0.2	86	0.000698
0.3	149	0.001209
0.4	216	0.001752
0.5	333	0.002701
0.6	418	0.003391
0.7	595	0.004827
0.8	745	0.006044
0.9	1003	0.008137
1	1183	0.009597
1.1	1299	0.010538
1.2	1503	0.012193
1.3	1745	0.014156
1.4	1914	0.015527
1.5	2076	0.016841
1.6	2286	0.018545
1.7	2359	0.019137
1.8	2643	0.021441
1.9	2645	0.021457
2	2738	0.022212
2.1	2807	0.022772
2.2	2904	0.023558
2.3	3216	0.026089
2.4	3176	0.025765
2.5	3344	0.027128
2.6	3508	0.028458
2.7	3471	0.028158
2.8	3531	0.028645
2.9	3638	0.029513
3	3422	0.027761
3.1	3373	0.027363
3.2	3451	0.027996
3.3	3409	0.027655
3.4	3442	0.027923
3.5	3549	0.028791
3.6	3784	0.030697
3.7	3746	0.030389
3.8	3783	0.030689

3.9	3959	0.032117
4	3857	0.03129
4.1	3840	0.031152
4.2	3684	0.029886
4.3	3550	0.028799
4.4	3281	0.026617
4.5	2982	0.024191
4.6	2627	0.021311
4.7	2190	0.017766
4.8	1739	0.014107
4.9	1300	0.010546
5	986	0.007999
5.1	673	0.00546
5.2	520	0.004218
5.3	315	0.002555
5.4	146	0.001184
5.5	66	0.000535
5.6	24	0.000195
5.7	3	2.43E-05
Total	**123268**	**1**

6.3.7 Derivation of Probabilities

The following steps are used to derive the exceedance probabilities of water levels in the TIZ using the Bivariate Distribution Method given that the probability density functions of H and Q are established from observed series, and the function $h(H,Q)$ is derived by either the Rating Curve Method or hydraulic modelling.

(i) Select an interval of Q to obtain $P(Q_i)$ the probability of Q_i from the discharge probability density function.

(ii) For each Z group, obtain the matching H_i level in the sea condition using $h(H,Q)$ and the corresponding probabilities of H, $P(H_i)$.

(iii) For each Z group compute the joint probability(JP) of occurrence of each H-Q pair as $P(Q_i) \cdot P(H_i)$ and sum up the joint probability as $P(Z=Z_j)$, the probability of $Z=Z_j$. This satisfies Equation 6-3.

(iv) Exceedance Probabilities of $Z_j = 1 - F_H = \{P(Z=Z_j) + P(Z=Z_j+1) + P(Z=Z_j+2) + ..+ P(Z>Z_n)\}$ where $Z>Z_n$ is the last Z group. The interval of Z is 1 metre for this case. F_H is defined in Equation 6-4.

(v) Calculate the return period using Equation 6-7. For the present purpose, a correction factor of 36.8, which is found in a later section (7.5.1.5), is used.

The example used in Section 6.3.6 is investigated using this method and the details of the computation can be seen in the spreadsheet as shown in **Appendix B3**. The summary of the exceedance probabilities obtained is shown in Table 6-5.

Table 6-5 Exceedance Probabilities and Return Period of Water Level Z at Siniawan (TIZ)

Z (m)	Exceedance Probability (1- F_H)	Return Period (Year)
7	0.060	0.07
>8	0.028	0.15

6.3.8 Limitations of the Bivariate Distribution Method

Discrete approximations are used in the computation of probabilities. This is only possible after selecting an arbitrary interval of Q in the probability derivation of Q, and this selection in turn predetermines the value of H to be entered for selecting the probability of H from the density function of H. The limitations of the Rating Curve

Method are also inherited when it is used in estimating $h(H, Q)$. This is reflected in the inability to define a much finer interval of Z, especially in the extreme upper tail end. It is in this upper region that flood quantiles are assigned to extreme observed values.

6.4 Chapter Summary

Three methods that can be used for estimating design floods in the TIZ have been presented and illustrated with examples in this chapter. The Direct Frequency Method can be used when there are long-term records available. The results can be obtained readily since the effects of the underlying physical causes of floods in the TIZ due to storms, surges, and tides are integrated into the record.

The Rating Curve Method can be used to produce regression curves that provide preliminary assessment of water level in the TIZ given a set of H and Q values. The curves may not be used in the final detailed design purposes since the regression curves are likely to be inaccurate with large residuals and unacceptably low R^2. However, the curves can be useful in the Bivariate Distribution Method.

The Bivariate Distribution Method produces bivariate probabilities given that the two underlying processes at work in the TIZ are independent and the surge effects are ignored. With relatively short records, the probability density functions of tides and discharges can be derived. The disadvantages of the method includes the limitation of the Rating Curve Method if the relationship of $h(H,Q)$ is used, and the arbitrary selection of correction factor. Deriving H-Q curves with a fine interval of Z, especially at very high levels, is a challenge.

Chapter 7 Fourier Method

The flood estimation method for the TIZ which has been briefly introduced as the *Fourier Method* in Section 4.5 is the main subject of this chapter. An idea that originated from the harmonic analysis of tidal river records by Goring (1984) is adopted and extended. The method is based on the premise that since the flooding events in the TIZ are caused by a certain combination of tides and fluvial discharges, the observed water level series should carry the signatures of these generating processes. The tide generation process is essentially deterministic because the tidal forces are identifiable. A number of unique frequencies related to the maximal effects of these forces are known. The identification of these frequencies provides the foundation for using a decomposition technique by digital filtering, a widely used tool in the field of electrical engineering.

7.1 Frequency Domain-Based Approach

An observed water level series in a TIZ, $Z(t)$, is regarded as a results of a mixture of tides and river flows. Since the tidal regime is relatively well behaved with harmonic characteristics, it can be separated using the frequency domain-based technique of Fourier

Transform. The flow-separation technique, which identifies the line spectra associated with the high frequency tidal components, is performed on the river stage series observed hourly at a station in the TIZ. The Fourier transforms identified for the low frequency components are used to synthesize a time series that resemble a *river series, R(t)* using the Inverse Fourier Transform(IFT). According to the schematic of the approach as shown in Figure 4-4, an empirical probability density function associated with $R(t)$ can be established. Similarly, performing IFT on the high frequency components can synthesize a time series that resemble a *tidal series, K(t)*. From this series, an empirical probability density function can also be derived for the tidal series. Exceedance probabilities are then derived from both probability density functions of the tides and river series using a joint probability method that has been developed for sea conditions (see Pugh and Vassie (1980), Tawn and Vassie (1989), Tawn (1992)). Flood quantiles, L_T, in terms of flood levels, are then estimated for engineering design purposes based on the joint probability.

7.2 Fourier Analysis

Fourier analysis is the process of extracting frequencies and amplitudes that are present in a time series. This powerful method is often used in the general process of frequency identification. In order to pursue analysis and separation of tides from river flow series in the interaction zone, the observed water level series $Z(t)$ in the TIZ's of Sarawak are first subjected to the Fourier analysis. $Z(t)$ values are transformed into a Discrete Fourier Transform (DFT) series $Z_n(\omega)$ in the frequency domain using the Fast Fourier Transform (FFT). Identifying the unique frequencies in an observed series is

made even easier by taking a power factor on the DFT series, producing a Fourier Power Spectrum.

7.2.1 Identifying Tidal Harmonics

Tidal harmonics theory as reviewed in Chapter 2 shows that each harmonic component has a unique frequency that correspond to a certain configuration of tidal forces. The water level H in a sea at any time t can be represented by

$$H(t) = H_0 + \sum_{k=1}^{R} A_k \cos(\sigma_k t - \alpha_k) \tag{7-1}$$

where H_0 is the mean sea level, A_k , σ_k, and α_k are the amplitude, radial frequency of constituent and phase of constituent k. The total number of constituent R varies according to the location on the earth. In most places the constituents of the lunar semi-diurnal ($\mathbf{M_2}$), solar semi-diurnal ($\mathbf{S_2}$) and the lunar elliptic semi-diurnal ($\mathbf{N_2}$) are most dominant. The signatures of these major constituents can be readily traced in the Fourier frequency domain.

7.2.2 Fourier Power Spectrum

The identification of the different frequencies is best performed by inspecting the power spectrum of the series. The spectrum estimation techniques available are categorized as nonparametric and parametric. The nonparametric methods, which are efficiently implemented by using the FFT, are the periodogram, the Barlett and Welch modified periodogram, and the Blackman-Tukey methods. On the other hand, the parametric approach uses parameters of an autoregressive model (AR) of the series to

derive the spectrum. The general name of Yule-Walker method is given to the AR method.

7.2.3 An Example of the Identification of Tidal Components

Siniawan is a station located in a TIZ, 45 km from the mouth of the Sarawak River, Sarawak. Figure 7-1 shows the Fourier power spectrum of about 5 months of hourly stages at Siniawan. There are dominant frequencies at around 29.0, 30.0, 15.0, and 13.9 degree/hour which correspond to the $\mathbf{M_2}$, $\mathbf{S_2}$, $\mathbf{K_1}$ or $\mathbf{S_1}$, and $\mathbf{O_1}$ tidal species respectively. The presence of tidal harmonics at around 43 and 58 degree/hour are due to shallow water effects.

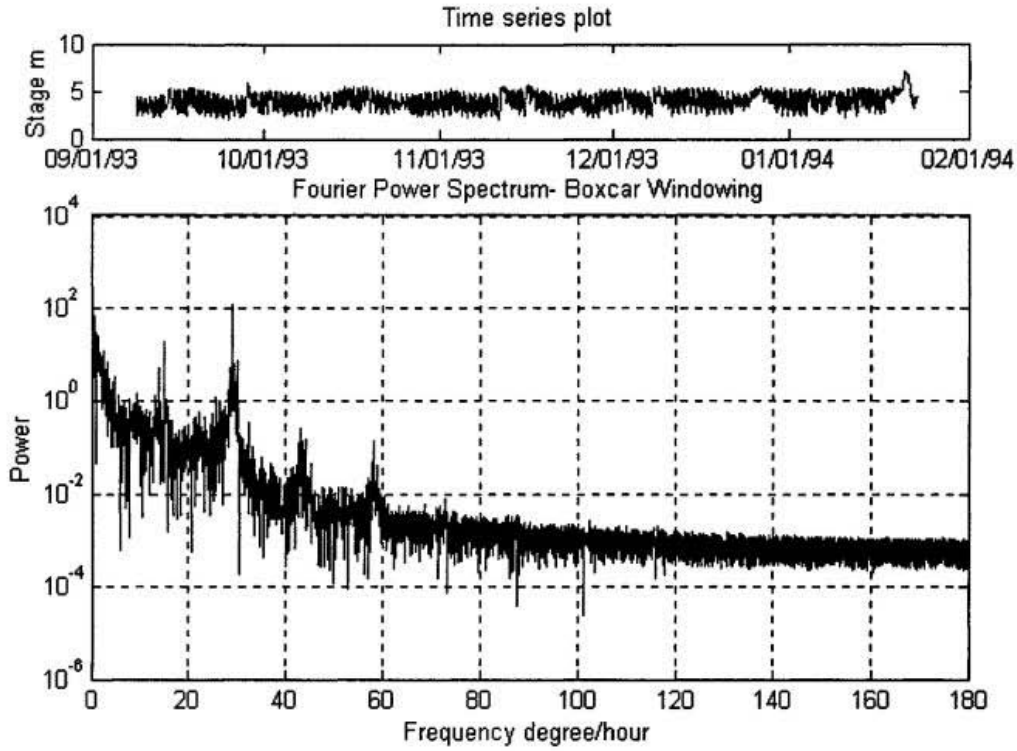


Figure 7-1 Fourier Power Spectrum Identifying the Major Tidal Frequencies at Siniawan

By using the definition of Equation 2-29 and the corresponding amplitudes of $\mathbf{M}_2$, $\mathbf{S}_2$, $\mathbf{K}_1$, and $\mathbf{O}_1$, the F factor is found to be 47/108 or 0.44. This indicates the tides are mixed, and are mainly semidiurnal. Although a longer series is required to determine each species with certainty, it is sufficient to note that that the major tidal frequencies are over 12 deg./hour.

7.3 Tidal Signature Removal

7.3.1 Components of Water Levels in the TIZ

The general observed water level $Z(t)$ in the tidal interaction zone are assumed to be a superposition of the different components as:

$$Z(t) = Z_0(t) + K(t) + R(t) + S(t) + I(t) \tag{7-2}$$

$Z_0(t)$ is the mean sea level which may be changing slowly with time, $K(t)$ is the component depicting the tidal dynamic processes, $S(t)$ is a meteorological surge component largely due to pressure variations in the atmosphere, $R(t)$ is a discharge related component due to the fluvial flows, and $I(t)$ is an interaction term due to the interaction of fluvial flows and tides. Wind set up can be a major factor in some places in the world. Equation 7-2 is an extension of the general representation of sea-level in the open sea, $H(t)$, used by Pugh(1987):

$$H(t) = Z_0(t) + K(t) + S(t) \tag{7-3}$$

where there is no fluvial flow effects.

In shallow rivers, the effect of atmospheric pressure changes can often be ignored due to the limited mass of water as compared to the sea conditions. Surge effects generated in the open seas, however, can propagate into the estuary at the same time as tides. In places with open seas that have small surge component, such as the case of the study area as found in Chapter 3, Equation 7-2 can be simplified to

$$Z(t) \approx Z_0(t) + K(t) + R(t) + I(t) \tag{7-4}$$

If all the readings are corrected to a mean sea level datum, or MSL datum, and the mean sea level remains constant, Equation 7-4 becomes

$$Z(t) \approx K(t) + R(t) + I(t) \tag{7-5}$$

In the Fourier Method, $I(t)$ is assumed to be present amongst the stochastic components of $R(t)$ and $K(t)$. Hence the Equation 7-5 is simplified to:

$$Z(t) \approx K(t) + R(t) \tag{7-6}$$

7.3.2 Methodologies in Tidal Signature Removal

In evaluating the surge in the sea conditions, the removal of tidal signature can be carried out using a tidal harmonic analysis routine to identify the amplitudes and phases of the various tidal components from the observed sea level series and to use them to reconstruct a pure tidal time series. However, this approach cannot be applied directly to the riverine conditions. Godin (1999) found that damping induced by bottom friction of the Upper Saint Lawrence River creates a non-linear distortion of the signal. Beside friction, the interaction effect of tidal and the outgoing fresh water also modify the tidal amplitudes. All these effects vary in accordance with the location in the tidal interaction

zones. Owing to these deficiencies, the direct application of the harmonic constants for tidal reconstructing in the TIZ is not used.

One of the other potential separation methods is the filtering method. It also makes use of the technique of identifying tidal harmonics. The following sub-sections describe the filtering aspects and the details of the method.

7.3.3 Digital Filters

A digital filter is defined as some sort of device that takes in an input signal $A(t)$ and produces an output signal $B(t)$. The usage of digital filters can be found in many fields in science and engineering. The simplest filter is a *linear filter*. The filter is considered linear and invariant if the output is related to the input by a convolution:

$$B(t) = U(t) * A(t) \tag{7-7}$$

The function $U(t)$ is an impulse function. In the Fourier or frequency domain, this becomes

$$\mathbf{F}B(\omega) = G(\omega)\mathbf{F}A(\omega) \tag{7-8}$$

where $G(\omega) = \mathbf{F}U(\omega)$ is the transfer function of the filter, and ω is the frequency.

An example of nonlinear filter is a *limiting filter*. Truncation of a signal is done whenever it goes below for a certain fixed minimum or exceeds a certain maximum value. Another class of filter is the *dynamic filter* that is designed to fit the expected noise behaviour by altering their characteristics in response to the signal. Although they can be very effective, they are not invariant by virtual of their definition.

A linear filter is still a preferred choice due to the preference for its invariant characteristic. *Mask filters* that let through some frequencies and stop others are the most common filters among the linear filter class, and this is the type that will be used for the purpose of tidal signature removal. The transfer function $G(\omega)$ takes the value of 0 and 1. The effect of a mask filter on the Fourier spectrum of a signal like the river stage is to remove certain frequency bands above f_r where $G(\omega) = 0$. The filter can be expressed mathematically as

$$G(\omega) = \begin{cases} 1 & (f \le f_r) \\ 0 & (f > f_r) \end{cases} \tag{7-9}$$

Hence a low-pass filter that lets through some frequencies below some critical frequency and block off the higher frequencies typical of tidal signature is the simplest of the mask filters that can be used. Other types of mask filters that are more robust are the band-pass filter (consisting of a low-pass and a high-pass filter), band-stop filter (opposite of band-pass), and comb filters (consist of multiple band-pass or band-stop filters).

7.3.4 Filters for Tidal Studies

Many filters are developed in the process of studying tides in sea conditions such as those proposed by Godin (1972), Walters and Heston (1982), and Dijkzeul (1984). In most of these cases, low frequency components were not their prime subject. However, Thompson (1983) proposed a set of low pass filters for studying low frequency motions in the oceans. Tides are considered a strong "high-frequency noise" in sea-level records. Goring (1984) used a low pass filter to separate tides from the river series at a few river

ports in New Zealand. The main objective was to obtain the amplitudes and phases at those locations for tidal forecasting.

7.3.5 Low-Pass Filter

The major tidal constituents are identified to be above 12 degrees/hour (Easton, 1977). By allowing low frequencies to be associated with the river flows and higher frequencies identified with the tides, the Fourier spectrum of the observed series in a TIZ can be split into two. Back transforming from the Fourier space by the inverse Fourier Transform allows two time series to be synthesized.

A MATLAB program is written as shown in **Appendix A3** that allows the processing and analyzing of the data series. The algorithm for the Low Pass Filter process is:

1. Assign the cut off frequency Fc =12 degrees/hour.
2. Read data series x and determine the length of series n.
3. Derive DFT of x using Fast Fourier Transform.
4. Compute the index of DFT, $ca = (n \times Fc)/360$, that coincide with the frequency Fc. The indices of DFT start with 1 at the lowest frequency.
5. Assign the DFT's with indices that are less than ca as due to the river flows.
6. Pad the rest of components with zeros to make up *ftr*, a sequence of DFT.
7. Invert *ftr* to get the filtered river series *ytr* using the Inverse Fast Fourier Transform.

8. Invert the time scale related to *ytr*.
9. Assign *ca* +1 until *n* as the number of DFT associated with tidal series.
10. Pad the component from 1 to *ca* with zeros to make up *ftt*, a sequence of DFT.
11. Invert *ftt* to get the tidal series *ytt* using the Inverse Fast Fourier Transform.
12. Invert the time scale related to *ytt*.
13. Save *ytr* and *ytt* as series related to river flows and tides for further analysis.

7.3.6 Applications of Low-pass Filter

Filtering was done on most of the data series available at stations found in the tidal interaction zones of Sarawak. These stations are listed in Table 7-1. Some of the filtered series are plotted as shown in **Appendix C1**. Figure 7-2 shows the result of applying the low pass filter to the same Siniawan series used in Section 7.2.3 and back transforming the split spectrum at around 12 degrees/hour. With all the available data at Siniawan separated, conventional probabilistic methods as described in the next section can be used to derive the associated flood quantiles.

Table 7-1 Stations in the Tidal Interaction Zones in Sarawak

Station Name	Corresponding Sea or Downstream Station	Corresponding Discharge Stations (Upstream)
Satok	Pulau Lakei	Kpg Git, Buan Bidi
Batu Kawa	Pulau Lakei, Satok	Kpg Git, Buan Bidi
Siniawan	Satok, Batu Kitang	Buan Bidi
Batu Kitang	Pulau Lakei, Satok	Kpg Git
Gedong	Pulau Lakei	Serian
Sg Salim	Pulau Lakei	Kapit, Ng Mukeh
Sg Merah	Pulau Lakei	Kapit, Ng. Mukeh
Marudi	Miri	Long Terawan, Long Jegan, Lio Matu

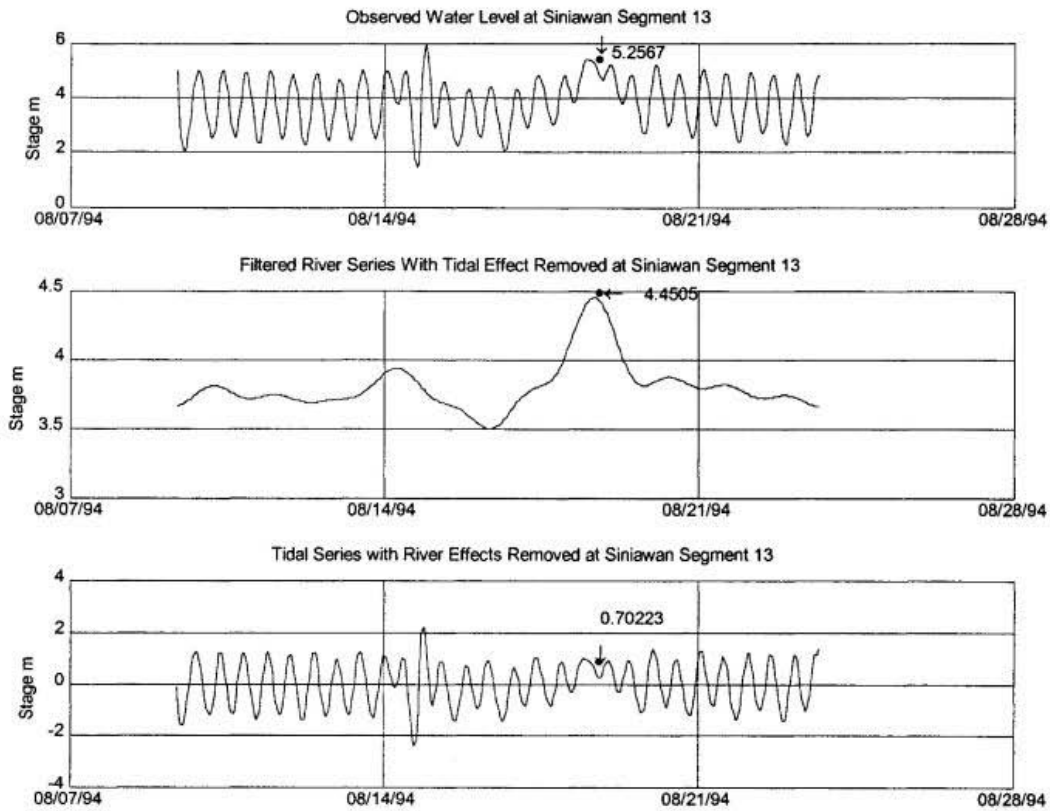


Figure 7-2 Results of Low Pass Filter Method at Siniawan

7.4 Extreme Level Frequency Analysis

Let $P_R(.)$ denote the probability density function of the filtered river series $R(t)$ and $P_K(.)$ represent the probability density function of the filtered tide series $K(t)$. Let r be the instantaneous value of the filtered river series $R(t)$. Since $Z(t) \approx K(t) + R(t)$, the tidal level would be Z-r and the probability density function for the total observed level Z in the TIZ is a combination of the probabilities of the filtered tides and river series expressed mathematically as:

$$P_Z(Z) = \int_{-\infty}^{\infty} P_K(Z-r) \cdot P_R(r) dr \tag{7-10}$$

The probability of exceedance for a particular water level η in the TIZ is the cumulative distribution function $F_Z(\eta)$

$$F_Z(\eta) = \int_{\eta}^{\infty} P_Z(Z) dZ \tag{7-11}$$

and the probability of non-exceedance of η is given by

$$G_Z(\eta) = \int_{-\infty}^{\eta} P(Z) dZ . \tag{7-12}$$

The entire probabilities have to be unity giving

$$F_Z(\eta) + G_Z(\eta) = 1 . \tag{7-13}$$

For the purposes of engineering design, the return period T of a level η in the TIZ is then defined as:

$$T(\eta) = \frac{1}{F_Z(\eta)} = \frac{1}{1 - G_Z(\eta)} \tag{7-14}$$

Derivation of the probability density functions, however, is dependent on the sampling interval of the maxima series used. Two intervals are conveniently chosen: hourly and yearly. The choice entails two separate sets of procedures in deriving the joint probability and quantiles estimations.

7.4.1 Joint probability of Annual Extreme Series

Annual maxima of the filtered river series and tidal series are treated as two separate series. Flood frequency analysis used in Chapter 5 and 6 to derive flood estimates on a single station basis can be adapted for the frequency analysis of the filtered series. The probability density functions $P_R(.)$ and $P_K(.)$ can be estimated by fitting a suitable extreme event distribution function to each of the annual maxima series. The distributions typically used for the river discharge series such as Log-normal, GEV, Generalized Pareto and Generalized Logistics are the possible candidates. Once these fitted distributions are available, the estimation of the distribution of annual extreme levels in the TIZ, $P_Z(Z)$, as in Equation 7-9, can be derived by a Monte Carlo (MC) Simulation procedure. If the distribution fitted to the river maxima series R_{max} is a Log-normal distribution, given by

$$LogR_{\max} = \mu_R + \sigma_R z \qquad (7\text{-}15)$$

and if a normal distribution fits well to the tidal series, $K(t)$,

$$K_{\max} = \mu_K + \sigma_K z \qquad (7\text{-}16)$$

then the following MC simulation procedure can be used:

(i) Generate a random number between 0 and 1. Obtain the corresponding normal variate z_1 from the cumulative normal probability distribution function.

(ii) Compute an extreme river level using $\exp(\mu_R + \sigma_R z_1)$

(iii) Generate another random number and obtain the corresponding normal variate z_2 as in (i).

(iv) Compute an extreme tidal level using $\mu_T + \sigma_T z_2$.

(v) Compute the sum of the total level Z by superposition.

(vi) Repeat step (i) to (v) for a large number of times, say 10000 times to obtain 10000 simulated years of combined levels.

(vii) Rank the simulated sequence and obtain the empirical frequency distribution of the combined levels.

The main obstacle with this approach is the difficulty of finding a sufficiently long observed water level series in the TIZ, as discussed in Chapter 3, such that an appropriate distribution can be fitted.

7.4.2 Joint probability of Hourly Series

Hourly series are becoming readily available from stations that are equipped with automatic recorders and hence using the hourly series is a natural option. It is noted in Chapter 2 that Pugh and Vassie (1980) used some hourly series of tides and surges in the sea conditions off U.K. to compute the joint probability of combined levels and derived the quantile estimates. The advantage of this approach is that a short but intensely sampled water level series can be used. However, the main hurdle of this approach lies

with the difficulties in associating the hourly extremes with the return period, which is often expressed in annual interval in most engineering design.

Considering hourly samples, and assuming that hourly values are independent, the distribution function of the annual maxima $G_Z(.)$ is related to $Fz(.)$ by

$$G_Z(\eta) = F_Z{}^N(\eta) \qquad (7\text{-}17)$$

where N is 8766, the number of hourly observation in a year (with adjustment for leap years). The return period T_Z of a level η in the TIZ is then given by:

$$T_Z(\eta) = \frac{1}{1-G_Z(\eta)} = \frac{1}{1-F_Z{}^N(\eta)} \approx \frac{1}{N(1-F_Z(\eta))} \qquad (7\text{-}18)$$

A correction factor 1- $q_1(\eta)/q_2(\eta)$ can be used to adjust for the dependent hourly data resulting in the expression:

$$T_Z(\eta) \approx \frac{1}{N(1-q_2(\eta)/q_2(\eta))(1-F_Z(\eta))} \qquad (7\text{-}19)$$

where $q_i(\eta)$ denotes the probability that i consecutive levels are greater than η. Pugh and Vassie (1980) found the correction factor if used resulting in marginal adjustment of the quantile levels as compared with the magnitude of the extreme levels.

Deriving an approximate $P_K(.)$ from the observed data series of sea tide levels has been discussed and applied in Chapter 3 and 6. The observed series are discretely sorted at certain fine intervals or bins for frequency or histogram analysis. If the data series are of sufficient length, recommended to be greater than 18.6 years to cover a full tidal cycle, the tidal density function can be approximated by the discrete empirical density function. If the tidal series is short, the series is fitted with the harmonic function and an 18.6-year series can be simulated. However, the latter method of simulation

applicable to the sea conditions is not strictly applicable to the filtered tide series obtained in the TIZ. From a hydraulic perspective, it is well known that the bed friction factor can dampen the propagation of tides up a relatively shallow river. An empirical tidal density function is thus preferred for the TIZ.

Deriving a probability density function from an observed hourly discharge series has been used in Chapter 6. This similar approach can be used for deriving an approximation of the $P_R(.)$ from the filtered river series $R(t)$. However, the independence assumption is not valid for the hourly river levels. The levels that are exceeded during the same extreme storm events are likely to be counted more than once. The correction factor $\theta(\eta)$ is the reciprocal of the mean time spent above a particular level η on each excursion above this level. Pugh and Vassie (1980) found that 1-dependent process is adequate to described the surge process in the seas, and the correction factor

$$\theta(\eta) = 1 - q_1(\eta)/q_2(\eta) \qquad (7\text{-}20)$$

was justified as most surges do not last more than one high water (of tides). However, storm runoffs in the rivers are often last longer. The river levels tend to stay above the high tidal levels even more frequently as the zone of concern shifted upstream. A procedure to estimate the average number of hours that surges in the sea exceed an extreme level can be found in Tawn and Vassie (1989). Equation 7-18 is generalized as:

$$T_Z(\eta) \approx \frac{1}{N\theta(\eta)(1 - F_Z(\eta))} \qquad (7\text{-}21)$$

This procedure may be adapted for analysing the effects of storm runoffs in the TIZ. Nevertheless, there are many issues yet to be resolved, and amongst them are:

(i) The number of storm per year to be used in abstracting extreme sequence; for the surge data in the sea, five storms per year was reported to be optimum by Smith (1986) and Tawn (1988).

(ii) The appropriate threshold level to estimate the correction factor $\theta(\eta)$; for the surge data in the sea, a rough guide provided by Tawn and Vassie (1989) is to choose the threshold such that the total number of hourly exceedances is equal to the product of the number of years, the number of independent extreme storms per year, and $1/\theta(\eta)$.

7.5 Example of the Extreme Level Frequency Analysis

Siniawan is one of the stations that was analyzed using the filtering steps producing two separate series that relate to the tides and the river flows. The complete plots of the series are shown in **Appendix C1**.

7.5.1 Annual Maxima Series

The method as described in 7.4.1 was applied to the Siniawan series. Table 7-2 shows the annual maximum of the two separate series as R_{max} and K_{max}. Although the series are short and far from the ideal length of 25, the methodology used is the main purpose of this illustration. Fitting of distributions were performed and the best was found for each of the filtered series.

7.5.1.1 River series

The best-fit distribution for the extreme filtered river series, in terms of the least standard errors of fit, is the Generalized Pareto distribution. The fit is accomplished by using the L-moment procedure. The quantiles are given by:

$$R_{\max}(T) = \varepsilon + \frac{\alpha}{k}\left(1 - T^{-k}\right) \qquad (7\text{-}22)$$

where T is the return period or $1/p$. The estimates of parameters ε, α and k are 4.2286, 3.6298 and 0.9815 respectively.

It is interesting to note that Generalized Pareto distribution was found earlier in Chapter 5 as the appropriate regional flood frequency distribution for the non-tidal zones in western Sarawak. The station Siniawan is downstream of the Sungai Sarawak, which is in the same homogeneous region.

Table 7-2 Annual Maximum Levels at Siniawan Based on Filtered Series

Year	K_{Max}	R_{Max}	Z_{Max}
1992	2.829	7.360	10.223
1993	1.547	5.305	7.215
1994	1.534	5.524	7.080
1995	2.601	7.278	10.099
1996	2.012	6.039	8.007
1997	1.205	5.011	6.224
1998	2.385*	6.286	8.698
1999	1.416*	4.626	6.077
2000	2.855*	7.115	9.971

* These are in the post-barrage condition and hence are not considered in the determination of the probability density functions.

7.5.1.2 Tide series

The best-fit distribution for the corresponding filtered tidal series is the GEV distribution using the MOM. The estimates of parameters ε, α and k are 1.6919, 0.5939 and 0.1552 respectively for

$$K_{\max}(T) = \varepsilon + \frac{\alpha}{k}\left(1 - (-\ln(1/T))^k\right) \qquad (7\text{-}23)$$

7.5.1.3 Monte Carlo Simulation

Monte Carlo simulation procedures as described in Section 7.4.1 were used, making use of the generic Equations 7-21 and 7-22. The annual probability of exceedance, p, was simulated by a random number between 0 and 1 which gave rise to a return period T associated with the series. A listing of the program used for this particular simulation is shown in **Appendix A4**.

After running 10,000 number of simulations, the results were abstracted and shown in Table 7-3. The ranges of the flood levels are within reasonable limits as compared with the observed series. The comparisons with other methods are to be discussed later.

7.5.1.4 ACF and PACF of Hourly Series

The autocorrelation function (ACF) and partial autocorrelation function (PACF) of the filtered river series are plotted in Figure 7-3. It can be deduced that the hourly filtered river series are highly dependent, as the ACF does not decay rapidly. The adjustment factor for dependence must be used.

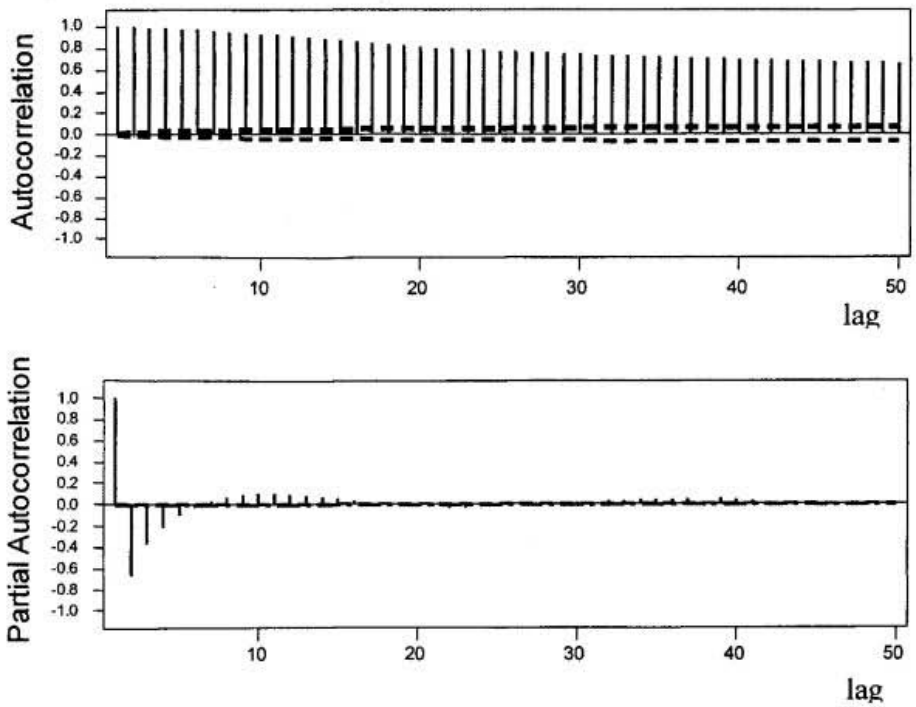


Figure 7-3 ACF and PACF of Filtered River Series at Siniawan

Table 7-3 Design Flood Levels at Siniawan Estimated Using Empirical Distribution Simulated by MC Using Distributions Fitted to Filtered Annual Maxima Series

Return Period T	i	Z_T
5	8001	9.136
10	9001	9.634
50	9801	10.464
100	9901	10.755
200	9951	11.050

i = Index of simulated annual maxima sequence(sorted in ascending order, number of years simulated N=10000)

The distribution of river series without any adjustment is shown in Table 7-4. The variability is relatively high in the upper extreme end as seen in the logarithmic plot of the pdf. Moreover, it does not allow the possibility that a flood level may exceed the maximum observed level.

7.5.1.5 Correction Factor and Smoothing Upper Tail of Distribution

A correction factor used to adjust for the dependence in hourly data was studied by Pugh and Vassie (1980) and Tawn and Vassie (1989) as noted earlier in Section 7.4.2. A similar technique is used to estimate the correction factor $\theta(\eta)$ for the Siniawan filtered river series. A program is written to scan the river series as shown in **Appendix A5**. The results are shown in Table 7-4. It can be seen that based on the criteria of 5 storms per year, based on surge statistics of Tawn(1989), the total number of hourly exceedance that approximately equal the products as shown in column (4) of Table 7-4 is at a threshold level η of 4.7 and $\theta^{-1}(\eta)$ of 30.3 hours. Column (4) is the product of the number of year (m), the number of storm per year (u), and $\theta^{-1}(\eta)$. However, based on the observation of the Siniawan series, the average number of storms per year with level exceeding the astronomical high water is around 20. Hence based on this new criterion, column (5) of Table 7-4 is computed leading to a threshold level η of 4.2m and 36.8 hours for θ^{-1} being chosen. The filtered river series is then subjected to a filtering process to obtain independent extremes by identifying storms that exceed the threshold level, 4.2 m in this example. The maximum within each individual storm is extracted to form a point in an extreme series. The extreme series is ordered with the r largest selected such that

Table 7-4 Correction Factor $\theta(\eta)$ for Different Threshold Level η

Threshold Level η(m) (1)	**Mean $\theta^{-1}(\eta)$ (hours)** (2)	**Total Number of Hourly Exceedance** (3)	**$m \times u \times \theta^{-1}(\eta)$ u = 5** (4)	**$m \times u \times \theta^{-1}(\eta)$ u = 20** (5)
4.0	53.3	9484	1866	7462
4.1	42.8	6640	1498	5992
4.2	**36.8**	**4854**	1288	**5152**
4.3	31.4	3267	1099	4396
4.4	33.2	2354	1162	4648
4.5	31.4	1851	1099	4396
4.6	32.4	1460	1134	4536
4.7	**30.3**	**1213**	**1061**	4242
4.8	31.1	995	1089	4354
4.9	31.4	847	1099	4396
5.0	32.9	724	1152	4606
5.1	34.3	652	1201	4802
5.2	34.7	590	1215	4858
5.3	32.3	517	1131	4522
5.4	35.6	463	1246	4984
5.5	32.0	416	1120	4480
5.6	41.2	371	1442	5768
5.7	38.2	344	1337	5348
5.8	44.0	308	1540	6160
5.9	38.6	270	1351	5404
6.0	29.4	206	1029	4116
6.1	39.5	158	1383	5530
6.2	34.8	139	1218	4872
6.3	39.0	117	1365	5460
6.4	33.7	101	1180	4718
6.5	24.0	72	840	3360
6.6	31.5	63	1103	4410
6.7	29.0	58	1015	4060
6.8	26.5	53	928	3710
6.9	24.0	48	840	3360
7.0	21.0	42	735	2940
7.1	17.0	34	595	2380
7.2	11.5	23	403	1610
7.3	5.0	5	175	700
7.4	0.0	0	0	0

Note: m = No. of year = 6 for record length
u = No. of storm per year

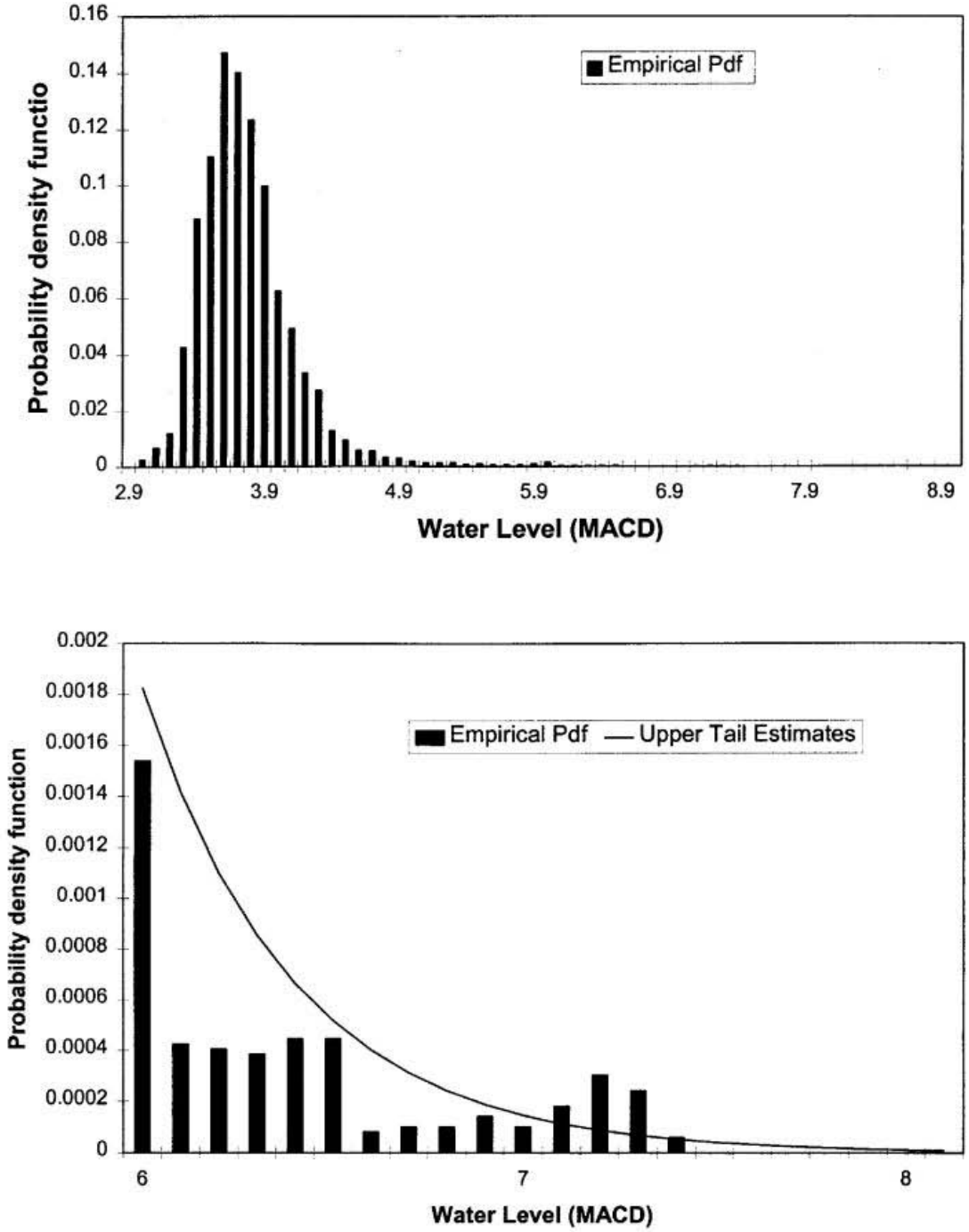


Figure 7-4 pdf of Hourly Filtered River Series Siniawan 1992-1997 and the Smoothing of the Upper Tail Using Gumbel Distribution

$$\eta_1 \geq \eta_2 \geq .....\eta_r \qquad (7\text{-}24)$$

with $r = u \times m$, where u is the number of storms per years, as recommended by Tawan and Vassie(1989). A full listing of the r largest extreme series is shown in **Appendix C2**.

Assuming that Gumbel extreme value distribution fit these r extreme values, the parameters of the river series R_{max}

can be estimated by

$$\hat{\sigma} = \frac{1}{r}\sum_{i=1}^{r}(\eta_i - \eta_r) \qquad (7\text{-}25)$$

and

$$\hat{\mu} = \hat{\sigma}\log r + \eta_r \qquad (7\text{-}26)$$

For the Siniawan filtered river series, the estimated parameters of σ and μ are 0.3965 and 6.0744 respectively. Using the same relation of hourly dependent sequence with respect to annual sequence, the fitted Gumbel distribution is related to the distribution of the hourly filtered river series $F_R(.)$ by

$$R_{\max}(\eta) = F_R^{\,Nm\theta}(\eta) \qquad (7\text{-}27)$$

where N = 8766, the number of hours in a year, m=6 for the record used, and θ=1/38.6. Explicitly for very extreme levels that are beyond the recorded level, i.e. for $\eta > \eta_r$, the cdf is given by

$$F_R(\eta) = R_{\max}^{1/Nm\theta}(\eta) \qquad (7\text{-}28)$$

Hence, the density function for very large η is approximated by

$$P_R(\eta) \approx \frac{1}{\sigma N m \theta} \exp\left[-\left(\frac{\eta-\mu}{\sigma}\right)\right] \qquad (7\text{-}29)$$

The fitted distribution allows a smoothing of the upper tail of the empirical pdf as shown in Figure 7-4. This allows the probabilities of extremes well beyond the recorded levels to be reasonably estimated.

7.5.2 Filtered Tide Series

Having derived the density function of filtered river series, the next step would be to derive the pdf for the filtered tides. The joint probability method outlined in Section 7.4.2 can then be used.

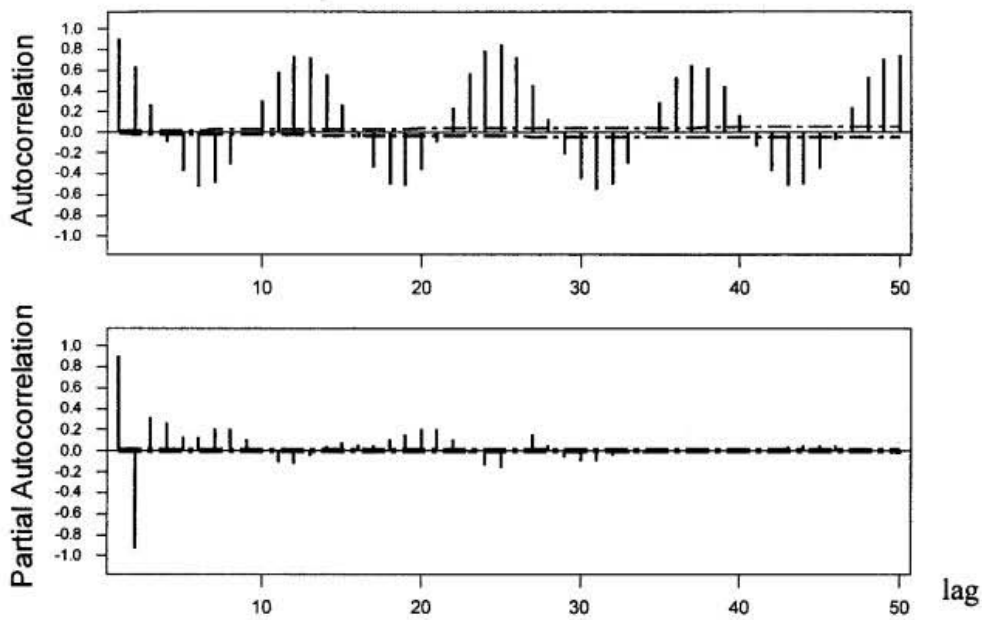


Figure 7-5 ACF and PACF of Filtered Tide Series at Siniawan

After filtering from the observed series by the Fourier low-pass filter, the series is checked for its correlation structure to confirm that the series is indeed behaving like a tidal series. The ACF plotted in Figure 7-5 shows a sinusoidal fluctuation with a cycle of 12 lags. The PACF has two initial significant spikes that finally decayed. These indicate that the tidal cycles with regular period are present.

The descriptive statistics of the hourly sequence $K(t)$ is shown below:

Descriptive Statistics: K(t)

Variable	N	Mean	Median	TrMean	StDev	SE Mean
K(t)	32737	0.00000	0.02878	-0.00266	0.76877	0.00425

Variable	Minimum	Maximum	Q1	Q3
K(t)	-2.62643	4.19576	-0.63788	0.62510

From the inspection of these statistics, it is postulated that a normal distribution may be suitable in describing the hourly data. A normality test based on Anderson-Darling test was carried out with the plot shown in Figure 7-6. The test showed a p-value of less than 0.0005, meaning that there is evidence to suggest that the data do not follow a normal distribution. In fact this test is not appropriate because most large samples such as this, with N=32737, would likely to fail the test. A plot of the empirical density function based on hourly-observed series is shown in Figure 7-7. The bimodal feature is typical of the tides in the sea as presented in Chapter 3. Apparently the tidal regime is very well preserved at Siniawan although it is located upstream at 45 km from the river mouth. Several other distributions types are fitted but none is found suitable. In this context, the empirical density function derived using hourly data are to be considered for a MC simulation at hourly interval. This is described in details in the next section.

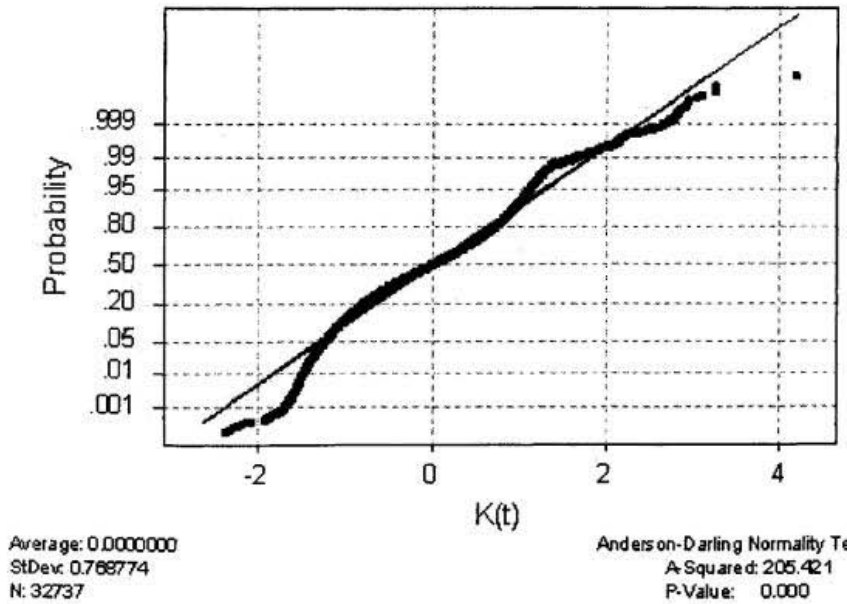


Figure 7-6 Normality Test for Distribution of Filtered Tides Series at Siniawan *K(t)*

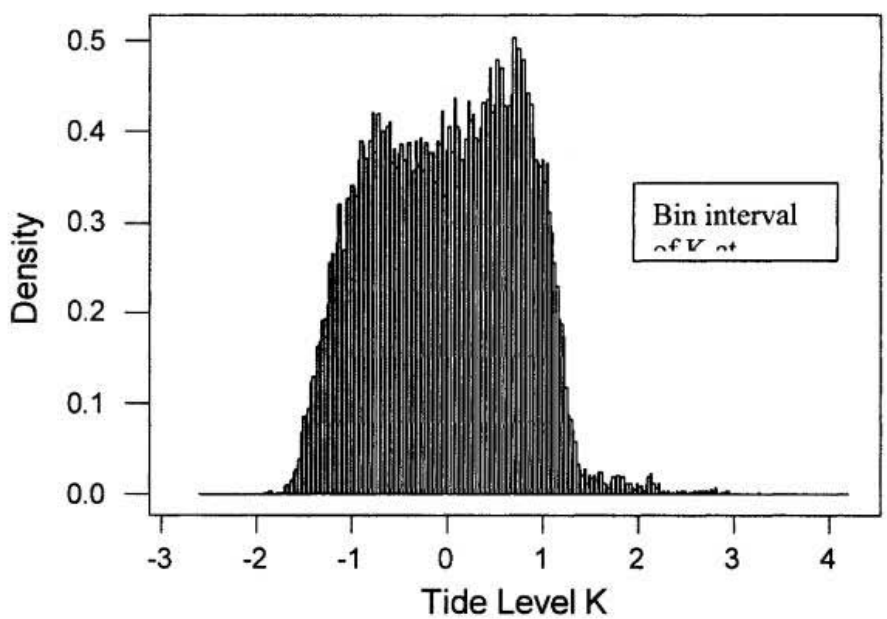


Figure 7-7 Empirical pdf of the Filtered Tide Levels K(t)

7.5.3 MC Simulation of Hourly Sequence

The pdf of both the tides and river series were derived using hourly data. Hence, it was appropriate that the simulations of sequences were carried out at hourly intervals. The process used is similar to the method used earlier in deriving the annual sequence except that empirically derived and adjusted pdf's were used instead of the fitted distribution.

Since the river series are governed by stochastic processes, the extension of the upper tail beyond the maximum observed level was achieved by fitting the Gumbel distribution to the r largest extreme series. The probabilities were adjusted, by insignificant amounts, to ensure that the total of probabilities sum to one. **Appendix C3** shows the empirical probability density function for both the tide and river series, which were used as lookup tables in the MC simulation. The listing of the program for carrying out this lengthy simulation is shown in **Appendix A6**. The results of the MC simulation involving 1000 years of hourly sequence are shown in Table 7-5. The results give similar but slightly lower quantiles as compared with the results of the annual extreme method given in Table 7-3.

7.5.4 Other Considerations

shows the method's ability to continuously obtain the river series at Siniawan despite the fact that a river barrage has been built at 23 km downstream of Siniawan. The barrage is described in detailed in Sharp and Lim(2000). Because the barrage is opened frequently for flushing purposes during low tides, some altered form of tidal series can be

resolved by the Fourier filtering method. For future flooding probabilities, sufficient length of the altered series can be used to derive the pdf of the tidal intrusion in post barrage condition. For the example series used, the filtered tidal series in the post barrage condition (after Febuary1998) is not used as the tidal regime has changed. On the other hand, the filtered river series in the post barrage condition can still be used as the water levels are affected mainly by the upstream stochastic processes that remain unchanged.

Table 7-5 Design Flood Levels at Siniawan Estimated by MC Simulation At Hourly Interval

Return Period T	i	Z_T
5	801	8.868
10	901	9.182
50	981	9.783
100	991	10.020
200	996	10.330

i = Index of simulated annual maxima sequence(sorted in ascending order, number of years simulated N=1000)

7.6 Discussions on the Fourier Method

In general, there are many deficiencies of the Fourier analysis, which have been widely discussed in the literature. In fact there were many attempts to alleviate these deficiencies. Window Fourier Transform or Short-Time Fourier Transform was one of the revised methods (Gabor,1946; Allen and Rabiner, 1977; Portnoff, 1980).

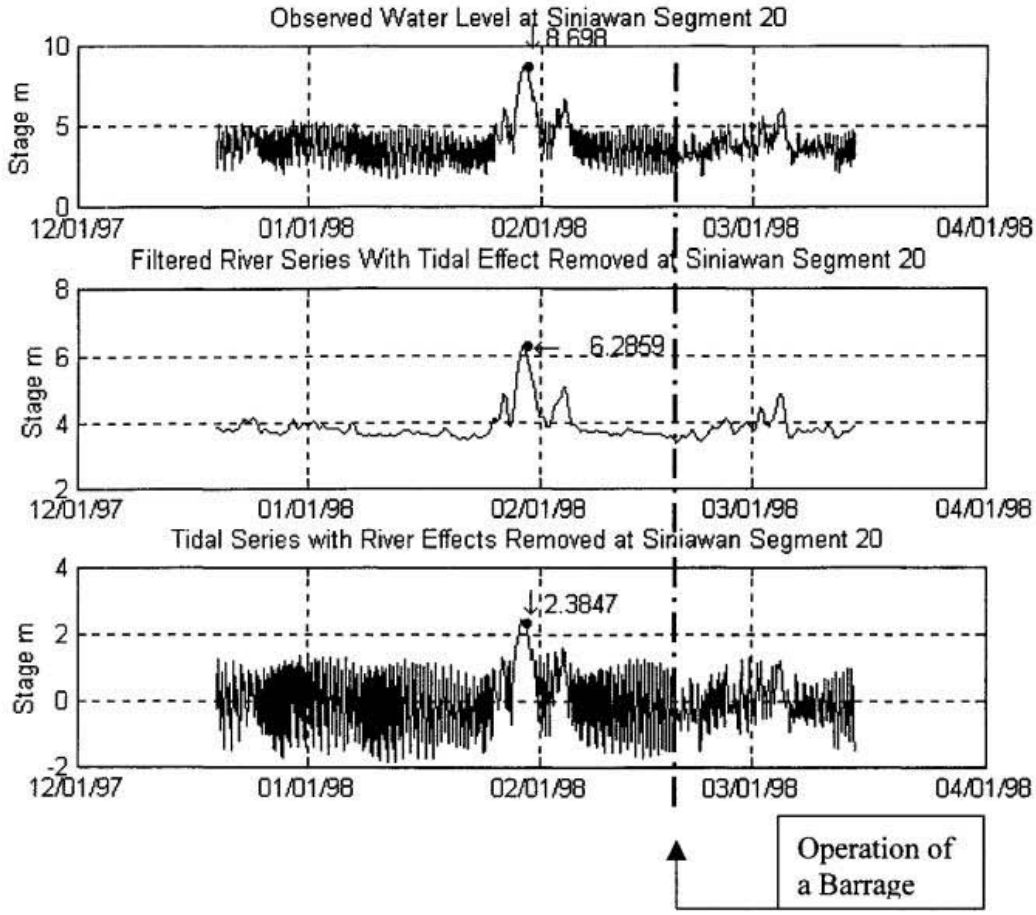


Figure 7-8 Continuous River Series in Post-Barrage Conditions

According to the Heisenberg Uncertainty Principle, time and frequency energy concentration are restricted. Fourier Transform deals with the frequency localization well and neglect the time aspects. While transforming into the frequency domain as a whole, the time information is lost. It is unable to provide much information about the occurrence of spike or discontinuities, which can be very crucial in time-resolution sensitive cases. Indeed, the inherent deficiencies in the Fourier Transform have motivated

the search for a general alternative method of data analysis – wavelet methods- which are to be discussed in the next chapter.

The filtering process carried out has been limited to the use of a low-pass filter. Further studies may be conducted to examine the sensitivity of refining the filter. Comb-filter may also be used if the specific site has sufficiently long record (18.8 years) with good time resolutions.

7.7 Chapter Summary

This chapter presents a new design flood estimation method which endeavours to separate the effects of the two underlying processes and assigning the probabilities of occurrence separately and then calculating the joint probability. The method is applied directly to assess the flooding frequencies at the location of interest without much prior knowledge about the physical conditions. The method makes use of hourly data and is very practical for situations where only a very short length of observed records exists, e.g., 6 years for the Siniawan series. MC simulation has been utilised to derive the joint probability of the two underlying regimes. The method has an advantage of keeping a long-term record of the filtered river series, which is not affected by the alteration of the tidal propagation processes in the downstream areas, e.g., construction of a barrage. Fourier analysis in general has the inherent disadvantage of poor time localization, and is unable to resolve the existence of spikes and discontinuities in time series very well. Despite these disadvantages, the method itself has laid the groundwork for yet another elegant, new method, the Wavelet Method, elaborated in the next chapter.

Chapter 8 Wavelet Method

8.1 An Exciting New Tool

The Fourier low-pass filter method as discussed in Chapter 7 provided a good start, but it is not satisfactory due to the lack of a good separation algorithm and robustness in considering the localization effects due to the non-stationary nature, in terms of frequencies, of tides that interact with river flows. There are many successes of practical applications using wavelet transform method as a replacement method for the traditional Fourier transform method in all fields of science. A wavelet analysis method is an excellent mathematical tool that emerged in the late 1980's, while the original idea of Fourier Transform that expresses a periodic function using an infinite sum of periodic complex exponential functions dated back to 1822. Being a new statistical and mathematical tool, the wavelet transform is a promising technique for use in the TIZ.

In this Chapter, a wavelet analysis method, applicable to flood frequency analysis in the TIZ, is developed and tested. The results, in terms of tidal separation, are very encouraging. The methodology proposed is a totally new approach in design flood estimation method for the TIZ.

8.2 Wavelets Transform

8.2.1 Time Series and Wavelet Analyses

The basics of wavelets in general were introduced in Chapter 2. A brief preview of the proposed Wavelet Method is found in Section 4.6. The application of wavelet methods in time series analysis is the focus of this section. The essential terms and symbols to be used in later parts of the Chapter are defined herein.

8.2.1.1 Wavelet Decomposition

Multilevel decomposition using wavelets allows signals or time series to be decomposed into coefficients of approximations and details. Essentially, the signal is analyzed at different frequencies with different resolutions. Not every spectral component is resolved equally. Multi-resolution analysis as such is designed to give good time resolution and poor frequency resolution at high frequencies and good frequency resolution and poor time resolution at low frequencies. This approach makes sense especially when the signal in hand has high frequency components for short durations and low frequency components for long durations

8.2.1.2 Scale

The *scale* parameter in wavelet analysis is similar to the scale used in geographic maps. As in the case of geographic maps, small or high scales correspond to a non-detailed global view, and large or low scales correspond to a detailed view. Similarly for wavelets, low frequencies (high scales) correspond to global information of a signal,

whereas high frequencies (low scales) correspond to a detailed information of a hidden pattern in the signal that usually lasts a relatively short time.

8.2.2 Comparison with Fourier Transform

There are two main reasons why Fourier Transform(FT) is losing its popularity in many areas of signal processing. The first reason is that FT provides the average frequency content over the entire signal, unacceptable to many applications. The other shortcoming is the inability of FT to handle non-stationary signals. Chui (1992) traces the attempt of Gabor (1946) to partially correct these deficiencies by introducing a windowed FT. The following subsection examined these two factors in details and explains how wavelets can overcome these deficiencies.

8.2.2.1 Localization

The deficiency of Fourier transform lies with the inherent poor localization characteristic. In transforming to the frequency domain, time information is lost, i.e., it is impossible to tell when a particular event took place. A wavelet transform draws its strength from its potential ability to localize the time-frequency coherent energy in a given signal in the time-scale plane (Teolis, 1998). By choosing a wavelet that is well localized in time and /or frequency, the corresponding wavelet transform will exhibit the same localizing ability. Time localization and frequency localization are in fundamental conflict, governed by an uncertainty principle well known in quantum mechanics called the Heisenberg Uncertainty Principle (Messiah, 1961). This principle has been used in

describing the incompatibility of frequency and time mathematically (Daubechies, 1990; Teolis, 1998). Hence a suitable wavelet can be selected based on the objective of the signal to be processed.

8.2.2.2 Stationarity

Stationarity in signal processing mostly refers to the state of the frequency in signals. If the frequency changes over time, the signal is said to be non-stationary. Fourier transformation is inadequate for dealing with non-stationary signals. Most signals contain many nonstationary or transitory characteristics: drift, trends, abrupt changes, and beginnings and ends of events. These characteristics are often the most important part of the signal, and Fourier analysis is not suited to detecting them. On the other hand, wavelet possesses the varying dilation characteristics and hence is able to handle the variation of frequencies in a signal.

8.3 Wavelet De-noising

One of the successful applications of wavelets is in the field of non-parametric statistical estimation (Bruce and Gao, 1996). Wavelet de-noising can be regarded as a major breakthrough in wavelet application pioneered by Donoho and Johnstone (1994). It is a powerful tool developed in the domain of wavelet analysis, providing a nonparametric estimation of the function f using orthogonal basis. A general suppression method is formulated based on the fundamental principal that noise is incoherent with respect to a set of wavelet functions. In general, a smoothed signal $f(t)$ is obtainable from a signal $s(t)$ by

$$s(t) = f(t) + \sigma\xi(t) \tag{8-1}$$

where $\xi(t)$ represents noise such as the white noise $N(0,1)$ and σ is the noise level.

8.3.1 De-noising by Wavelet Shrinkage

The principle of wavelet shrinkage used effectively in wavelet de-noising was developed by Donoho and Johnstone (1994, 1995). The method called *WaveShrink* consists of three steps: applying discrete wavelet transform, removing noise by shrinking wavelet coefficients towards zero, and finally applying the inverse discrete wavelet transform. The method shows great promise and further developments are carried out by Nason (1995) using cross-validation; and Ogden and Parzen (1996) developed a threshholding scheme which is known to be data-dependent.

For a signal s, the wavelet transform with J levels produces wavelet detailed and approximation coefficients d_1, d_2, ...d_J, a_J. The detail coefficients at different scale levels are shrinked towards zero via a function $\delta c(x)$ such that the new detail coefficients becomes $\delta_{\lambda_1\sigma_1}(d_1), \ldots..\delta_{\lambda_j\sigma_j}(d_j)$ where $c = \lambda\sigma$. λ is the threshold while σ is an estimate of the scale of the noise. These detail coefficients are reverse transformed to obtain the smooth estimate.

There are certain choices of the threshold λ that can achieve optimal estimate of $f(t)$ with a minimum number of assumptions about the underlying nature of the true function F. In statistical terms, there is a *minimax risk* over a broad class of the underlying functions. The method also possesses a locally adaptive bandwidth such that the

WaveShrink estimates are provided based on smaller bandwidth for the localized abrupt changes and wider band for smooth portion of a signal.

There are thus four major adjustment allowed to execute the de-noising of a signal:

1. type of wavelet for analysis
2. shrinkage function $\delta c(x)$
3. rule in computing the threshold λ_j
4. rule in estimating the scale of the noise or noise level σ_j

Each of the four adjustments are further discussed below.

8.3.1.1 Type of Wavelet

The selection of the wavelet has to be selected to match the characteristics of the signal to be analysed. There are no hard and fast rules for selecting a wavelet to use for an analysis (Bruce and Gao, 1996). The main factors to watch out for are the smoothness and spatial localization of the wavelet, which are determined primarily by the support of the wavelet. There are also other trade-off factors such as vanishing moments, frequency localization, symmetry, and orthogonality. In general, a wavelet with a higher number of vanishing moments can better represent higher degree polynomial signals; a smoother wavelet has better frequency localization properties.

8.3.1.2 Shrinkage Function

There are soft and hard shrinkage functions defined respectively by

$$\delta^{S}{}_{\lambda}(x) = \begin{cases} 0 & \text{if } |x| \le \lambda \\ sign(x)\{|x| - \lambda\} & \text{if } |x| > c \end{cases} \tag{8-2}$$

$$\delta^{H}{}_{\lambda}(x) = \begin{cases} 0 & \text{if } |x| \le \lambda \\ x & \text{if } |x| > \lambda \end{cases} \tag{8-3}$$

Statistically the soft shrinkage function is preferred as it is continuous, shrinking values above the threshold. It concurs very well with the principle that noise affects all wavelet coefficients. However, in a situation when reducing the bias in estimation is of great concern, the hard shrinkage function is used. Bruce and Gao (1995) reviewed the differences and provided a semisoft shrinkage to remedy the drawbacks of both.

8.3.1.3 The Threshold Rule

A universal threshold λ_j is one of the option that can be used for thresholding the coefficients. This is given by:

$$\lambda_j = \sqrt{2\log(n)} \tag{8-4}$$

where n is the sample size. The universal threshold is good where a very high degree of smoothness is desired with largest thresholds. The other choice is minimax threshold for hard and soft shrinkage computed by Dohono and Johnstone (1994) for a range of sample sizes. The method minimizes the upper bound on the asymptotic risk, giving smaller threshold than an universal method for a given sample size, and resulting in less smoothing.

8.3.1.4 Scale of the Noise

Three rules can be specified for selecting the scale of the noise σ_j:

1. A single scale factor using the finest scale detail coefficients d_1 such

$$\sigma_j = \hat{\sigma}(d_1) \tag{8-5}$$

2. A single scale factor using all the detail coefficients $d_1,\ldots d_J$ such that

$$\sigma_j = \hat{\sigma}(d_1, d_2, \ldots d_J) \tag{8-6}$$

Although the estimate of scale can be more accurate since all the coefficients are use, it may be influenced by the signal (Bruce and Gao, 1996).

3. A separate scale factor for each level such that

$$\sigma_j = \hat{\sigma}(d_J) \quad \text{for } j = 1,2,\ldots J. \tag{8-7}$$

Conceptually this approach is good for characterizing signals with noise that varies from level to level.

8.3.1.5 Other Controls

There are other controls that may be used:

1. Setting all the coefficients at fine scale resolution to zero
2. Providing an individual choice of the vector of vertical thresholds and noise scales

8.4 De-noising Flows Series in TIZ

The process of river flows at any time is stochastic in nature while the tidal fluctuations are relatively deterministic based on the knowledge of the periodic lunar-solar effects. The superposition of the tides and river flow series is not linear due to the hydraulic factors in mixing processes. However, the width of the window is changed as the wavelet transform is computed for every single spectral component, which is probably the most significant characteristic of the wavelet transform, would enable the dynamic factor to be explained.

Wavelet analysis method is known to be excellent in handling localization and discontinuities. Streamflows are typically varying at a rate slower than the tidal fluctuation. The tidal fluctuations at higher frequency, typically evident at lower water level, are treated as 'noise'. The tidal series in the open seas are coherent as it was seen in Chapter 3. It has an almost Gaussian distribution with some fixed periodicities. However, as the tide propagates upstream it begins to posses some irregularities due to lag effects caused by channel frictions and the interaction effects due to the river flows. Based on the assumption that the tidal effects in the TIZ are non-coherent just like noise, the smoothed or approximation component can be treated as the underlying function.

The wavelet de-noising approach, namely the WaveShrink method, is applied to most of the water level series observed in the TIZ of the study area. A program written in MATLAB is shown in **Appendix A7**. Various options as described in Section 8.3 are to be chosen by the user.

8.4.1 Preliminary Assessment of Wavelet De-noising

Figure 8-1 shows the results of a wavelet de-noising performed on an hourly water level series observed at a station called Batu Kitang situated in a TIZ. Casual inspection of the series reveals that the tidal range is in the order of 4.5 to 5 metres. Some bumps can be observed which ride well above the fluctuating tidal noise. These bumps located at around 2400th hour (from the beginning of series) are attributed to the river flows. A smooth function obtained by wavelet de-noising is plotted over the signal. The actual peaks well above the tide levels are perfectly replicated while the undulating portion is estimated. It can be seen that the tidal component is not significant at the peak of the extreme events. The interaction effects by tides near the peaks are negligible. One therefore attribute those peaks to the river discharges. With the same rigor, the method is also able to separate those intermediate peaks with magnitude closer to the maximum high tides such as one around 550th hour. The smooth function represents a de-noised series, which can be used for selecting the peak events at a preset interval, say one year as in conventional flood frequency methods.

8.4.2 Assessment of De-noising Performance

The performance of the de-noising can be assessed based on three criteria:

1. the goodness of fit in the range >Hmax
2. the distribution of the tidal noise derived from the noise series
3. the correlation of the de-noised series with a flow series at an upstream gauged station

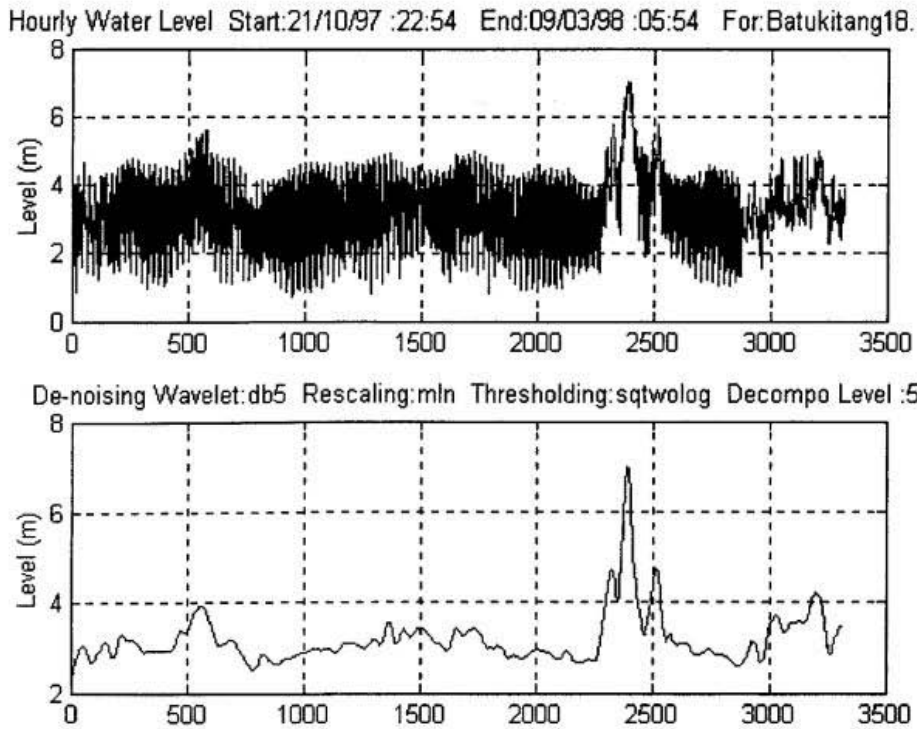


Figure 8-1 Results of Wavelet De-noising with Distinct Peak, Hmax ~ 6.0m

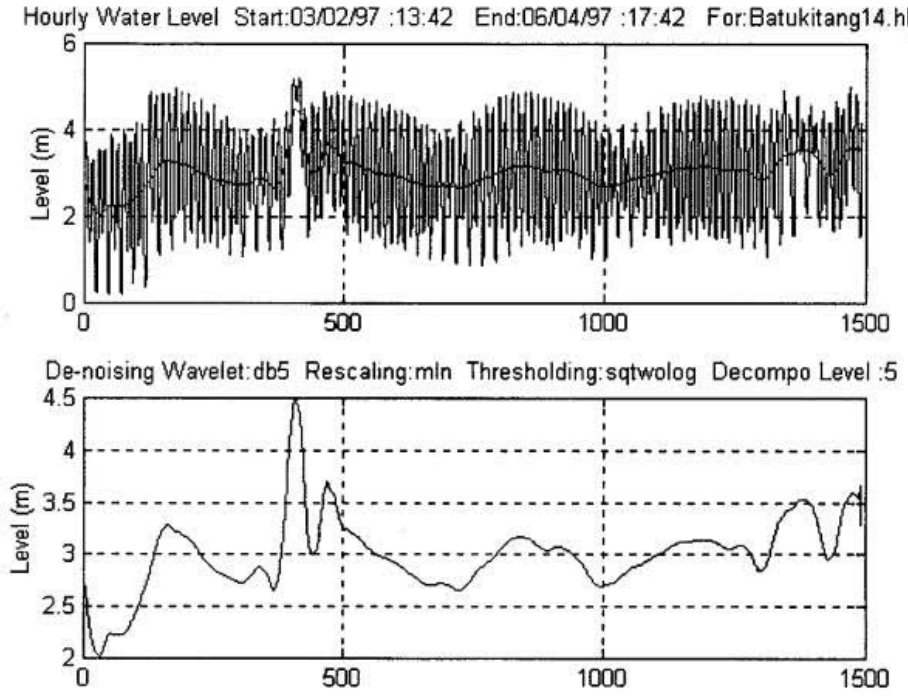


Figure 8-2 Results of Wavelet De-noising Without Distinct Peak

Heuristic inspection, which can be very effective for eliminating some obviously poor de-noising, is used as an initial screening process. The method is akin to the plotting of functional fits in regression model.

Another segment of the same series is shown in Figure 8-2. For this case the tides and river effects are almost the same. The de-noising algorithm is also able to yield a reasonable estimate of the underlying smooth function based on an initial visual inspection.

The wavelet method certainly out-performs other methods of smoothing algorithms such as LOESS, double exponential smoothing, etc. Figure 8-3 show an example of LOESS while Figure 8-4 shows a double exponential smoothing. The LOESS cannot distinguish the peak while the exponential smoothing is fitting a line through most points.

8.4.3 Application of Wavelet De-noising on a Full Series

Siniawan series used in Chapter 7 was also subjected to a wavelet de-noising procedure. The procedure with the best de-noising options, resulted in the de-noised series shown in **Appendix D1**, were established after several exploratory attempts. Owing to the importance of de-noising options selection, the topic is discussed fully in Section 8.7 where sensitivity analyses are also dealt with.

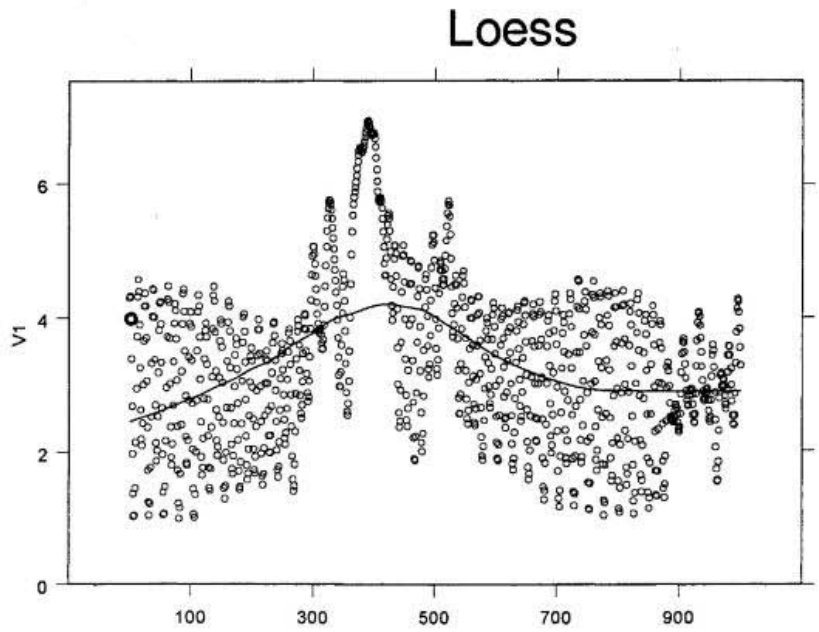


Figure 8-3 LOESS of Batu Kitang Data Segment 18:2000-2999 hr (using SPLUS)

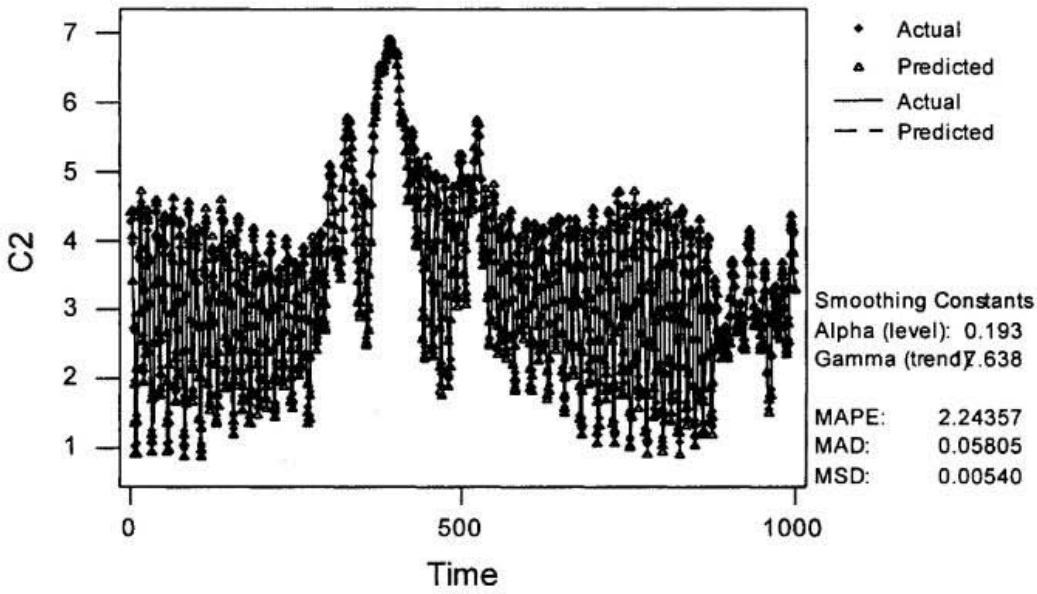


Figure 8-4 Double Exponential Smoothing of Batu Kitang Data Segment (C2) 18:2000-2999 hr (using MINITAB)

The selected options used for the de-noising are as follows:

- wavelet type: Daubechies 4 (Db4)
- decomposition level: 4
- thresholding: soft
- thresholding rule: universal
- scale of noise: separate scale rule for each level (or mln)

8.4.3.1 Multi-Level Decomposition

The ability of wavelet to decompose a signal into multiple levels is unique. Figure 8-5 through Figure 8-8 shows the de-noised series for various levels of decomposition for a segment of the hourly water level series at Siniawan (close to 3 months) using Daubechies 4. A major flood was recorded in the January month.

At Level 3, the de-noised series maintains some significant tidal signatures, especially within the tidal range of about 1 to 6m (see Figure 8-5). Level 6 can be ruled out, as the peaks are not reproduced well enough (see Figure 8-8). Hence among the options available, Level 4 and 5 are the closest in reproducing the peaks.

The peak values shown are all with reference to the time of the peak values found for the segment of de-noised series. It may not correspond to the peak of the original series.

Other types of wavelets and level of decomposition are also investigated. The results are shown in **Appendix D9** for comparison purposes. The same typical segment of data from Siniawan is used in all the cases.

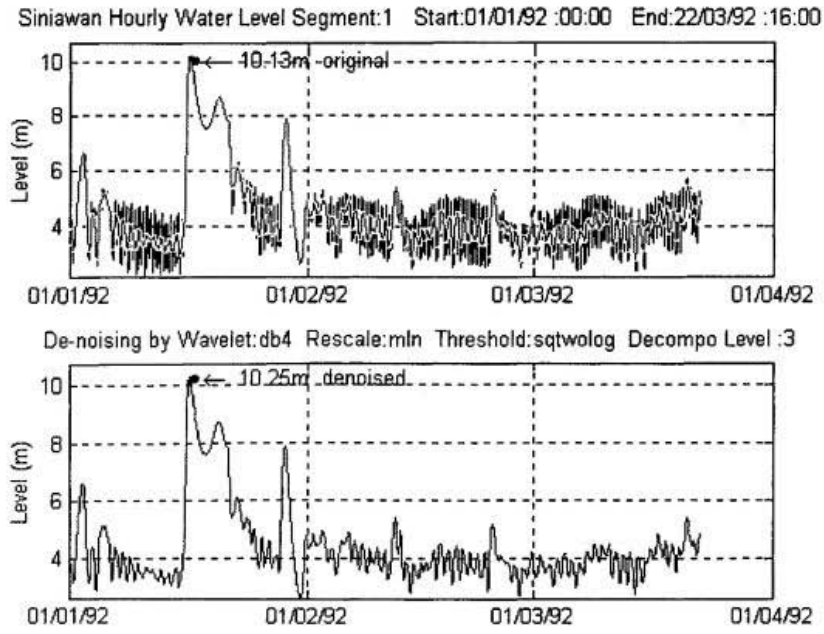


Figure 8-5 Wavelet De-noising Using Decomposition at Level 3, Universal Soft Threshold, and Rescaling by Level-dependent Estimation of Noise

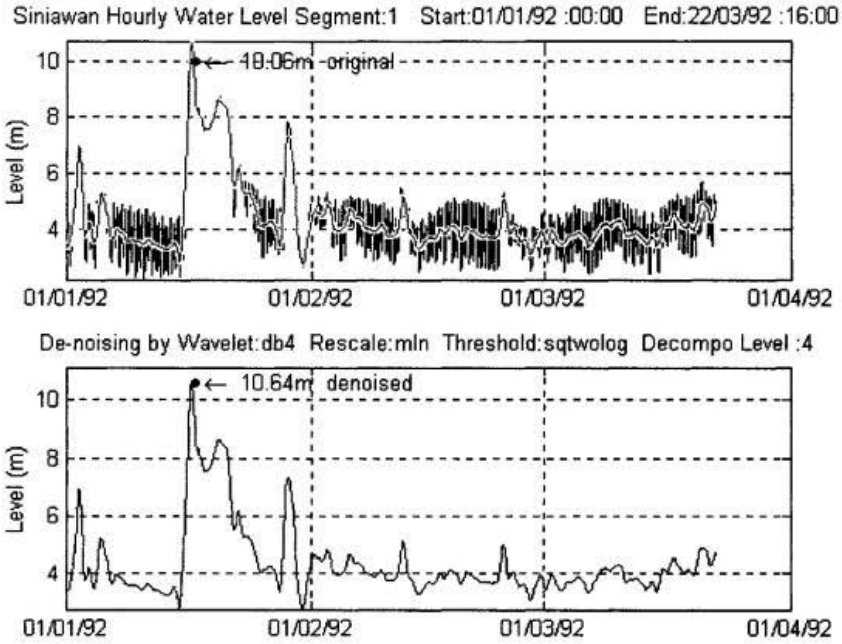


Figure 8-6 Wavelet De-noising Using Decomposition at Level 4, Universal Soft Threshold, and Rescaling by Level-dependent Estimation of Noise

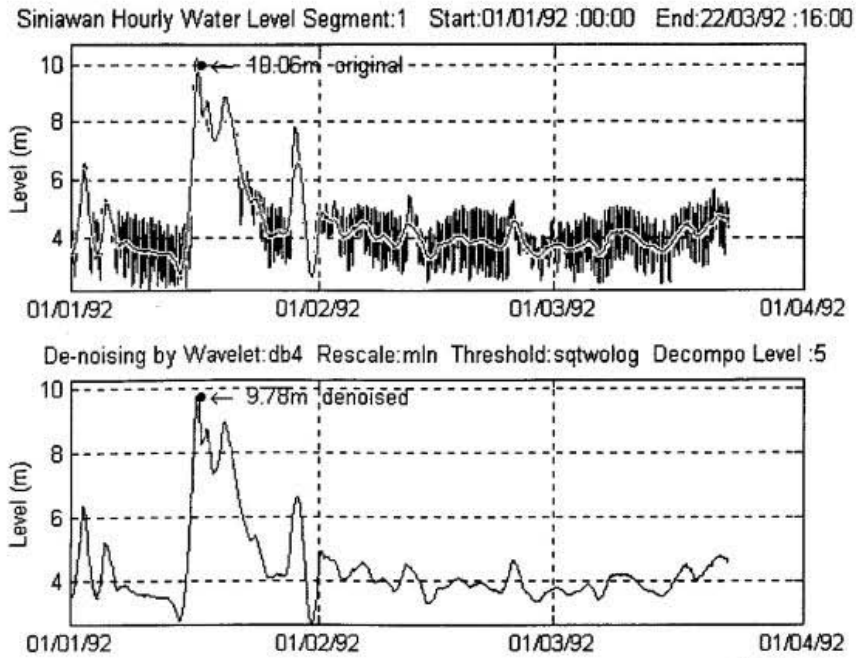


Figure 8-7 Wavelet De-noising Using Decomposition at Level 5, Universal Soft Threshold, and Rescaling by Level-dependent Estimation of Noise

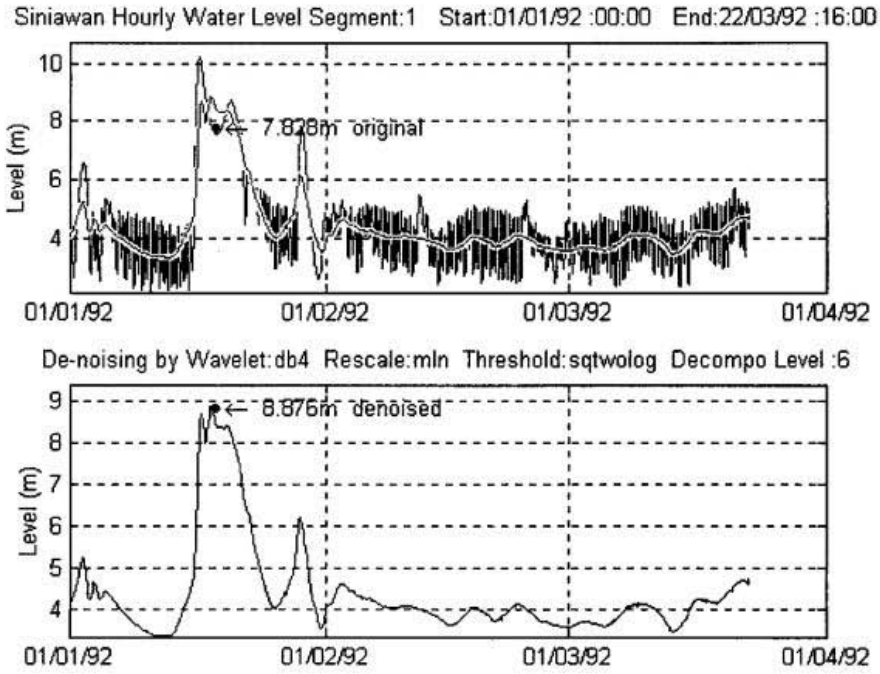


Figure 8-8 Wavelet De-noising Using Decomposition at Level 6, Universal Soft Threshold, and Rescaling by Level-dependent Estimation of Noise

8.5 Distribution of De-noised Series

Once the series has been satisfactorily de-noised, the classical methods of flood frequency analysis as discussed in Chapter 5 are then applied to the series. There are two components to be considered, the river series and the tidal series, and the combinatory effects can be assessed similarly by the joint probability approach as discussed in Chapter 6 and 7. However, there are some differences that warrant some specific treatments. The foremost concern is the varying levels of tidal noise.

Other pertinent issues are the distribution of tidal noise and the maximum range at which tides become ineffective. The distribution of the de-noised river series is treated as described in Chapter 7, considering the hourly dependence aspects while deriving the pdf and cdf from the hourly series.

8.5.1 Tidal Noise Distributions

In Chapters 3 and 7, it was noted that the tidal amplitudes in the sea condition off the coast of Sarawak are distributed almost normally. The tidal noise series T_n has very different characteristic of distribution when compared with the filtered tidal series found using the Fourier method. The interaction effects of tides and river flows have certainly contributed towards a diminishing tidal magnitude when the hydraulic momentum effect of river flow is very prominent as in a high flow or peak discharge event.

8.5.1.1 Statistics of T_n

Some basic statistics about the distribution of T_n is as follows:

Variable	N	Mean	Median	TrMean	StDev	SE Mean
Tn	46780	0.00001	0.00962	0.00265	0.72559	0.00335
Variable	Minimum	Maximum	Q1	Q3		
Tn	-2.06608	2.22576	-0.60844	0.61072		

A box plot in Figure 8-9 shows that T_n has a symmetric distribution with a mean of zero and a quartile range of ± 0.61. Hence it is likely that the noise is similar to the Gaussian white noise.

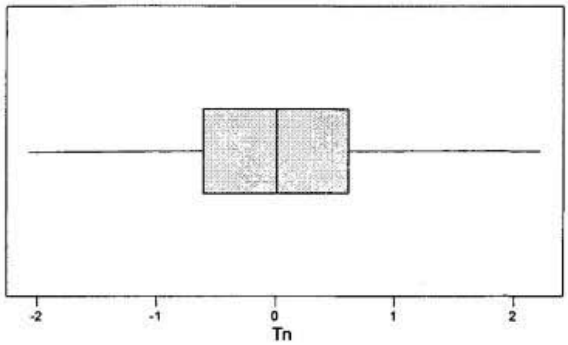


Figure 8-9 Box Plot of T_n

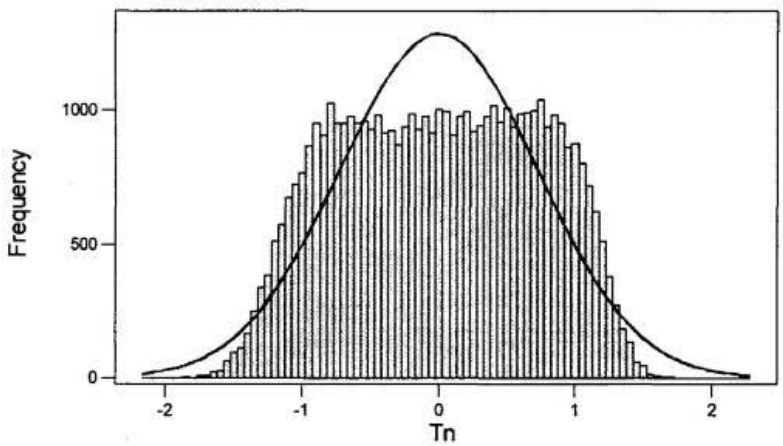


Figure 8-10 Histogram of T_n, with Normal Curve Superimposed

However, a histogram with a normal curve superimposed in Figure 8-10 shows that the overall distribution of T_n is not exactly Gaussian. There still exists a bimodal characteristic typical of the semidiurnal tides.

8.5.1.2 Distribution of Noise with Respect to the De-noised River Series

Figure 8-11 shows a typical distribution of the tidal noise T_n at different levels of the de-noised river series R_d. As expected, the noise series is scattered most around the lower de-noised levels where tidal effects are felt most. The plot also shows high variance at lower de-noised level and a trend of diminishing tidal noise at the upper levels. It is obvious that the excursion into the extreme high level is not as frequent as the lower level, typical of river flows.

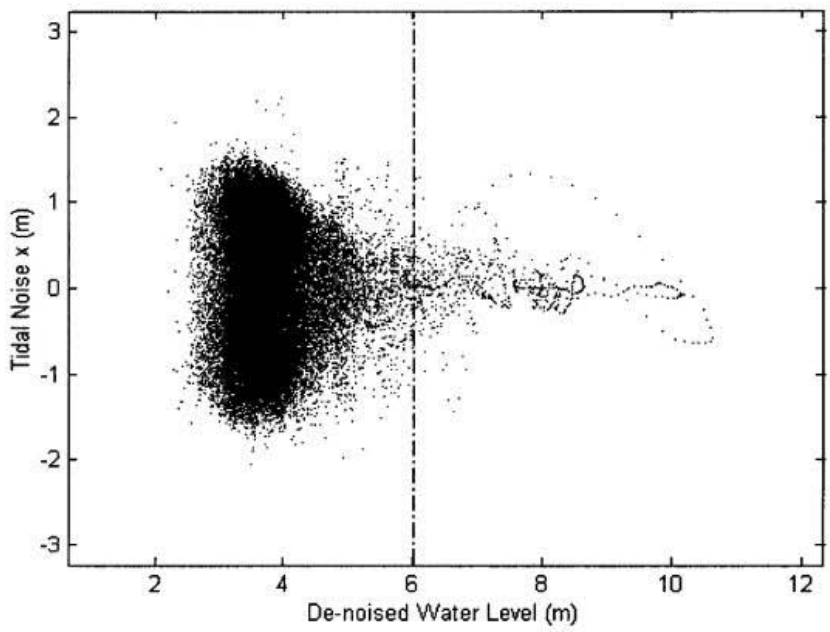


Figure 8-11 Distribution of Tidal Noise With Respect to De-noised Levels at Siniawan

For estimating the magnitude of tidal noise, a simplistic approach whereby a certain maximum level, called tidal threshold H_{max}, was assumed beyond which the tidal effect has diminished. For the example of the de-noised series obtained from Siniawan, H_{max} was estimated at 6.0m. A more elaborate approach was to investigate the noise distribution at each discrete de-noised level. For the example, a discrete interval of 1 m was used for the series, which ranges from 2.3 to 10.6m, making a total of 9 intervals. For each discrete de-noised level interval, a normal distribution was fitted to the associated noise levels. The results of the fits are shown as in **Appendix D3**. It can be seen that at some of the discrete intervals, the normal distribution fits very well. The best fit occurs at around 5 m, the most frequent de-noised level. The fits are poor at the extreme ends where sample sizes are small. A tendency of the tidal noise levels distributed as Gaussian is likely as more points are observed. Figure 8-12 shows the mean of T_n for different discrete de-noised level R_d. In the most frequent range of 3 to 6m, the mean is zero within the 95% confident interval. The highest level of R_d at 11m shows the greatest deviation from zero. Figure 8-13 shows a distinct exponential decrease in standard deviation moving from the low de-noised level to the highest level. The detailed values of the means and standard deviation at each discrete de-noised level are tabulated in Table 8-1.

8.5.2 Distribution of De-noised River Series

An empirical cdf of the de-noised river series R_d can be derived from the distribution of the series at hourly interval. Deriving an appropriate empirical cdf is the prime objective of this section.

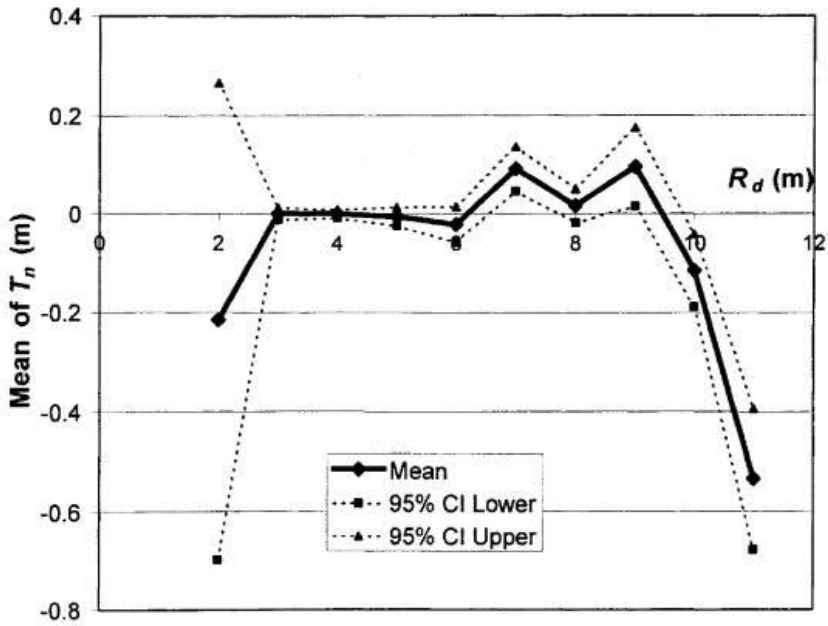


Figure 8-12 Mean Tidal Noise Per Discrete De-noised Level

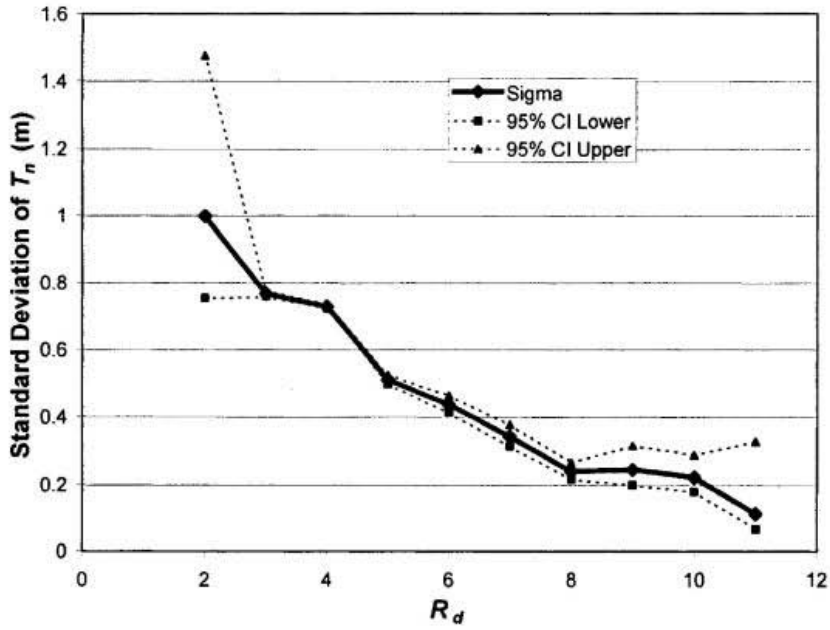


Figure 8-13 Standard Deviation of Tidal Noise (Sigma) at Discrete De-noised Levels

Table 8-1 Mean and Standard Deviation at Various De-noised Level

Rd	Mean	Standard deviation
2	-0.2151	0.999
3	0.000518	0.7675
4	0.000106	0.7295
5	-0.0054	0.5113
6	-0.0215	0.4376
7	0.0919	0.3431
8	0.0168	0.2391
9	0.0961	0.2457
10	-0.1149	0.2223
11	-0.5356	0.1145

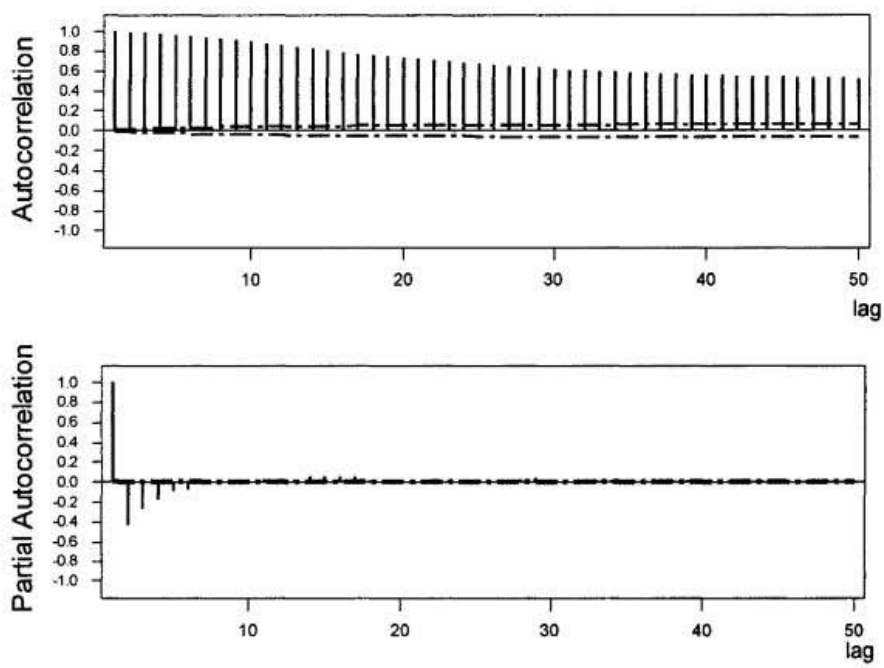


Figure 8-14 ACF and PACF of R^d

8.5.2.1 Time Series Diagnostic

Figure 8-14 shows the ACF and PACF of the de-noised river series. ACF shows that the hourly series are correlated, as it does not decay rapidly. PACF shows there are significant spikes at the first few lags. These are typical features of the hourly river series as noted in Chapter 7.

8.5.2.2 Theta and the *r* largest Storm

The treatment of the dependency in filtered hourly river series R was shown in Chapter 7. A correction factor θ and the number of storms per year were required in an adjustment technique applied to the de-noised river series R_d. The r largest storms were chosen after studying the average number of storms per year shown in **Appendix D4**. A full listing of the r largest storms is shown in **Appendix D5**.

Figure 8-15 shows the plot on which the threshold level for the r largest storms can be selected. The crossing-over of the plot of the number of hourly exceedance and the factor of $m \times u \times \theta^{-1}$ marks the threshold level for the r largest storms, which occurs at about 4.0m in this example. The mean number of storm per year, u, approaches a constant value of 37 at the crossing-over point.

A listing of a program that is used for scanning the de-noised series for deriving the factor of θ^{-1} and the number of storm per year is shown in **Appendix A8**.

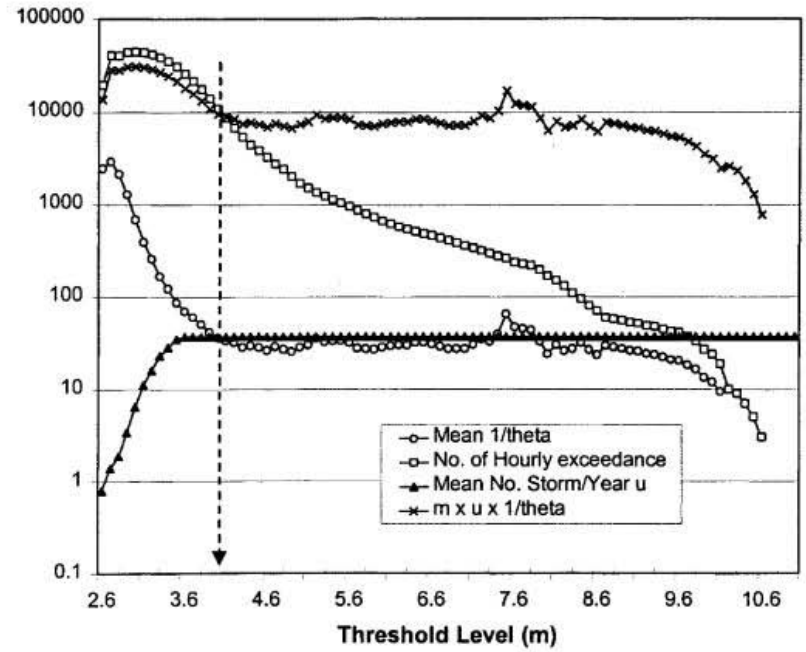


Figure 8-15 Selection of Threshold Level for the *r* Largest Storm

(y axis shows the number of storm)

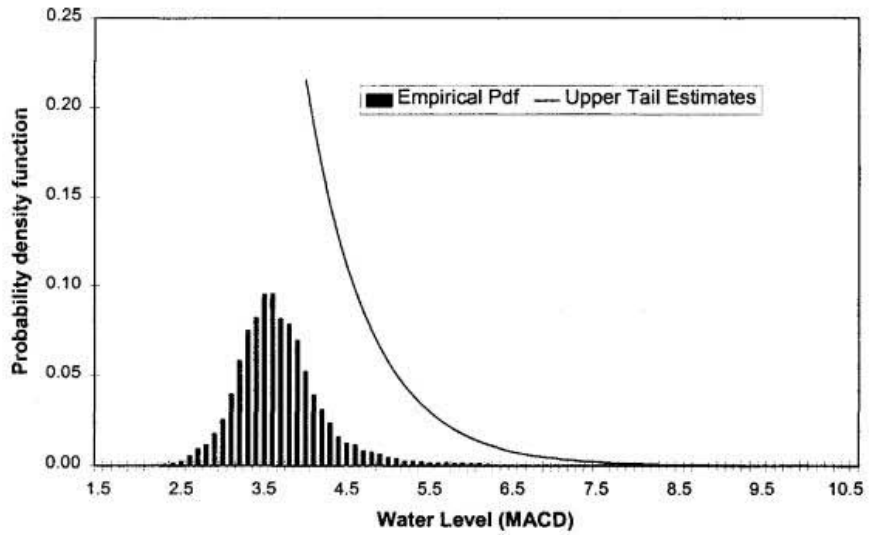


Figure 8-16 Approximation of Upper Tail by Gumbel Distribution

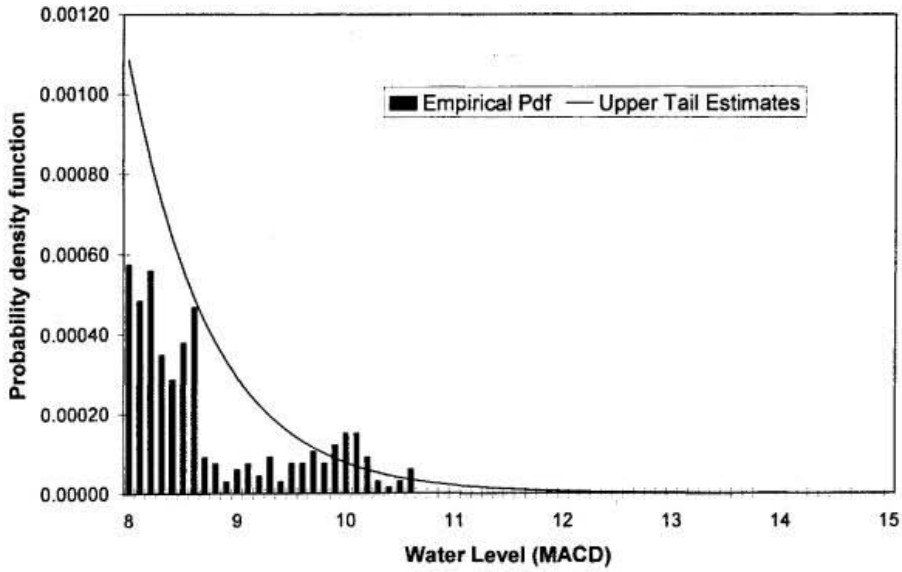


Figure 8-17 Upper Tail Extension in Details

8.5.2.3 Approximation of Upper Tail of the pdf of R_n

An approximation of the upper tail of pdf of R_n based on the fitting a distribution to the r largest storms was shown in Chapter 7. The same technique was used in the derivation of cdf from the de-noised river series. Figure 8-16 and Figure 8-17 show the plot of the approximation of the upper tail using a fitted Gumbel distribution to the r largest storms. A cdf of R_n was derived using this approximation for extending the upper tail beyond the observed range of R_n. The complete tabulation for the pdf and cdf are shown in **Appendix D6**. The tabulation of cdf are used as a lookup table for MC simulations to be described in the next section. The sum of the ordinates of the extended pdf is adjusted marginally to unity. A heuristic judgment was made in limiting the extension of the upper tail. For consistent results, the cdf in all other cases were extended such that the probability of non-exceedance in the tabulation was in the order of at least 1×10^{-7}.

8.6 Extreme Level Frequency Analysis

Once the characterization of the de-noised series associated with river flows R_d and the noise series associated with the tides T_n were completed, the next step was to determine the flood quantiles.

8.6.1 Models for Deriving Flood Quantiles

Two models are deemed possible in characterizing the tidal noise with the objective of deriving flood frequency analysis. Model A assumes an exponential decrease in the tidal noise as the de-noised level increases while Model B assumes that there exists a threshold level H_{max} above which the tidal noise is negligible. The rationale for both models are based partly on the pattern of observed data as seen in Figure 8-13 and Figure 8-13. Model B is based on the priori that tidal dynamic influence will be physically diminished beyond a particular water level elevation. The algorithms applicable to both of the models are presented in the following sections.

8.6.1.1 Model A: Exponential Decrease in Noise Level

1. Compute the means and variances for tidal noise at each discrete interval of the de-noised level R_d. It is assumed that the noise is normally distributed at each of the de-noised level with varying means and variances (and standard deviations).
2. Derive the pdf and cdf estimates of the de-noised levels from the observed hourly series. Adjustments are made to the hourly dependence effects.

3. Perform MC simulations for n years with the following steps:
    a. Randomly generate a random number between 0 and 1 for each simulated hour.
    b. Search the table of cdf of the de-noised level R_d and obtain a generated value for the hour.
    c. Search for the mean μ and standard deviation σ at the discrete R_d level by using the lookup table of mean and standard deviation derived in step 1.
    d. Generate a noise level using a normal distribution with the selected μ and σ.
    e. Obtain the i^{th} hourly simulation Z_i by the addition of the noise level obtained in step (d) with the de-noised level obtained at step (b).
    f. Repeat step (a) to (e) until $i = 8766 \times n$ are accomplished.

4. Sort the generated hourly data and derived the flood quantiles based on the empirical cdf of Z.

The above algorithm is represented in a flow chart as shown in Figure 8-18. A MATHLAB program that uses the algorithm is shown in **Appendix 9**.

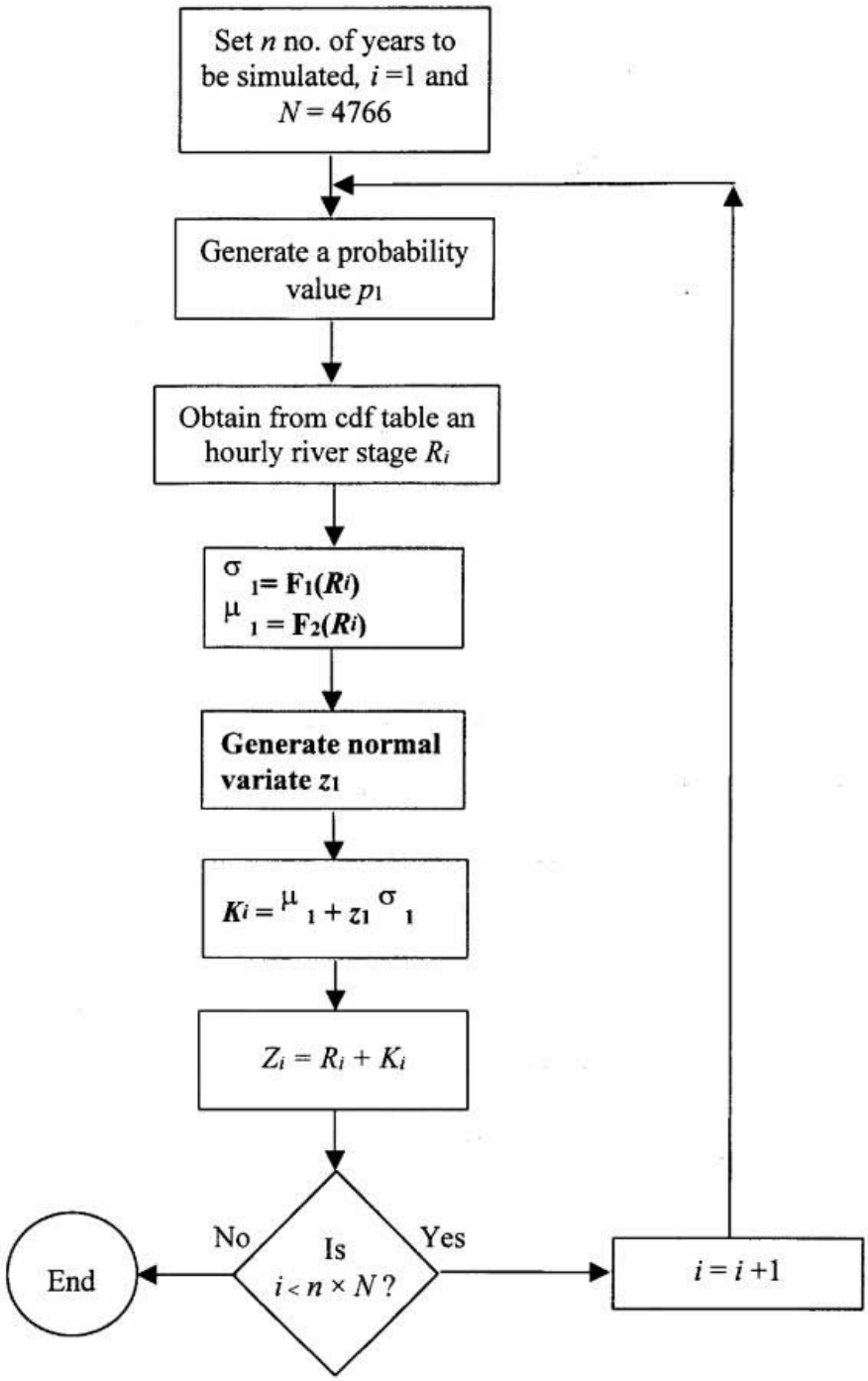


Figure 8-18 Flow Chart for MC in Model A

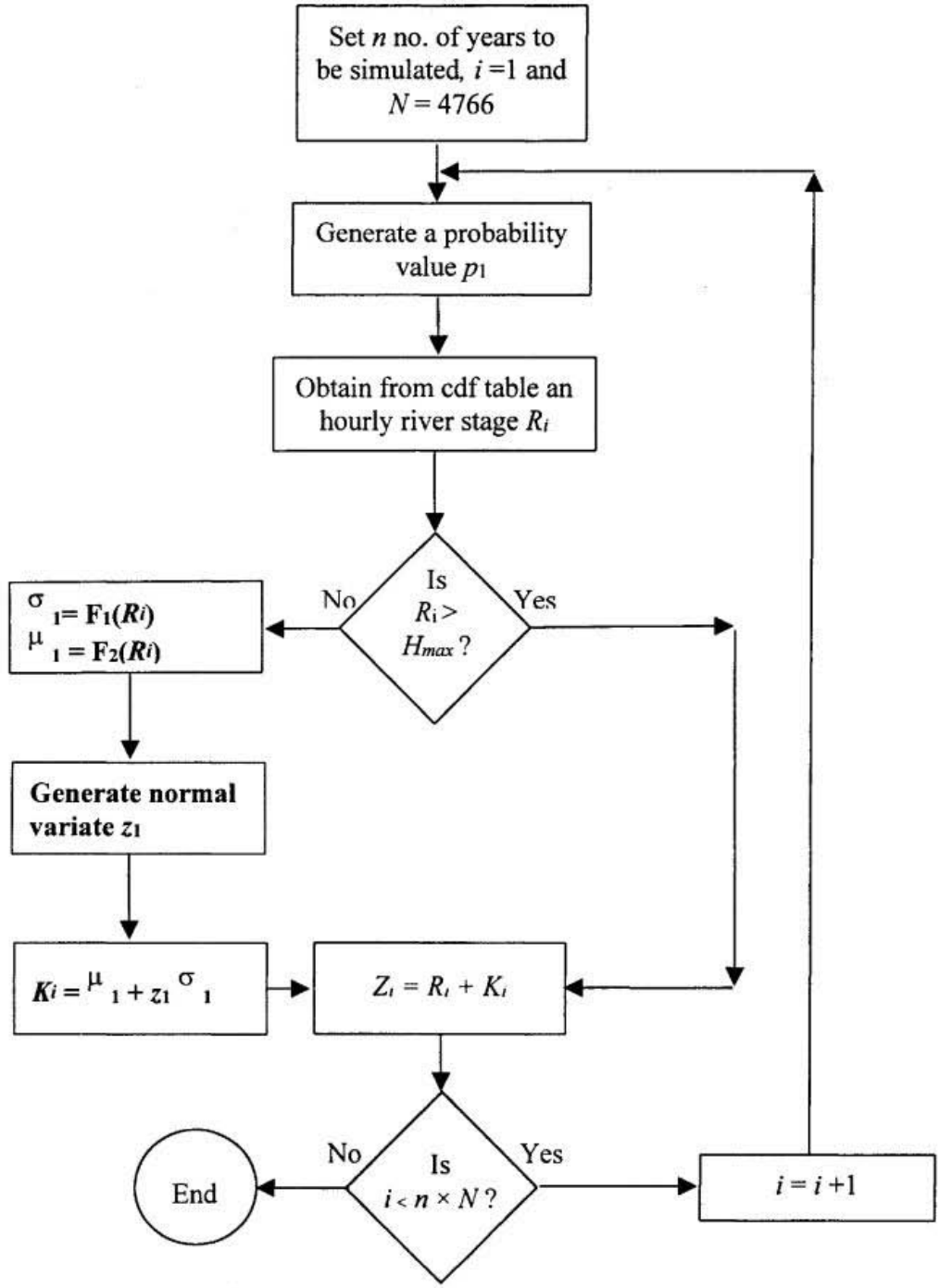


Figure 8-19 Flow Chart for MC in Model B(2)

8.6.1.2 Model B: A Threshold Level H_{max} Exists

This model can be applied in two alternative ways:

Model B(1): Using the model to derive an annual maximum series and apply traditional flood frequency analysis

Model B(2): Using the model in a MC simulation to derive an hourly sequence for many years and estimate flood quantiles using the simulated empirical cdf.

For Model B (1), the following algorithm is proposed:

1. Estimate H_{max} the maximum tidal amplitude based on the probability density function of the tidal noise or scatter plot of the tidal noise.
2. Search the de-noised series R_d for an annual peaks H_p and check with H_{max}.
3. If $H_p > H_{max}$, the peak is accepted as the extreme level for the year since the interaction effect is weak and negligible.
4. If $H_p < H_{max}$, the interaction effect is significant. Obtain the corresponding tidal amplitude to be added from the noise series.
5. Abstract the annual maximum tidal amplitude A_m corresponding to H_p level.
6. Perform traditional flood frequency analysis on the adjusted peak series:
    a. Fit a distribution to the annual maximum series of H_p.
    b. Fit a distribution to the annual tidal maximum amplitude A_m.
    c. Perform a joint probability computation by using MC simulation making use of the fitted distributions. Annual results for a large number of years are hence derived.

d. Derive flood quantiles from the simulated annual peak series.

For the Model B (2), an algorithm similar to the Model A except that H_{max} is used. The proposed algorithm is as follows:

1. Estimate H_{max} the maximum tidal amplitude based on the probability density function of the tidal noise or scatter plot of the tidal noise.
2. Perform computation of mean and standard deviation for tidal noise at discrete interval of the de-noised level R_d below H_{max}. It is assume that the noise is normally distributed at each de-noised level with varying values of mean μ and standard deviation σ.
3. Derive pdf and cdf estimates of R_d from observed hourly data. Adjustments are made to the hourly dependence effects.
4. Perform MC simulation for n years with the following steps:
    a. Generate a random number p_1 between 0 and 1 as the probability of non-exceedance for each simulated i^{th} hour.
    b. Search the cdf of the de-noised level for a level that correspond to the probability value p_1.
    c. If Rd > Hmax, assign $Z_i = R_d$ and skip to step (g). otherwise generate another random number p_2 between 0 and 1.
    d. Search for the mean and standard deviation at the discrete R_d level using the lookup table of mean and standard deviation derived in step (2).

e. Generate a noise level using a normal distribution with the selected mean and standard deviation.

f. Obtain the i^{th} hourly simulation Z_i by the addition of the noise level obtained in step (e) with the de-noised level obtained at step (b).

g. Repeat step (a) to (f) until $i = 8766 \times n$ are accomplished.

5. Sort the generated hourly data and derived the flood quantiles based on the empirical cdf of Z.

Between the two approaches applicable, the latter method of Model B(2) involving MC simulation is preferred as it is suitable for short observed records. The method of Model B(1) is only appropriate when data series are of considerable length, often recommended to be more than 18.6 years for any analysis of tidal series (Pugh, 1987). Model B(2) is referred later in this Chapter simply as Model B. A MATLAB program that uses Model B is shown in **Appendix A10**.

8.6.2 Application of Flood Frequency Analysis

Both the de-noised series of Sinawan water level series and the tidal noise series were subjected to the techniques of flood frequency analysis developed in Section 8.6.1. The results of characterizing the distribution of the series in Section 8.5 were used. Various models of generating flood series using MC were examined using the cdf's of R_d, the means and variances of T_n, and a threshold level for the non-tidal range H_{max}. Table 8-2 shows the tabulation of the flood quantiles obtained, using Model A and B. Model B gives slightly higher flood quantiles as compared with Model A. Ratios of the flood

quantile levels to the maximum recorded level in a six year period is also shown in the table. The ratios are not high which are typical characteristics of floods found in the tidal zone.

Table 8-2 Flood Quantiles L_T at Siniawan Derived by the Wavelet Method

Return Period (Years)	L_T Using Model A (mean, variance) (m)	L_T Using Model B (Hmax=6m) (m)	L_T/L_{max} Model A	L_T/L_{max} Model B
2	11.04	11.54	1.08	1.13
5	11.90	12.41	1.16	1.21
10	12.57	12.93	1.23	1.26
50	13.91	14.06	1.36	1.37
100	14.17	14.30	1.38	1.40
200	14.26	14.40	1.39	1.41

Note:

1. Option used in wavelet de-noising of the series was by Daubechies 4 at decomposition level 4, soft thresholding using universal rule and scale factors for noise at different level.
2. L_{max} was the maximum observed level, at 10.24m, of the Siniawan series with a record length of 6 years.

8.7 Sensitivity Analysis of De-noising Factors

One of the main drawbacks in any mathematical simulation that produces results with no direct physical observations for comparison, is the validation process. And this is the case for comparing the results of the estimated flood quantiles derived by using simulation methods. In fact, even the traditional flood frequency analysis in the non-tidal

zone using statistical and probabilistic inferences, too, suffers from the same problem. To alleviate this drawback, a sensitivity analysis is performed to check the major factors that influence the outcomes. All the factors discussed in Section 8.3.1 concerning wavelet de-noising factors influence the tidal noise. Factors that are considered in de-noising process are the type of wavelet, the shrinkage function, the thresholding options, and the scale factor on noise, among others. The sensitivity of the results to these factors was therefore assessed and described in the following sub-sections.

8.7.1 Type of Wavelet and Decomposition Levels

The type of wavelet and the decomposition level are the most important factors found in the initial graphical inspections in Section 8.4. The type of wavelet has been limited to the Daubechies family of wavelets as these are the type suitable for studying responsive series such as tides. Table 8-4 shows all the basic statistics of various de-noised series S_d obtained by using different options of wavelet type and decomposition levels which are in the suitable range as found in the initial plots in Section 8.4. The statistics of the observed series at Siniawan are also listed for comparison purposes.

All the plots of the statistics, with the various types of wavelet used, are shown in **Appendix D7**. The most significant statistic affected by the type of wavelet used is the range, maximum and minimum. The maximum statistics is the most important factor for flood estimation purposes. The plot of maximum versus various type of wavelet used is highlighted and shown as Figure 8-20. It is observed that the option at level 4 produces the highest possible maximum level while moving from level 3 to 6, with the exception of

wavelet Daubechies Type 5. The other plots shown in **Appendix D7** also show that the means, variances and inter quartile ranges are not sensitive to the type of wavelet being used. However, these later statistics may not be meaningful as the number of sampling points is large, at 46780. The scatter plots of tidal noise are shown in **Appendix D8**.

Table 8-3 Statistics of Various De-noised Series Obtained by Different Options as Compared with the Observed Series at Siniawan

Wav.	L	Mean	Var.	Q1	Median	Q3	IQR	Mini.	Maxi.	Range	Skew	Kurt
Observed series		3.7789	0.963	3.039	3.780	4.441	1.403	0.98	10.24	9.26	0.78	2.75
Db3	3	3.7787	0.544	3.351	3.699	4.080	0.730	1.39	10.33	8.93	2.28	11.92
Db3	4	3.7789	0.441	3.410	3.669	3.982	0.571	2.08	10.86	8.78	3.22	18.76
Db3	5	3.7792	0.407	3.426	3.664	3.979	0.553	2.26	10.75	8.50	3.27	19.48
Db3	6	3.7795	0.364	3.434	3.690	3.980	0.545	2.56	9.96	7.40	3.23	19.76
Db4	3	3.7789	0.531	3.358	3.698	4.074	0.715	1.79	10.25	8.46	2.39	12.60
Db4	4	3.7789	0.437	3.409	3.664	3.971	0.562	2.10	10.64	8.54	3.25	18.91
Db4	5	3.7789	0.405	3.423	3.674	3.969	0.547	2.30	9.78	7.48	3.22	18.70
Db4	6	3.7799	0.358	3.436	3.689	3.991	0.556	2.55	8.88	6.32	3.08	17.47
Db5	3	3.7788	0.523	3.362	3.695	4.060	0.698	1.32	10.42	9.10	2.46	13.02
Db5	4	3.7787	0.437	3.410	3.661	3.975	0.565	1.68	10.06	8.38	3.25	18.67
Db5	5	3.7792	0.405	3.424	3.671	3.980	0.556	1.90	10.44	8.54	3.30	20.11
Db5	6	3.7798	0.362	3.438	3.684	3.994	0.556	2.13	9.43	7.31	3.09	17.52
Db6	3	3.7788	0.516	3.367	3.692	4.051	0.684	1.67	10.13	8.46	2.53	13.46
Db6	4	3.7789	0.437	3.411	3.659	3.977	0.566	1.76	10.77	9.01	3.29	19.26
Db6	5	3.7790	0.408	3.420	3.665	3.979	0.559	2.00	10.48	8.48	3.29	20.05
Db6	6	3.7789	0.364	3.434	3.691	3.982	0.548	2.02	9.77	7.75	3.20	19.49

Note: Sample size N=46780, Wav.= Type of Wavelet, L= Decomposition Level

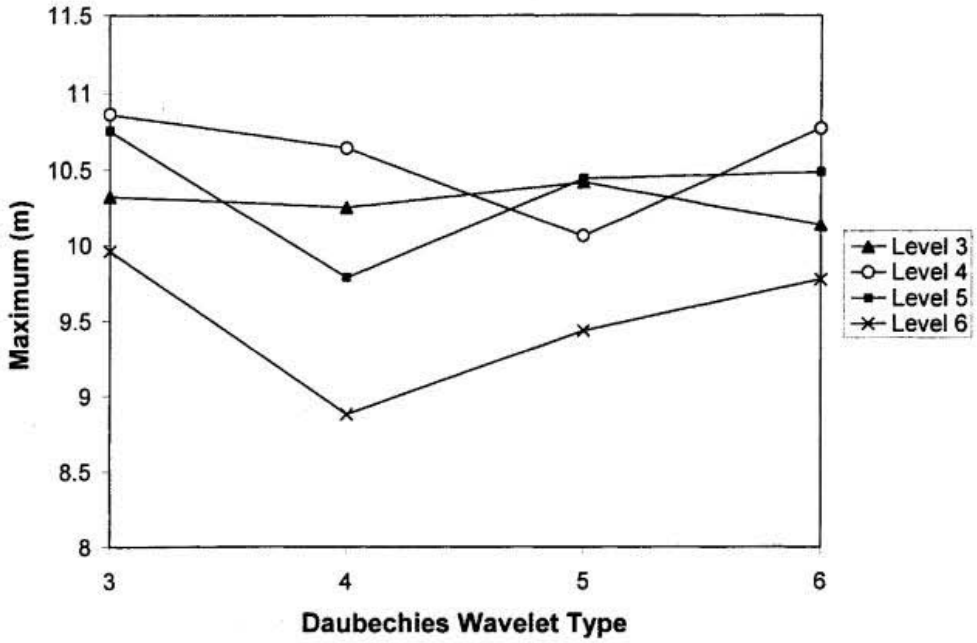


Figure 8-20 Maximum Statistics of De-noised Siniawan Level Series by Different Types of Wavelet and Decomposition Levels

8.7.2 Flood Quantiles

The foregoing section of looking at the basic statistics of de-noised series may not yield much insight as far as the flood estimation methods are concerned. The direct approach of looking at the flood quantiles is more meaningful, although more elaborate efforts are required. Beside the type of wavelet and decomposition level factors, two modelling methods for deriving the flood quantiles from the two separated series of S_d and T_n are also considered.

From the observation in Section 8.4, Daubechies 4 and 5 are the closest contenders in de-noising the water level series in TIZ. Decomposition at level 3, 4 and 5

are amongst the best choices. Investigation on the effects of these factors is summarized in Table 8-4. For each option, a MC simulation of hourly data up to 1000 years are carried out, which takes approximately 14 hours to run on a computer equipped with a CPU of Pentium II 450 MHz and RAM of 192MB.

Table 8-4 Comparision of Flood Quantiles Obtained by Various Wavelet and Frequency Modelling Options

Option	D4L4 A	D4L4 B	D4L3 A	D4L3 B	D3L4 A	D3L4 B	D5L4 A	D5L4 B	D5L5 A	D5L5 B	D4L5 A	D4L5 B
Dau	4	4	4	4	3	3	5	5	5	5	4	4
Level	4	4	3	3	4	4	4	4	5	5	5	5
Model	A	B	A	B	A	B	A	B	A	B	A	B
T=2	11.04	11.54	10.16	10.10	10.37	10.80	10.75	10.72	11.30	11.17	11.51	11.27
T=5	11.90	12.41	10.27	10.26	10.92	11.45	11.49	11.49	12.08	11.87	12.27	12.11
T=10	12.57	12.93	10.34	10.29	11.92	11.92	11.97	11.97	12.59	12.53	12.76	12.63
T=50	13.91	14.06	10.85	10.85	13.01	13.01	13.10	13.10	13.75	13.53	13.89	13.82
T=100	14.17	14.30	11.05	11.05	13.29	13.29	13.39	13.39	14.02	13.86	14.24	14.13
T=200	14.26	14.40	11.15	11.15	13.41	13.41	13.52	13.52	14.13	14.10	14.37	14.26

Note: Dau = Daubechies Wavelet type, Level = Decomposition level, Model = Frequency model, Model B with Hmax=6m, T =Return period

A quick comparison of the flood quantiles shows that the options using Daubechies 4 wavelets with decomposition up to level 4 and 5 give the highest values while the option of Daubechies 4 with decomposition level 3 gives the lowest. This can be seen clearly as in Figure 8-21and Figure 8-22. The option Daubechies 3 with level 4 is

in between. It is also observed that the decomposition level plays a greater role as compared to the type of wavelet. The choice of Model A or B is not important except when Daubechies 4 with decomposition levels of 4 and 5 are used.(see Figure 8-23). Overall, the option of Daubechies 4 with decomposition level of 4 gives the most conservative results. In flood frequency analysis, there is often an affinity to opt for the methods that give more conservative results. Without exception, the Daubechies 4 with decomposition level 4 is the preferred wavelet de-noising option to be used for flood frequency analysis in the tidal interaction zone.

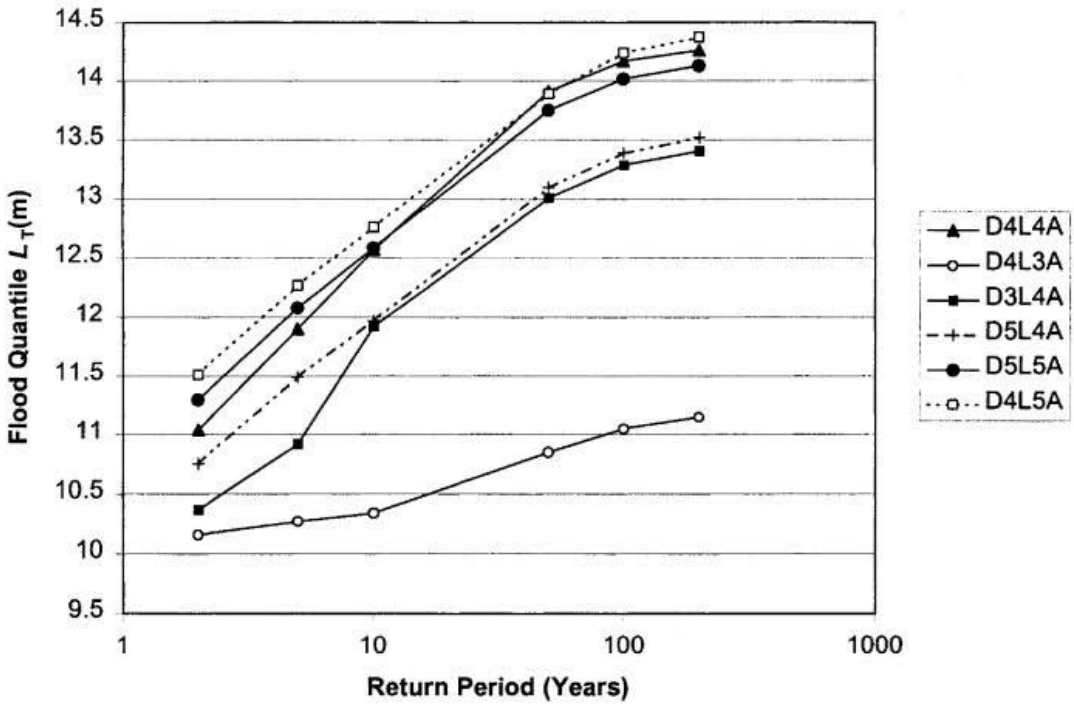


Figure 8-21 Flood Quantiles for Various Options Using Simulation Model A

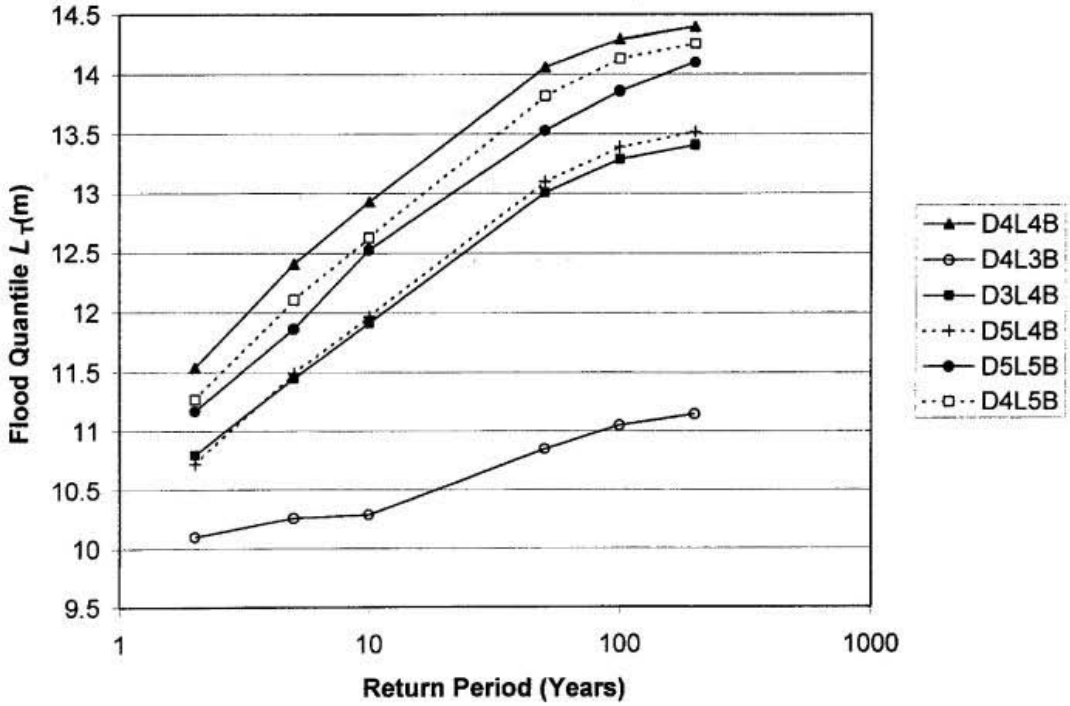


Figure 8-22 Flood Quantiles for Various Options Using Simulation Model B

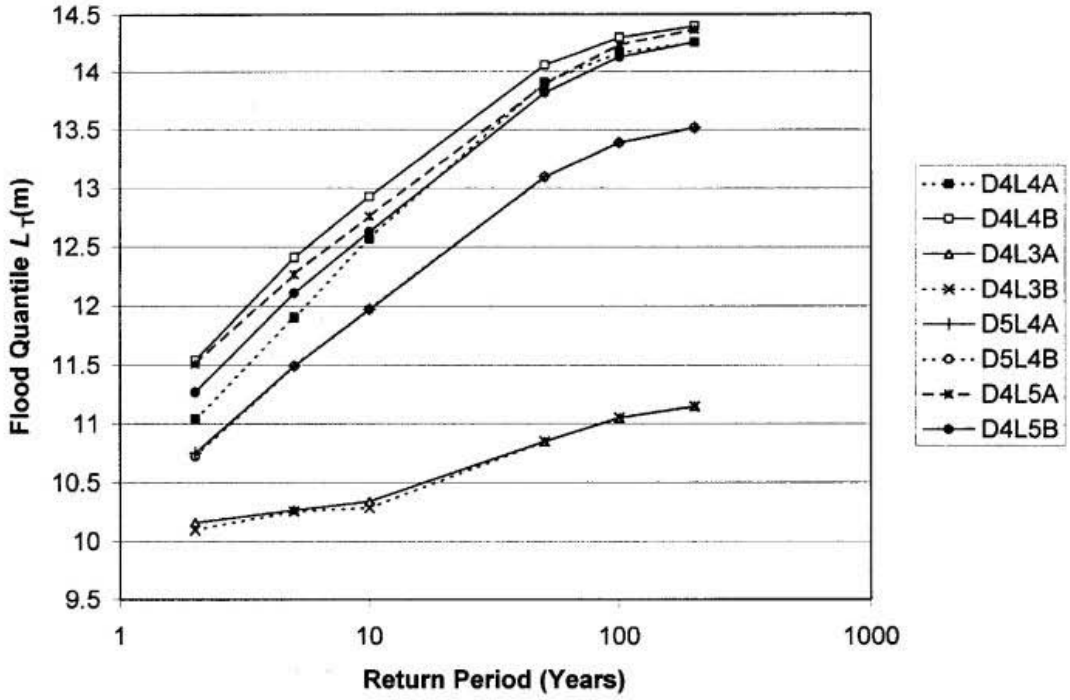


Figure 8-23 Comparing Model A and Model B by Selected Options

8.8 Chapter Summary

A new approach in for estimating design floods in a TIZ has been presented. It is based on a nonparametric wavelet de-noising method. Initial inspection of the plots of de-noised series using various wavelet types, thresholding methods and options in de-noising has pointed to the Daubechies 4 decomposed up to the fourth level as the most suitable combination in solving the problem of separating tidal signatures from river flow series. Subsequent treatments of the de-noised series to derive flood quantiles were also developed in close association with the appropriate characterization of the noise and de-noised series. Intensive computation was used in the MC simulations. Sensitivity analyses were performed giving some insights into the effects of using various major wavelet de-noising options. The flood quantiles derived by the most suitable wavelet de-noising options were compared. Daubechies 4 with decomposition at level 4 gives the most conservative results.

Chapter 9 Summary

This thesis has reviewed and analyzed six different methods that can be used to estimate design floods in tidal interaction zones (TIZ). In addition, subsidiary issues related to the study area such as the regional flood frequency analysis for the non-tidal zones, and the characterization of tidal phenomena in the estuaries or open sea areas are also examined. This chapter summarizes and consolidates the major findings of the thesis. The final conclusions and recommendations are presented in the next chapter.

9.1 Design Flood Estimation Methods in TIZ

Design flood estimation methods for use in tidal interaction zones are assessed and presented. Some of the methods are modified from earlier works while others are newly developed in this thesis. These methods are

1. Hydrodynamic Modelling Method (Reviewed herein)
2. Direct Frequency Analysis (Reviewed herein)
3. Rating Curve Method (Reviewed herein)
4. Joint probability Method (Reviewed herein)
5. Fourier Method (Further developed herein)
6. Wavelet De-noising Method (New method developed herein)

9.2 Hydrodynamic Modelling Method

The Hydrodynamic Modelling Method is one of the major methods used in many major river engineering projects. It centres on the use of the St. Venant equation to describe the hydrodynamic processes, which are modelled progressively at a certain fixed time step. In cases where the flood flows are confined in the main channel, and lateral flows are negligible, one-dimensional modelling is sufficient. Two-dimensional modelling is necessary for wide flood plains. Even though it is a deterministic approach which requires a lot of resources such as in data collection, surveying, data processing and analysis, the results obtained are normally attached with some probabilistic assumptions due to the many combined cases of input factors. The method can be very efficient when flood simulation on a basin-wide scale is demanded. On the contrary, if the quantile of an isolated location is required, especially for many small projects, this method is cumbersome, and it is often beyond the resources of the project. For this study, this method is not further elaborated, as there are extensive references to the subject in the literature.

9.3 Direct Frequency Method

It has been shown that the Direct Frequency Method can be effective when applied in the upper reaches and as well as out in the open seas, where the flow conditions or levels are purely influenced by a single phenomenon. Annual maximal flood series are used as in a traditional flood frequency analysis, which involves the fitting of a suitable probability distribution. The application of this method in the TIZ is not recommended

since the basis of the assumption is not well founded. The method is suitable for a single process and is not appropriate for a two-process joint event as in the TIZ, which involves meteorological forcing events in the river and the astronomical forcing events in the open sea. However, for a quick baseline check, this method does provide some initial answers.

9.4 Rating Curve Method

The Rating Curve Method involves fitting regression lines to relate the discharge, Q, at the upstream inflow locations with the sea levels, H, in the estuaries or the open sea. Superficially this method may sound simple and good for a quick estimation, however, it suffers from very poor correlations between H and Q. The large variance may be due to unaccounted factors, such as the tributary effects, time synchronization problem and lag effects. If it is properly established, the results can be superseded by the Joint probability Method or JPM. The simple relationship makes it useful in simple real time flood forecasting purposes. Although it is not recommended for any serious design application, the method may serve its purpose for initial checking of quantile estimations.

9.5 Bivariate Distribution Method or JPM

The Bivariate Distribution or Joint Probability Method is good for the estuaries, where the effects of river flow is minimum. It relies heavily on the availability of data at three locations: upstream, downstream, and the TIZ. By using the pdf's of the distributions of discharges at the upstream inflow location and water levels at the estuaries or at the open sea, the probability of non-exceedance associated with the

discharges and water levels at the boundaries can be estimated. By using further information obtained elsewhere, such as the H-Q relationship from the rating curve method, a joint probability distribution can be computed for the TIZ by using the marginal distributions. Theoretically this method is well defined, but in practice it still depends on the correlation of H-Q to define the final joint distribution. The correlation is usually ill-established as discussed earlier.

9.6 Fourier Method

The Fourier Method is adapted and developed further in this study for the TIZ. It is based on the knowledge of the characteristics of tides which have higher frequencies than the river flows. A low pass filter in the Fourier domain is used to separate the slow frequency components from the higher frequencies. A newly developed computation procedure is used to derive the flood quantiles from the filtered series involving the use of a correction factor to adjust for hourly data. By adopting a set of Monte Carlo procedure, an ensemble of annual peaks can be simulated using an extended cdf of the filtered river flow series. Although it is less precise about the interaction effects of the tides and fluvial discharges by assuming the superposition of the tides and river series, the method has its merits in terms of the principle of separating the tidal from the river flow series. It also has the advantage that it can be applied directly to an isolated location of interest. This method lays the foundation for the next method using wavelets since the techniques used for flood quantile determination are adopted in the wavelet method.

9.7 Wavelet De-noising Method

The wavelet De-noising Method is a completely new approach for design flood estimation in the TIZ. The water level series is subjected to a set of decomposition techniques, which effectively filter out the tidal influence as noise. The de-noising technique consistently detects the time evolution of the frequency content of data at all time scales. It separates tidal and river flows, similar to the Fourier Method, but the main difference is that it uses a wavelet, which has the ability to shrink or expand, as opposed to the fixed sine or cosine function available in Fourier domain. The distribution of both two series is derivable after some adjustment for the correlation effects of hourly data.

Two models are developed to extract flood quantiles using MC simulations. One of the advantages of this method is that it is applied to the point of interest and does not require great efforts in gathering vast amount of data and information. The method is promising because of the superiority of wavelets in handling non-stationary tidal effects at the TIZ.

9.8 Comparison of the Methods

After the detailed assessment of the various methods in Chapter 5 through 8, it is now possible to compare several of these using the same water level series at Siniawan, in the TIZ of Sarawak, Malaysia. The comparison of the results using several methods considers other factors beyond just the flood quantiles. This is because in a flood estimation exercise one may never know the true flood quantiles exactly, especially when the data series available is short.

9.8.1 Comparison in Terms of Flood Quantiles

A straight comparison of various methods by the flood quantiles estimated using various methods is possible, although some choices have to be made pertaining to the selection of the best option available within each method. Table 9-1 shows all the flood quantiles at Siniawan derived using various methods. A plot of the results is shown in Figure 9-1. Two historical flood levels are obtained by level surveying after the floods in 1963 and 1976. The frequency analysis on the rainstorms that caused the extreme floods showed that the return periods for the January 1963 and January 1976 rainfalls are of the order of 100- and 25- year respectively. The observed flood peak levels at Siniawan in these two floods are also plotted in Figure 9-1. It can be seen that the wavelet method produces flood quantiles within close range of these great floods. The method of Wavelet (D4L4*) gives the most conservative results. D4L3 and D4L4 refer to the case of wavelet de-noising using the Daubechies wavelet Type 4 with the decomposition level set at 3 and 4 respectively, as described in Section 8.7.1. The Model B (denoted by *) is used in obtaining the flood quantiles from the de-noised series. It is interesting to note that the Fourier Method gives the lowest flood quantiles and hence it can be used as a check on the lower limits of the quantile estimates for other methods. Overall, the Wavelet Method appears to perform well in flood quantile derivation, although some further fine tuning may be required.

Table 9-1 Flood Quantiles at Siniawan Derived Using Various Methods

Method	T=2	T=10	T=50	T=100	T=200
Direct Frequency	8.49	10.21	10.84	10.99	11.10
Fourier (Annual)	9.14	9.63	10.46	10.76	11.05
Fourier (Hourly)	8.87	9.18	9.78	10.02	10.33
Wavelet (D4L3*)	10.16	10.34	10.85	11.05	11.15
Wavelet (D4L4*)	11.54	12.93	14.06	14.30	14.40
Historical Flood Caused by Return Period Rainstorms					
Flood of 1963 ($T \approx 100$)	13.71				
Flood of 1976 ($T \approx 25$)	11.32				

Note: (1)Flood quantiles are presented as the water levels in MACD. (2) * refers to Model B. (3) The return periods T (years) of historical floods are estimated by the frequency analysis of rainfall.

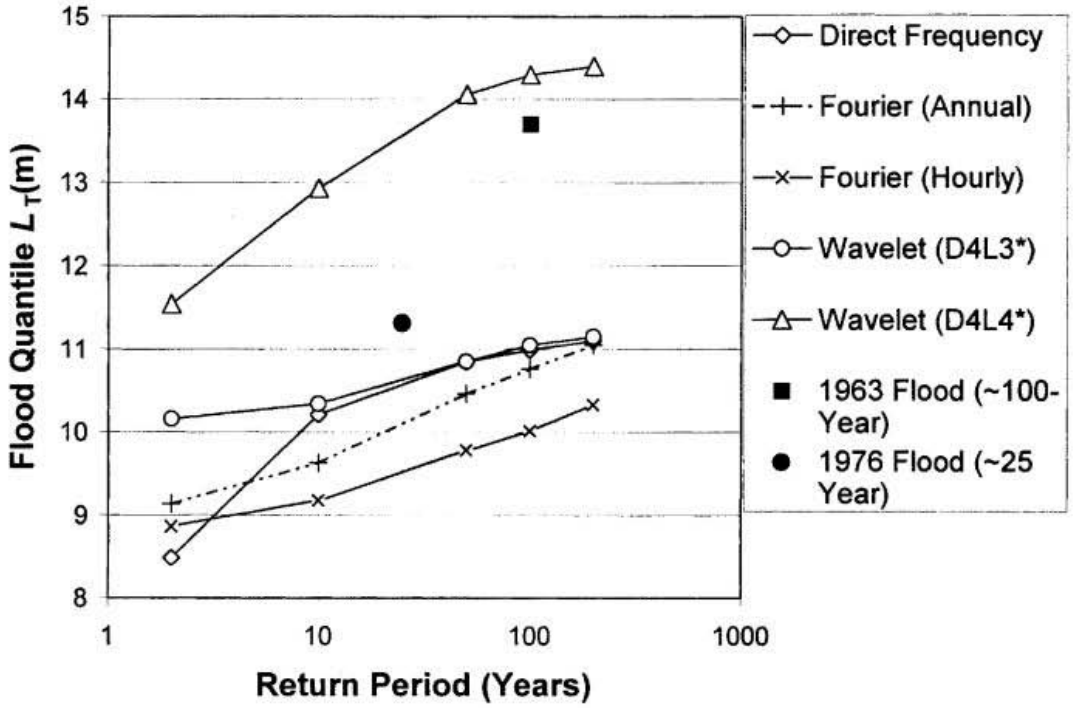


Figure 9-1 Plot of Flood Quantiles at Siniawan (Water Level MACD) Derived Using Various Methods

9.8.2 *Comparison in terms of Effort Required*

The effort required for a flood estimation exercise can be broadly defined as the time and resources required in deriving the best possible estimates. These include the processes involved in gathering of all the available information and data to satisfy the needs of the method. However, it does not include conventional data gathering over a long-term basis, as it is normally the jurisdiction of a particular agency or a government department to conduct that over a long-term basis. The broad comparison on the data and information requirement for various methods can be seen in Table 9-2.

Table 9-2 Comparison of Data and Information Required for Various Method

Method	Sea Station WL	TIZ Station WL	Upstream Station Q	Others
Direct Frequency		Required		
Hydrodynamic	Required	Required	Required	River morphology, cross-sectional and flood plains survey
Rating Curve	Required	Required	Required	
Joint probability	Required	Required	Required	
Fourier		Required		
Wavelet		Required		

Note: WL= Water level series; Q= Discharge series

9.8.3 Assumptions of the Methods

The interaction effect of tidal and fluvial discharges is one of the fundamental factors for consideration in the comparison. There is a varying degree of assumption about the interaction effects as can be seen listed in Table 9-3. On the other hand, the stationary effects of tides are assumed with the exception of the wavelet method. The stationary assumption about tides is not valid whenever the river fluctuations affect the tide propagation into the estuary. In addition, the stationary assumption about the mean tide level is not well founded in the context of the current perceived global warming phenomenon, which may result in the rise of the mean sea level (and hence the mean tide levels).

Table 9-3 Comparison of Major Assumptions for Various Method

Method	Stationary Tidal Series	Interaction Effects
Direct Frequency	Assumed	No interaction
Hydrodynamic	Assumed	Full interaction effect considered by solving St. Venant equation
Rating Curve	Assumed	Effects indirectly taken care of by the regression of H-Q at various Z levels
Joint probability	Assumed	Superposition of extreme events by likelihood, using rating curves
Fourier	Assumed	Superposition of extreme events by likelihood, using simple additive model
Wavelet	Not assumed	Superposition of extreme events by likelihood, using varying level of noise at various de-noised levels

9.8.4 Ideal Record-Length of Data Series

The length of the water level series or discharge series available is crucial in choosing the appropriate flood estimation method. For satisfying statistical coherency, the failure to meet the lowest limit may prevent the usage of a particular method. Table 9-4 shows the ideal record-length of data series to be used in various methods. In order to cover a full tidal cycle, 18.6 years is recommended in Chapter 2. For the frequency analysis of annual peak discharge series in the non-tidal zone, it is recommended to have series of at least 25 years. From the tabulation, it can be seen that the Joint Probability, the Rating Curve, and the Hydrodynamic Modelling Methods are more demanding than the Direct Frequency, Fourier and Wavelet Methods are less demanding as far as the ideal record-length of data series or implicitly, data availability, is concerned.

Table 9-4 Ideal Data Length for Various Methods

Method	Sea Station WL	TIZ Station WL	Upstream Station Q
Direct Frequency		18.6 years, hourly	
Hydrodynamic	18.6 years, hourly	18.6 years, hourly	25 years
Rating Curve	18.6 years, hourly	18.6 years, hourly	25 years
Joint probability	18.6 years, hourly	18.6 years, hourly	25 years
Fourier		18.6 years, hourly	
Wavelet		18.6 years, hourly	

9.8.5 *Special Comparison between Fourier and Wavelet Methods*

The differences between Fourier and Wavelet Methods can be seen graphically by comparing the results of filtering performed on a data series. Figure 9-2 show a typical set of results obtained by applying both methods.

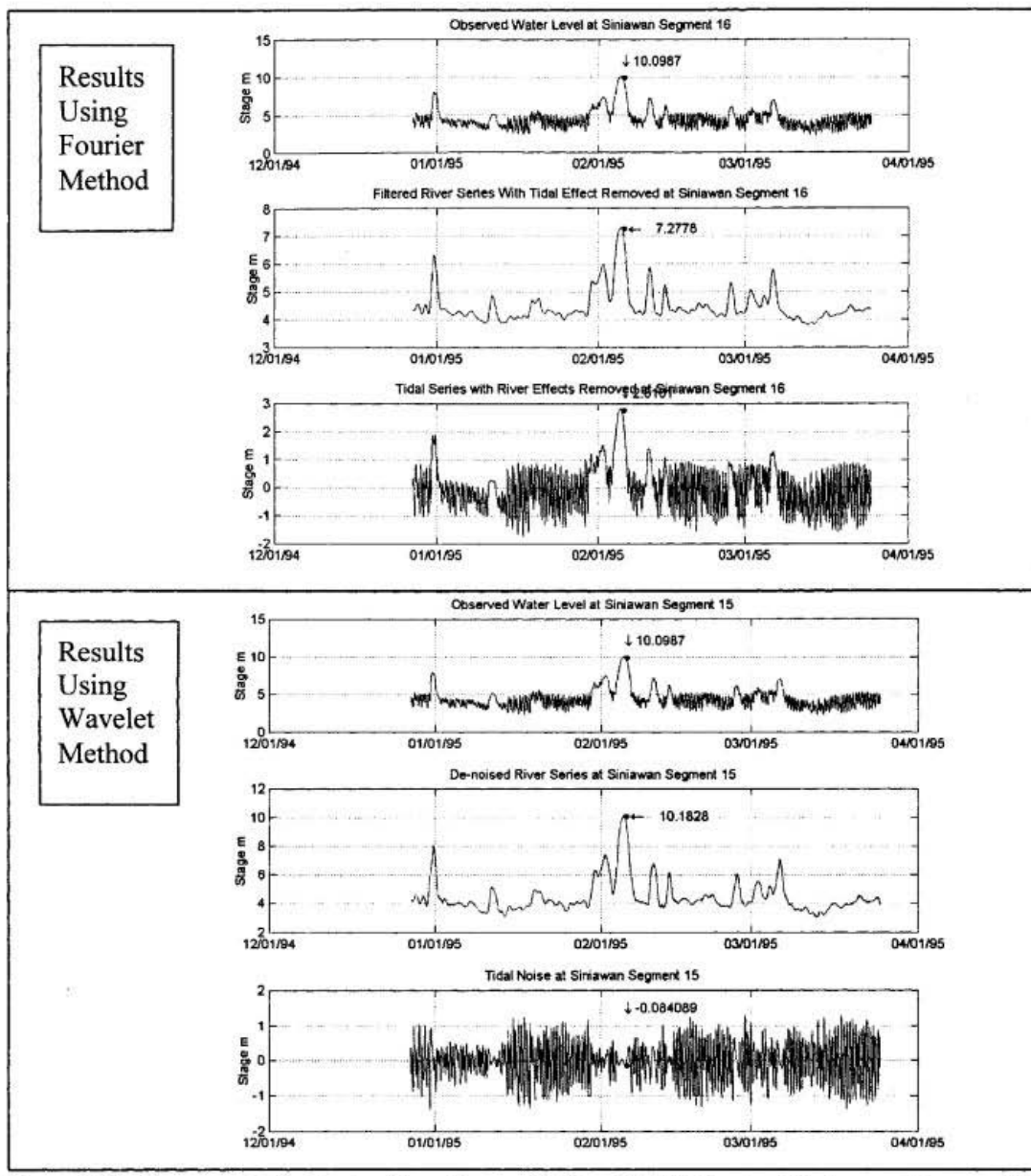


Figure 9-2 Processed Series by Fourier and Wavelet Methods: Siniawan Segment 15

The most significant difference is the plots of the tidal series or noise. Fourier Method gives results that are not realistic at the periods of high fluvial flows. The tidal noise series as derived by the Wavelet Method, however, shows progressive reduction in tidal noise as the fluvial flows increases. This concurs very well with the physical expectation that the tidal effect diminishes as the fluvial process increases. In fact this major difference in results presents the single most important reason why the Wavelet Method is the preferred method.

A complete set of similar filtering results using the two methods are shown in Appendix C1 and D1.

9.9 Recommended Method

The Wavelet Method as presented is recommended as the best based on the following attributes:

1. More robustness due to the good time-frequency localization characteristics as compared to the Fourier Method. The ability to recognize the tidal effects and separate them from the streamflow stage hydrograph is much more robust.
2. More consistency in output – the tidal noise series is consistently separated out while the variation of distribution over the entire range of de-noised level has implicitly taken care of the interaction effects of tidal and fluvial discharge,
3. Fast algorithms – fewer coefficients are needed in wavelet decompositions than in Fourier transform; this can be important for the decomposition of a lengthy data record.

4. The wavelet method can handle non-stationary series – the constant effect of some series is not a problem with the wavelet analysis. There is no assumption made on the stationary nature of the tidal series, which may be changing due to the rise of the mean sea levels.

9.10 Limitations of the Recommended Method

There are some known limitations concerning the recommended methods as listed below.

1. The new methods developed, mainly the Fourier and the Wavelet Methods, are logically sound although an absolute verification of the quantiles estimated may be impossible (the true quantiles are never known). Comparison of the results of the various methods and sensitivity analysis may help to reduce the uncertainty.
2. The local flow condition at a site may be changed drastically, such as the imposition of a certain major change to the river morphology in proximity, e.g., a river barrage. .This may affect the results using methods that rely on the analysis of historical data series. Care must be exercised in selecting the data series that are free from such disturbances.
3. The data series available in the study area are relatively short and some series do not fully satisfy the ideal data record-length criteria. Another unfavourable factor is that some discontinuities are found in the records. For practical purposes, the short gaps are filled with the mean level of the observed series to minimize the number of discontinuity.

9.11 Upstream and Downstream Boundary Conditions

While the main objective of this thesis is to compile and develop the methods for the estimation of design floods in the TIZ, the upstream and the downstream boundary conditions are required in order to carry out some of the methods. The upstream requirement is the design flow, determined from the regional flood frequency analysis. The downstream requirement is the characterization of the tidal phenomena.

9.11.1 Regional Flood Frequency Analysis for the Non-Tidal Zones

Two homogeneous regions were found for Sarawak based on the cluster analysis initially and the subsequent derivation of heterogeneity statistics using the L-moment approach. Monte Carlo simulation tests have indicated that the Generalized Logistic and the Generalized Pareto distributions are appropriate for Region A and B respectively. However, the next closest fit by the Generalized Extreme Value distribution for Region B is adopted instead of the Generalized Pareto distribution, which has an upper bound. A regional growth curve is developed for each of the region, which enables the derivation of flood quantiles for any ungauged basin in the non-tidal zone of Sarawak.

9.11.2 Characterization of Tidal Phenomena

The flow conditions in the sea or estuaries of all the rivers in Sarawak are predominantly tidal, which has caused the rivers to have extensive tidal interaction zones. The tidal species is mainly semidiurnal in the coastal waters of western Sarawak. Progressively, it becomes a mixture of semidiurnal and diurnal in the sea off Brunei, to northeastern Sarawak. The distribution of tidal levels is not exactly normal due to the

bimodal characteristics of the semidiurnal tides. This implies that the empirical distributions are preferred in simulation purposes.

In the TIZ, the characteristics of tides are not changed drastically as compared with the state in the sea or estuaries. The bimodal characteristics of the semidiurnal tides are still noticeable in the distribution of the filtered tidal series obtained by Fourier Method and the tidal noise series obtained by the Wavelet De-noising Method. Indirectly, this serves as an indication of the success in separating the tidal components from the observed water level series in the TIZ.

Chapter 10 Conclusions and Recommendations

10.1 Conclusions

The estimation of design floods, both in the non-tidal and tidal zones, remains an important element in the planning and design processes of many civil engineering projects. The Hydrodynamic Modelling Method, which is based on simulation of flow conditions in a TIZ, is comprehensive but it does not assign probabilities of events by itself. Relational methods, such as the Rating Curves Method and Bivariate Method have been used, but generally both suffer from the deficiencies of poor regression fits. The Direct Frequency Method is not strictly applicable as it is meant for the univariate case. It may be used when hourly data series longer than 18.6 years are available. It is shown that the Wavelet De-noising Method, a new approach utilizing wavelet theory and Monte Carlo simulation, permits the derivations of flood quantiles in a modest and consistent manner. The Fourier Method is another new method, which, too, gives the direct

solutions to the specific site in question and employs Monte Carlo simulation, but the results are lower in magnitude than those obtained from using other methods.

From the compilation of the existing methods, the development of two new methods, and the subsequent confirmatory studies carried out in this study, the following specific conclusions can be made:

1. There have been no comprehensive compilation and streamlining of the flood estimation methods applicable to the TIZ. This thesis has addressed this specific area in a comprehensive manner.
2. A new methodology using wavelet method is developed. It has the advantage of providing flood estimates at the point of interest provided the observed series in the vicinity is available. In comparison with the hydrodynamic method, it requires much less efforts and avoids the weak assumptions about joint probability at different sites. Theoretically, it is superior to the simple additive model such as the Fourier method. Using an extended pdf, an approach by Monte Carlo simulation has made the derivation of flood quantiles possible.
3. Type 4 of the Daubechies wavelet family with the decomposition level set to 4 is found to be an acceptable combination that serves to perform de-noising on the data series found in the study area.
4. The Fourier method is also developed and it is recommended as a check on the lower bound of flood quantiles. Being an additive model, it does not consider the interaction effects explicitly. As compared with the direct frequency method, this method is an improvement as it considers the tidal and fluvial processes

separately. It tends to give less conservative results as compared with the wavelet method.

5. The Direct Frequency Method performs well in the upper reaches, where the fluvial flooding is dominant. It requires the least effort and can be used as a quick check on the magnitudes of the other methods. In order to have credible results, at least 18.6 years of data series is required - a difficult criterion to meet in Sarawak.
6. The Joint Probability Method used for the sea level analysis by coastal engineers is extended to the TIZ. It has a good theoretical basis but in practice it suffers from the difficulties in obtaining a credible *H-Q* relationship to allow the computation of the joint probability. Synchronous series at three locations is required for this method- another difficult criterion to be met.
7. For a single river channel with no major side tributaries, the Rating Curve Method is applicable provided there are synchronized series available at three locations. One caveat to watch is that the large scatters of the *Q-H* plot may prevent the derivation of the loci of *Z* at the TIZ distinctively.
8. The installation of a long-term observatory network for water levels at major towns and cities in the TIZ should be promoted. Data in the TIZ are scarce or of poor quality. Most of the methods for flood estimation in the TIZ use water level data at hourly or finer time intervals. The urge to know the long term effects of climate change on the coastal waters and estuaries including the TIZ will certainly demand these data sets for accurate assessment. A current shift in the attention of some researchers on global climate change to quantify the change based on a

micro-environment such as a river basin instead of the global circulation model is a sign that more data in the TIZ will be in great demand.

9. Hydrodynamic models have been used in large scale studies and engineering design. Continuous simulation of the inflows or rainfall requires some assumptions of the combinatory effects. It also has the disadvantage of demanding great efforts in data collection. For large-scale studies, this method will continue to be used.
10. The regional flood frequency analysis of the non-tidal zones in Sarawak has concluded that two hydrological homogeneous regions, with a possible third mountainous region, exist in Sarawak. The best distributions to describe the annual flood series for the two identified regions are the Generalized Logistic and the Generalized Extreme Value distributions.
11. The tides in the coastal waters of Sarawak are mainly semi-diurnal. The bimodal characteristic of the tides can be seen in the filtered tidal series obtained by the Fourier Method and tidal noise series obtained by the Wavelet De-noising Method. This can be regarded as an indirect indication of the success in the separation of the tidal component using the two methods.

Overall, the alternative methods for design flood estimation in the TIZ are developed and illustrated by applying most of the methods to some data series obtained from Sarawak, Malaysia. Some existing methods are modified. The whole approach in the design flood estimation in the TIZ is streamlined, based on the availability of data and requirements. Comparison of some of the methods is carried out, and sensitivity analysis

is performed whenever possible. The methods and results presented are thus very useful for estimating design floods in any other river basins with similar extensive tidal interaction zones.

10.2 Recommendations

The use of wavelet method is recommended for estimating design floods in the TIZ of Sarawak, especially for isolated locations where data series are short. Hydrodynamic models are to be used in relatively large scale projects. The characteristics of the boundaries conditions, presented as the regional flood frequency analysis and tidal distributions in Sarawak, are recommended to be used in association with the method chosen.

During the course of this research, certain topics that were found to be of interest that warrant some further in-depth studies. These are listed and briefly discussed as follows:

1. **Choice of Wavelet Configurations**

   There are many choices of wavelet configurations and each has its unique characteristics. A specific wavelet can be tailored purely for the de-noising of tides in the TIZ for the analysis of tides. In order to narrow down the ideal wavelet options, one may employ the design of experiment (DOE) approach. For assessing the performance of each wavelet option, a relation of water level in the TIZ with the discharges upstream as done by Jay and Flinchem(1997) may be explored and modified.

2. **Selection of Distribution for Noise**

The noise distribution over the de-noised levels may not follow the normal distribution. The data series available were too short to confirm with certainty the underlying distribution. Further investigation can be performed on other river series, such as the tidal rivers in Canada, US and Europe, which has much longer records.

3. **Design of Filter for Fourier Method**

The filter used for separating the tidal series from the river series in the Fourier Method is based on a low-pass filter. There are more elaborate filters, which may be tailored to refine the separation technique.

4. **Wavelet Method and Hydrometric Network**

The use of network analysis method may be employed in tandem with the wavelet method to study the spatial relationship between a network of water levels stations that are located in proximity. The results may assist in the planning of an optimum network for the purpose of long-term water level observation in the TIZ.

5. **Sea-level Rise and Estimating Design Floods**

Sea-level rise has been a hot issue, as it is perceived as a direct effect of the global warming phenomenon. The marginal rise in the sea level may affect the rise in the flood levels in the TIZ. The effect may be amplified. However, without solid research in this area it is not possible to quantify with certainty the increased risk in flooding. Using the methods as developed in this thesis, coupled with a concerted effort in gathering data series, a new research area can

be developed to study the effects on the increased risk of civil structures located in the TIZ.

REFERENCES

Abbott, M. B., and Basco, D. R. (1990). *Computational Fluid Dynamics*. John Wiley, New York, NY.

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

Acres Consulting Services Limited. (1987). *Hydrotechnical Study – Stephenville Crossing and Black Duck Siding Areas*. A report prepared for Environment Canada under the Canada-Newfoundland Flood Damage Reduction Program.

Allen, J.B. and Rabiner, L.R. (1977). A unified approach to Short-Time Fourier Analysis and Synthesis. *Proceeding of the IEEE*, 65(11):1558-1564.

Antoniadis, A. and Oppenheim, G. (1995). *Wavelets and Statistics*. Springer-Verlag, New York, NY.

Bardsley, W.E., Mitchell, W.M. and Lennon, G.W. (1990). Estimating future sea level extremes under conditions of sea level rise. *Coastal Engineering*, **14**: 295-303.

Basco, D.R. (1983). Computation of rapidly varied, unsteady free-surface flow. *U. S. Geological Survey Report* WRI 83-4284.

Ben-Zvi, A. and Azmon, B. (1997). Joint use of L-moment diagram and goodness-of-fit test: a case study of diverse series. *Journal of Hydrology*, **198**: 245-259.

Bhaskar, N.R. and O'Connor, C.A. (1989) Comparison of method of residuals and cluster analysis for flood regionalization. *Journal of Water Resources Planning & Management*, American Society of Civil Engineers, **115**(6): 793-808.

Bobée, B. and Rasmussen, P.F. (1995). Recent advances in flood frequency analysis, U.S. National Report to International Union of Geodesy and Geophysics 1991-1994. *Review of Geophysics, Supplement*, 1111-1116.

Bowden, K.F. (1983). *Physical Oceanography of Coastal Waters*. Ellis Horwood Ltd., West Sussex, U.K.

Bradley, A. A. and Potter, K.W. (1991). Flood frequency analysis for evaluating watershed conditions with rainfall-runoff models. *Water Resources Bulletin*, **27**(1): 83-91.

Bradley, A. A. and Potter, K.W. (1992). Flood frequency analysis of simulated flows. *Water Resources Research*, **28**(9): 2375-2385.

Bras, R.L. , Gaboury, D.R., Grossman, D.S., and Vicen, G.J. (1985). Spatially varying rainfall and flood risk analysis. *Journal of Hydraulic Engineering*, **111**(5): 754-773.

Brath, A., Castellarin, A., Franchini, M. and Galeati, G. (2001) Estimating the index flood using indirect methods. *Hydrological Sciences Journal*, **46**(3): 399-418.

Bruce, A. and Gao, H.Y. (1996). *Applied Wavelet Analysis with S-Plus*. Springer, N.Y.

Burn, D.H. (1989) Cluster analysis as applied to regional flood frequency. *Journal of Hydrology*, **115**(5): 567-582.

Burn, D.H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resources Research*, **26**(10): 2257-2265.

Burn, D.H. (1991). An approach to the rationalization of stream data collection networks. *Journal of Hydrology*, **122**(1991): 71-91.

Burn, D.H. and Goel, N.K. (2000). The formation of groups for regional flood frequency analysis. *Hydrological Sciences Journal*, **45**(1): 97-112.

Caissie, D. and El-Jabi, N. (1990). A stochastic study of floods in Canada: frequency analysis and regionalization. *Canadian Journal of Civil Engineering*, **18**:225-236.

Calver, A., Lamb, R. and Morris, S.E. (1999). River flood frequency estimation using continuous runoff modelling. *Proceedings of the Institution of Civil Engineers, Water Maritime and Energy*, **136**(4): 225-234.

Cameron, D.S., Beven, K.J., Tawn, J., Blazkova, S., and Naden, P.(1999). Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). *Journal of Hydrology*, **219**(3-4): 169-187.

Caselton, W.F. and T. Husain (1980). Hydrologic networks: information transmission. *Journal of the Water Resources Planning and Management Division, ASCE*, **106**: WR2. Proc. Paper 15572, 503-520.

Chow, V. T. 1959. *Open-Channel Hydraulics*. McGraw-Hill, New York, NY.

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

Chow, V.T. Maidment, D.R. and Mays, L.W. (1988). *Applied Hydrology*. McGraw-Hill, New York.

Chui, C. K. (1992). An Introduction to Wavelets. Academic Press, Inc.

Cooley, J.W. and Tukey, J.W. (1965). An algorithm for machine calculation of complex Fourier series. *Math. Comput.*, **19**: 297-301.

Cunge, J., Holly, F.M. and Verwey, A. (1980). Practical Aspects of Computational River Hydraulics. Pitman, London.

Cunnane, C. (1987). Review of statistical methods for flood frequency estimation, in *Hydrologic Frequency Modeling*, edited by V.P. Singh, pp. 49-95, D. Reidel, Dordrecht.

Cunnane, C. (1988). Methods and merits of regional flood frequency analysis. *Journal of Hydrology*, **100**: 269-290.

Cunnane, C. (1989). *Statistical Distributions for Flood Frequency Analysis*, Operational Hydrology Report No. 33, World Meteorological Organization No. 718, Geneva, Switzerland, 1989.

Dalrymple, T. (1960). Flood frequency analyses. *Water Supply Paper*, 1543-A, U.S. Geological Survey, Reston, VA.

Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, 36(5):961-1005.

De Michele, C. and Rosso, R. (2001) Uncertainty assessment of regionalized flood frequency estimates. *Journal of Hydrologic Engineering,* **6**(6): 453-459.

DID (1963-1998). Sarawak Hydrological Yearbook Series. Hydrology Branch, Department of Irrigation and Drainage, Kuching, Malaysia.

Dijkzeul, J.C.M. (1984). Tidal filters. Journal of Hydraulic Engineering, 110(7):981-987

Donoho, D.L. and Johnstone, I.M. (1994). Ideal Spatial Adaptation via Wavelet shrinkage. *Biometrika*, **81**: 425-455.

Donoho, D.L. and Johnstone, I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of American Statistical Association*, **90** (432): 1200-1224.

D'Onofrio E.E., Fiore M.M.E. and Romero S.I.(1999). Return period of extreme water levels estimated for some vulnerable areas of Buenos Aires. *Continental Shelf Research*, **19**: 1681-1693.

Dronkers, J.J. (1964). *Tidal Computations in Rivers and Coastal Waters*. North-Holland Publishing Company, Armsterdam.

Dubreuil, P.L. (1986). Review of relationships between geophysical factors and hydrological characteristics in the tropic. *Journal of Hydrology,* **87**: 201-222.

Easton, A.K. (1977). *Selected Programs for Tidal Analysis and Prediction*. Computing Report No.9, Flinders Institute for Atmospheric and Marine Sciences, The Flinders University of South Australia, Adelaide, Australia.

El-Jabi, N., Ashkar, F. and Hebabi, S. (1998). Regionalization of floods in New Brunswick (Canada). *Stochastic Hydrology and Hydraulics*, **12**: 65-82.

Fill, H.D. and Stedinger, J.R. (1998). Using regional regression within index flood procedure and an empirical Bayesian estimator. *Journal of Hydrology*, **210**: 128-145.

Filliben, J.J. (1975). The probability plot correlation test for normality. *Technometrics*, 17(1), 111-117.

Fovell, R.G. and Fovell, M.Y.C. (1993). Climatic zones of the conterminous United States defined using cluster analysis. *Journal of Climate*, **6**: 2103-2135.

Franz, D.D., Kraeger, B.A., and Linsley, R.K. (1989). *A system for generating long streamflow records for study of floods of long return period.* Report NUREG/CR-4496, U.S. Nuclear Regulatory Commission, Washington, D.C.

Gan, T.Y. (2001). Precipitation of Western Canada - Wavelet, Scaling, and Multifractal Analysis and Teleconnection to Large-Scale Climate Anomalies. *Proceedings of 15th Hydrotechnical Specialty Conference, Canadian Society of Civil Engineers*, Victoria, 2001.

Gabor, D. (1946). Theory of communication. *Journal of IEEE*, **93**:429-457.

Glaves, R. and Waylen, P.R. (1997). Regional flood frequency analysis in Southern Ontario using L-moments. *The Canadian Geographer,* **41**(2): 178-193.

Godin, G. (1972). *The Analysis of Tides*. University of Toronto Press and Liverpool University Press, Toronto and Buffalo.

Godin, G. (1985). Modification of river tides by the discharge. *Journal of Waterway, Port, Coastal and Ocean Engineering*, **111**(2): 257-274.

Godin, G. (1997). *Tides*. Anadyomene Edition, Ottawa.

Godin, G. (1999). The propagation of tides up rivers with special considerations on the Upper Saint Lawrence River. *Estuarine, Coastal and Shelf Science*, **48**: 307-324.

Goring, D.G. (1984). Analysis of tidal river records by a harmonic regressive technique. *Journal of Hydrology*, **73**: 21-37.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979) Probability weighted moments:distribution and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, **15**(5): 1049-1054.

Harmancioglu, N.B. (1981). Measuring the information content of hydrological processes by the entropy concept. *J. Civ. Engrg. Facu., Spec. Issue for the Centennial of Ataturk's Birth*, Edge University, Izmir, Turkey, 13-40.

Harmancioglu, N.B. and Alpaslan, M.N. (1992). Water quality monitoring network design: a problem of multi-objective design making. *Water Resources Bulletin*, **28**(1): 179-192.

Heiler, T.D. and Hong, C.H. (1974). *Magnitude and Frequency of Floods in Peninsula Malaysia*. Hydrological Procedure No.4. Ministry of Agriculture and Fisheries, Malaysia.

Henderson, F.M. (1966). *Open Channel Flow*. Macmillan, New York, NY.

Heo, J.H., Boes, D.C. and Salas, J.D. (2001). Regional flood frequency analysis based on a Weibull model: part 1. estimation and asymptotic variances. *Journal of Hydrology*, **242**: 157-170.

Hill J.T. and Goring, D.G. (1998). Smoothing hydrological data: a neural network approach. *Journal of Hydrology(NZ)*, **37**(2): 81-94.

Ho, F.P. and Myers, V.A. (1975). *Joint Probability Method of Tide Frequency Analysis Applied to Apalachicola Bay and St. George Sound, Florida*. U.S. Department of Commerce, National Oceanic and Atmospheric Administration.

Hosking, J.R.M.(1990). L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of Royal Statistical Society, Service B*, **52**(1): 105-124.

Hosking, J. R. M. (1996). Fortran routines for use with the method of L-moments, version 3. *Research Report RC 20525*, IBM Research Division, Yorktown Heights, NY.

Hosking, J.R.M , Wallis, J.R. and Wood, E.F. (1985). Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, **27**: 251-61.

Hosking, J.R.M. and Wallis, J.R. (1990). Regional flood frequency analysis using L-moments. *IBM Research Report* RC15658, IBM Research Division, Yorktown Heights, N.Y.

Hosking, J.R.M and Wallis J.R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research,* **29**(2): 271-281.

Hosking, J.R.M. and Wallis, J.R. (1997). *Regional Frequency Analysis – An Approach Based on L-moments*. Cambridge University Press, Cambridge, U.K.

Husain T. (1989). Hydrologic uncertainty measure and network design. *Water Resources Bulletin, American Water Resources Association*, **25**(3): 527-534.

Jacob, D., Reed, D.W. and Robson, A.J. (1999) Choosing a pooling-group, Chapter C6. In: *Flood Estimation Handbook*, vol. 3, Statistical Procedures for Flood Estimation. Institute of Hydrology, Wallingford, UK.

Jay, D. A. and Flinchem, E.P. (1997). Interaction of fluctuating river flow with a barotropic tide: a demonstration of wavelet tidal analysis methods. *Journal of Geophysical Research*, **102**: 5705-5720.

Jay, D. A. and Flinchem, E.P. (1999). A comparison of methods for analysis of tidal records containing multi-scale non-tidal background energy. *Continental Shelf Research*, **19**: 1695-1732.

Jay, D.A. and Flinchem, E.P. (1995). Wavelet transform analyses of non-stationary tidal currents. *Proceedings of The IEEE Fifth Working Conference on Current Measurement*, A. E. Williams, ed., pp. 101-106.

Kachroo R.K. and Mkhandi, S.H. (2000). Flood frequency analysis of Southern Africa: I. Delineation of homogeneous regions. *Hydrological Sciences Journal*, **45**(3): 437-447.

Kalkstein, L.S., Tan, G. and Skindlov, J.A. (1987). An evaluation of three clustering procedures for use in synoptic climatological classification. *Journal of American Meteorological Society*, 26, 717-730.

Kendall, M.G. and Stuart, A. (1963). *The Adanced Theory of Statistics, Vol. 1. Distribution Theory*. Griffin, London.

Kirby, W. (1974). Algebraic boundness of sample statistics. *Water Resources Research*, **10**: 220-222.

Kite, G.W. (1977). *Frequency and Risk Analysis in Hydrology*. Water Resources Publications. Fort Collins, Colorado.

Kumar, R., Singh, R.D. and Seth, S.M. (1999). Regional flood formulas for seven subzones of Zone 3 of India. *Journal of Hydrologic Engineering*, **4**(3): 240-244.

Lai, C. (1986). Numerical modeling of unsteady open-channel flow. *Advances in Hydroscience,* **14**: 161-333.

Lamb, R. (1999). Calibration of a conceptual rainfall-runoff model for flood frequency estimation by continuous simulation. *Water Resources Research*, **35**(10): 3103-3114.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1978). Some comparisons of flood statistics in real and log space. *Water Resources Research*, **14**(5): 902-920.

Landwehr, J.M., Matalas, N.C. and Wallis, J.R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles. *Water Resources Research*, **15**(5): 1055-1064.

Liggett, J. A. (1975). Basic equations of unsteady flow. In: *Unsteady flow in open channels*, edited by K. Mahmood and V. Yevjevich, vol.1, Chap. 2, Water Resources Publications, Port Collins, Co.

Lim, Y.H. and Lye, L.M. (2000). Selection of a regional flood frequency distribution for western Sarawak, East Malaysia. *Proceedings of CSCE 28th Annual Conference* (on CD-ROM), Canadian Society of Civil Engineering, London, Ontario, June 7-10, 2000.

Lim, Y.H. and Lye, L.M. (2001). Developing temporal patterns of design storms for a rainforest city. *Proceedings of CSCE 29th Annual Conference* (on CD-ROM), Canadian Society of Civil Engineering, Victoria, British Columbia, May 30-June 2, 2001.

Lim, Y.H. and Lye, L.M. (2002). De-noising of streamflow series affected by tides using wavelet methods. *Proceedings of CSCE 30th Annual Conference* (on CD-ROM), Canadian Society of Civil Engineering, Montreal, Quebec, June 5- 8, 2002.

Lin Y. and Lye, L.M. (1994). Modelling long-term dependence based on cumulative departures of annual flow series. *Journal of Hydrology*, **160**:105-121.

Lye, L.M. (1993). A technique for selecting the box-cox transformation in flood frequency analysis. *Canadian Journal of Civil Engineering*, **20**: 760-766.

Lye, L.M. (2000). Applicability of a general-purpose approximate goodness-of-fit test for flood frequency analysis. *Proceedings of CSCE 28th Annual Conference* (on CD-ROM), Canadian Society of Civil Engineering, London, Ontario, June 7-10, 2000.

Madsen H. and Rosbjerg, D. (1997). The partial duration series method in regional index-flood modelling. *Water Resources Research*, **33**(4): 737-746.

Mallet S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, CA.

Messiah, A (1961). *Quantum Mechanics*. North-Holland, Amsterdam.

Middelton, J.F. and Thompson, K.R. (1986). Return period of extreme sea levels from short records. *Journal of Geophysical Research*, **91** (C10): 11707-11716.

Mkhandi, S.H., Kachroo, R.K. and Gunasekara, T.A.G. (2000). Flood frequency analysis of Southern Africa: II Identification of regional distributions. *Hydrological Sciences Journal*, **45**(3), 449-464.

Mosley, M. P. (1981). Delimitation of New Zealand hydrological regions. *Journal of Hydrology*, **49**:173-192.

Moss, M.E.. (1982). *Concepts and techniques in hydrological network design*. WMO No. 580, Operational Hydrology Report No.19, WMO, Geneva, Switzerland.

Moss, M.E. (1997). On the proper selection of surrogate measure in the design of data collection networks. *Integrated Approach to Environmental Data Management Systems, NATO ASI Series, 2, Environment*. Harmancioglu, N.B., Alpaslan, M.N., Ozkul S.D. and Singh, V.P.(eds) , Vol. 31, Kluwer Academic Publishers, 79-88.

Murty, T.S. (1984). Storm surges- meteorological ocean tides. *Canadian Bulletin of Fisheries and Aquatic Sciences* No. 212, Department of Fisheries and Oceans, Ottawa, Canada.

Nason G.P. (1995). Wavelet function estimation using cross-validation. In: Antoniadis A. and Oppenheim G., editors, *Wavelets and Statistics*, pp261-280. Springer-Verlag, New York, 1995.

Natural Environment Research Council (1975). *Flood Studies Report, Vol. 1*. Natural Environment Research Council, London.

Ogden R.T. and Parzen, E. (1996). Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Computational Statistics and Data Analysis*, **22**:53-70.

Ogden R.T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhauser, Boston, MA.

Ozkul, S., Harmancioglu, N.B. and Singh, V.P. (2000). Entropy-based assessment of water quality monitoring networks. *Journal of Hydrologic Engineering*, **5**(1):90-100.

Pandey, G.R. and Nguyen, V.-T.-V (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, **225**: 92-101.

Parker, B.B. The relative importance of the various nonlinear mechanisms in a wide range of tidal interactions (Review), in *Tidal Hydrodynamics*, edited by Parker, B.B., pp 237-268, John Wiley & Son, New York, 1991.

Pearson, C. P. (1991a). New Zealand regional flood frequency analysis using L-moments. *Journal of Hydrology (New Zealand)*, **30**: 53-64.

Pearson, C. P. (1991b). Regional flood frequency for small New Zealand basins, 2: Flood frequency groups. *Journal of Hydrology (New Zealand)*, **30**: 77-92.

Pearson, C. P. (1993). Application of L moments to maximum river flows. *The New Zealand Statistician*, **28**: 2-10.

Pearson, C.P. (1991a). New Zealand regional flood frequency analysis using L-moments. *Journal of Hydrology (N.Z.)*, **30**(2): 53-64.

Pearson, C.P. (1991b). Regional flood frequency analysis for small New Zealand basins 2. Flood frequency groups. *Journal of Hydrology (N.Z.)*, **30**(2):77-91.

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

Polonsky, V.F. (1996). Physiostatistical Approach to river delta hydrology. *Journal of Hydraulic Engineering*, **122**(6): 333-340.

Portnoff, M.R. (1980). Time frequency representation of digital signals and systems based on Short-Time Fourier Analysis. *IEEE Transaction on Acoustic, Speech, Signal Processing*, 28:55-69.

Pugh, D.T. (1987). *Tides, Surges and Mean Sea-Level*. John Wiley & Sons, Chichester.

Pugh, D.T. and Vassie, J.M (1980). Applications of the joint probability method for extreme sea level computations. *Proceedings of the Institution of Civil Engineers*, Part 2, **69**: 959-975.

Rao, A.R. and Hamed, K.H. (2000). *Flood Frequency Analysis*. CRC Press, Boca Raton, Florida.

Riggs, H.C. (1968). Frequency curves. *Techniques of Water-Resources Investigations of The United States Geological Survey*, Hydrologic analysis and investigation, Book 4, Chapter A2.

Peel, M.C., Wang, Q.J., Vogel, R.M., McMahon, T.A.(2001) The utility of L-moment ratio diagrams for selecting a regional probability distribution. *Hydrological Sciences Journal*, 46(1), 147-155.

Robson, A.J. and Reed, D.W. (1999) *Flood Estimation Handbook, Volume 3: Statistical Procedures For Flood Frequency Estimation*. Institute of Hydrology, Wallingford, UK.

Sankarasubramanian, A and Srinivasan, K. (1999) Investigation and comparison of sampling properties of L-moments and conventional moments. *Journal of Hydrology*, **218**: 13-34.

See, L. and Openshaw, S. (1999). Applying soft computing approaches to river level forecasting. *Hydrological Sciences Journal*, **44**(5): 763-778.

Sen, Z. (1991). Notes on the cyclic features in cumulative departures of annual flows series. Journal of Hydrology, **125**: 47-54.

Shannon, C.E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, Illinois.

Sharp, J.J. and Lim, Y.H. (2000). The Sarawak River Barrage – hydrotechnical and geotechnical aspects. *Water and Maritime Engineering, Proc. ICE,* **142:** 87-96.

Sikora, J.H. and Steffen, J.P. (1991). The use of continuous simulating and unsteady flow modelling in DuPage County, Illinois. In: *Urban Hydrology*, edited by Jennings, M.E., American Water Resources Association, Bethesda, MD.

Simonoff, J.S. (1996). *Smoothing Methods in Statistics*. Springer-Verlag, New York, pp338.

Singh, V.P. (1997). The use of entropy in hydrology and water resources. *Hydrological Processes*, **12**: 587-626.

Smith, L.C., Tureotte, D.L., and Isacks, B. (1998). Streamflow characterization and feature detection using a discrete wavelet transform. *Hydrological Processes*, **12**: 233-249.

Smith, R.L. (1986). Extreme value theory based on the *r* largest annual events. *Journal of Hydrology*, **86**: 27-43.

Spongberg M.E. (2000). Spectral analysis of baseflow separation with digital filters. *Water Resources Research*, **36**(3): 745-752.

Stedinger, J.R., Vogel, R.M., and E. Foufoula-Georgiou, Frequency analysis of extreme events, in *Handbook of Hydrology*, chapter 18, edited by D.A. Maidment, McGraw-Hill, New York, NY, 1992.

Stephens, M.A. (1986) Tests based on EDF statistics. In: D'Agostino, R.B. and M.A. Stephens(Editors), *Goodness-of-Fit Techniques*, Dekker, New York, pp97-193.

Stoker, J. J. (1957). *Water Waves*, Wiley Interscience, New York, NY.

Strelkoff, T. (1969). One-dimensional equations of open-channel flow. *Journal of Hydraulics Division*, American Society for Civil Engineers, 95, HY3:861-66.

Tawn, J.A. (1988). An extreme value theory model for dependent observations. *Journal of Hydrology*, **101**: 227-250.

Tawn, J.A. (1992). Estimating probabilities of extreme sea-levels. *Applied Statistics*, **41**(1): 77-93.

Tawn, J.A. and Vassie, J.M. (1989). Extreme sea levels: the joint probability method revisited and revised. *Proceeding of the Institution of Civil Engineers*, Part 2, **87**: 429-442.

Teolis A.(1998). *Computational Signal Processing with Wavelets*. Birkhauser, Boston, MA.

Teolis, A. (1998). *Computational Signal Processing with Wavelets*. Birkhäuser, Boston.

The Math Works, Inc. (1999). *Wavelet Toolbox for Use with MATLAB*. The Math Works, Inc., Natick, MA.

The Royal Malaysian Navy (1987). *Tide Tables 1988 – Malaysia, Singapore and Brunei, Vol 2*. The Hydrographer, Royal Malaysian Navy.

Thompson, R.O.R.Y. (1983). Low pass filters to suppress inertial and tidal frequencies. *Journal of Physical Oceanography*, **13**(6):1077-1083.

Torrence, C. and Compo, G.P. (1998). A practical guide to wavelet analysis. *Bulletin of The American Meteorological Society*, **79**(1): 61-78.

Ulrych, T.J., Velis, D.R., Woodbury, A.D. and Sacchi, M.D. (2000) L-moments and C-moments. *Stochastic Environmental Research and Risk Assessment*, **14**: 50-68.

van der Made, J.W. (1969). Design Levels in the Transition Zone Between the Tidal Reach and the River Regime Reach. *Symposium on the Hydrology of Deltas*, IASH/AIHS- Unesco, Bucharest, 6-14 May, 1969.

Vogel, R. M. and Fennessey, N.M. (1993). L-moment diagrams should replace product-moment diagrams. *Water Resources Research*, **29**: 1745-52.

Vogel, R.M. (1986). The probability plot correlation coefficient test for the Normal, Lognormal and Gumbel distributional hypothesis. *Water Resources Research*, **22**(4): 587-590, (corrections: *Water Resources Research*, **23**(10): 2013).

Wallis, J.R.(1981). Risk and uncertainties in the evaluation of flood events for the design of hydraulic structures. In *Piene e Siccità*, edited by E. Guggino, G. Rossi, and E. Todini, pp. 3-6. Fondazione Politecnica del Mediterrano. Catania, Italy.

Wallis, J.R.(1982). Hydrologic problems associated with oilshale development. In *Environmental Systems and Management*, edited by S. Rinaldi, pp.85-102. North-Holland, Amsterdam.

Walters, R.A. and Heston, C. (1982). Removing tidal-period variations from time-series data using low-pass digital filters. Journal of Physical Oceanography, 12:112-115.

Watt, W.E., et al. (1989). *Hydrology of Floods in Canada: A guide to Planning and Design*. Associate Committee on Hydrology, National Research Council of Canada.

WMO (1981) *Guide To Hydrological Practices*, Vol. I, 4th Edition, WMO No. 168.

WMO (1988). *Hydrological Aspects of Combined Effects of Storm Surges and Heavy Rainfall on River Flow, Operational Hydrology Report No 30 (WMO No704)*. Secretariat of the World Meteorological Organization, Geneva, Switzerland, 1988.

Wong, M.H. (1991). *The Distribution, Characteristics and Agricultural Utilisation of Peat in Sarawak*. Department of Agriculture, Sarawak, Malaysia.

Yang Y. and Burn, D.H. (1994). An entropy approach to data collection network design. *Journal of Hydrology*, **157**: 307-324.

Yen, B.C. (1973). Open-channel flow equations revisited. *Journal Engineering Mechanics Division*, American Society for Civil Engineers, **99**, EM 5:979-1009.

Zahel W. (1992). Ocean tides. In: *Tidal Phenomena*, by Helmut Wilhelm, W. Zurn and Wenzel H.G.

Appendix A1

Processing Water Level Data

Appendix A1

Processing Water Level Data

1. Data Format from DID Sarawak

The data obtained from DID Sarawak consisted of two main types:

1. Time-dependent data
2. Data in tables

The water level data belongs to the time dependent type. These data are generated from digitising programs that required the data processing personnel to identify all the maxima and minima, together with any other major points on the automatic water level recorder charts. The data points are event-based rather than at fixed time intervals.

2. Water Level Text File Processing Using EXCEL

Text file obtained from DID Sarawak are required to be processed prior to be used by the analysis programs. Due to the column oriented nature of the text files, Microsoft EXCEL is used in the initial recognition and checking purposes. Editing works are done such as the removal of header lines, any empty cell representing '00' in time is added, '00' in the year column is changed to "2000", etc.

3. Gap Treatment and Splitting of Series

DatareadSection30.m

This program is written to remove the unwanted gap marking code of "-2700000" intended for the processing programs installed at DID Sarawak. Individual segment of data are saved and renamed by the program. A file extension of "OUT" is used.

```
% DATAREADSECTION30.M
% READING FILE OF DID WATER LEVEL AFTER TREATMENTS IN EXCEL
% Input file name format: file in tab-delimited text
%    fname+xls.txt -use Excel for convert, changing
%    99 to 1999, 00 to 2000, __ to 0 (hour, minute)
% Ouput file name format
%   1. fname-i-.out   in -Ascii for each segment
%   2. fname.sse      file segments' starting and ending date-time
% Output figures on screens
%
clear all
%  Define input file name
fname='Satok'
in='.txt'
fname1=strcat(fname,in)

%  READ COLUMN DATA AS WATER LEVEL, YEAR, MONTH, DAY, HOUR, MINUTE, SECOND
[w,y,mo,d,h,m,s]=textread(fname1,'%f %u %u %u %u %u %u');
N=datenum(y,mo,d,h,m,s);                          % data now in date N and water level w
n=0;                                              % number of no good lines
c=length(w);

% COUNT THE NUMBER OF NO GOOD LINES
% Not counting data line that are marked -27000  for missing gap
for i=1:c
  if w(i)<0;
     n=n+1;
     fprintf('1 0 gaps mark at %f \n',i)
  elseif w(i)>200000 & w(i+1)>200000;
     n=n+2;
     i=i+1;
     fprintf('2 gaps mark at %f \n',i)
 elseif w(i)>200000
     n=n+1;
     fprintf('1 gaps mark at %f \n',i)
  end
end
```

```
% ERASE ROWS OF w AND N WITH MISSING DATA
k=c-n;                                  % no of rows of good data
i=1;                                    %counter of rows of original data
E(1,1)=1;EV(1,1)=N(1)
j=1;                                    % counter of E and EV

% DEFINE
% E = Error index, EV =Error Time Value
% E(1,.) EV(1,.) begin of segment
% E(2,.) EV(2,.) end of segment

while i<=c
   if w(i)<0;
     j=j+1;
     E(2,j-1)=i-1; EV(2,j-1)=N(i-1)          % to mark end of last good data
     E(1,j)=i+1;EV(1,j)=N(i+1)    % to mark begin of next good data
     i=i+1;
  elseif w(i)>200000 & w(i+1)>200000;
     j=j+1;                                  % to mark index of gap
     E(2,j-1)=i-1; EV(2,j-1)=N(i-1) % to mark end of last good data
     E(1,j)=i+2;EV(1,j)=N(i+2)    % to mark begin of next good data
     i=i+2;
  elseif w(i)>200000;
     j=j+1; % to mark index of gap
     E(2,j-1)=i-1; EV(2,j-1)=N(i-1) % to mark end of last good data
     E(1,j)=i+1;EV(1,j)=N(i+1)    % to mark begin of next good data
     i=i+1;
  else
     i=i+1;
  end
end
E(2,j)=i-1; EV(2,j)=N(i-1)          % to mark very end of good data

% MISSING GAPS BETWEEN EV=ERROR VALUE E=INDEX OF ERROR
for p=1:length(E)
fprintf('%3.0f From: %-20s  to  %-20s \n',p,datestr(EV(1,p)),datestr(EV(2,p)));
end

%  WRITE SEGMENT STARTS AND ENDS
fname3=strcat(fname,'.sse')
fid=fopen(fname3,'w')
fprintf(fid,' Data Segment Starting and Ending Date-time at %-20s \n\n',fname);
for p=1:length(E)
fprintf(fid,'%3.0f From: %-20s  to  %-20s \n',p,datestr(EV(1,p)),datestr(EV(2,p)));
```

```
end
fclose(fid)

% PLOTTING 4 ON 1 FIGURE
% If no plotting required set pting=0
pting=1;
if pting==1;
k1=length(E);
k2=(k1/4);                          %round to the no of 4x fig
k3=fix(k2);
k4=(k2-k3)/0.25;
if k4==0;
   k5=k3;
else
k5=k3+1;
end
m=0
for q=1:k5
   if q<k5;
      k6=4;
   else
      k6=k4;
   end
   figure(q)
   for r=1:k6
      m=m+1
      subplot(2,2,r)
      plot(N(E(1,m):E(2,m)),w(E(1,m):E(2,m)))
      hold on
      title([fname ':' int2str(m)]);
      datetick('x',26);
   end
end
figure(q+1);                                %plot overall
N9=N(E(1,1):E(2,1));
W9=w(E(1,1):E(2,1));
for i=2:k1;
   N9=[N9;N(E(1,i):E(2,i))];
   W9=[W9;w(E(1,i):E(2,i))];
end
plot(N9,W9)
xlim([0 1])
datetick('x',26);                           % 26 for y/m/d
title(['Observed Water Level at ' fname]);
ylabel('Stage in mm above chart datum');
```

```
end

% SAVE INDIVIDUAL SECTION IN ASCII FILES
for i=1:k1;
   fname4=strcat(fname,int2str(i),'.out')
   N1=(N(E(1,i):E(2,i)))';
   W1=(w(E(1,i):E(2,i)))';
   Y = [N1;W1];
   fid = fopen(fname4,'w')
   fprintf(fid,'%f %f\n',Y);
   fclose(fid)
   clear Y
end
end
```

4. Hourly Interpolation by Spline Fitting

Interhour30.m

This program allows batch processing of "OUT" files containing the original data in split segments of hourly data. A spline-fitting function is used. The output files are given the file extension "HLY".

```
% Interhour30.m
% USE INTERPOLATION30 TO GET HOURLY DATA AND
% following datareading program
% N for datenum w for corresponding water level
% z index for each hour
% W interpolated hourly water level
% Output hourly data stored in *.hly text file
%   variable: file name :fname change no1
%             number of interval: i change no2
% Input files are of fname-i-.out of 2 columns
%   <datenum> <water level>
% Ouput files are of fname-i-.hly
%   <datenum> <water level>
clear all
for i=1:59                              % to be changed no1 <= number of segments
```

```
fname='Siniawan'                    % to be changed no2 <= name of station
in1='.out'
out1='.hly'
fn1=strcat(fname,int2str(i),in1)
fn2=strcat(fname,int2str(i),out1)
[N,w]=textread(fn1,'%f %f');
% Set z index for hourly
z=N(1):1/24:N(end);
'running spline fitting'

% SPLINE FITTING OUTPUT HOURLY VALUES
for j=1:length(z)
    W(j)=interp1(N,w,z(j),'spline'); % spline fitting command
end
'interp end'
% Saving file, W and z are row vector
%  Y is a single vector with alternate values of W and z
Y=[W;z];
   fid = fopen(fn2,'w');
   fprintf(fid,'%f %f\n',Y);
   fclose(fid)
'saving ended'

clear N w z Y W;
end
```

5. Further Analysis

Once processed, the hourly data are used in various analysis programs such as Checking Surge, Fourier Method, and Wavelets Method.

Appendix A2

MATLAB Program for Checking Surge

Appendix A2

MATLAB Program for Checking Surge

```
% Checking Surge Effects at Pulau Lakei - READTIDElakei.m
% To read ASCII files of Malaysian National Mapping Department
% and read in the predicted tide from text file
% U Reading dummy matrix
% w Tidal waterlevel
% i no of day
% j line count for each day
% l hour cummulative
% k hour of day
% N number associated with hour-time
clear


%  READING COLUMN TEXT FILE OF PREDICTED TIDES IN SINGLE COLUMN
[p]=textread('lakeipredict88.txt','%f');

%  READING RECORDED TIDES IN ASCII FILE LINE BY LINE
%  The text file has repetitions of 26 lines of daily data
%  1st and last lines are not hourly data and are to be discarded
U=dlmread('lakeiobserve88raw.txt',',');
l=1
c=length(U)
for i=1:c/26
  j=2;
    for k=1:24;
      m=(i-1)*26+j;
      w(l)=U(m,1);
      j=j+1;
      l=l+1;
    end
end
N=1:l-1;

% OBS ARE THE OBSERVED TIDE LEVELS AND PRE ARE THE PREDICTED TIDE LEVELS
Obs=w'./100;                    % adjust for unit from cm to metre
Pre=p;
Diff=Obs-Pre;
```

```
% SAVING FILES
save ('plak88','Obs','-ASCII');
save ('Diffplak88','Diff','-ASCII');

% PLOT RESIDUALS, STATISTICS
figure(2)
plot(N,Diff);
ylabel('Residual Level (Metre)')
xlabel('Hours')
title('Residual Levels at Pulau Lakei in 1988 ')
stdev=std(Diff)
skew=skewness(Diff)
kurt=kurtosis(Diff)
me=mean(Diff)
figure(3)
histfit(Diff,25);
xlabel('Residual Level (Metre)')
title('Frequency Distribution of Residual Levels at Pulau Lakei in 1988 ')
```

Appendix A3

MATLAB Program for Low-Pass Filtering in Fourier Method

Appendix A3

MATLAB Program for Low-Pass Filtering in Fourier Method

```
% FOURIER METHOD -SPECTRUM, FILTERING AND DISTRIBUTION
% Run Filtering after DFT and harmonics
% Magnitude and Phase of Transformed Data
% Need Hourly data series

clear all

%  INPUT NUMBER OF SEGMENTS(nf) TO PROCESS FOR STATION NAME(ns)
nf=24
ns='Siniawan'

%  READING UP TO nf OF DATA SEGMENTS
for f1=1:nf;
fname=strcat(ns,int2str(f1))
in='.hly'
fname1=strcat(fname,in)

%  ADJUST FOR CUT OFF FREQUENCY IN DEGREES/HOUR
Fc=12

%  READING INPUT FILE WITH TIME SERIES X
[x,t]=textread(fname1,'%f %f');

% TO ADJUST FOR UNIT OF MM TO METRE
x=x./1000;

%  SET LIMIT OF PLOT FOR FREQ
w1=0;
w2=180;
n=length(x)

% PLOT ORIGINAL TIME SERIES ON THE TOP OF FIGURE 1
figure(1)
subplot('position',[0.1 0.8 0.85 0.125])
plot(t,x);
ylabel('Stage m')
title('Time series plot')
```

```
datetick('x',2);                    % Replace x-axis ticks with 6 digit date labels
hold off

% PLOT PLOTTING FOURIER POWER SPECTRUM AT BOTTOM OF FIGURE 1
[p,v]=periodogram(x);

% CONVERT ANGULAR SPEED FROM RAD/HOUR TO DEGREE/HOUR
vd=v*360/(2*pi);
subplot('position',[0.1 0.1 0.85 0.6])
semilogy(vd,p)
title('Fourier Power Spectrum- Boxcar Windowing')
xlabel('Frequency degree/hour')
ylabel('Power')
grid on
win = hann(n);
[Pxx,w]=periodogram(x,win);
wd=w*360/(2*pi);

% FILTERING OF DFT COMPONENT
% REMOVE PARTS DUE TO TIDES AT FREQUENCY >CA OF THE DFT
ft=fft(x);                          % DFT of x
ca=round(length(x)*Fc/360);
ftr=ft(1:ca);                       % only for the 1st ca components
z=zeros(1,length(x)-ca);            % pad with zeros
ftr=[ftr' z];                       % concatenate

% INVERT DFT TO GET FILTERED SERIES AS RIVER SERIES
ytr1=real(ifft(ftr));

% INVERT TO CORRECT TIME SCALE (BACK TO FRONT)
for i=1:length(ytr1);
   ytr(i)=ytr1(length(ytr1)+1-i);
end

% PLOT ORIGINAL SERIES ON TOP OF IN FIGURE 12
figure(12);
subplot('position',[0.1 0.60 0.85 0.35])
plot(t,x); %plot original time series
ylabel('Stage m')
title('Original Time Series Plot')
datetick('x',2);% Replace x-axis ticks with 6 digit date labels
grid on
hold off

% PLOT TRANSFORMED BACK RIVER SERIES IN BOTTOM OF FIGURE 12
```

```
subplot('position',[0.1 0.05 0.85 0.45])
plot(t,ytr)
str1=['River Series with Tides Removed at ',fname1];
title(str1);
datetick('x',2);
ylabel('Stage m')
grid on

% GET FILTERED TIDAL SERIES
ftt=ft(ca+1:length(x));                 % only for after ca components
zt=zeros(1,ca);                         % pad with zeros
ftt=[zt ftt'];                          % appended

% INVERT DFT TO GET FILTERED TIDAL SERIES
ytt1=real(ifft(ftt));

% INVERT TO CORRECT TIME SCALE (BACK TO FRONT)
for i=1:length(ytt1);
   ytt(i)=ytt1(length(ytt1)+1-i);
end

% PLOT ORIGINAL SERIES ON TOP OF FIGURE(13);
subplot('position',[0.1 0.60 0.85 0.35])
plot(t,x);                              %plot original time series
ylabel('Stage m')
title('Original Time Series Plot')
datetick('x',2);                        % Replace x-axis ticks with 6 digit date labels
grid on
hold off

% PLOT TRANSFORMED BACK TIDAL SERIES IN BOTTOM OF FIGURE 13
subplot('position',[0.1 0.05 0.85 0.45])
plot(t,ytt)
str1=['Tidal Series with River Effects Removed at ',fname1];
title(str1);
datetick('x',2);
ylabel('Stage m')
grid on

% CHECK DISTRIBUTION BY PLOTTING IN FIGURE 14
figure(14)
subplot('position',[0.1 0.6 0.85 0.35])
histfit(ytr,25);
str1=['Frequency Distribution of "Discharge Level" at ',fname1];
title(str1);
```

```
subplot('position',[0.1 0.1 0.85 0.35])
histfit(ytt,25);
str1=['Frequency Distribution of "Tidal Level" at ', fname1];
title(str1)'
xlabel('Level Above Chart Datum (m)')

%  SAVE FILE
%  ytr and ytt are row vector, t a column vector
%  Y1 a single vector with alternate values of ytr and t

Y1=[ytr;t'];
Y2=[ytt;t'];
out1='.ytr';
out2='.ytt';
fname1=strcat(fname,out1);
fname2=strcat(fname,out2);
   fid = fopen(fname1,'w');
   fprintf(fid,'%f %f\n',Y1);
   fclose(fid)
   fid = fopen(fname2,'w');
   fprintf(fid,'%f %f\n',Y2);
   fclose(fid)

'saving ended'
clear x t Y1 Y2 ytr ytt
end

%THE FOLLOWINGS ARE BRIEF NOTES ON FOURIER TRANSFORM RELATED COMMANDS
% PERIODOGRAMS
%[Pxx,w] = periodogram(x) returns the power spectral density (PSD) estimate
%Pxx of the sequence x using a periodogram. The power spectral density is
%calculated in units of power per radians per sample. The corresponding vector of
%frequencies w is computed in radians per sample, and has the same length as
%Pxx. A real-valued input vector x produces a full power one-sided (in frequency)
%PSD (by default)

%FFT
%In general, the length N of the FFT and the values of the input x determine the length
%of Pxx and the range of normalized frequencies. For this syntax, the (default) length N
%of the FFT is the larger of 256 and the next power of two greater than the length of x.
%Real-valued Input
%Length of Pxx=(N/2)+1
%Range of the Corresponding Normalized Frequencies= [0, Pi]

%SPECIFYING THE LENGTH OF THE FFT WITH THE INTEGER NFFT.
```

```
%If nfft is empty vector [], it takes the default value for N listed above
%The length of Pxx and the frequency range for w depend on nfft and the values of
%the input x.
%                nfft      length of Pxx      Range of w
%Real-valued     Even      (nfft/2 + 1)       [ 0,pi ]
%Real-valued     Odd       (nfft + 1)/2       [ 0,pi ]
%If nfft < length of x, x.*window is wrapped modulo nfft.
%If nfft > length of x, x.*window is zero-padded to compute the FFT.
%[Pxx,f] = periodogram(x,window,nfft,fs,'range') or
%[Pxx,w] = periodogram(x,window,nfft,'range') specifies the range of
%frequency values to include in f or w. This syntax is useful when x is real.

%WINDOW CAN BE:
%bartlett, blackman(preferred parametric), chebwin, hamming,
%hann(same as hanning as mentioned in Brigham for leakage),
%kaiser, triang

%'RANGE' DETERMINING THE FREQUENCY RANGE, CAN BE EITHER:
%'onesided': Compute the one-sided PSD over the frequency ranges
%specified for real x. This is the default for real-valued x.
%'twosided': Compute the two-sided PSD over the frequency range [0,fs).
%This is the default for complex-valued x.
%If fs is empty vector, [ ], the frequency range is [0,1).
%If no fs specified, the frequency range is [0, 2pi).
```

Appendix A4

MATLAB Program For MC Simulation of Joint Probabilities for Annual Extreme Series

Appendix A4

MATLAB Program For MC Simulation of Joint Probabilities for Annual Extreme Series

```
% MC SIMULATION FOR JOINT PROBABILITY OF TIDES AND RIVER LEVELS
% 1. Filtered river follows GPA for Siniawan
% 2. Filtered tidal series follows GEV for Siniawan
e1=4.2286;
a1=3.6298;
k1=0.9815;
e2=1.6919;
a2=0.5939;
k2=0.1552;
n=10000;
for i=1:n;
p1=rand;
p2=rand;
R(i)=e1+(a1/k1)*(1-(1/p1)^-k1);
K(i)=e2+(a2/k2)*(1-(-log(p2))^k2);
Z(i)=R(i)+K(i);
end
Z=sort(Z);

% SAMPLE/RETURN PERIOD
p5=n/5;
p10=n/10;
p50=n/50;
p100=n/100;
p200=n/200;

% QUANTILE FROM EMPIRICAL DIST
Z5=Z(n-p5+1)
Z10=Z(n-p10+1)
Z50=Z(n-p50+1)
Z100=Z(n-p100+1)
Z200=Z(n-p200+1)
fname1='Sin-jpm-four.out1';
   fid = fopen(fname1,'w');
   fprintf(fid,'%f \n',Z);
   fclose(fid)
```

Appendix A5

MATLAB Program For Deriving Correction Factor In Hourly River Series (Fourier Method)

Appendix A5

MATLAB Program for Deriving Correction Factor

In Hourly River Series (Fourier Method)

```
% EVALUATING THETA: CORRECTION FACTOR FOR HOURLY STORM DATA
% theta=reciprocal of the mean time spent above a level eta
% on each independent excursion above this level
clear all

% READ INPUT FILE
[R]=textread('siniawanRiv.txt','%f');          % reading column text file
k=0

% LOOP #1 FOR ALL POSSIBLE EXTREME ETA=T
for t=4:0.1:7.5;
k=k+1;
j=0;
c=0;
x=0;
theta=0;
for i=1:length(R)
   if R(i)>=t
      theta=theta+1;
      x=1;                                     % in exceedance mode
   elseif x==1
      c=c+1;                                   % no of exceedance
      j(c)=theta;
      p(c)=max(R(i-theta:i))                   % maxima in each storm
      x=0;
      theta=0;
   else
      x=0;
      theta=0;
   end
end
me(k)=mean(j);
ne(k)=me(k)*c;
end
```

Appendix A6

MATLAB Program for MC Simulation of Joint Probabilities for Hourly Series

Appendix A6

MATLAB Program for MC Simulation of Joint Probabilities for Hourly Series

```
% MC SIMULATION FOR JOINT PROBABILITY OF TIDES AND RIVER LEVELS
% Siniawan With Hourly data
% 1. Filtered river pdf smoothed by Gumbel extreme in annual
% 2. Filtered tidal series follows empirical hourly
clear

% READ CDF TABLES
[pt vt]=textread('th-cdf-value.txt','%f %f');
[pr vr]=textread('rh-cdf-value.txt','%f %f');
cdft=[pt vt];
cdfr=[pr vr];

% GENERATE FOR N YEARS
n=1000;
for j=1:n

% GENERATE FOR M HOURS IN EACH YEAR
m=8766;
for i=1:m;
p1=rand;
p2=rand;
R(i) = interp1(cdfr(:,1),cdfr(:,2),p1);        % lookup table
K(i) = interp1(cdft(:,1),cdft(:,2),p2);
Zh(i)=R(i)+K(i);
end
Z(j)=max(Zh)                                   % abstract maximum of year
'one year done'
end
Z=sort(Z);

% SAMPLE/RETURN PERIOD n YEAR
p5=n/5;
p10=n/10;
p50=n/50;
```

```
p100=n/100;
p200=n/200;

% QUANTILE FROM EMPIRICAL DIST
Z5=Z(n-p5+1)
Z10=Z(n-p10+1)
Z50=Z(n-p50+1)
Z100=Z(n-p100+1)
Z200=Z(n-p200+1)
fname1='Sin-jpm-four-hour.out1';
   fid = fopen(fname1,'w');
   fprintf(fid,'%f \n',Z);
   fclose(fid)
```

Appendix A7

MATLAB Program for Wavelet De-noising to Determine River Flow and Tidal Noise Series

Appendix A7

MATLAB Program for Wavelet De-noising to Determine River Flow and Tidal Noise Series

```
% PROGRAM DENOISING FOR MULTIRESOLUTION DECOMPOSITION OF S
% and de-nosing using wavelets transforms
% obtain: approximation at x level
% input: hourly level
% adjust: different wavelets and level of decomposition
% output: plots of de-noised series, files on series, stat,

clear all

% DEFINE FILE SYSTEM
ext1='.hly' % input file extension
out1='.bmp' % coefficient output file1
out2='.ydr' % denoised river series
out3='.yno' % tidal noise series
out10='.sta' % statistical ouput file 10
out4='.ssd' % total sequence of denoised series
out5='.sno' % total sequence of denoised tides series
out6='.sss' % total sequence of original series
fname='Siniawan'   % to be changed no 3

%PRINT DETAILS YES=1 NO=0
pdetail=1

% TO BE CHANGED
ns=18                    % number of signal segment, see for i=1:ns  statement
ctr=0                    % counter

% NAME FOR STAT FILE
fname10=strcat(fname,out10);
   fid10 = fopen(fname10,'w');

% ALL MATRICES ASSIGN
Noi=[];Ssd=[];Sss=[];

% FOR EACH DATA SEGMENT
for i=1:ns
```

```
ctr=ctr+1
fn1=strcat(fname,int2str(i),ext1) % input file name
if i==1

% PRINTING HEADINGS
   fprintf(fid10,' Segment   == Max == ==== MaxTime ====  ==Mean ==  ==StDev.==
==Skew== ==Auto1== ==size==\n\n')
   fprintf(fid10,'\n\nResults of Denoising\n');
   fprintf(fid10,'  Station:  %s\n\n',fname);
end

% READ INPUT FILE, S SIGNAL, N TIME IN TIME CODE
[s,N]=textread(fn1,'%f %f');
ls=length(s);
st=strcat(datestr(N(1),20),' :',datestr(N(1),15));
et=strcat(datestr(N(ls),20),' :',datestr(N(ls),15));

% CONVERT TO METRES
s=s./1000;

% DE-NOISING PROCESS OF A ONE-DIMENSIONAL SIGNAL USING WAVELETS
th='sqtwolog'                    % <<===== Thresholding by:
%'rigrsure' use the principle of Stein's Unbiased Risk(quadratic loss function). One gets an
%       estimate of the risk for a particular threshold value t.
%       Minimizing the risks in t gives a selection of the threshold value.
%'sqtwolog' for universal threshold. uses a fixed-form threshold yielding minimax performance
%       multiplied by a small factor proportional to log(length(X)).
%'heursure' is an heuristic variant of the two previous options. As a result, if the signal
%       to noise ratio is very small, the SURE estimate is very noisy.
%       If such a situation is detected, the fixed form threshold is used.
%'minimaxi' for minimax thresholding. uses a fixed threshold chosen to yield minimax performance
%       for mean square error against an ideal procedure. The minimax principle is used in
%       statistics in order to design estimators. Since the de-noised signal can be
%       assimilated to the estimator of the unknown regression function, the minimax
%       estimator is the one that realizes the minimum of the maximum mean square error
%       obtained for the worst function in a given set.
%   Threshold selection rules are based on the underlying model y = f(t) + e
%   where e is a white noise N(0,1). Dealing with unscaled or non-white noise can be
%   handled using rescaling output threshold

% HARD OR SOFT THRESHOLD
sorh='s'                         % <<===== Thresholding hard or soft

% SOFT OR HARD THRESHOLDING SEE <WTHRESH>
scal='mln'                       % <<===== Multiplicative threshold scaling:
```

```
%'one' for no rescaling
%'sln' for rescaling using a single estimation of level noise based on first-level coefficients
%'mln' for rescaling done using level-dependent estimation of level noise

% LEVEL OF DECOMPOSITION
x=6                                % <<===== Level x of decomposition

% WAVELET TYPE
wt='db5'                           % <<===== Wavelet type 5 chosen
%  Daubechies:'db1' or 'haar', 'db2', ... , 'db45'
%  Coiflets:'coif1', ... , 'coif5'
%  Symlets: 'sym2', ... , 'sym45'
%  Discrete Meyer: 'dmey'
%  Biorthogonal:'bior1.1', 'bior1.3', 'bior1.5','bior2.2', 'bior2.4', 'bior2.6', 'bior2.8'
%           'bior3.1', 'bior3.3', 'bior3.5', 'bior3.7','bior3.9', 'bior4.4', 'bior5.5', 'bior6.8'
%  Reverse Biorthogonal: 'rbio1.1', 'rbio1.3', 'rbio1.5','rbio2.2', 'rbio2.4', 'rbio2.6', 'rbio2.8'
%        'rbio3.1', 'rbio3.3', 'rbio3.5', 'rbio3.7', 'rbio3.9', 'rbio4.4', 'rbio5.5', 'rbio6.8'

% FIND FIRST VALUE IN ORDER TO AVOID EDGE EFFECTS
deb=s(1);

% DE-NOISNG
[sd,CXD,LXD]=wden(s-deb,th,sorh,scal,x,wt);
% denoise using soft fixed form thresholding, scaling 'no'
% and unknown noise (non-white) option, level x, wavelet wt
sd=sd+deb; % adjust back the first s(1)

% NOISE- TIDES- ORIGINAL
noise= s-sd;
C1=[N noise];
C2=[N sd];
C3=[N s];

% APPEND SERIES TO FORM TOTAL SERIES
Noi=[Noi;C1];
Ssd=[Ssd;C2];
Sss=[Sss;C3];

% IDENTIFY MAXIUM OF SD AND FIND CORRESP OF OTHERS
[sdm m9]=max(sd);
tsm=N(m9); % time corr to max of original signal
sm=s(m9);
sdm99=[' ' num2str(sdm,4) 'm  denoised'];
sm99=[' ' num2str(sm,4) 'm  original'];
```

```
% PLOT COMPARE ORIGINAL AND DE-NOISED SIGNAL
if pdetail==1
figure(i)
pt1=strcat([fname ' Hourly Water Level'], ' Segment:',int2str(i),'  Start:',st,'  End:',et)
pt2=strcat('De-noising by Wavelet:',wt,'  Rescale:',scal,'  Threshold:',th,'  Decompo Level :',int2str(x))
subplot(2,1,1);plot(N,s); title(pt1);ylabel('Level (m)');grid on;hold on;
plot(N,sd,'-w','LineWidth',2.5);plot(N,sd,'-r','LineWidth',1);% white line and black on top
axis([min(N),max(N),-Inf,max(s)+0.5]);
datetick('x',20);
text(tsm,sm,['\bullet']);
text(tsm,sm,['  \leftarrow' sm99]);
hold off;
subplot(2,1,2);plot(N,sd); title(pt2);ylabel('Level (m)');
axis([min(N),max(N),-Inf,max(sd)+0.5]);
datetick('x',20);
text(tsm,sdm,['\bullet']);
text(tsm,sdm,['  \leftarrow' sdm99]);
grid on;
end

% SAVE PLOTS OF SERIES IN BMP
fn2=strcat(fname,int2str(i),'_',wt,'_',int2str(x),'_',sorh,'_',th,'_',scal,out1)
saveas(gcf,fn2);

% SAVING DENOISED SD AND NOISE SERIES
Y1=[N';sd'];
Y2=[N';noise'];
printtofile =1; % no printing =1
if printtofile==0;
fn2=strcat(fname,int2str(i),'_',wt,'_',int2str(x),'_',sorh,'_',th,'_',scal,out2);
fn3=strcat(fname,int2str(i),'_',wt,'_',int2str(x),'_',sorh,'_',th,'_',scal,out3);
   fid2 = fopen(fn2,'w');
   fid3 = fopen(fn3,'w');
   fprintf(fid2,'%f %f\n',Y1);
   fprintf(fid3,'%f %f\n',Y2);
   fclose(fid2); fclose(fid3);
end
'saving ended'

% REGISTER STATISTICS
n=length(sd);
n4=round(n/4);
rk=xcov(sd,n4,'coeff');
r1sd=rk(n4);
```

```
% PRINT SAVE STATISTICS IN FILE *.STA
fprintf(fid10,' %s \n', int2str(i));
fprintf(fid10,' Origin    %6.2f  %s %6.2f   %6.2f\n',...
  sm,datestr(tsm,0),mean(s),std(s));
fprintf(fid10,' Denoised  %6.2f  %s %6.2f   %6.2f    %6.2f    %6.2f    %6.2f\n\n\n',...
   sdm,datestr(tsm,0),mean(sd),std(sd),skewness(sd),r1sd,n);
end
fclose(fid10)

% SCATTER PLOT OF NOISE WITH DISTN OF i TO ns SERIES
figure(500)
[H,AX,BigAx,P] = plotmatrix(Ssd(:,2),Noi(:,2));
set(gcf,'CurrentAxes',AX(1,1))
xlabel('De-noised Water Level (m)')
ylabel('Tidal Noise x (m)');
fn8=strcat(fname,'Fig500.bmp');
saveas(gcf,fn8);
%[H,AX,BigAx,P] = plotmatrix(...) returns a matrix of handles to the objects
%created in H, a matrix of handles to the individual subaxes in AX, a handle to a big
%(invisible) axes that frames the subaxes in BigAx, and a matrix of handles for the
%histogram plots in P. BigAx is left as the current axes so that a subsequent title,
%xlabel, or ylabel commands are centered with respect to the matrix of axes.
% set(gcf,'CurrentAxes',axes_handle) both make the
% axes identified by the handle axes_handle the current axes.

% PLOT CUMULATIVE DISTN OF i TO ns SERIES
figure(501)
[h,stats] = cdfplot(Noi(:,2))
hold on;
xx = -3:.1:3;
plot(xx,normcdf(xx),'r--');
pt0=strcat([fname ' Tidal Noise cdf'],'  Dash=Normal');
title(pt0);
xlabel('Tidal Noise x (m)');
hold off;
fn9=strcat(fname,'Fig501.bmp');
saveas(gcf,fn9);

% SAVE TOTAL JOINT SERIES
fn11=strcat(fname,out4);% denoised
fn12=strcat(fname,out5);% noise
fn13=strcat(fname,out6);% original
   f11 = fopen(fn11,'w');
      f12 = fopen(fn12,'w');
```

```
      f13 = fopen(fn13,'w');
fprintf(f11,'%f %f\n',Ssd');
   fprintf(f12,'%f %f\n',Noi');
      fprintf(f13,'%f %f\n',Sss');
fclose(f11)
   fclose(f12)
      fclose(f13)
```

Appendix A8

MATLAB Program For Deriving Correction Factor for Hourly Data, Number of Storm per Year, *r* Largest Storm and Fitting of Distribution

Appendix A8

MATLAB Program For Deriving Correction Factor for Hourly Data, Number of Storm per Year, *r* Largest Storm and Fitting of Distribution

```
% EVALUATING THETA: CORRECTION FACTOR FOR HOURLY STORM DATA
% theta= reciprocal of the mean time spent above a level eta
% on each independent excursion above this level
clear all
[N, R]=textread('siniawan.ssd','%f %f');    % reading column text file
k=0;

% FITTING PROCESS OPTION
% after choosing storm value and t, fit distribution by assigning ps=1
ps=0                                        % assign 1 for fitting distribution or 0 for none

if ps=0
% PROCESSING ALL POSSIBLE EXTREME ETA=t
for t=2.6:0.1:11.0;                         % assign for all ranges
k=k+1;
j=0;
c=0;
x=0;
theta=0;
for i=1:length(R)
   if R(i)>=t
      theta=theta+1;
      x=1;                                  % in exceedance mode
   elseif x==1
      c=c+1;                                % no of exceedance
      Ns(c)=N(i);                           % date-time of storm
      j(c)=theta;
      p(c)=max(R(i-theta:i));               % maxima in each storm
      x=0;
      theta=0;
   else
      x=0;
      theta=0;
   end
end
me(k)=mean(j);                  % mean of theta-1
```

```
ne(k)=me(k)*c;                  % total no hour exceeding=theta-1 x no of storm

% STORM PER YEAR
[Y,M,D,H,MI,S] = datevec(Ns);       % get date-time of storms
SD=hist(Y);                         % no. of storm in each year at this level
msl(k)=mean(SD);                    % mean no storm at each level k
end
me'
ne'
msl'
end

if ps==1                        % start fitting process

% FOR CHOSEN t VALUE
t=4.1                               % assign t=4.0
k= 1;
j=0;
c=0;
x=0;
theta=0;
for i=1:length(R)
   if R(i)>=t
      theta=theta+1;
      x=1;                          % in exceedance mode
   elseif x==1
      c=c+1;                        % no of exceedance
      Ns(c)=N(i);                   % time of storm
      j(c)=theta;
      p(c)=max(R(i-theta:i));       % maxima in each storm
      x=0;
      theta=0;
   else
      x=0;
      theta=0;
   end
end
me(k)=mean(j);                  % mean of theta-1
ne(k)=me(k)*c;                  % total no hour exceeding=theta-1 x no of storm

% STORM PER YEAR
[Y,M,D,H,MI,S] = datevec(Ns);       % get year etc back
SD=hist(Y);                         % no. of storm in each year at this level
msl(k)=mean(SD);                    % mean no storm at each level k
me'
```

```
ne'
msl'


% FITTING r EXTREME VALUE DISRTIBUTION
% Gumbel dist
u=21;                                   % assign # storm per year
m=6;                                    % assign no of data year
r=u*m;
s=length(p);
p=sort(p);
for q=1:s;
   pp(q)=p(s-q+1);
end
pm=pp(1:r);
for h=1:r
si(h)=(1/r)*(pm(h)-pm(r));
end
pm'
sigma=sum(si)
mu=sigma*log(r)+pm(r)
end
```

Appendix A9

MATLAB Program For MC Simulation and Flood Quantiles Determination by Model A Using Characteristics of Wavelet De-noised Series and Tidal Noise

Appendix A9

MATLAB Program For MC Simulation and Flood Quantiles Determination by Model A Using Characteristics of Wavelet De-noised Series and Tidal Noise

```
% MC SIMULATION AFTER WAVELET DE-NOSING- FOR JOINT PROBABILITY OF
% TIDES AND RIVER LEVELS
% Siniawan With Hourly data
% Simulation assume
% 1. De-nosied river pdf smoothed by Gumbel extreme in annual
% 2. Tidal Noise series generated by varying normal distribution
clear

% READ CDF TABLES
[pr vr]=textread('cdf_Rn.txt','%f %f');
[Rn Mu Sigma]=textread('meansigma.txt','%f %f %f');
cdfr=[pr vr];
parak=[Rn Mu Sigma];

% GENERATE FOR N YEARS
n=1000;
for j=1:n

% GENERATE FOR m HOURS IN EACH YEAR
m=8766;
for i=1:m;
p1=rand;
p2=normrnd(0,1);                    % random normal variate (mean 0, std 1)
R(i) = interp1(cdfr(:,1),cdfr(:,2),p1);    % lookup table
Rr=round(R(i));
mu = interp1(parak(:,1),parak(:,2),Rr);
sigma = interp1(parak(:,1),parak(:,3),Rr);
K(i)=mu+sigma*p2;
Zh(i)=R(i)+K(i);
end
Z(j)=max(Zh) % abstract maximum of year
j
'one year done'
end
Z=sort(Z);
```

```
% SAMPLE/RETURN PERIOD N YEAR
p2=n/2;
p5=n/5;
p10=n/10;
p50=n/50;
p100=n/100;
p200=n/200;

% QUANTILE FROM EMPIRICAL DIST
Z2=Z(n-p2+1)
Z5=Z(n-p5+1)
Z10=Z(n-p10+1)
Z50=Z(n-p50+1)
Z100=Z(n-p100+1)
Z200=Z(n-p200+1)

% SAVE FILES
fname1='SinModelAexp.out1';
   fid = fopen(fname1,'w');
      fprintf(fid,'  Results of Model A exponential decrease simulation T=2,5,10,50,100,200 \n\n');
   fprintf(fid,'%f %f %f %f %f %f\n',Z2,Z5,Z10,Z50,Z100,Z200);
   fprintf(fid,'  Results of simulation T=2,5,10,50,100,200 \n\n');
    fprintf(fid,'%f \n',Z);
fclose(fid)
```

Appendix A10

MATLAB Program For MC Simulation and Flood Quantiles Determination by Model B Using Characteristics of Wavelet De-noised Series and Tidal Noise

Appendix A10

MATLAB Program For MC Simulation and Flood Quantiles Determination by Model B Using Characteristics of Wavelet De-noised Series and Tidal Noise

```
% MODEL B MC SIMULATION AFTER WAVELET DE-NOSING- FOR JOINT PROBABILITY OF
% WITH HMAX THRESHOLDING TIDES AND RIVER LEVELS
% Siniawan With Hourly data
% Simulation assume
% 1. De-nosied river pdf smoothed by Gumbel extreme in annual
% 2. Tidal Noise series generated by varying normal distribution

clear
% DECLARE THE THRESHOLDING LEVEL
Hmax=6.0

% READ CDF TABLES
[pr vr]=textread('cdf_Rn.txt','%f %f');
[Rn Mu Sigma]=textread('meansigma.txt','%f %f %f');
cdfr=[pr vr];
parak=[Rn Mu Sigma];

% GENERATE FOR N YEARS
n=1000;
for j=1:n

% GENERATE FOR m HOURS IN EACH YEAR
m=8766;
for i=1:m;
p1=rand;
R(i) = interp1(cdfr(:,1),cdfr(:,2),p1); % lookup table
if R(i)>Hmax;
   Zh(i)=R(i);
else
Rr=round(R(i));
p2=normrnd(0,1); % random normal variate (mean 0, std 1)
mu = interp1(parak(:,1),parak(:,2),Rr);
sigma = interp1(parak(:,1),parak(:,3),Rr);
K(i)=mu+sigma*p2;
Zh(i)=R(i)+K(i);
end
```

```
end
Z(j)=max(Zh)                    % abstract maximum of year
j
'one year done'
end
Z=sort(Z);

% FILE OUTPUT
fname1='B-Sin-hour.out1';
   fid = fopen(fname1,'w');

% SAMPLE/RETURN PERIOD N YEAR
p2=n/2;
p5=n/5;
p10=n/10;
p50=n/50;
p100=n/100;
p200=n/200;

% QUANTILE FROM EMPIRICAL DIST
Z2=Z(n-p2+1)
Z5=Z(n-p5+1)
Z10=Z(n-p10+1)
Z50=Z(n-p50+1)
Z100=Z(n-p100+1)
Z200=Z(n-p200+1)

  fprintf(fid,'  Results of Model B simulation T=2,5,10,50,100,200 \n\n');
  fprintf(fid,'%f %f %f %f %f %f\n',Z2,Z5,Z10,Z50,Z100,Z200);
  fprintf(fid,'  Results of simulation T=2,5,10,50,100,200 \n\n');
  fprintf(fid,'  Z\n');
   fprintf(fid,'%f \n',Z);
   fclose(fid)
```

Appendix B1

Direct Distribution Fitting for Annual Maxima Series in the TIZ

Table B1-1 Direct Distribution Fitting for Annual Maxima Series at Satok

Distribution Type	No. of Para.	Method	Standard Error	Rank by SE	Method Reference
Normal MOM	2	MOM	72.00	6	
Normal LM	2	LM	73.74	7	Hosking
Lognormal (2) MOM	2	MOM	71.88	5	
Lognormal (2) LM	2	LM	77.94	10	Hosking
Lognormal (3) MOM	3	MOM	76.60	9	
Lognormal (3) LM	3	LM	83.08	15	Hosking
Gumbel or EV1(2) MOM	2	MOM	83.44	17	
Gumbel or EV1(2) LM	2	LM	89.16	21	Hosking
LogGumbel (2) MOM	2	MOM	86.79	20	
LogGumbel (2) PWM Dir	2	PWM	92.65	23	Heo
LogGumbel (2) LM Ind	2	LM	93.67	24	Hosking
LogGumbel (3) MOM	3	MOM	83.59	18	Heo
LogGumbel (3) PWM	3	PWM	78.11	12	Heo
Pearson III MOM	3	MOM	76.51	8	
Pearson III LM	3	LM	81.51	13	Hosking
Log Pearson III MOM Dir	3	MOM	127.58	25	Rao
Log Pearson III MOM Ind	3	MOM	70.04	4	Rao
Log Pearson III PWM	3	PWM	83.27	16	Rao
Generalized Pareto MOM	3	MOM	58.65	1	
Generalized Pareto LM	3	LM	63.62	2	Hosking
GEV MOM	3	MOM	69.95	3	
GEV LM	3	LM	78.07	11	Hosking
Logistic MOM	2	MOM	82.87	14	
Logistic LM	2	LM	85.18	19	Hosking
Generalized Logistic LM	3	LM	90.55	22	Hosking

Quantile Estimates

Distribution Type and No. of Parameter		Method	Return Period				Rank by SE
			2	10	50	100	
Generalized Pareto MOM	3	MOM	5750	6065	6159	6175	1
Generalized Pareto LM	3	LM	5742	6076	6194	6217	2
GEV MOM	3	MOM	5764	6038	6189	6236	3
Log Pearson III MOM Ind	3	MOM	5757	6036	6225	6295	4
Lognormal (2) MOM	2	MOM	5764	6032	6199	6259	5
Normal MOM	2	MOM	5768	6030	6188	6244	6
Normal LM	2	LM	5768	6040	6203	6261	7
Pearson III MOM	3	MOM	5759	6034	6217	6283	8
Lognormal (3) MOM	3	MOM	5759	6034	6214	6281	9
Lognormal (2) LM	2	LM	5764	6041	6214	6276	10
GEV LM	3	LM	5891	6171	6370	6444	11
LogGumbel LM	3	LM	5746	6054	6275	6356	12

Table B1-2 Direct Distribution Fitting for Annual Maxima Series at Batu Kawa

Distribution Type	No. of Para.	Method	Standard Error	Rank by SE	Method Reference
Normal MOM	2	MOM	106.77	10	
Normal LM	2	LM	108.91	11	Hosking
Lognormal (2) MOM	2	MOM	117.59	14	
Lognormal (2) LM	2	LM	125.12	18	Hosking
Lognormal (3) MOM	3	MOM	127.68	19	
Lognormal (3) LM	3	LM	99.14	7	Hosking
Gumbel or EV1(2) MOM	2	MOM	164.94	20	
Gumbel or EV1(2) LM	2	LM	174.87	21	Hosking
LogGumbel (2) MOM	2	MOM	183.99	22	
LogGumbel (2) PWM Dir	2	PWM	191.29	23	Heo
LogGumbel (2) LM Ind	2	LM	198.96	24	Hosking
LogGumbel (3) MOM	3	MOM	94.71	1	Heo
LogGumbel (3) PWM	3	PWM	95.93	3	Heo
Pearson III MOM	3	MOM	102.08	8	
Pearson III LM	3	LM	98.52	4	Hosking
Log Pearson III MOM Dir	3	MOM	214.42	25	Rao
Log Pearson III MOM Ind	3	MOM	98.87	5	Rao
Log Pearson III PWM	3	PWM	98.91	6	Rao
Generalized Pareto MOM	3	MOM	119.53	16	
Generalized Pareto LM	3	LM	121.53	17	Hosking
GEV MOM	3	MOM	111.97	12	
GEV LM	3	LM	95.92	2	Hosking
Logistic MOM	2	MOM	116.38	13	
Logistic LM	2	LM	118.60	15	Hosking
Generalized Logistic LM	3	LM	103.07	9	Hosking

Quantile Estimates

Distribution Type and No. of Parameter		Method	Return Period				Rank by SE
			2	10	50	100	
LogGumbel (3) MOM	3	MOM	6209	6706	6964	7048	1
GEV LM	3	LM	6500	6818	6933	6960	2
LogGumbel (3) PWM	3	PWM	6232	6712	6886	6926	3
Pearson III LM	3	LM	6231	6696	6911	6976	4
Log Pearson III MOM Ind	3	MOM	6201	6713	6991	7083	5
Log Pearson III PWM	3	PWM	6232	6693	6897	6955	6
Lognormal (3) LM	3	LM	6232	6702	6928	6998	7
Pearson III MOM	3	MOM	6204	6707	6981	7071	8
Generalized Logistic LM	3	LM	6227	6694	6983	7087	9
Normal MOM	2	MOM	6183	6723	7048	7163	10
Normal LM	2	LM	6183	6738	7072	7190	11
GEV MOM	3	MOM	6175	6739	7051	7148	12

Table B1-3 Direct Distribution Fitting for Annual Maxima Series at Batu Kitang

Distribution Type	No. of Para.	Method	Standard Error	Rank by SE	Method Reference
Normal MOM	2	MOM	268.90	6	
Normal LM	2	LM	272.15	8	Hosking
Lognormal (2) MOM	2	MOM	318.83	15	
Lognormal (2) LM	2	LM	334.57	16	Hosking
Lognormal (3) MOM	3	MOM	316.38	14	
Lognormal (3) LM	3	LM	267.35	4	Hosking
Gumbel or EV1(2) MOM	2	MOM	365.15	17	
Gumbel or EV1(2) LM	2	LM	383.88	21	Hosking
LogGumbel (2) MOM	2	MOM	449.91	24	
LogGumbel (2) PWM Dir	2	PWM	439.64	23	Heo
LogGumbel (2) LM Ind	2	LM	486.73	25	Hosking
LogGumbel (3) MOM	3	MOM	365.30	18	Heo
LogGumbel (3) PWM	3	PWM	261.10	3	Heo
Pearson III MOM	3	MOM	273.07	9	
Pearson III LM	3	LM	268.03	5	Hosking
Log Pearson III MOM Dir	3	MOM	427.76	22	Rao
Log Pearson III MOM Ind	3	MOM	272.13	7	Rao
Log Pearson III PWM	3	PWM	278.76	10	Rao
Generalized Pareto MOM	3	MOM	366.73	19	
Generalized Pareto LM	3	LM	368.90	20	Hosking
GEV MOM	3	MOM	283.63	13	
GEV LM	3	LM	261.09	2	Hosking
Logistic MOM	2	MOM	280.09	11	
Logistic LM	2	LM	282.91	12	Hosking
Generalized Logistic LM	3	LM	259.22	1	Hosking

Quantile Estimates

Distribution Type and No. of Parameter		Method	Return Period				Rank by SE
			2	10	50	100	
Generalized Logistic LM	3	LM	6686	7712	8354	8588	1
GEV LM	3	LM	7284	7990	8250	8311	2
LogGumbel (3) PWM	3	PWM	6696	7754	8143	8234	3
Lognormal (3) LM	3	LM	6696	7732	8235	8392	4
Pearson III LM	3	LM	6695	7720	8201	8346	5
Normal MOM	2	MOM	6593	7783	8500	8753	6
Log Pearson III MOM Ind	3	MOM	6639	7769	8352	8537	7
Normal LM	2	LM	6593	7806	8537	8795	8
Pearson III MOM	3	MOM	6641	7747	8350	8548	9
Log Pearson III PWM	3	PWM	6711	7682	8064	8162	10
Logistic MOM	2	MOM	6593	7718	8586	8946	11
Logistic LM	2	LM	6593	7766	8672	9047	12

Table B1-4 Direct Distribution Fitting for Annual Maxima Series at Siniawan

Distribution Type	No. of Para.	Method	Standard Error	Rank by SE	Method Reference
Normal MOM	2	MOM	405.26	10	
Normal LM	2	LM	426.98	12	Hosking
Lognormal (2) MOM	2	MOM	539.08	17	
Lognormal (2) LM	2	LM	587.90	18	Hosking
Lognormal (3) MOM	3	MOM	507.55	16	
Lognormal (3) LM	3	LM	380.62	8	Hosking
Gumbel or EV1(2) MOM	2	MOM	623.21	19	
Gumbel or EV1(2) LM	2	LM	679.74	21	Hosking
LogGumbel (2) MOM	2	MOM	831.46	23	
LogGumbel (2) PWM Dir	2	PWM	831.84	24	Heo
LogGumbel (2) LM Ind	2	LM	950.13	25	Hosking
LogGumbel (3) MOM	3	MOM	623.55	20	Heo
LogGumbel (3) PWM	3	PWM	330.74	3	Heo
Pearson III MOM	3	MOM	386.14	9	
Pearson III LM	3	LM	372.02	7	Hosking
Log Pearson III MOM Dir	3	MOM	719.54	22	Rao
Log Pearson III MOM Ind	3	MOM	350.14	5	Rao
Log Pearson III PWM	3	PWM	361.47	6	Rao
Generalized Pareto MOM	3	MOM	296.23	1	
Generalized Pareto LM	3	LM	326.15	2	Hosking
GEV MOM	3	MOM	412.62	11	
GEV LM	3	LM	330.83	4	Hosking
Logistic MOM	2	MOM	485.25	14	
Logistic LM	2	LM	505.65	15	Hosking
Generalized Logistic LM	3	LM	429.74	13	Hosking

Quantile Estimates

Distribution Type and No. of Parameter		Method	Return Period				Rank by SE
			2	10	50	100	
Generalized Pareto MOM	3	MOM	8548	10103	10245	10255	1
Generalized Pareto LM	3	LM	8548	10194	10350	10360	2
LogGumbel (3) PWM	3	PWM	8492	10205	10843	10993	3
GEV LM	3	LM	9441	10590	11018	11119	4
Log Pearson III MOM Ind	3	MOM	8364	10200	11207	11538	5
Log Pearson III PWM	3	PWM	8486	10141	10844	11034	6
Pearson III LM	3	LM	8490	10150	10937	11175	7
Lognormal (3) LM	3	LM	8492	10169	10989	11245	8
Pearson III MOM	3	MOM	8416	10110	11017	11312	9
Normal MOM	2	MOM	8331	10174	11285	11677	10
GEV MOM	3	MOM	8305	10232	11296	11625	11
Normal LM	2	LM	8331	10284	11461	11876	12

Table B1-5 Direct Distribution Fitting for Annual Maxima Series at Sungai Merah

Distribution Type	No. of Para.	Method	Standard Error	Rank by SE	Method Reference
Normal MOM	2	MOM	97.19	23	
Normal LM	2	LM	95.87	19	Hosking
Lognormal (2) MOM	2	MOM	90.52	18	
Lognormal (2) LM	2	LM	98.76	25	Hosking
Lognormal (3) MOM	3	MOM	58.17	11	
Lognormal (3) LM	3	LM	50.80	4	Hosking
Gumbel or EV1(2) MOM	2	MOM	61.52	13	
Gumbel or EV1(2) LM	2	LM	61.79	14	Hosking
LogGumbel (2) MOM	2	MOM	56.03	7	
LogGumbel (2) PWM Dir	2	PWM	56.58	8	Heo
LogGumbel (2) LM Ind	2	LM	56.81	9	Hosking
LogGumbel (3) MOM	3	MOM	54.25	6	Heo
LogGumbel (3) PWM	3	PWM	63.11	16	Heo
Pearson III MOM	3	MOM	57.60	10	
Pearson III LM	3	LM	51.76	5	Hosking
Log Pearson III MOM Dir	3	MOM	63.01	15	Rao
Log Pearson III MOM Ind	3	MOM	75.15	17	Rao
Log Pearson III PWM	3	PWM	49.62	3	Rao
Generalized Pareto MOM	3	MOM	97.44	24	
Generalized Pareto LM	3	LM	60.97	12	Hosking
GEV MOM	3	MOM	96.90	22	
GEV LM	3	LM	39.45	2	Hosking
Logistic MOM	2	MOM	96.81	21	
Logistic LM	2	LM	96.68	20	Hosking
Generalized Logistic LM	3	LM	38.25	1	Hosking

Quantile Estimates

Distribution Type and No. of Parameter		Method	Return Period				Rank by SE
			2	10	50	100	
Generalized Logistic LM	3	LM	4040	4389	4920	5252	1
GEV LM	3	LM	4177	4640	5289	5662	2
Log Pearson III PWM	3	PWM	4040	4407	4818	5009	3
Lognormal (3) LM	3	LM	4032	4425	4931	5190	4
Pearson III LM	3	LM	4039	4403	4793	4966	5
LogGumbel (3) MOM	3	MOM	4047	3875	3842	3838	6
LogGumbel (2) MOM	2	MOM	4067	4419	4753	4902	7
LogGumbel (2) PWM Dir	2	PWM	4065	4425	4767	4920	8
LogGumbel (2) LM Ind	2	LM	4066	4425	4766	4918	9
Pearson III MOM	3	MOM	4051	4426	4781	4931	10
Lognormal (3) MOM	3	MOM	4061	4418	4761	4915	11
Generalized Pareto LM	3	LM	4027	4447	4897	5100	12

Table B1-6 Direct Distribution Fitting for Annual Maxima Series at Sungai Salim

Distribution Type	No. of Para.	Method	Standard Error	Rank by SE	Method Reference
Normal MOM	2	MOM	120.07	11	
Normal LM	2	LM	123.05	14	Hosking
Lognormal (2) MOM	2	MOM	128.04	15	
Lognormal (2) LM	2	LM	142.24	17	Hosking
Lognormal (3) MOM	3	MOM	155.47	19	
Lognormal (3) LM	3	LM	110.75	9	Hosking
Gumbel or EV1(2) MOM	2	MOM	162.27	20	
Gumbel or EV1(2) LM	2	LM	171.74	21	Hosking
LogGumbel (2) MOM	2	MOM	174.29	22	
LogGumbel (2) PWM Dir	2	PWM	181.36	23	Heo
LogGumbel (2) LM Ind	2	LM	186.88	24	Hosking
LogGumbel (3) MOM	3	MOM	102.64	5	Heo
LogGumbel (3) PWM	3	PWM	94.61	4	Heo
Pearson III MOM	3	MOM	113.30	10	
Pearson III LM	3	LM	106.73	8	Hosking
Log Pearson III MOM Dir	3	MOM	213.27	25	Rao
Log Pearson III MOM Ind	3	MOM	104.34	6	Rao
Log Pearson III PWM	3	PWM	105.52	7	Rao
Generalized Pareto MOM	3	MOM	75.97	1	
Generalized Pareto LM	3	LM	85.24	2	Hosking
GEV MOM	3	MOM	122.58	13	
GEV LM	3	LM	94.50	3	Hosking
Logistic MOM	2	MOM	139.25	16	
Logistic LM	2	LM	142.98	18	Hosking
Generalized Logistic LM	3	LM	121.15	12	Hosking

Quantile Estimates

Distribution Type and No. of Parameter		Method	Return Period				Rank by SE
			2	10	50	100	
Generalized Pareto MOM	3	MOM	5117	5372	5384	5385	1
Generalized Pareto LM	3	LM	5136	5372	5380	5381	2
GEV LM	3	LM	5280	5434	5471	5478	3
LogGumbel (3) PWM	3	PWM	5112	5389	5456	5467	4
LogGumbel (3) MOM	3	MOM	5072	5396	5553	5604	5
Log Pearson III MOM Ind	3	MOM	5065	5406	5576	5629	6
Log Pearson III PWM	3	PWM	5104	5369	5449	5466	7
Pearson III LM	3	LM	5106	5369	5453	5472	8
Lognormal (3) LM	3	LM	5109	5385	5490	5519	9
Pearson III MOM	3	MOM	5070	5398	5557	5606	10
Normal MOM	2	MOM	5042	5422	5651	5731	11
Generalized Logistic LM	3	LM	5102	5383	5525	5570	12

Table B1-7 Direct Distribution Fitting for Annual Maxima Series at Marudi

Distribution Type	No. of Para.	Method	Standard Error	Rank by SE	Method Reference
Normal MOM	2	MOM	171.12	5	
Normal LM	2	LM	177.82	6	Hosking
Lognormal (2) MOM	2	MOM	165.89	3	
Lognormal (2) LM	2	LM	188.60	12	Hosking
Lognormal (3) MOM	3	MOM	178.90	8	
Lognormal (3) LM	3	LM	209.09	20	Hosking
Gumbel or EV1(2) MOM	2	MOM	181.04	9	
Gumbel or EV1(2) LM	2	LM	199.31	16	Hosking
LogGumbel (2) MOM	2	MOM	206.30	18	
LogGumbel (2) PWM Dir	2	PWM	225.02	24	Heo
LogGumbel (2) LM Ind	2	LM	237.03	25	Hosking
LogGumbel (3) MOM	3	MOM	181.22	10	Heo
LogGumbel (3) PWM	3	PWM	199.20	15	Heo
Pearson III MOM	3	MOM	178.31	7	
Pearson III LM	3	LM	196.50	14	Hosking
Log Pearson III MOM Dir	3	MOM	213.99	21	Rao
Log Pearson III MOM Ind	3	MOM	183.98	11	Rao
Log Pearson III PWM	3	PWM	218.91	22	Rao
Generalized Pareto MOM	3	MOM	135.83	1	
Generalized Pareto LM	3	LM	170.94	4	Hosking
GEV MOM	3	MOM	164.50	2	
GEV LM	3	LM	196.23	13	Hosking
Logistic MOM	2	MOM	199.78	17	
Logistic LM	2	LM	206.82	19	Hosking
Generalized Logistic LM	3	LM	223.39	23	Hosking

Quantile Estimates

Distribution Type and No. of Parameter		Method	Return Period				Rank by SE
			2	10	50	100	
Generalized Pareto MOM	3	MOM	4520	5370	5687	5753	1
Log Pearson III MOM Ind	3	MOM	4537	5296	5872	6102	2
GEV MOM	3	MOM	4581	5287	5678	5798	3
Lognormal (2) MOM	2	MOM	4564	5279	5763	5944	4
Generalized Pareto LM	3	LM	4484	5418	5875	5991	5
Normal MOM	2	MOM	4590	5266	5674	5818	6
Normal LM	2	LM	4590	5301	5729	5881	7
Pearson III MOM	3	MOM	4555	5285	5793	5982	8
Lognormal (3) MOM	3	MOM	4555	5283	5786	5977	9
Gumbel or EV1(2) MOM	2	MOM	4504	5279	5958	6245	10
LogGumbel (3) MOM	3	MOM	4503	5278	5959	6247	11
Lognormal (2) LM	2	LM	4557	5321	5842	6038	12

Appendix B2

Plots of Best Distributions For Annual Series in the TIZ

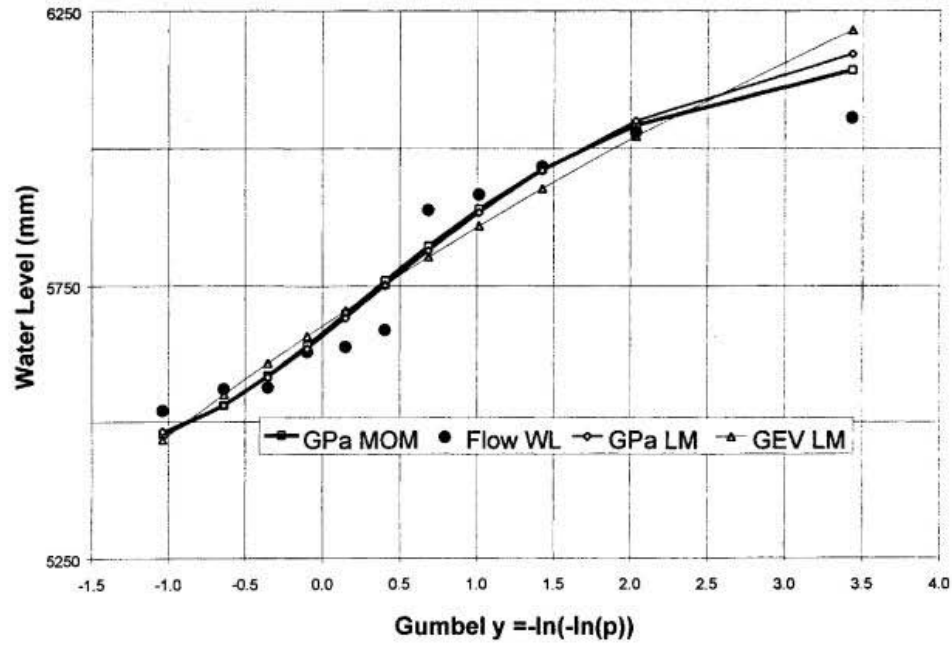


Figure B2-1 Best Distributions for Annual Extremes at Satok

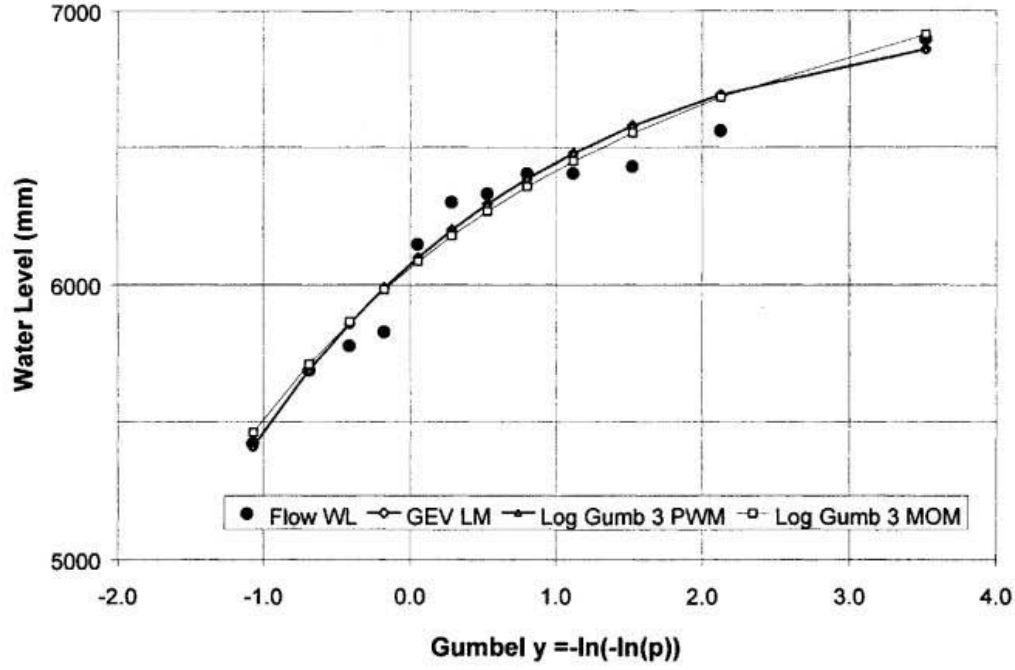


Figure B2-2 Best Distributions for Annual Extremes at Batu Kawa

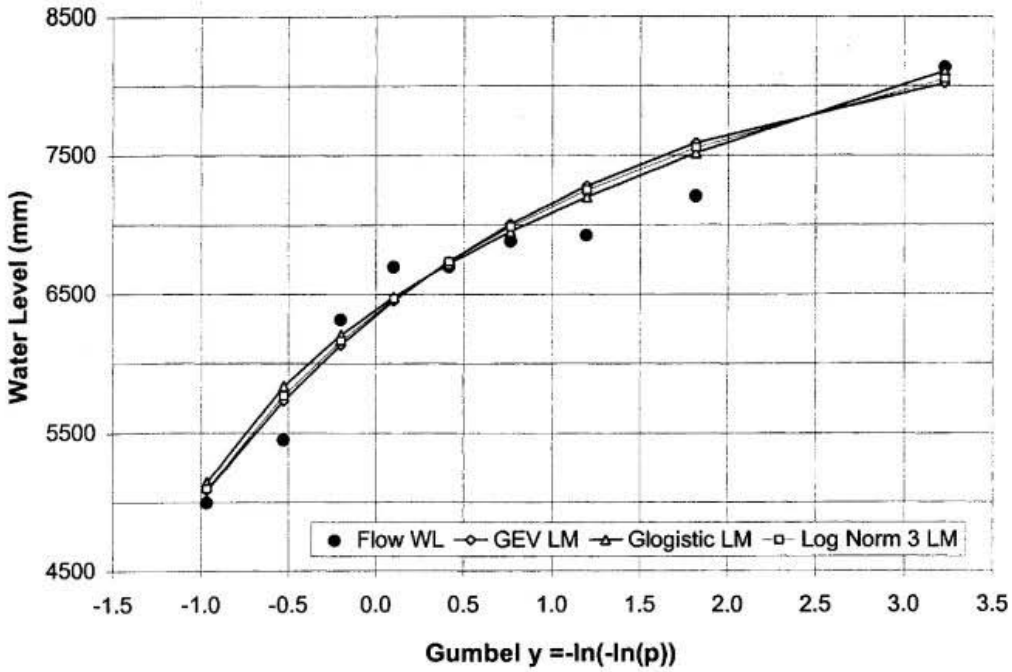


Figure B2-3 Best Distributions for Annual Extremes at Batu Kitang

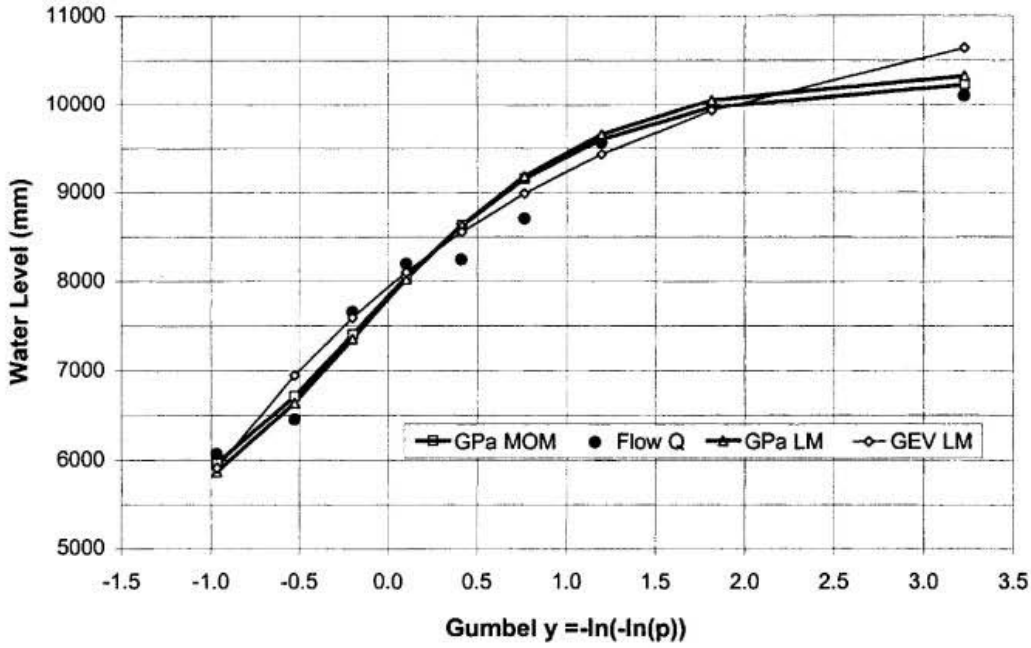


Figure B2-4 Best Distributions for Annual Extremes at Siniawan

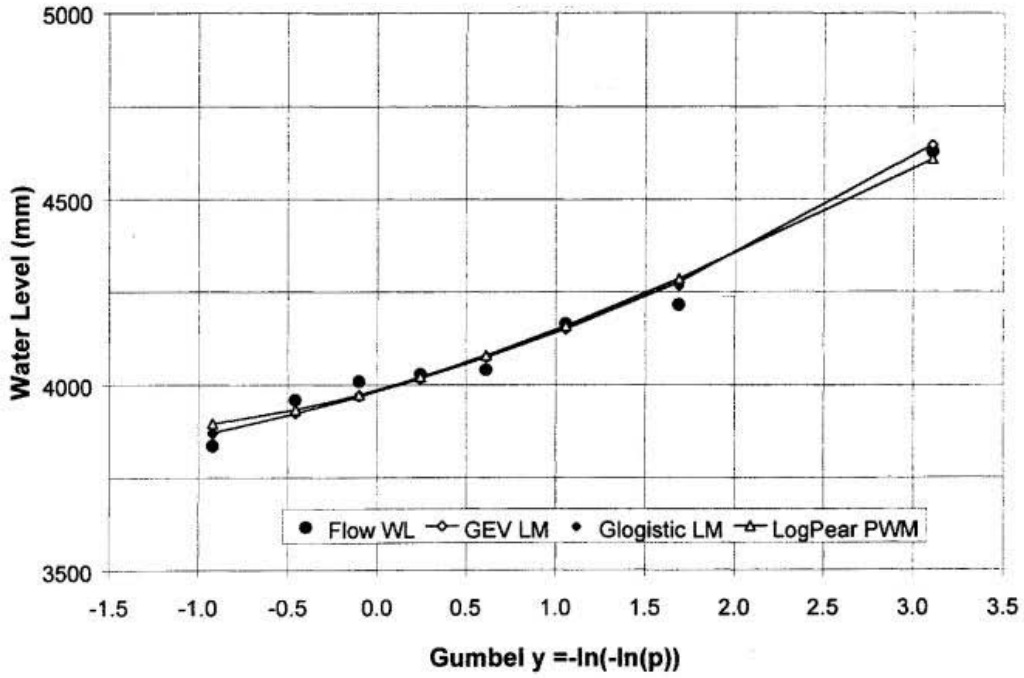


Figure B2-5 Best Distributions for Annual Extremes at Sungai Merah

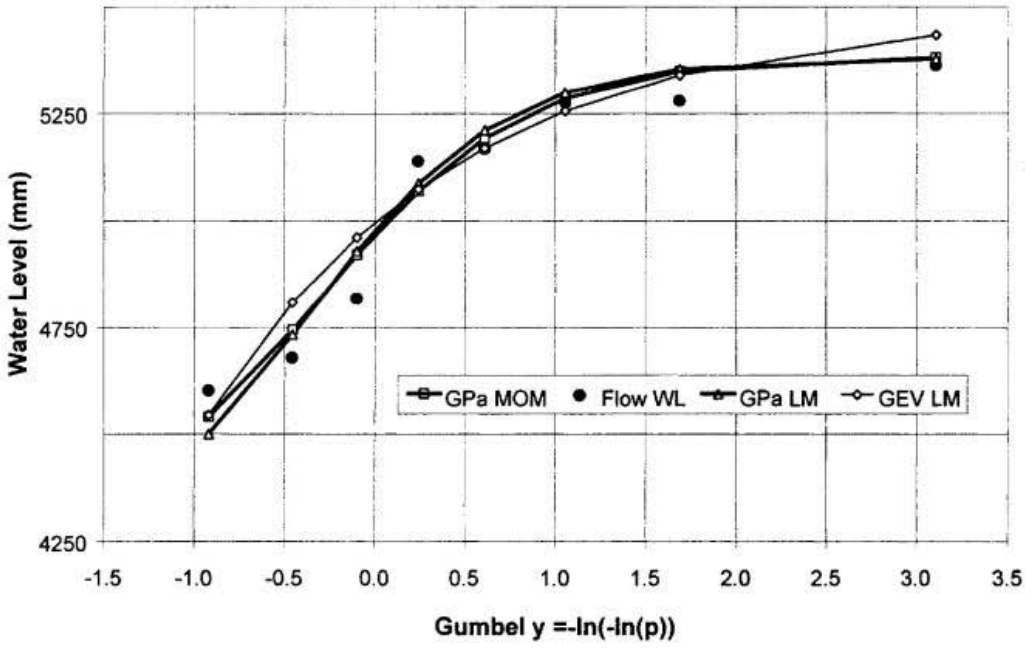


Figure B2-6 Best Distributions for Annual Extremes at Sungai Salim

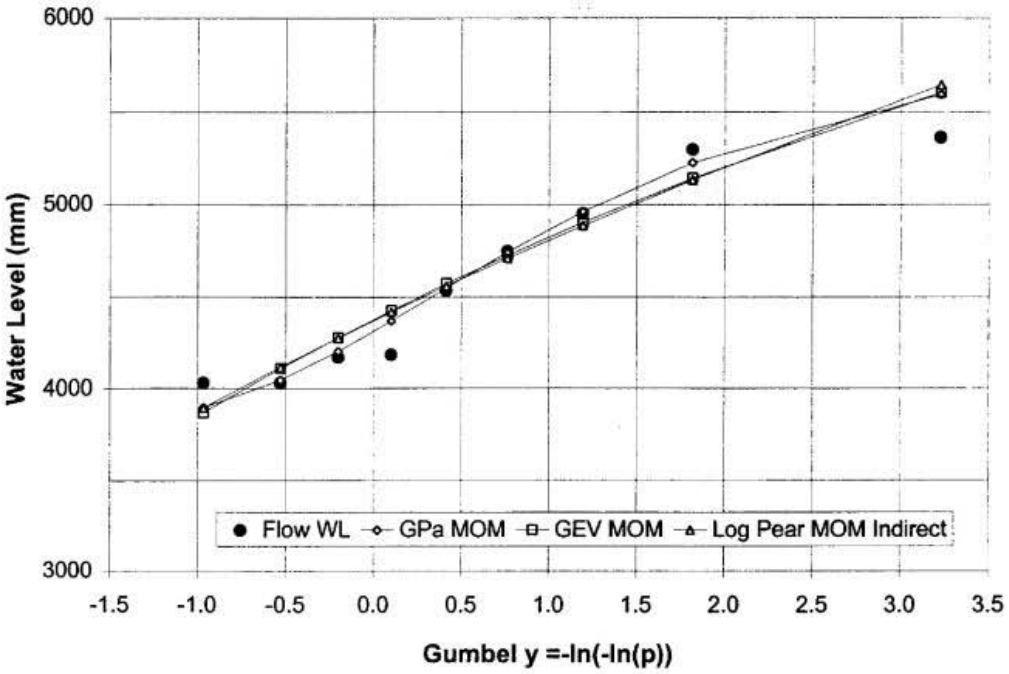


Figure B2-7 Best Distributions for Annual Extremes at Marudi

Appendix B3

Probability Calculation Using the Bivariate Distribution Method

Appendix B3
Probability Calculation Using the Bivariate Distribution Method

Buan Bidi Qi	Probability of Q P(Qi)	Group of Z, Water Level in TIZ					
		Z=4			Z=5		
		H = 3.4717e-0.0043Q			H = 4.1839e-0.0029Q		
				$R^2 = 0.059$			$R^2 = 0.1$
		Hi	P(Hi)	JP	Hi	P(Hi)	JP
5	0.10521	3.4	0.0279	0.002938	4.1	0.0312	0.0032774
10	0.60512	3.3	0.0277	0.016735	4.1	0.0312	0.0188503
20	0.15212	3.2	0.0280	0.004259	3.9	0.0321	0.0048855
30	0.05106	3.1	0.0274	0.001397	3.8	0.0307	0.0015671
40	0.02472	2.9	0.0295	0.000730	3.7	0.0304	0.0007513
50	0.01529	2.8	0.0286	0.000438	3.6	0.0307	0.0004692
60	0.00932	2.7	0.0282	0.000263	3.5	0.0288	0.0002685
70	0.00664	2.6	0.0285	0.000189	3.4	0.0279	0.0001854
80	0.00456	2.5	0.0271	0.000124	3.3	0.0277	0.0001262
90	0.00371	2.4	0.0258	0.000096	3.2	0.0280	0.0001039
100	0.00304	2.3	0.0261	0.000079	3.1	0.0274	0.0000832
110	0.00228	2.2	0.0236	0.000054	3.0	0.0278	0.0000632
120	0.00190	2.1	0.0228	0.000043	3.0	0.0278	0.0000529
130	0.00189	2.0	0.0222	0.000042	2.9	0.0295	0.0000558
140	0.00162	1.9	0.0215	0.000035	2.8	0.0286	0.0000464
150	0.00141	1.8	0.0214	0.000030	2.7	0.0282	0.0000397
160	0.00125	1.7	0.0191	0.000024	2.6	0.0285	0.0000355
170	0.00129	1.7	0.0191	0.000025	2.6	0.0285	0.0000366
180	0.00113	1.6	0.0185	0.000021	2.5	0.0271	0.0000305
190	0.00096	1.5	0.0168	0.000016	2.4	0.0258	0.0000249
200	0.00100	1.5	0.0168	0.000017	2.3	0.0261	0.0000261
210	0.00087	1.4	0.0155	0.000013	2.3	0.0261	0.0000226
220	0.00072	1.3	0.0142	0.000010	2.2	0.0236	0.0000169
230	0.00059	1.3	0.0142	0.000008	2.1	0.0228	0.0000135
240	0.00053	1.2	0.0122	0.000006	2.1	0.0228	0.0000121
250	0.00049	1.2	0.0122	0.000006	2.0	0.0222	0.0000110
260	0.00028	1.1	0.0105	0.000003	2.0	0.0222	0.0000063
270	0.00022	1.1	0.0105	0.000002	1.9	0.0215	0.0000048
280	0.00016	1.0	0.0096	0.000002	1.9	0.0215	0.0000034
290	0.00009	1.0	0.0096	0.000001	1.8	0.0214	0.0000019
300	0.00009	1.0	0.0096	0.000001	1.8	0.0214	0.0000019
310	0.00010	0.9	0.0081	0.000001	1.7	0.0191	0.0000019
320	0.00012	0.9	0.0081	0.000001	1.7	0.0191	0.0000024
330	0.00009	0.8	0.0060	0.000001	1.6	0.0185	0.0000016
340	0.00011	0.8	0.0060	0.000001	1.6	0.0185	0.0000021
350	0.00001	0.8	0.0060	0.000000	1.5	0.0168	0.0000002
Sum of Probabilities				**0.028**			**0.031**

Appendix B-3 Continue ..

Group of Z, Water Level in TIZ								
Z=6			Z=7			Z>8		
H = 3.4927e-0.0011Q			H = 3.902e-0.0013Q			H = -0.0598Ln(Q) + 3.4287		
		$R^2 = 0.028$			$R^2 = 0.060$			$R^2 = 0.002$
Hi	P(Hi)	JP	Hi	P(Hi)	JP	Hi	P(Hi)	JP
3.5	0.0288	0.0030290	3.9	0.0321	0.0033790	3.3	0.0277	0.0029095
3.5	0.0288	0.0174218	3.9	0.0321	0.0194345	3.29	0.0280	0.0169407
3.4	0.0279	0.0042475	3.8	0.0307	0.0046683	3.25	0.0280	0.0042586
3.4	0.0279	0.0014259	3.8	0.0307	0.0015671	3.23	0.0280	0.0014296
3.3	0.0277	0.0006837	3.7	0.0304	0.0007513	3.21	0.0280	0.0006921
3.3	0.0277	0.0004227	3.7	0.0304	0.0004645	3.19	0.0274	0.0004183
3.3	0.0277	0.0002579	3.6	0.0307	0.0002862	3.18	0.0274	0.0002552
3.2	0.0280	0.0001859	3.6	0.0307	0.0002039	3.17	0.0274	0.0001817
3.2	0.0280	0.0001278	3.5	0.0288	0.0001314	3.17	0.0274	0.0001249
3.2	0.0280	0.0001039	3.5	0.0288	0.0001068	3.16	0.0274	0.0001015
3.1	0.0274	0.0000832	3.4	0.0279	0.0000850	3.15	0.0274	0.0000832
3.1	0.0274	0.0000623	3.4	0.0279	0.0000635	3.15	0.0274	0.0000623
3.1	0.0274	0.0000521	3.3	0.0277	0.0000527	3.14	0.0274	0.0000521
3.0	0.0278	0.0000525	3.3	0.0277	0.0000523	3.14	0.0274	0.0000518
3.0	0.0278	0.0000450	3.3	0.0277	0.0000448	3.13	0.0274	0.0000443
3.0	0.0278	0.0000391	3.2	0.0280	0.0000395	3.13	0.0274	0.0000386
2.9	0.0295	0.0000369	3.2	0.0280	0.0000350	3.13	0.0274	0.0000342
2.9	0.0295	0.0000380	3.1	0.0274	0.0000352	3.12	0.0274	0.0000352
2.9	0.0295	0.0000332	3.1	0.0274	0.0000308	3.12	0.0274	0.0000308
2.8	0.0286	0.0000276	3.0	0.0278	0.0000268	3.11	0.0274	0.0000264
2.8	0.0286	0.0000287	3.0	0.0278	0.0000278	3.11	0.0274	0.0000274
2.8	0.0286	0.0000248	3.0	0.0278	0.0000240	3.11	0.0274	0.0000237
2.7	0.0282	0.0000202	2.9	0.0295	0.0000212	3.11	0.0274	0.0000196
2.7	0.0282	0.0000167	2.9	0.0295	0.0000175	3.10	0.0274	0.0000162
2.7	0.0282	0.0000150	2.9	0.0295	0.0000157	3.10	0.0274	0.0000146
2.7	0.0282	0.0000139	2.8	0.0286	0.0000142	3.10	0.0278	0.0000137
2.6	0.0285	0.0000081	2.8	0.0286	0.0000081	3.10	0.0278	0.0000079
2.6	0.0285	0.0000063	2.7	0.0282	0.0000063	3.09	0.0278	0.0000062
2.6	0.0285	0.0000046	2.7	0.0282	0.0000045	3.09	0.0278	0.0000045
2.5	0.0271	0.0000023	2.7	0.0282	0.0000024	3.09	0.0278	0.0000024
2.5	0.0271	0.0000023	2.6	0.0285	0.0000025	3.09	0.0278	0.0000024
2.5	0.0271	0.0000027	2.6	0.0285	0.0000028	3.09	0.0278	0.0000027
2.5	0.0271	0.0000034	2.6	0.0285	0.0000035	3.08	0.0278	0.0000034
2.4	0.0258	0.0000022	2.5	0.0271	0.0000023	3.08	0.0278	0.0000024
2.4	0.0258	0.0000029	2.5	0.0271	0.0000030	3.08	0.0278	0.0000031
2.4	0.0258	0.0000003	2.5	0.0271	0.0000003	3.08	0.0278	0.0000003
		0.029			**0.032**			**0.028**

Note: P=Probability of occurrence, JP= Joint probability of occurrence

Appendix C1

Filtered Series of Siniawan Using the Fourier Method

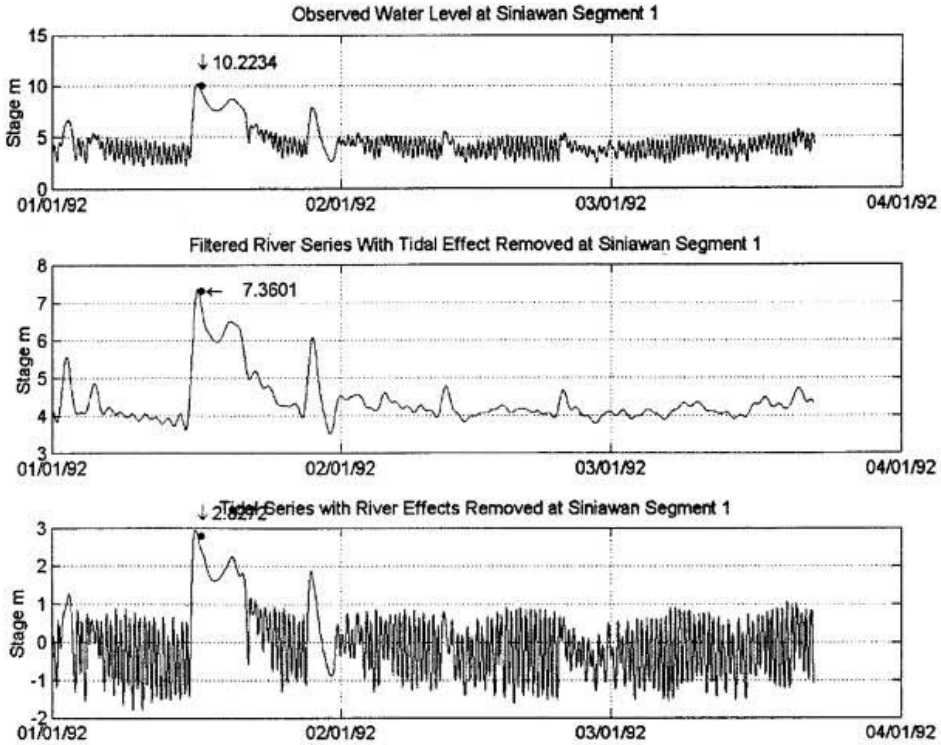


Figure C1-1

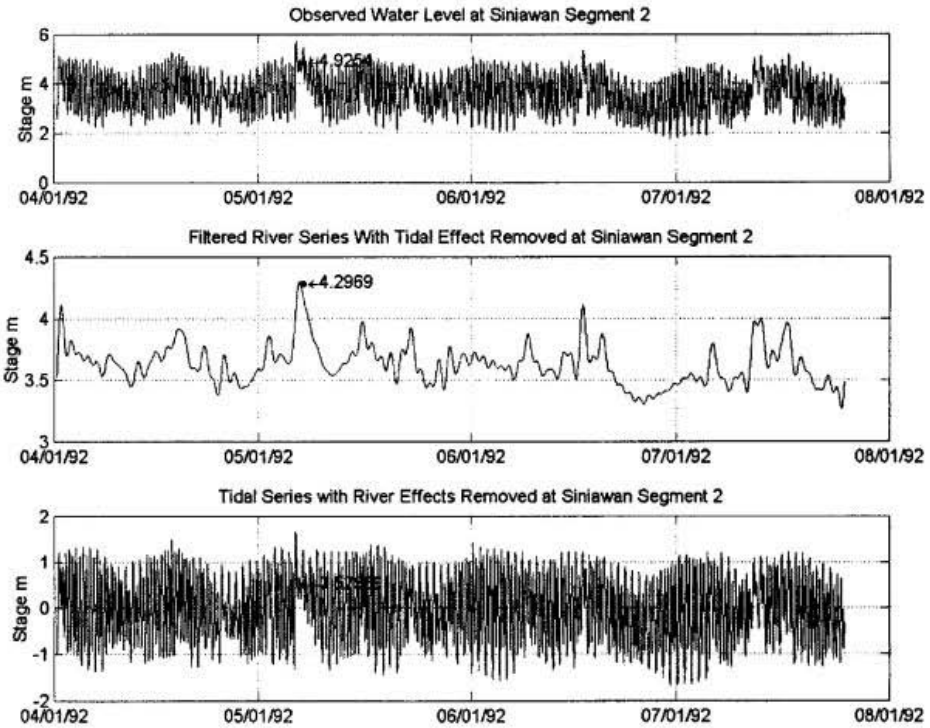


Figure C1-2

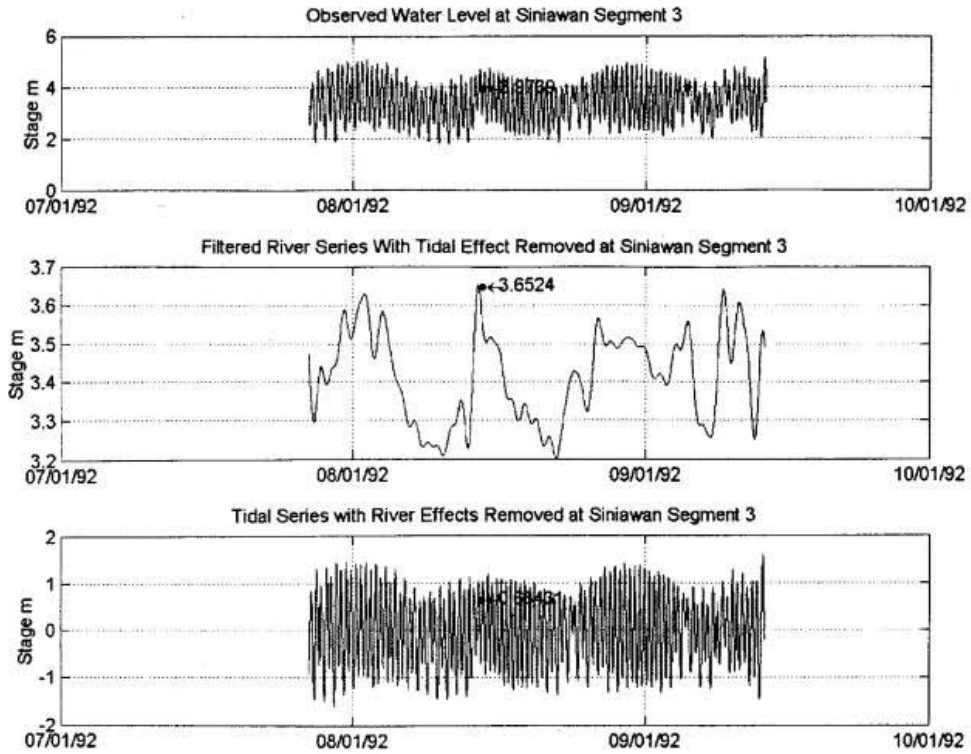


Figure C1-3

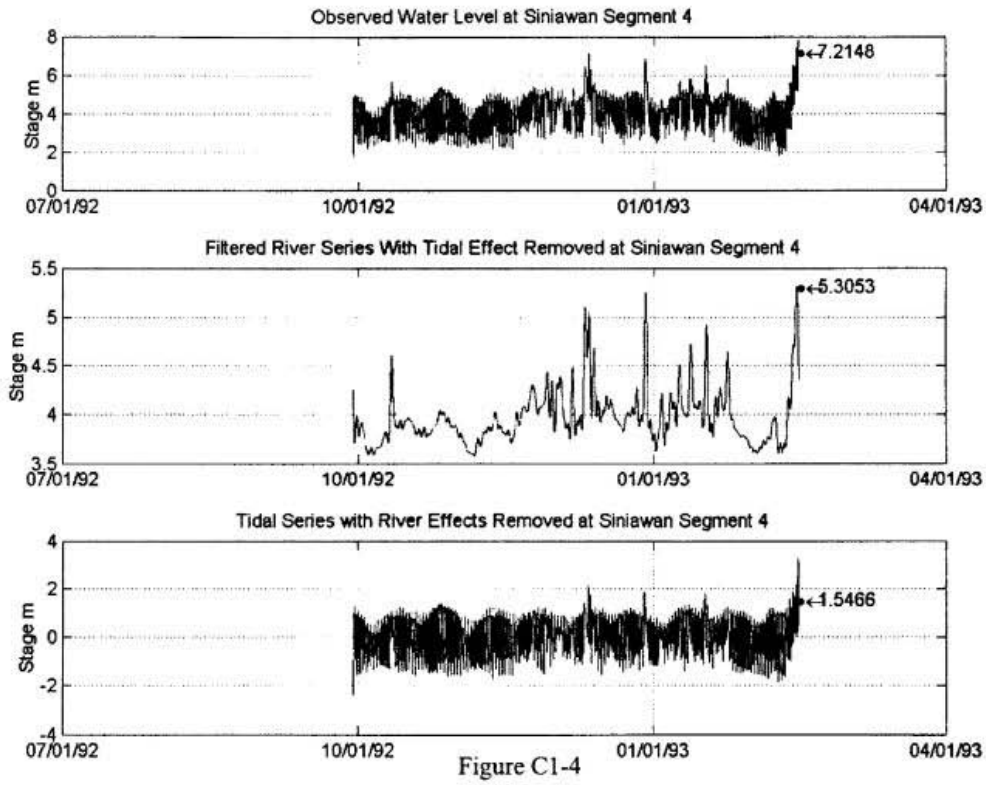


Figure C1-4

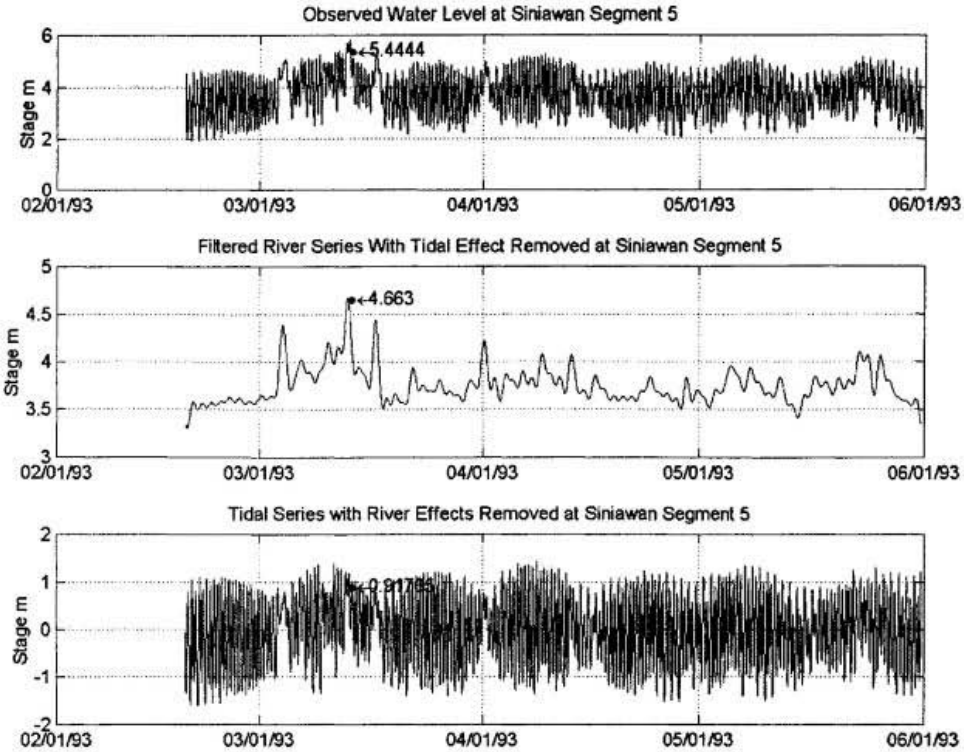


Figure C1-5

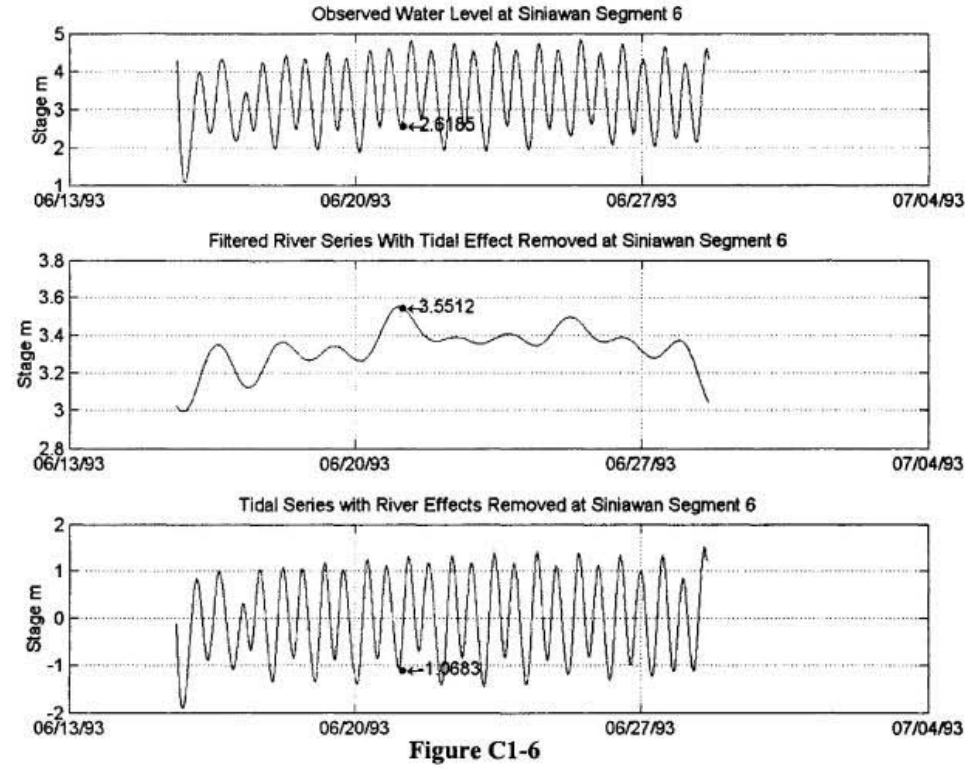


Figure C1-6

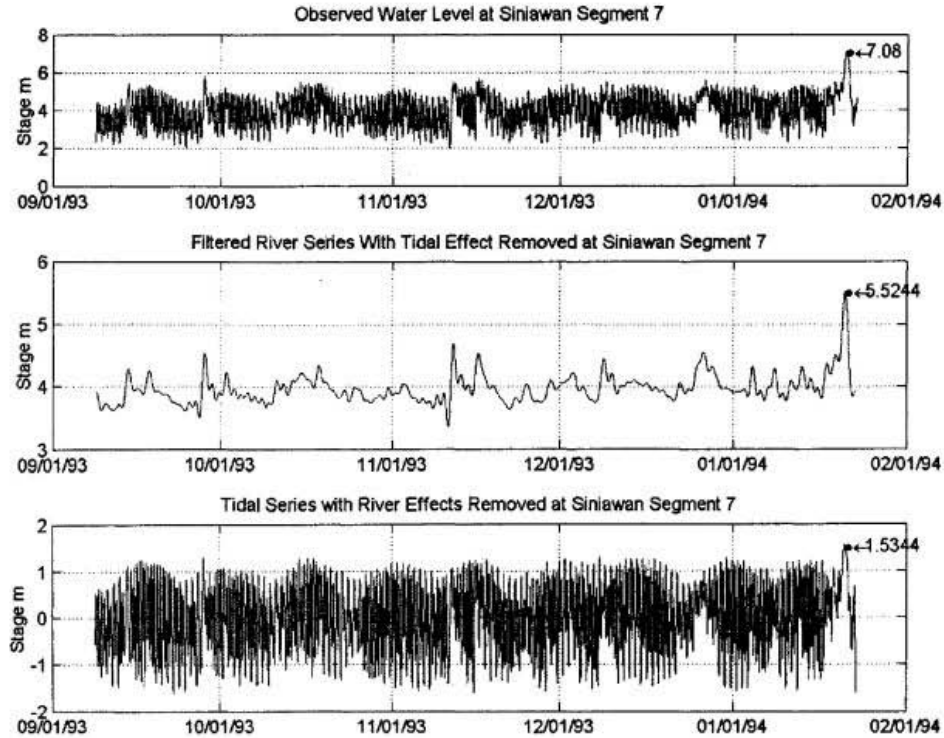


Figure C1-7

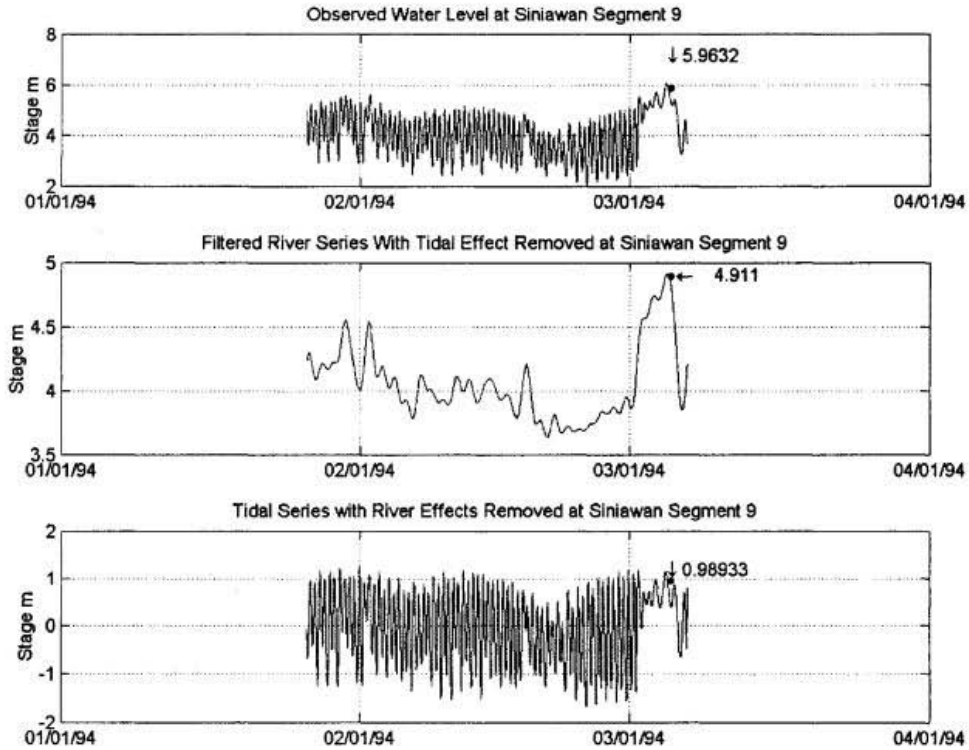


Figure C1-8

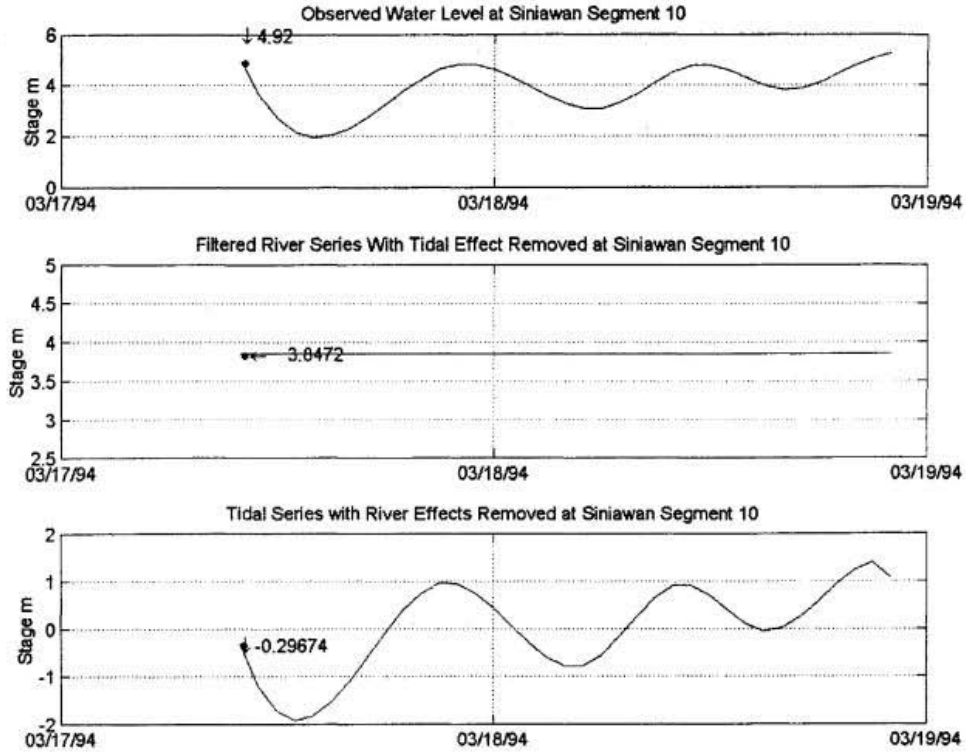


Figure C1-9

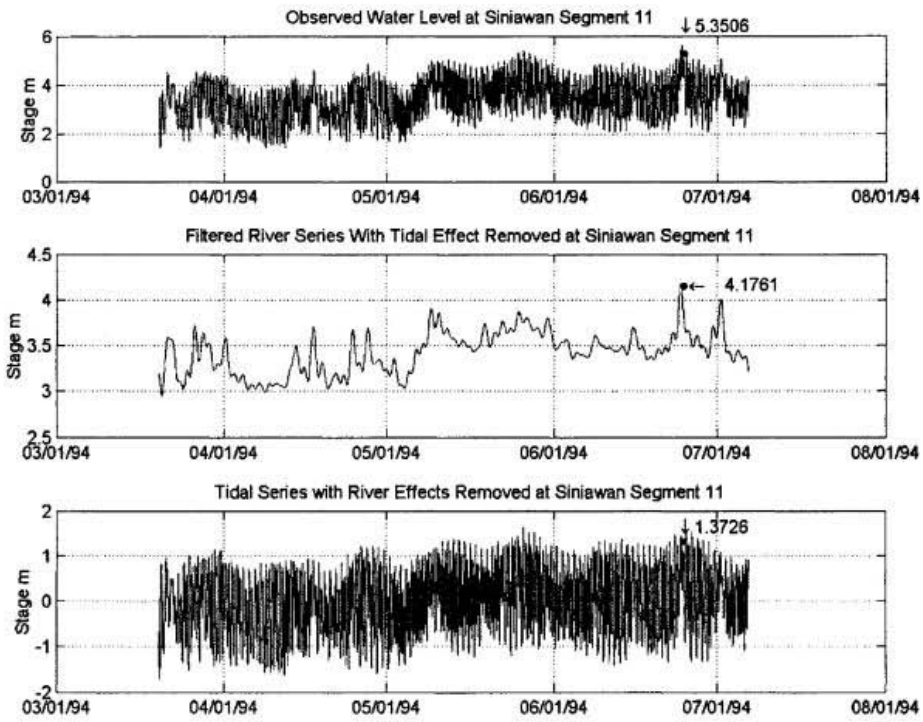


Figure C1-10

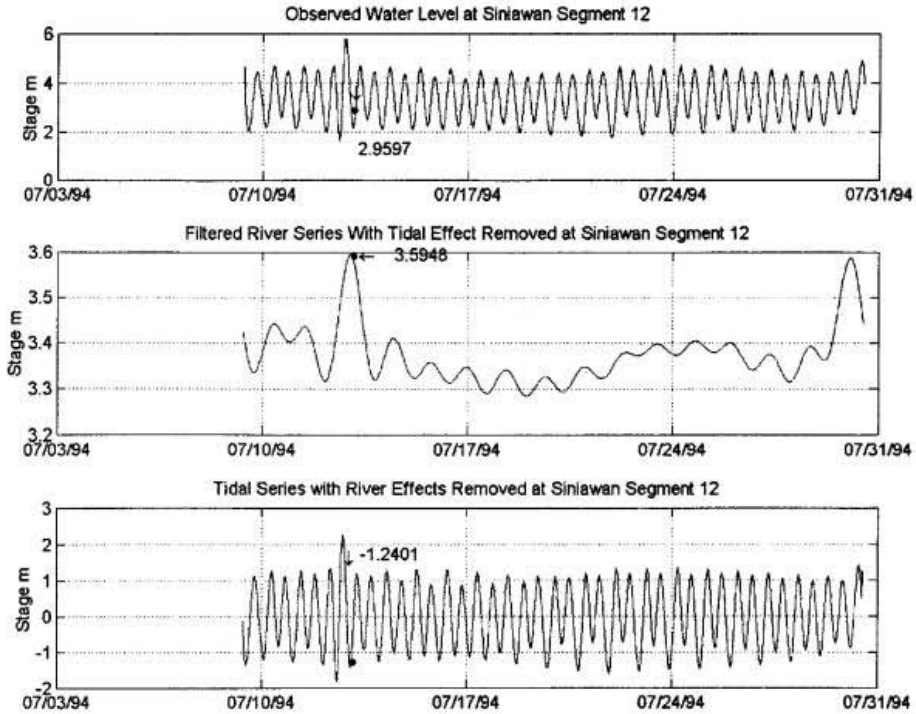


Figure C1-11

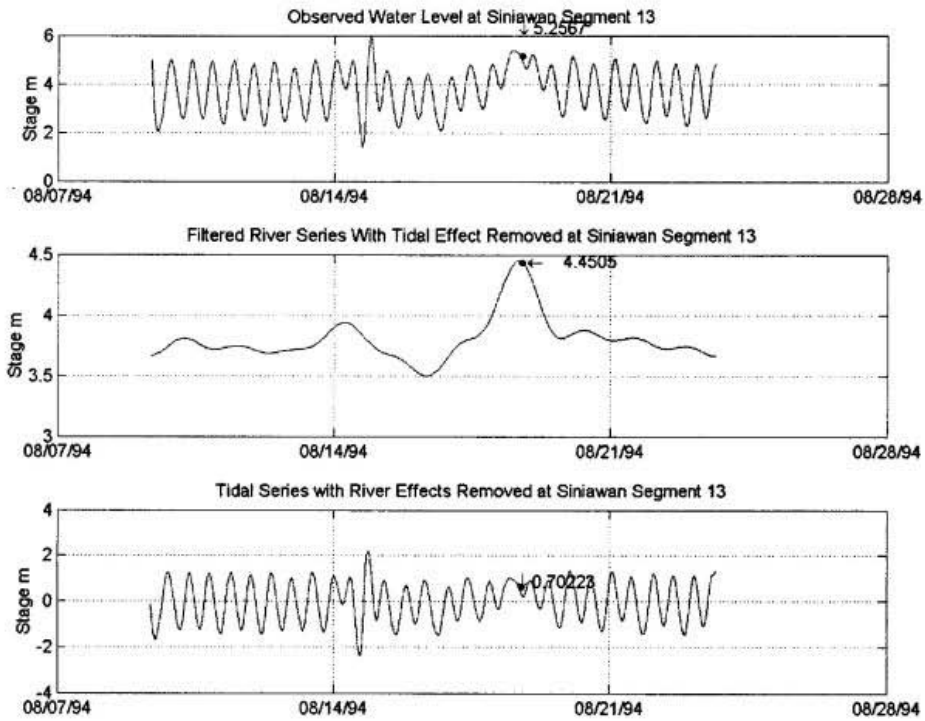


Figure C1-12

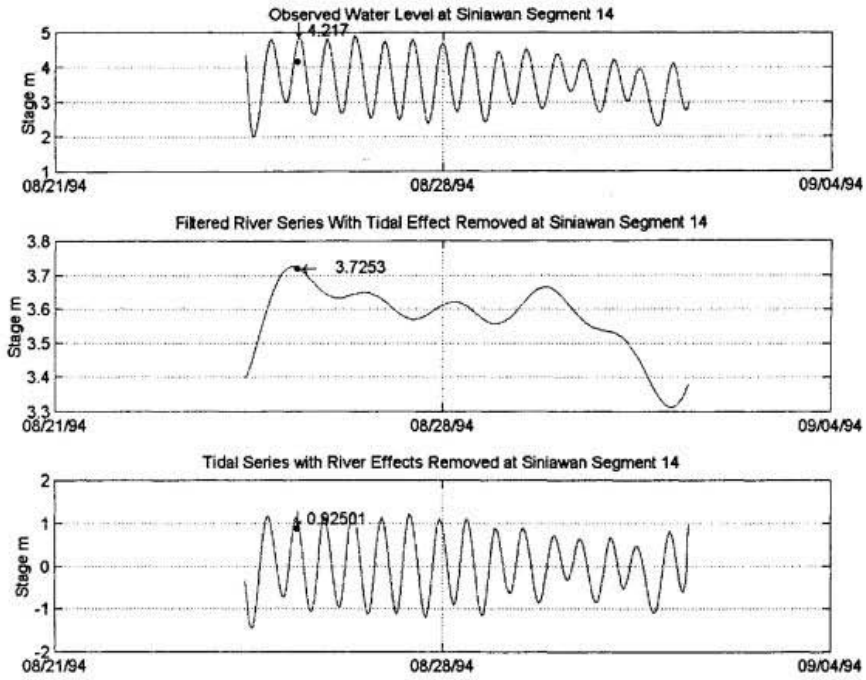


Figure C1-13

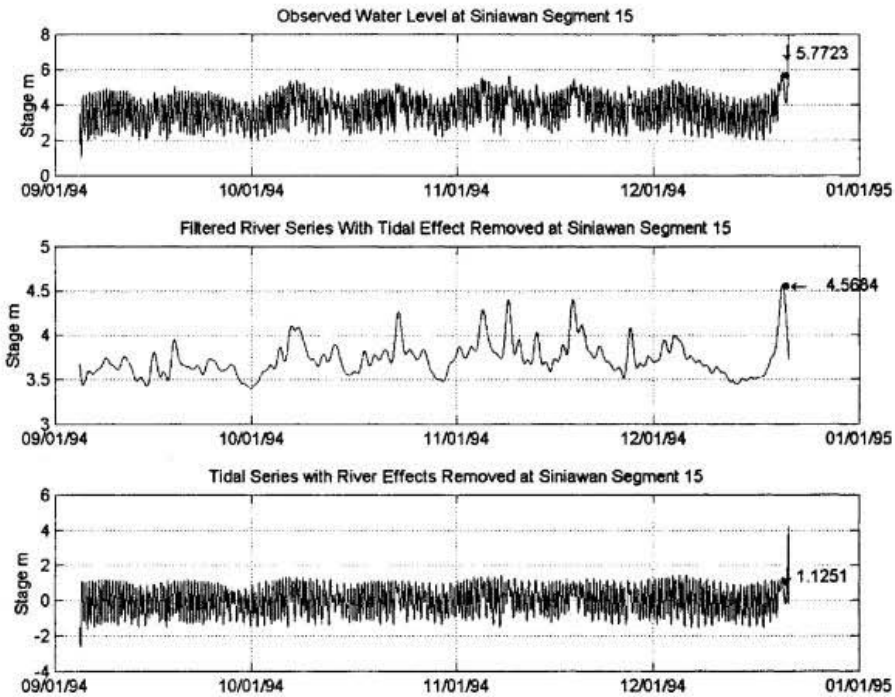


Figure C1-14

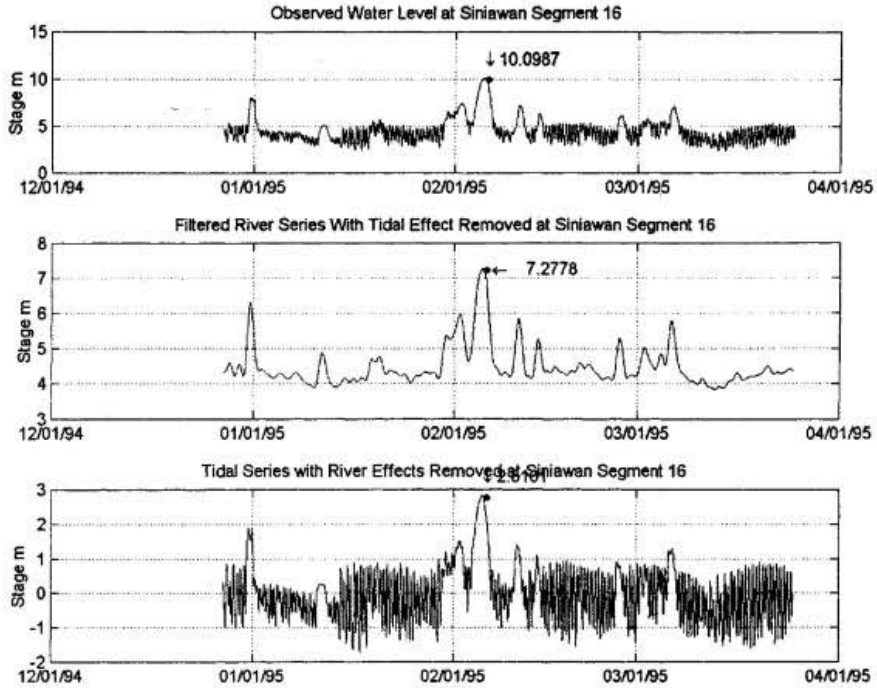


Figure C1-15

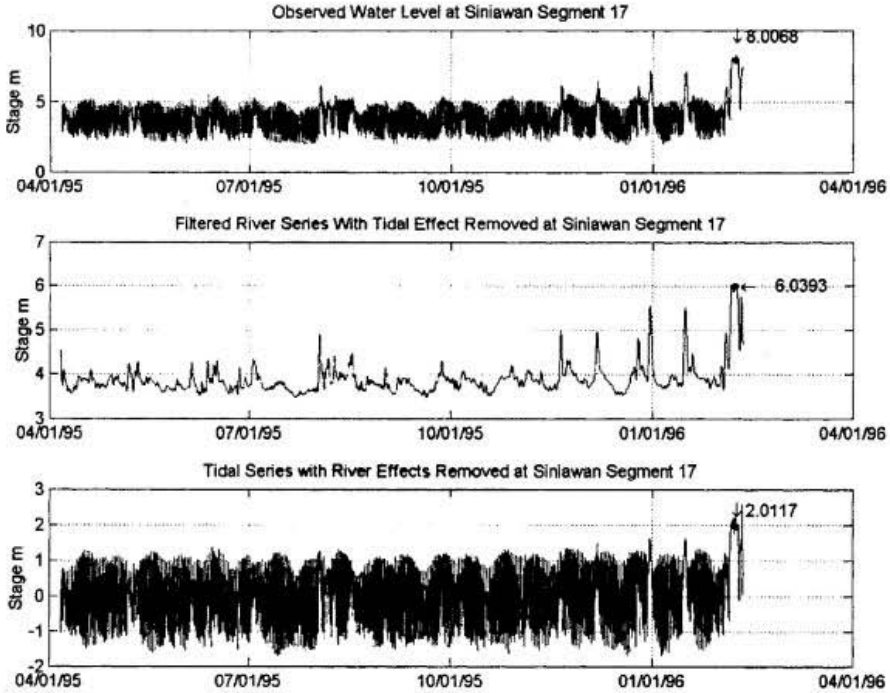


Figure C1-16

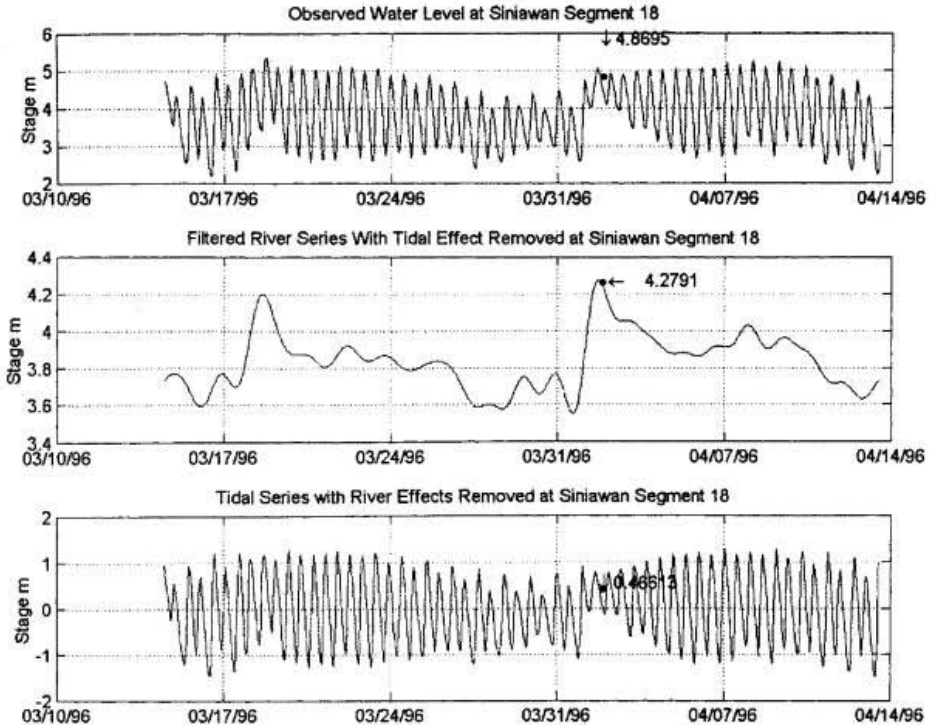


Figure C1-17

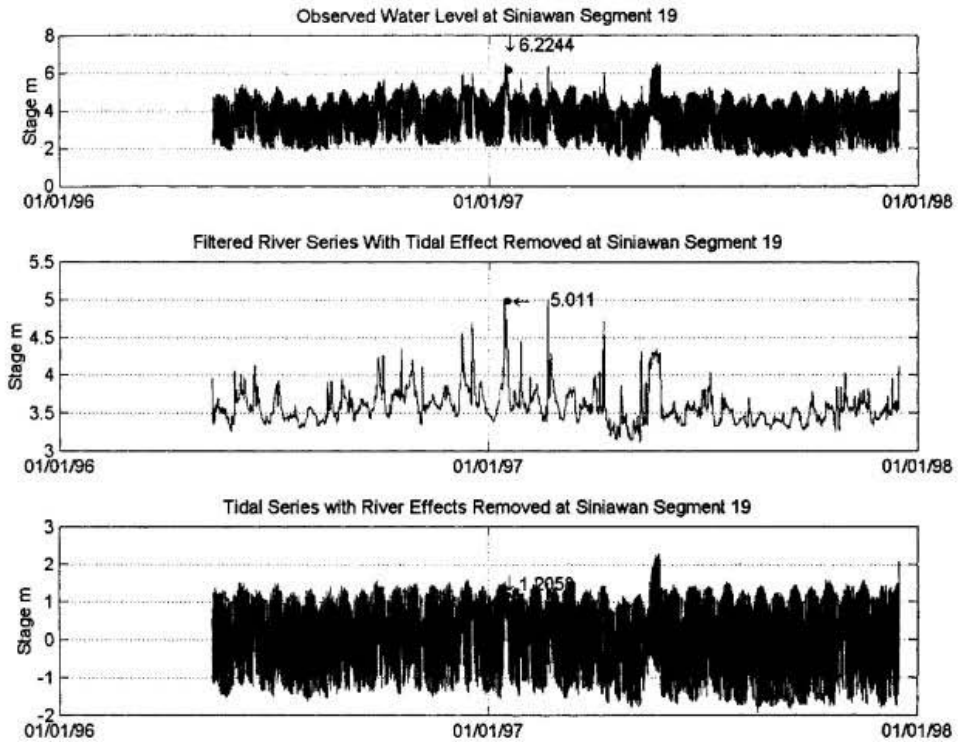


Figure C1-18

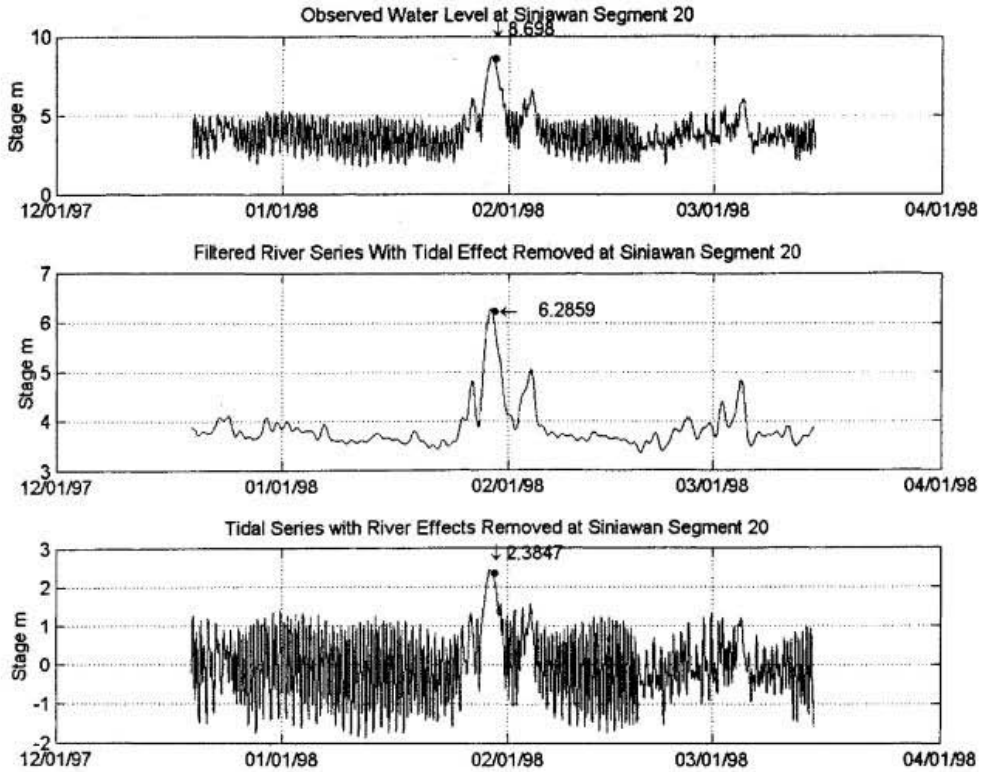


Figure C1-19

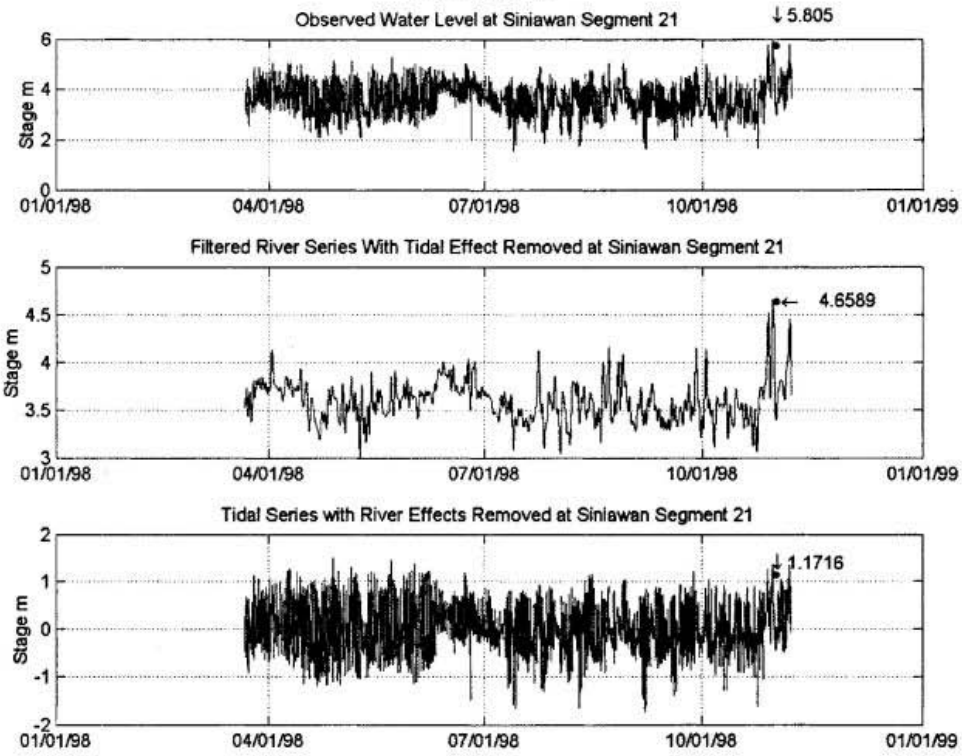


Figure C1-20

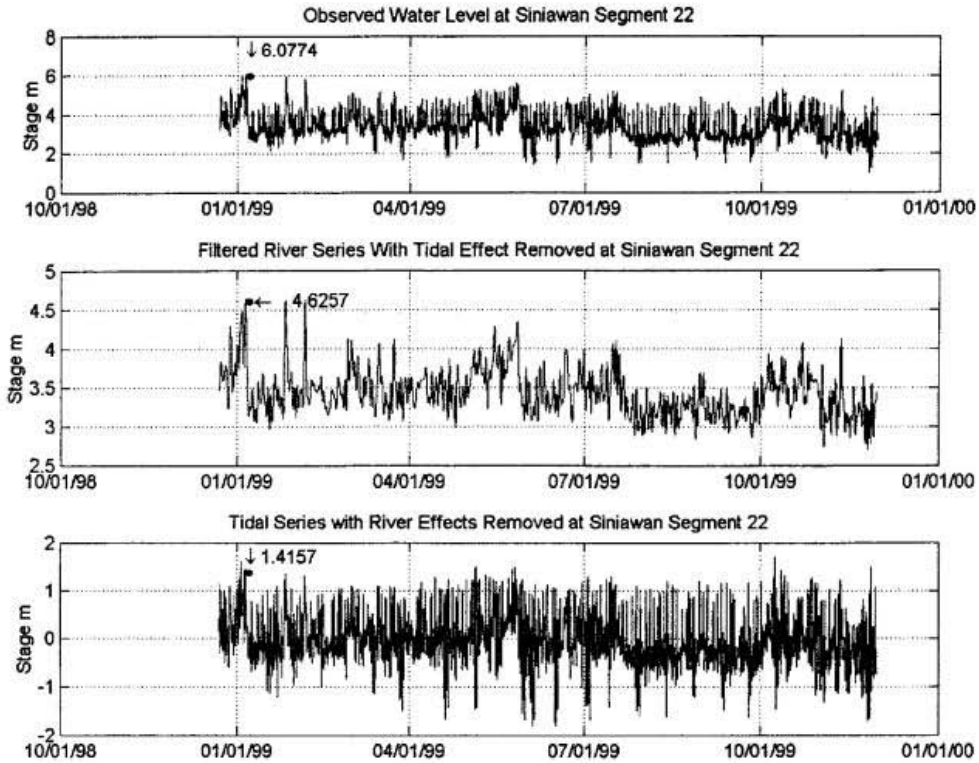


Figure C1-21

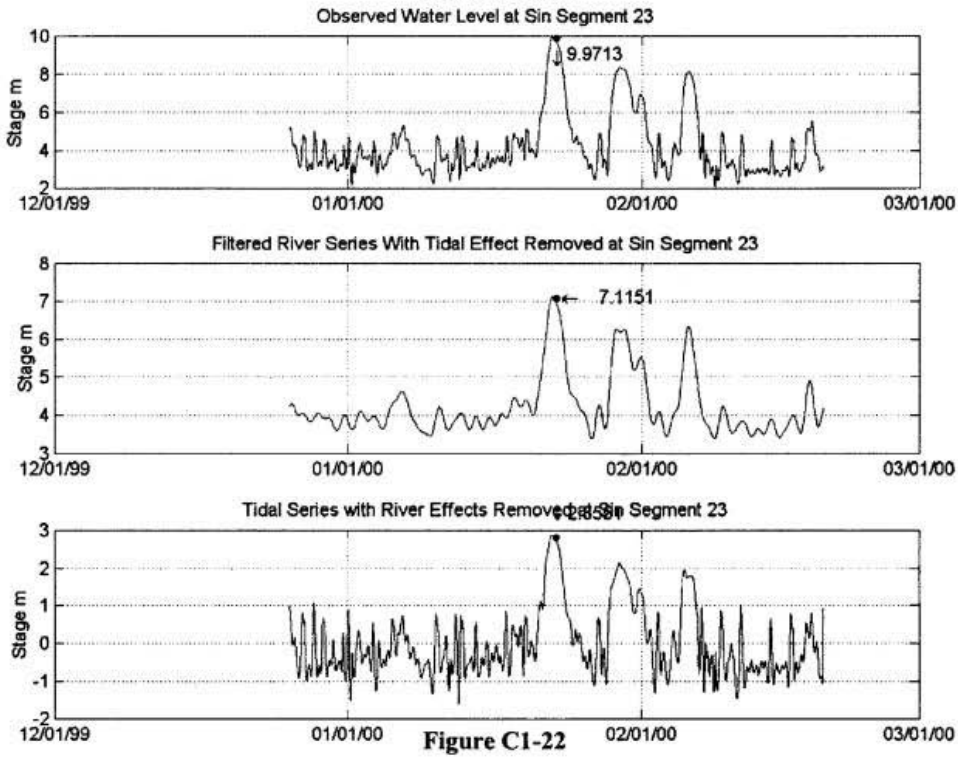


Figure C1-22

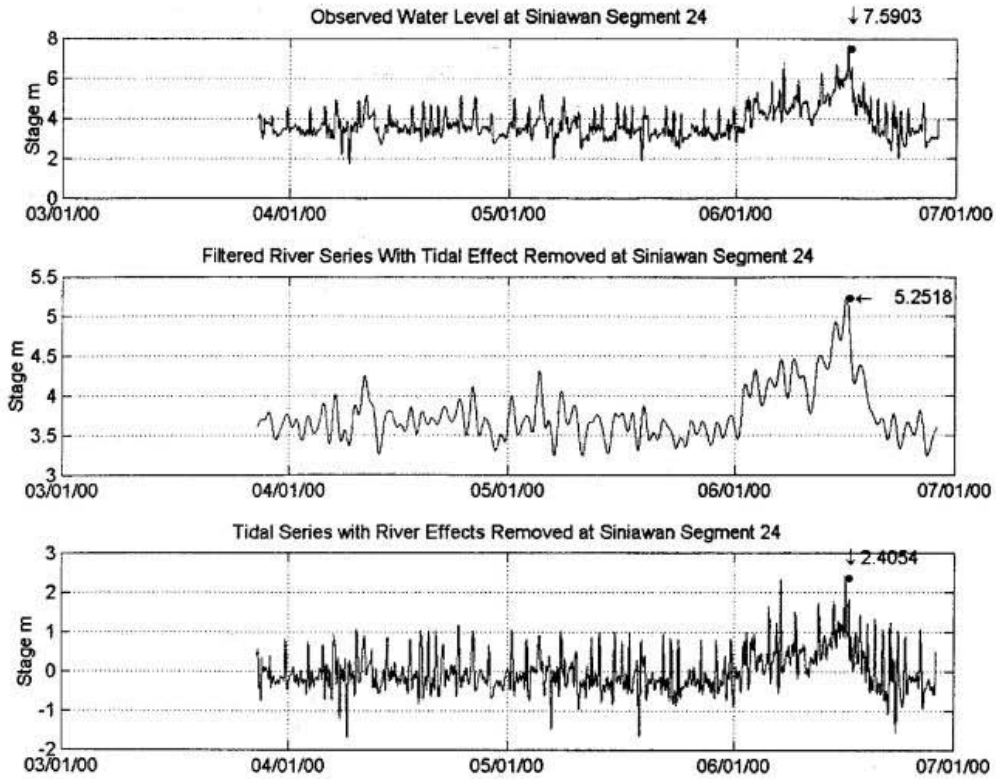

Observed Water Level at Siniawan Segment 24
↓7.5903
Stage m
8
6
4
2
0
03/01/00
04/01/00
05/01/00
06/01/00
07/01/00
Filtered River Series With Tidal Effect Removed at Siniawan Segment 24
5.2518
Stage m
5.5
5
4.5
4
3.5
3
03/01/00
04/01/00
05/01/00
06/01/00
07/01/00
Tidal Series with River Effects Removed at Siniawan Segment 24
↓2.4054
Stage m
3
2
1
0
-1
-2
03/01/00
04/01/00
05/01/00
06/01/00
07/01/00

Figure C1-23

Appendix C2

r Largest Extreme Series Abstracted from Filtered River Series at Siniawan

r Largest Extreme Levels of Filtered River Series at Siniawan Sorted From Independent Storms

$m = 7$
$u = 20$
$r = u \times m$

7.3601	4.8641	4.5427	4.3610	4.2867	4.2215	4.1761
7.2778	4.8432	4.5411	4.3552	4.2831	4.2199	4.1754
6.2972	4.7725	4.5313	4.3531	4.2817	4.2190	4.1648
6.0765	4.7460	4.5219	4.3508	4.2804	4.2189	4.1622
6.0393	4.7253	4.5056	4.3445	4.2791	4.2168	4.1597
5.5589	4.7242	4.4936	4.3414	4.2744	4.2139	4.1540
5.5493	4.7128	4.4934	4.3393	4.2743	4.2131	4.1518
5.5244	4.7004	4.4681	4.3363	4.2710	4.2114	4.1494
5.5215	4.6874	4.4505	4.3325	4.2676	4.2102	4.1436
5.3053	4.6826	4.4498	4.3317	4.2676	4.2100	4.1426
5.2482	4.6630	4.4497	4.3210	4.2655	4.2092	4.1410
5.1007	4.6627	4.4367	4.3163	4.2652	4.2077	4.1372
5.0110	4.6386	4.4346	4.3094	4.2620	4.2059	4.1326
4.9923	4.6067	4.4332	4.3051	4.2603	4.1995	4.1316
4.9820	4.5901	4.3971	4.3042	4.2591	4.1941	4.1309
4.9521	4.5684	4.3958	4.3026	4.2500	4.1940	4.1304
4.9135	4.5684	4.3958	4.2969	4.2366	4.1903	4.1295
4.9110	4.5558	4.3957	4.2946	4.2280	4.1861	4.1270
4.8970	4.5537	4.3957	4.2880	4.2271	4.1824	4.1180
4.8661	4.5435	4.3939	4.2867	4.2259	4.1765	4.1150

Estimates of

$\sigma =$ 0.3965

$\mu =$ 6.0744