University of the Western Cape

Faculty of Science

Department of Statistics

Fitting extreme value distributions to the Zambezi River flood water levels recorded at Katima Mulilo in Namibia (1965-2003)

UNIVERSITY of the
WESTERN CAPE

By
Innocent Silibelo Kamwi
Student number: 2400163

MSc 2005

University of the Western Cape

Faculty of Science

Department of Statistics

Fitting extreme value distributions to the Zambezi River flood water levels recorded at Katima Mulilo in Namibia

(1965-2003)

UNIVERSITY of the
WESTERN CAPE

By
Innocent Silibelo Kamwi
Student number: 2400163

Research project submitted in partial fulfilment of the requirements for the Degree of Master of Science
Department of Statistics
2005

Supervisor: Prof. C. Koen
Co-supervisor: Prof. R. J. Blignaut

Keywords:

Extreme value theory

Generalized extreme value distribution

QQ plots

Limit to flood levels

Annual maximum flood level

Maximum likelihood estimation

Empirical information matrix

Likelihood ratio testing

Weibull distribution

Zambezi River

UNIVERSITY of the
WESTERN CAPE

ABSTRACT

This study sought to identify and fit the appropriate extreme value distribution to flood data, using the method of maximum likelihood. To examine the uncertainty of the estimated parameters and evaluate the goodness of fit of the model identified. The study revealed that the three parameter Weibull and the generalised extreme value (GEV) distributions fit the data very well. Standard errors for the estimated parameters were calculated from the empirical information matrix. An upper limit to the flood levels followed from the fitted distribution.

UNIVERSITY of the
WESTERN CAPE

Declaration

I declare that *Fitting extreme value distributions to the Zambezi River flood water levels recorded at Katima Mulilo in Namibia (1965-2003)* is my own work, that it has not been submitted for any degree or examination in any other university, and that all sources I have used or quoted have been indicated and acknowledged by complete references.

Full Name: Mr. I.S. Kamwi Date: 11 November 2005.

Signed:………………………………..

UNIVERSITY of the
WESTERN CAPE

Acknowledgements

I wish to record my deepest gratitude and appreciation to my supervisor, Prof. C. Koen and my co-supervisor Prof. R. J. Blignaut, for their able guidance, patience and reassurance during the preparation of this work. Without their marked assistance and understanding this project work would not have been accomplished.

My singular gratitude goes to my family, particularly my wife, Hileni Namene Kamwi for her untiring and valuable support during my continued long absence from home and to son Richard Kamwi Silibelo for his tolerance of my absenteeism from home.

I am further indebted to many people especially to Mr Guido van Langenhove of the hydrology division in the Ministry of Agriculture, Water and Rural Development without whose data, the attainment of my academic goal would have been unattainable.

Last but not least, I would like to thank the Africa-America Institute and the Namibia Government Scholarship and Training Program under whose scholarship this study became a reality.

UNIVERSITY of the
WESTERN CAPE

Table of Contents

UNIVERSITY of the WESTERN CAPE

UNIVERSITY of the
WESTERN CAPE

List of Tables and Figures

List of Tables

List of Figures

UNIVERSITY of the
WESTERN CAPE

Chapter 1
Introduction

1.1 Introduction

Increases in rare events both natural and human in nature are being observed in recent years all over the world. An example of that is the flooding in Venezuela in 1999 due to extreme rainfall [Coles; 2001]. The increased interest in studying these extreme natural events is to mitigate their impact on humans and properties. Statistics which is a science of decision making based on data is one of the fields involved in studying extreme events. "Extreme value theory is that part of statistics concerned with the probabilistic and statistical questions related to these very low or very high values in a sequence of random variables and in stochastic processes" [Smith; 2004]. Extreme value theory has long been applied to the study of these rare events and has been proven to be reliable in fitting models to historical data. In particular, in hydrology the question of return periods of severe floods is always answered using extreme value distributions. The application of the extreme value theory has been used in diverse fields such as finance, environmental studies, economics and meteorology. In Namibia extreme value theory has been applied in the environmental fields of hydrology. This study, applies extreme value theory to the observed annual maximum flood height for the Zambezi River at Katima Mulilo.

UNIVERSITY of the
WESTERN CAPE

1.2 Problem statement

Namibia is a semi arid country with more than half of its land as a desert, and the other being a wet area, prone to floods that can be either negative or positive for the north eastern parts of the country. Flooding is viewed as negative as it always destroys maize crops cultivated in the fertile flood plains. The positive aspect of flooding is that with more water it means more fish will be caught after the flood and fish is an important part of the diet for the people living in the Caprivi region. The Zambezi River is one of the largest perennial rivers in Southern Africa and serves as a natural border between Namibia and Zambia. The origin of this river is located in Zambia and its catchments can be found in both Zambia and Angola. This river is characterized by seasonal floods due to rainfall in the upper catchments and its drainage area covers much of eastern Caprivi therefore making the area prone to seasonal floods that always damage crops and properties in the region.

The aim of this research project is to estimate parameters for the distribution of annual maximum flood levels for the Zambezi River at Katima Mulilo. The estimation of parameters will be done by using the maximum likelihood method. The study aims to explore data of the Zambezi's annual maximum flood heights at Katima Mulilo by means of fitting the Gumbel, Weibull and the generalized extreme value distributions and evaluating their goodness of fit. Extreme value theory can be a tool that can be used to study the distribution of droughts (or minimum flood level) which is the opposite to studying the maximum flood heights. Therefore extreme value theory can be one of the tools that can be constantly utilised in order to improve planning for the alleviation of problems due to both the abundance and scarcity of water. The understanding of the form of the distribution of the observed flood water level can lead to better estimation and forecasting of future flood levels of the Zambezi River.

1.3 Importance and benefits

UNIVERSITY of the
WESTERN CAPE

The results obtained in this study will be very useful for policy-makers in the fields of hydrology and water management, especially with respect to the estimated distribution function of annual maximum floods for the Zambezi River. The study can serve as a bench mark for comparison with similar studies based on other models not covered in this analysis. The study will also lay the foundation for future research on the subject and can be expanded as more data becomes available.

1.4 Research objectives

1. The study aims to contribute to knowledge about the underlying distribution of observed floods for the Zambezi River.
2. Use statistical techniques such as extreme value theory and model selection to fit the data of annual maximum flood water level for the Zambezi River. The following distributions will be considered: Gumbel, Weibull and the generalized extreme value distribution (GEV).

3. Estimate parameters for the model by method of maximum likelihood (ML) and evaluate the goodness of fit of the models and standard errors of the estimated parameters.
4. The study aims to lay a foundation for future research on maximum flood water levels for the Zambezi River at Katima Mulilo.

1.5 Research design and analysis

Firstly, the general concepts of extreme value theory will be discussed. Secondly, the quantitative analysis of data will be carried out. The time series data on water levels for the Zambezi River was collected from the Ministry of Agriculture and Water and Rural Development's department of Hydrology. The Matlab software and a Microsoft excel spreadsheet was used for data analysis.

1.6 Data

For the purposes of this study, annual maximum flood water levels covering the time period from 1965 to 2003 will be investigated. This is an uninterrupted thirty-nine year period. For the observed data set to be assumed independent and identically distributed (iid), the block maxima method will be used [Smith;1984]. The block maxima method provides for samples to be taken from blocks of one year and assumed to be iid. Further discussion of block maxima is given in Chapter 2. The Zambezi River records contain some missing data for the entire period of observation [1935-2003], hence the choice to start from 1965.

1.7 Limitations

In this research project only the block maximum method of extreme value analysis could be explored. In a more comprehensive thesis or mini-thesis additional analyses could be explored. The sample size of 39 measurements of annual maximum flood levels also limited the type of conclusions that could be drawn from this study. The non-inclusion of covariates in the study also limited the optimal use of other information for a better understanding of the effects of other external factors on flooding over the Zambezi River at Katima Mulilo.

UNIVERSITY of the WESTERN CAPE

Chapter 2

Literature Review

2.1 Introduction

This chapter starts by defining concepts used in this research project. It also discusses in short the theories related to the fitting of extreme value distributions and the underlying assumptions for modelling these distributions. Understanding of the concepts discussed in this chapter is imperative in the appreciation of the whole document and the steps used in this study.

2.2 Definition of concepts

2.2.1 Extreme value theory (EVT) [Kotz S. *et al.*; 1985:608]

The brief historical development of extreme value is given in the encyclopedia edited by Kotz, S. *et al.* states that the basic theory of extreme value was first developed by Fréchet in 1927 and by Fisher & Tippet in 1928 but was formalized by Gnedenko in 1943.

UNIVERSITY of the
WESTERN CAPE

Suppose there exists an independent and identically distributed (iid) sequence of random variables $X_1, X_2, ...$ whose cumulative distribution function (CDF) is:

$F(x) = \Pr\{X_i < x\}$

also $M_n = \max(X_1, ..., X_n)$ which is the n^{th} sample maximum of the process and M_n has a CDF: $\Pr\{M_n \le x\} = [F(x)]^n$ ……………….(i)

Equation (i) states that for any fixed *x* for which $F(x) < 1$, we have $\Pr\{M_n \le x\} \to 0$ as $n \to \infty$ which is not useful. However sequences of constants a_n, b_n exist such that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \le x\right\} = [H(a_n x + b_n)]^n \to H(x)$$

is independent of *n*. According to extreme value theory *H(x)* must be one of the three possible forms of distributions. The importance of this result is that irrespective of what the original distribution F is, the asymptotic distribution of $X_{(n)}$ is any of the three forms of the extreme value distribution. This theory is analogous to the asymptotic normality of sample means, invariant with respect to the underlying population.

In their simplest forms the three types of the extreme value distribution are:

1. $F(x) = \exp(-e^{-x}), \quad \text{all } x$

2. $F(x) = \begin{cases} 0, & x < 0 \\ \exp(-x^{-\alpha}), & x > 0,\ \alpha > 0 \end{cases}$

3. $F(x) = \begin{cases} \exp(-|x|^{\alpha}), & x < 0,\ \alpha > 0 \\ 1, & x > 0 \end{cases}$

where equation (1) corresponds to the Gumbel distribution, (2) is called the Fréchet distribution and (3) is the Weibull distribution.

2.2.2 Generalized extreme value distribution (GEV) [Smith; 2004:8]

WESTERN CAPE

The three types of the extreme value distribution mentioned above have been combined by Von Mises and Jenkinson into a single distribution:

$F(x) = \exp\left\{-\left(1+\beta\frac{x-\lambda}{\delta}\right)^{\frac{1}{\beta}}\right\}$, where λ is the location parameter, δ is the scale parameter and β is the shape parameter. When the limit $\beta = 0$ the GEV corresponds to the Gumbel distribution, $\beta < 0$ corresponds to the Fréchet distribution and $\beta > 0$ corresponds to the Weibull distribution and has finite upper limit.

2.3 Overview of extreme value theory

The following literature review on the applications of extreme value theory gives a summary of the use of extreme value theory by researchers around the world. The usefulness of this theory is in studying data that seem to contain outliers that are real observed values that may normally be removed from the set as they appear to distort the results. Extreme value theory is currently applied in at least two forms, the oldest form of block maxima and the recent method of peaks over threshold method (POT). The difference between these two methods lies in that the block maxima depends on grouping the series of observed data into blocks according to time. In this case it can be a day or month but the natural, and often used, block in hydrology is a year. A single maximum value represents a series of observed values for analysis. The peak over threshold method relies on first setting a level such that all measurements above the earmarked threshold will constitute the sample to be studied. Both methods have their pros and cons. One of the advantages of block maxima is that the chosen sample values can be assumed to be independent. The advantage of the POT method is that more efficient use of available information is made as more cases will be included in the sample. The disadvantage of the block maxima method is that it restricts the scope of inferences that can be made from such a study, as the only inference that can be made relates to the variation between blocks only and nothing can be said about the variations within the blocks. This is especially of concern when dealing with environmental phenomena which are mostly affected by differences due to seasonality. The POT method's disadvantages are due to the subjectivity in choosing a threshold level. This can result in two different conclusions on the same observed data depending on the level chosen.

UNIVERSITY of the
WESTERN CAPE

This research project is based on the block maxima principle, because this method has a more meaningful explanation in hydrological practice, where the annual maximum flood level can be assumed to be from an independent and identical distribution. Though this cannot be fully justified, models based on this method are more close to the definition that extreme value distributions has been designed to follow [Kotz, *et al.*; 1985]. The interest is in studying the year to year variation in the flood levels measured at Katima Mulilo. The distribution will be fitted based on past data on flood water levels. The unknown distribution can be fitted based on past data using the method of maximum likelihood estimation. Though there are many factors that influence flooding, the fitted model does not claim to cover all, as this is just an idealized way to study some natural random event such as maximum flood levels.

The question of fitting distributions to flood water levels by hydrologists arises from the need to find such distributions which could be used as tools in decision making. Suppose the mean flood level for the past six years was 45 meters, the question that might be asked by the hydrologist is: what is the expected next flood level? Fitting of distributions could help to answer such a question. The following quote from Pericchi and Rodriguez-Itube summarizes the reasons for the interest in applying extreme value theory in engineering practice. "The concern of civil engineers lies in the largest or the smallest values which a design variable may take during a certain length of time" [Pericchi & Rodriguez-Itube; 1985]. The work of the hydrologist of fitting distributions forms part of planning in some engineering designs such as the building of dams, bridges and flood protection walls. Fitting of extreme value distributions is an exercise that according to Pericchi and Rodriguez-Itube has a number of uncertainties which seems to be ignored in practice. The uncertainties as outlined by Pericchi and Rodriguez-Itube are:

1. Natural uncertainty, the uncertainty in the random process that is generating the occurrence of the extreme event.
2. Parameter uncertainty, the uncertainty related to the estimation of parameters of the model of the stochastic process due to limited data.
3. Model uncertainty, the lack of certainty that a particular probabilistic model of the stochastic process is true.

UNIVERSITY of the
WESTERN CAPE

The first of the three issues cannot be reduced in practice, but the last two can be reduced by judicious choice of methods of parameter estimation and the family of distributions to which the data are believed to belong to. The method of maximum likelihood parameter estimation is one of the methods where uncertainty in the estimated parameters of the distribution can be quantified. The method of comparing the fitted distribution to the expected quantiles based on the observed data also gives an indication as to how uncertain the model is in modelling the event of interest, based on historic data.

Research studies on the application of extreme value theory in the environmental fields in Namibia exists, but studies conducted in other countries will also be used. In the book edited by Finkenstad and Rootzen, Smith writes "extreme value theory is concerned with the probabilistic and statistical questions related to very high and very low values in sequences of random variables and in stochastic processes" [Smith; 2004:2]. The application of extreme value theory therefore could be of benefit to the private sector engineers, financial institutions and government agencies in Namibia tasked with planning and designing infrastructure and management systems in the country. Observed random extreme events such as the occurrence of extreme floods, heavy rainfall, the value-at-risk (VaR) of stocks, are examples which can be modelled well by extreme value distributions with relevant results for Namibia.

A report appeared in the Namibian newspaper of 19 January 2004, about flooding in one of the suburbs in Windhoek due to heavy rainfall which caused much damage to property and endangered people's lives. The story highlights one example where authorities and engineers were caught off guard with regards to planning for rare events such as these. The concerned suburb which was flooded is known to be built in a low lying area near a riverbed and the developers for this housing project seem to have built without regards to the possibility that the houses built were below the flood level. This oversight cost the developers a lot of money. This situation went by without any mention of calculating the chances of the flood being predicted based on the analysis of historical rainfall figures of the area.

UNIVERSITY of the
WESTERN CAPE

The hydrological flood study relating to Namibia is a regional study, which combines annual maximum flood data from different sites according to proximity or spatial measures [Mkhandi & Kachroo; 1996, Ware & Lad; 2003]. One of the perceived advantages of these regionalized studies is that the combining of data from many sites help to increase the sample size. The problem of small samples in hydrology is due to the changing nature of physical processes. An extreme event today might no longer be as extreme if a more extreme event takes place the next day, leading to a need to change all past inferences. The disadvantage of regional studies is that the more heterogeneous a region is, the less reliable the estimate derived by such methods. On the other hand these studies have more appeal in practice due to scarcity of data, due to the low availability of recording stations for sites of interest. Estimates based on regional studies can be used for un-gauged sites by simple transformation of the regional parameters to estimate parameters for the location of interest.

The study where the fitting of models to the distributions of floods in Namibia is that of Mkhandi and Kachroo [1996]. This study recommended a Pearson-three model with probability weighted moments and the Log-Pearson (gamma three) model with method of moment parameter estimation to be used. The study of Mkhandi and Kachroo used moment based estimation methods which are considered unreliable due to poor sampling properties of the second and higher order sample moments [Ware and Lad; 2003].

This study will use the maximum likelihood method to estimate parameters, as this method meets the property of efficiency and consistency [Kotz; 1985:611]. Extreme value theory is being applied extensively in the hydrology field in other parts of the world and hence the need to start exploring its use in Namibia as well.

UNIVERSITY of the
WESTERN CAPE

Chapter 3

Model estimation

3.1 Introduction

This chapter will outline the steps followed in model estimation by the maximum likelihood (ML) method. The maximum likelihood method is discussed further in Section 3.3. In practice everyone analyzing time series data should construct time plots. Time plots are important tools used for checking obvious patterns in the data. For example, Figure 3.1 is a plot of the annual maximum flood water levels recorded at Katima Mulilo, a town in north-eastern Namibia, over the period from 1965 to 2003. The plot does not show any pronounced systematic changes or patterns over the period recorded. From such data it might be possible to obtain an estimate of the maximum flood level that is likely to happen at Katima Mulilo over the next 10 or 100 years. In order to answer these questions one needs to fit a probability distribution to the observed data for the particular river. The next section starts with the initial step in fitting probability distributions which entails identifying a probability model that fits the observed data.

UNIVERSITY of the
WESTERN CAPE

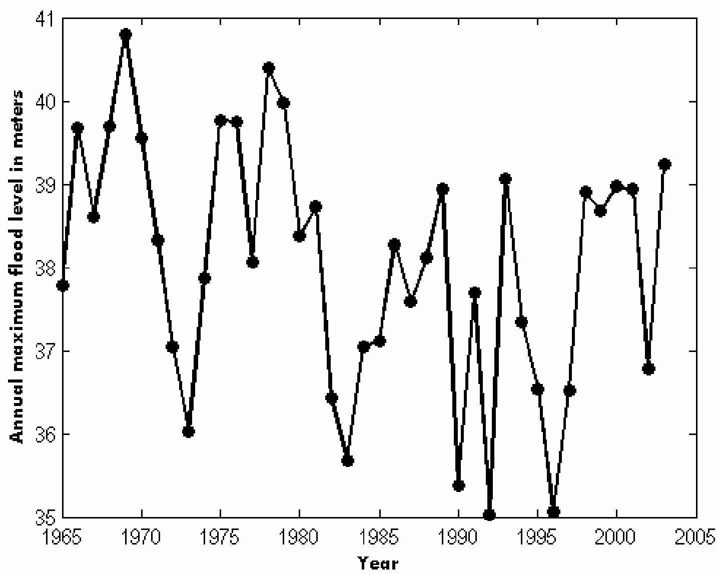

Figure 3.1 Annual maximum flood levels at Katima Mulilo, Namibia.

3.2 Model diagnostics

One of the methods advocated by the literature on extreme value theory for model identification is that of using the Gumbel QQ (Quantile-Quantile) plots. This standard method is a useful way of choosing among the three types of extreme value distributions. The method works as follows: let $X_{(j)}$ be the set of ordered observations. Plotting $X_{(j)}$ against –log [-log (j-0.5/n)], it is expected that the resulting graph will be a straight line if the Gumbel distribution is a good fit. If the plot shows a downward curvature then the Weibull distribution will be a better fit for the data, otherwise if the curvature is upward then the data follows the Fretchet distribution [Smith; 1984:445].

The motivation for the plot is now discussed for the Gumbel distribution. Suppose we have a sample of ordered values $X_{(1)}, X_{(2)}, ..., X_{(n)}$ from some distribution. The standard Gumbel Cumulative Distribution Function (CDF) is defined as [Smith; 1984:438],

$$H(x) = \exp\{-\exp(-x)\}, \quad -\infty < x < \infty$$

UNIVERSITY of the
WESTERN CAPE

Taking the logarithm on both sides of the above expression,

$$\log H(x) = -\exp(x)$$

Taking the negative on both sides,

$$-\log H(x) = \exp(x)$$

Taking the logarithm again,

$$\log[-\log H(x)] = -x$$

Multiplying by negative on both sides,

$$-\log[-\log H(x)] = x$$

Using the estimator

$$H[x_{(j)}] = \frac{j-0.5}{n},$$

$$-\log\left[-\log\left(\frac{j-0.5}{n}\right)\right] = x_{(j)} \ldots\ldots\text{(ii)}$$

Quantities given by the left hand side of this equation are referred to as "Reduced values". When the ordered values $x_{(j)}$ are plotted against the reduced values the resulting graph is expected to be a straight line if the data follows the Gumbel distribution. Similar transformations of the standard CDFs for the Weibull distribution are done and plotted for visual inspection. Figure 3.2 shows the QQ plot for the Zambezi annual maximum flood water levels at Katima Mulilo. From the plot it is evident that the Gumbel distribution is not a good fit for the Zambezi flood height data as the resulting QQ plots shows a downward curvature, which suggests a Weibull distribution. The only good fit for the Zambezi annual maximum flood water level is the Weibull distribution as the QQ plot is approximately linear over most of its domain as shown in Figure 3.3. This suggests fitting the Weibull distribution to the data.

UNIVERSITY of the
WESTERN CAPE

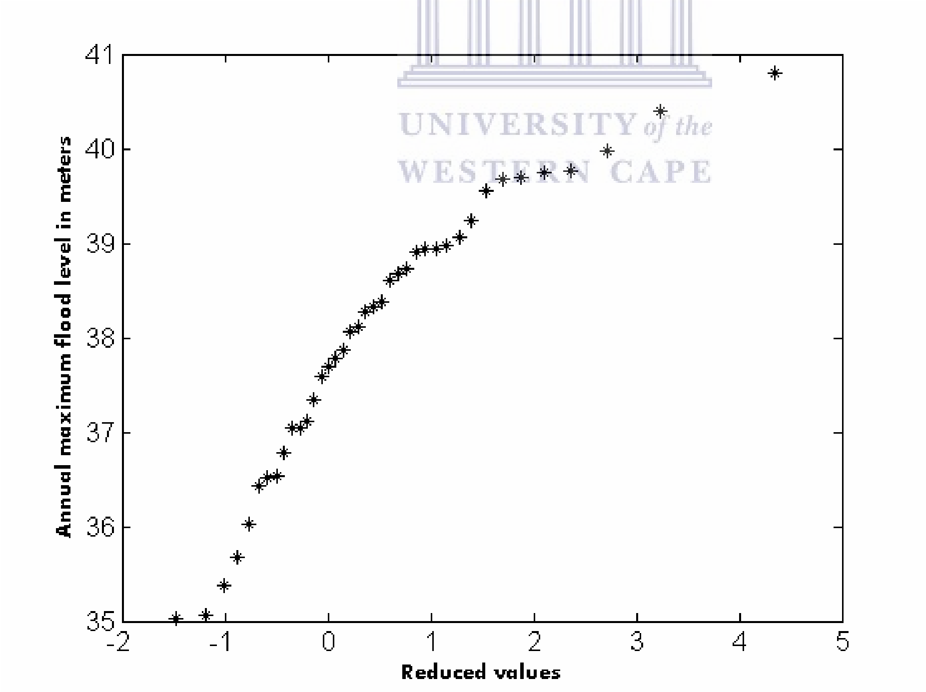

Figure 3.2 Gumbel QQ plot for the Zambezi annual maximum flood water level at Katima Mulilo

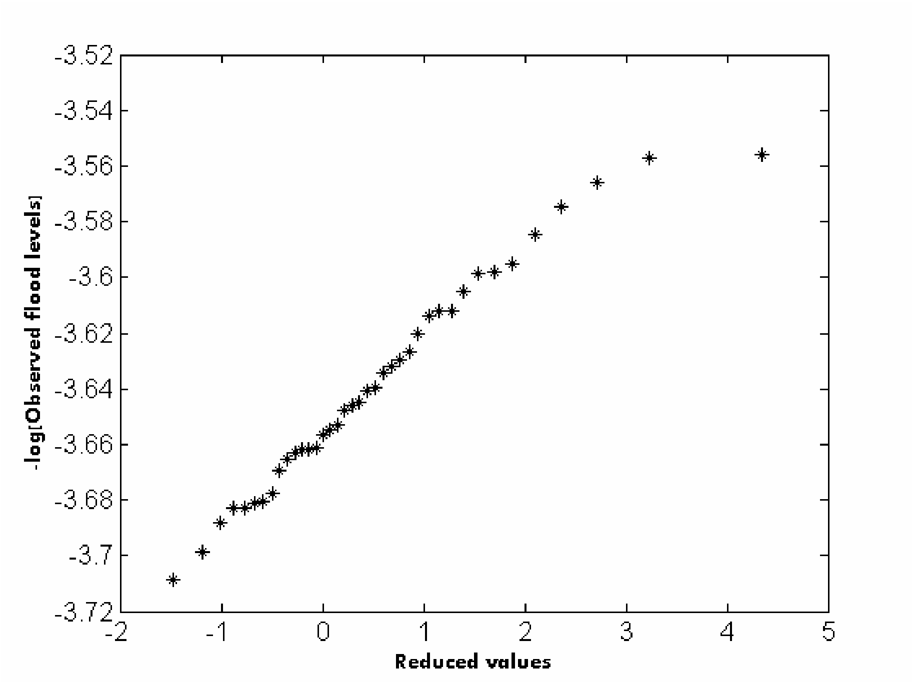

Figure 3.3 Weibull QQ plot for the Zambezi annual maximum flood water level at Katima Mulilo

UNIVERSITY of the
WESTERN CAPE

3.3 Maximum likelihood estimation with observed information matrix

Once it is known which of the three distributions is a good representation of our observed data, the task of estimation of the parameters for the probability distribution follows. This section will show the steps involved in using the method of maximum likelihood (ML). The method of maximum likelihood is one of the types of estimation methods for unknown parameters, when fitting a model to observed data. The method is preferable due to its adaptability to model change [Coles; 2001:3]. Though there is software, it is essential to convince oneself as to how the software is able to find the solution and this involves understanding the programme doing the computation, and verifying that the output is indeed appropriate. The following are the steps taken before using the computer software Matlab in finding the ML estimate.

3.3.1 Weibull maximum likelihood estimation

The three parameter Weibull CDF is given by [Castillo; 2004:201];

$$F(x) = \exp[-(\frac{\lambda - x}{\delta})^{\beta}] \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(iii)}$$

where $\lambda \geq 0, \delta > 0$ and β are the position, scale and shape parameters respectively.

The Weibull PDF is defined as follows after taking the derivative of the CDF function given above;

$$f(x) = \frac{\beta}{\delta}\left(\frac{\lambda - x}{\delta}\right)^{\beta - 1} * \exp\left(-\left(\frac{\lambda - x}{\delta}\right)^{\beta}\right) \ldots\ldots\ldots\ldots\ldots\ldots \text{(iv)}$$

In the present case the unit of time is one year and x_s represents the annual maximum value for each of the n years. The corresponding negative log likelihood is:

UNIVERSITY of the
WESTERN CAPE

$$L_x(\lambda, \delta, \beta) = -n\log\beta + n\beta\log\delta - (\beta - 1)\sum_i \log(\lambda - x_i) + \sum_i \left(\frac{\lambda - x_i}{\delta}\right)^{\beta} \ldots\ldots \text{(v)}$$

where $x \leq \lambda,\ \delta > 0,\ \beta > 0$.

Equation (v) is the one used to solve for the unknown parameters, by using Matlab script files to minimize this non linear function. See Appendix F for the Weibull script file. The script file works by first specifying the function given in equation (v) above and then using the Matlab's built in function called "fminsearch" to minimize L_x. The use of the negative log likelihood is dictated by the availability of the minimization facility "fminsearch": standard Matlab does not have a maximization routine.

One of the advantages of likelihood-based estimation is that the method facilitates the calculation of the standard errors of the estimated parameters. [Smith; 2004:17]. The standard errors are a measure of precision and can be derived from the observed information matrix. The information matrix is an $n \times n$ matrix derived by taking the second partial derivatives of the log likelihood function with respect to the parameters estimated. The true information matrix is approximated by the empirical information matrix. This is accomplished by replacing expected values of random variables by sample values. The first partial derivatives of equation (v) are:

$$D_\lambda = -(\beta-1)\sum\frac{1}{(\lambda-x_i)} + \sum\frac{\beta}{\delta}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta-1}$$

$$D_\delta = \frac{n\beta}{\delta} - \sum\beta\left(\frac{\lambda-x_i}{\delta^2}\right)\left(\frac{\lambda-x_i}{\delta}\right)^{\beta-1}$$

$$D_\beta = -\frac{n}{\beta} + n\log\delta - \sum\log(\lambda-x_i) + \sum\log\left(\frac{\lambda-x_i}{\delta}\right)\left(\frac{\lambda-x_i}{\delta}\right)^{\beta} .$$

The second partial derivatives from equation (v) follow;

$$D_\lambda^2 = (\beta-1)\sum\frac{1}{(\lambda-x_i)^2} + \sum\frac{\beta(\beta-1)}{\delta^2}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta-2}$$

$$D_\beta^2 = \frac{n}{\beta^2} + \sum\left(\frac{\lambda-x_i}{\delta}\right)^{\beta}\left[\log\left(\frac{\lambda-x_i}{\delta}\right)\right]^2$$

$$D_\delta^2 = -\frac{n\beta}{\delta^2} + \sum\left\{\frac{\beta(\beta-1)}{\delta^2}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta} + \frac{2\beta}{\delta^2}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta}\right\} = \frac{-n\beta}{\delta^2} + \sum\frac{\beta(\beta+1)}{\delta^2}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta}$$

$$D_{\lambda\delta}^2 = -\sum\left\{\frac{\beta(\beta-1)}{\delta^2}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta-1} + \frac{\beta}{\delta^2}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta-1}\right\} = -\sum\frac{\beta^2}{\delta^2}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta-1}$$

$$D_{\beta\delta}^2 = \frac{n}{\delta} - \sum\left\{\frac{1}{\delta}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta} + \frac{\beta}{\delta}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta}\log\left(\frac{\lambda-x_i}{\delta}\right)\right\}$$

$$D_{\lambda\beta}^2 = -\sum\frac{1}{(\lambda-x_i)} + \sum\left\{\frac{\beta}{\delta}\left(\frac{\lambda-x_i}{\delta}\right)^{\beta-1}\log\left(\frac{\lambda-x_i}{\delta}\right) + \left(\frac{\lambda-x_i}{\delta}\right)^{\beta-1}\frac{1}{\delta}\right\}$$

UNIVERSITY of the
WESTERN CAPE

The variance-covariance matrix is a symmetric matrix from the equations above,

$$I_O = \begin{pmatrix} D^2_{\lambda} & D^2_{\lambda\delta} & D^2_{\lambda\beta} \\ D^2_{\delta\lambda} & D^2_{\delta} & D^2_{\delta\beta} \\ D^2_{\beta\lambda} & D^2_{\beta\delta} & D^2_{\beta} \end{pmatrix},$$

where the inverse of I_O is the observed information matrix. The square roots of the diagonal elements form the standard errors for estimated parameters. The Matlab script file for calculating the empirical information matrix as given in Appendix F. Similar steps are followed in fitting the GEV and Gumbel distributions shown in Section 3.3.2 and 3.3.3.

3.3.2 GEV maximum likelihood estimation

Figure 3.2 indicates that the data are probably best fitted by a Weibull distribution. This can be verified by first fitting a GEV distribution, which encompasses all three extreme value distribution families, to the data.
The CDF for the GEV is given [Smith; 2004:16] by:

$$H(x) = \exp\left[-\left[1+\beta\left(\frac{x-\lambda}{\delta}\right)\right]^{-1/\beta}\right], \quad 1+\beta\left(\frac{x-\lambda}{\delta}\right) \geq 0, \ \beta \neq 0.$$

UNIVERSITY of the
WESTERN CAPE

The PDF of the GEV follows after taking the derivative of CDF above,

$$h(x) = -\frac{1}{\delta}\left(1+\beta\left(\frac{x-\lambda}{\delta}\right)^{\frac{-1}{\beta}-1}\right)\exp\left\{-\left[1+\beta\left(\frac{x-\lambda}{\delta}\right)^{\frac{-1}{\beta}}\right]\right\}, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(vi).$$

The negative log likelihood for the GEV distribution is

$$L = n\log\delta + \left(1+\frac{1}{\beta}\right)\sum_i \log\left[1+\beta\left(\frac{x_i-\lambda}{\delta}\right)\right] + \sum_i\left[1+\beta\left(\frac{x_i-\lambda}{\delta}\right)\right]^{-1/\beta}, \ldots\ldots\ldots\ (vii)$$

provided that, $1+\beta\left(\frac{x_i-\lambda}{\delta}\right) > 0, \quad \text{for } i = 1,\ldots\ldots, n.$

Equation (vii) is the function that needs to be maximized with respect to the parameters (λ, δ, β). It is quite complicated to work out the first and second partial derivatives of the log likelihood for the GEV. The results as given in Castillo *et al.* (2004) are therefore simply quoted in Appendix A.

3.3.3 Gumbel maximum likelihood estimation

For completeness, results for the Gumbel distribution are also repeated.

The CDF for the two parameter maximal Gumbel distribution is [Castillo; 2004:16]:

$$G_x = \exp\left[-\exp\left(-\left(\frac{x-\lambda}{\delta}\right)\right)\right], \quad -\infty < x < \infty.$$

The PDF is,

$$g(x) = \frac{1}{\delta}\exp\left(-\left(\frac{x-\lambda}{\delta}\right)\right)\exp\left(-e^{-\left(\frac{x-\lambda}{\delta}\right)}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(vii)}.$$

The negative log likelihood follows,

$$L = n\log\delta + \sum_i \exp\left[-\left(\frac{x_i-\lambda}{\delta}\right)\right] + \sum_i\left(\frac{x_i-\lambda}{\delta}\right), \ldots\ldots\ldots\ldots\ldots \text{(viii)}.$$

The variance-covariance matrix is,

$$I_g = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} ,$$

where $$m_{ij} = -\frac{\partial^2}{\partial\theta_i\partial\theta_j}L \qquad \theta_1 = \lambda \qquad \theta_2 = \delta$$

UNIVERSITY of the
WESTERN CAPE

$$m_{11} = \sum_i \frac{1}{\delta^2} \exp\left[-\left(\frac{x_i - \lambda}{\delta} \right) \right],$$

$$m_{22} = -\frac{n}{\delta^2} + \sum_i 2\left(\frac{x_i - \lambda}{\delta^3} \right) - \sum_i 2\exp\left[-\left(\frac{x_i - \lambda}{\delta} \right) \right]\left(\frac{x_i - \lambda}{\delta^3} \right) + \sum_i \exp\left[-\left(\frac{x_i - \lambda}{\delta} \right) \right] \frac{(x_i - \lambda)^2}{\delta^4},$$

$$m_{12} = m_{21} = \frac{n}{\delta^2} + \frac{1}{\delta^3} \sum \exp\left[-\left(\frac{x_i - \lambda}{\delta} \right) \right] (x_i - \lambda - \delta)$$

In the case of the Gumbel distribution equation (viii) is to be minimized. The inverse of I_g is the empirical information matrix. The square roots of the diagonals entries of I_g^{-1} are approximate standard errors for the estimated parameters.

3.4 Matlab results

The results found when using Matlab script files for minimizing the negative log likelihood functions with respect to unknown parameters in equations (v), (vii) and (viii) are shown in Table 3.1. The table shows the parameter estimates together with the estimated standard errors in parentheses. Since the GEV is a re-parametrisation of the three standard distributionsal form (Gumbel, Frechet and Weibull), transformed GEV parameters suitable for comparison with the estimated Weibull parameters are also given. The agreement is, as could be expected, excellent. The steps used to calculate the standard errors of the transformed GEV parameters are given in Appendix A.

UNIVERSITY of the
WESTERN CAPE

Once the results for the parameters are found, there is still a need for confirming which one of the three distributions does the observed flood water levels for the Zambezi at Katima Mulilo fit. This can be done by observing the sign of the shape parameter for the GEV distribution or by using the hypothesis testing method of likelihood ratio testing.

Table 3.1 shows that the estimate for GEV's shape parameter beta is greater than zero, indicating that our data set can be modelled well by the Weibull distribution. This confirms the result found earlier from the QQ plots, that the Weibull is a better fit for the Zambezi data. The Weibull distribution is known to have a finite upper limit [Castillo; 2004]. The value of lambda for the Weibull distribution in Table 3.1 is the estimated upper limit for the maximum flood that can be reached in any year for the Zambezi River flood level. The values for the parameter estimates in Table 3.1 are given as point estimates. Table 3.2 provides 95% confidence intervals for the parameter estimates given in Table 3.1.

Table 3.1 Matlab results of parameter estimates and standard errors obtained from fitting the Zambezi flood level data using the ML method

PARAMETER DISTRIBUTION	*LAMBDA*	*DELTA*	*BETA*
GEV **STD error**	37.6373 (0.2662)	1.5659 (0.1952)	0.4418 (0.0686)
Weibull **STD error**	41.1819 (0.4956)	3.5446 (0.6267)	2.2636 (0.5302)
Transformed GEV parameters **STD error**	41.1817 (0.2399)	3.5444 (0.3813)	2.2635 (0.3512)
Gumbel **STD error**	37.2851 (0.3985)	1.4820 (0.2752)	-

UNIVERSITY of the
WESTERN CAPE

The formula for computing the approximate confidence intervals is,

$$\hat{\phi} \pm z_{\frac{\alpha}{2}} * STDerror \text{, where } \hat{\phi} = (\hat{\lambda}, \ \hat{\delta}, \ \hat{\beta}).$$

The example of the GEV 95% confidence interval for lambda based on the formula above is:

Lower limit = 37.6373-1.96*0.2662=37.116
Upper limit = 37.6373+1.96*0.2662=38.159

Therefore the 95% CI for lambda is (37.116, 38.159).

The results in Table 3.2 below show that the confidence intervals for the GEV are narrower than the ones for Weibull distribution. This is due to the difference in the standard errors which are smaller for the GEV than for the Weibull distribution. The confidence interval for beta of the GEV distribution does not include zero which is an indication that the data fits the Weibull distribution.

To confirm the adequacy of the fitted model for the observed flood over the Zambezi we use residual plots, i.e order statistics against the expected values $\hat{F}^{-1}\left(\frac{i}{n}\right)$. Figure 3.4 shows the residual plot based on the Weibull distribution. The residual plot, based on the estimated parameters, is close to linear, with a single outlier at the lowest flood level. This confirms that the statistical model is satisfactory.

Table 3.2 95% confidence intervals for the parameter estimates

PARAMETER DISTRIBUTION	*LAMBDA*	*DELTA*	*BETA*
GEV	(37.116, 38.159)	(1.183, 1.948)	(0.307, 0.576)
Weibull	(40.211, 42.153)	(2.316, 4.773)	(1.224, 3.303)
Transformed GEV parameter	(40.711, 41.652)	(1.951, 4.292)	(1.575, 2.952)
Gumbel	(26.504, 38.066)	(0.9426, 2.0214)	-

UNIVERSITY of the
WESTERN CAPE

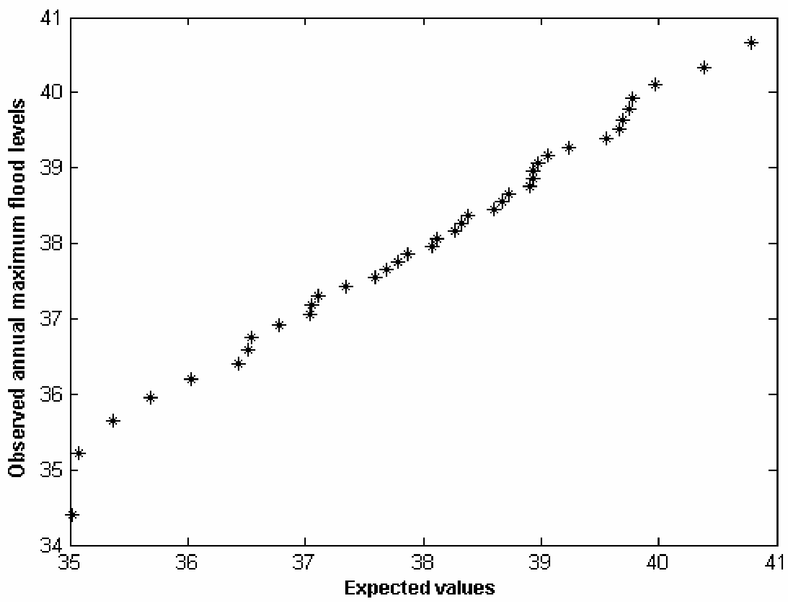

Figure 3.4 Residual plot for the Zambezi River flood water level based on the Weibull distribution

3.5 Likelihood ratio testing

Likelihood ratio testing is a hypothesis testing technique which can be used to assess the suitability of one of the three types of extreme value distribution. This test is based on comparing the likelihoods evaluated at parameter estimates for the distributions being tested.

$H_o : \beta = 0$ The observed data follows the Gumbel distribution;

$H_1 : \beta > 0$ The observed data follows the Weibull distribution;

Test statistics: D=2{74.29-69.25} =10.08

where the values 74.29 and 69.25 are the negative log-likelihoods associated with the Gumbel and Weibull distributions respectively.

Since D> $\chi_1^2(0.01) = 6.635$, the null hypothesis is rejected at the 1% level of significance, therefore the Weibull distribution is preferred.

UNIVERSITY of the
WESTERN CAPE

Chapter 4

Summary of results and conclusion

4.1 Introduction

This last chapter attempts to suggest some policy measures to strengthen Namibia's preparedness for extreme events in the light of empirical evidence. The recommendations are based on the findings in the previous chapter.

4.2 Measures to adopt on extreme value distribution modeling

The results in Chapter 3 indicated that the distribution of annual maximum flood measurements for the Zambezi River follows the Weibull distribution. The result of fitting the residual plot indicates that the three parameter Weibull distribution makes a reasonable fit. Though the Weibull distribution does not fit very well at the lowest point it remains a good fit at the upper levels. The difference between the QQ plot of the data (Figure 3.3) and the residual plot (Figure 3.4) is due to the number of parameters being fitted: the QQ plot is based on two parameters while the residual is based on the three estimated parameters. Therefore there is a need to review the statistical models targeted at modelling of extreme flood over the Zambezi River at Katima Mulilo to be compared with the following results to see if they are still appropriate. It is vital to compare these results as the methodology has improved. It is observed that the flood level possible for the Zambezi River at Katima Mulilo can be as high as 42.153 meters. Although, no flood as high as this has been observed so far for the Zambezi River, this does not exclude the chances of it happening hence the need to be prepared for such a high level of inundation of water. The ability of existing structures able to withstand such levels should be verified.

4.3 Conclusion

This paper set out to fit extreme value distributions to the observed flood water level data over the Zambezi at Katima Mulilo in Namibia. A comparison of the method used in this project to those being currently used in modelling the Zambezi River flood distributions is needed. Methods that include covariates such as the amount of rainfall received in the upper catchments, and other factors that seem to influence the distribution of flood water levels on the Zambezi River, is advocated.

Bibliography

Castillo, E., Hadi, A.S., Balakrishnan, N. & Sarabia, J.M. 2004. Extreme value and related models with applications in engineering and science. New York: John Wiley.

Coles, S. 2004. 'The use and misuse of extreme value models in practise', in Barbel Finkenstadt and Holger Rootzen (eds.) Extreme Values in Finance, Telecommunications and the Environment. London: Chapman and Hall.

Coles, S. 2001. An introduction to statistical modeling of extreme values. London: Springer-Verlag.

Gilchrist, W.G. 2000. Statistical modeling with quantile functions. London: Chapman & Hall.

Gumbel, E. J. 1958. Statistics of Extremes. New York: Columbia University Press.

Katz, R. W., Parlange, M.B. & Naveau, P. 2002. Statistics of extremes in hydrology [online].Available: http://www.isse.ucar.edu/HP_rick/pubsrick.html [2005 November 1].

Kotz, S., Johnson, N.L. & Read, C.B. 1985. Encyclopadea of Statistical sciences.Vol. 3. New York: Wiley.

Smith, R.L. 1984. 'Extreme Value Theory' in Ledermann, W. (Chief Editor). Handbook of Applicable Mathematics: Supplement. Chichester: John Wiley & Sons.

MATLAB is a registered trademark of The Mathworks, Inc. MATLAB website: http://www.mathworks.com [2005 November 1].

Mkhandi, S.H & Kachroo, R.K. 1996. Regional flood frequency analysis for Southern Africa[online].Available: http://www.cig.ensmp.fr/~hydro/ACT/ORSTOMXII/VENDREDI/MKANDI/MKANDI.PDF [2005 November 1]

Perrichi, L.R. & Rodriguez-Itube, I. 1985. 'On the statistical analysis of floods', in Atkinson A C and Fienberg E S (eds.) A celebration of Statistics, the ISI centenary volume, New York: Springer-Verlag.

Smith, R. L. 2004. 'Statistics of Extremes, with Applications in Environment, Insurance, and Finance', in Barbel Finkenstadt and Holger Rootzen (eds.) Extreme Values in Finance, Telecommunications and the Environment. London: Chapman and Hall.

Ware, R & Lad, F. 2003. Flood frequency analysis of the Waimakariri River [Online]. Available: http://www.math.canterbury.ac.nz/php/research/abstracts/abstract(2003-17).php [2005 November 1].

Appendices

Appendix A

Step for calculation of the standard errors of the transformed GEV parameter estimates in Table 3.1

Newton's approximation formula is given as

$$f(\lambda,\delta,\beta) \approx f(\lambda_0,\delta_0,\beta_0) + \frac{\partial f}{\partial \lambda}(\lambda-\lambda_0) + \frac{\partial f}{\partial \delta}(\delta-\delta_0) + \frac{\partial f}{\partial \beta}(\beta-\beta_0)$$

$$[f(\lambda,\delta,\beta) - f(\lambda_0,\delta_0,\beta_0)]^2 \approx [\frac{\partial f}{\partial \lambda}(\lambda-\lambda_0) + \frac{\partial f}{\partial \delta}(\delta-\delta_0) + \frac{\partial f}{\partial \beta}(\beta-\beta_0)]^2$$

$$\text{var}[f(\lambda,\delta,\beta)] \approx \left(\frac{\partial f}{\partial \lambda}\right)^2 \text{var}(\lambda) + \left(\frac{\partial f}{\partial \delta}\right)^2 \text{var}(\delta) + \left(\frac{\partial f}{\partial \beta}\right)^2 \text{var}(\beta) + 2\left(\frac{\partial f}{\partial \lambda}\right)\left(\frac{\partial f}{\partial \delta}\right)\text{cov}(\lambda,\delta) + 2\left(\frac{\partial f}{\partial \lambda}\right)\left(\frac{\partial f}{\partial \beta}\right)\text{cov}(\lambda,\beta) + 2\left(\frac{\partial f}{\partial \delta}\right)\left(\frac{\partial f}{\partial \beta}\right)\text{cov}(\delta,\beta)$$

UNIVERSITY of the
WESTERN CAPE

Decoding by λ^*, δ^* and β^* the GEV parameters, the transformed parameters are given by

1. $\lambda = \lambda^* + \frac{\delta^*}{\beta^*}$ 2. $\delta = \frac{\delta^*}{\beta^*}$ 3. $\beta = \frac{1}{\beta^*}$

For example, applying the approximation formula above to the function in number two:

let $\delta = \delta^*/\beta^*$ and $f(\delta,\beta) = \frac{\delta^*}{\beta^*}$ then ,

$$\text{var}[\delta] = \text{var}[\frac{\delta^*}{\beta^*}] \approx \left(\frac{1}{\beta^*}\right)^2 \text{var}(\delta^*) + \left(-\frac{\delta^*}{(\beta^*)^2}\right)^2 \text{var}(\beta^*) + 2\left(\frac{1}{\beta^*}\right)\left(\frac{-\delta^*}{(\beta^*)^2}\right)\text{cov}(\delta^*,\beta^*)$$

Similar steps were followed for the functions in 1 and 3 above.

Parameter estimates from the GEV and the empirical information matrix are used to complete the calculations.

Appendix B

Elements of the Fisher information matrix of the GEV distribution, from Castillo et *al.* (2004).

$$I_G = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} ,$$

$$m_{11} = E\left(-\frac{\partial^2 L}{\partial \lambda^2}\right) = \frac{n}{\delta^2} p,$$

$$m_{22} = E\left(-\frac{\partial^2 L}{\partial \delta^2}\right) = \frac{n}{\delta^2 \beta^2}\{1 - 2\Gamma(2-\beta) + p\}$$

$$m_{33} = E\left(-\frac{\partial^2 L}{\partial \beta^2}\right) = \frac{n}{\beta^2}\left\{\frac{\pi^2}{6} + \left(1 - \gamma - \frac{1}{\beta}\right)^2 + \frac{2q}{\beta} + \frac{p}{\beta^2}\right\}$$

$$m_{12} = m_{21} = E\left(-\frac{\partial^2 L}{\partial \lambda \partial \delta}\right) = \frac{n}{\delta^2 \beta}\{p - \Gamma(2-\beta)\}$$

$$m_{13} = m_{31} = E\left(-\frac{\partial^2 L}{\partial \lambda \partial \beta}\right) = -\frac{n}{\delta \beta}\left(q + \frac{p}{\beta}\right)$$

$$m_{23} = m_{32} = E\left(-\frac{\partial^2 L}{\partial \delta \partial \beta}\right) = \frac{n}{\delta \beta^2}\left[1 - \gamma - \frac{\{1 - \Gamma(2-\beta)\}}{\beta} - q - \frac{p}{\beta}\right],$$

where

$$\Gamma(u) = \int_0^\infty y^{u-1} e^{-y} dy \qquad \text{is the Gamma function;}$$

$$\psi(u) = \frac{d \log \Gamma(u)}{du}, \qquad \text{is the Psi function;}$$

$$p = (1-\beta)^2 \Gamma(1-2\beta),$$

$$q = \Gamma(2-\beta)[\psi(1-\beta) - \frac{(1-\beta)}{\beta}],$$

and $\gamma = 0.5772157$ is Euler's constant.

UNIVERSITY of the
WESTERN CAPE

Appendix C

Database of maximum flood water level of the Zambezi River at Katima Mulilo: Namibia, 1965-2003[1].

Year	Flood level (Meters)
1965	37.78
1966	39.67
1967	38.6
1968	39.69
1969	40.79
1970	39.55
1971	38.32
1972	37.04
1973	36.03
1974	37.87
1975	39.77
1976	39.75
1977	38.07
1978	40.39
1979	39.97
1980	38.38
1981	38.72
1982	36.43
1983	35.68
1984	37.05
1985	37.11
1986	38.27
1987	37.59
1988	38.12
1989	38.93
1990	35.37
1991	37.69
1992	35.02
1993	39.06
1994	37.35
1995	36.54
1996	35.07
1997	36.52
1998	38.9
1999	38.67
2000	38.97
2001	38.93
2002	36.78
2003	39.24

UNIVERSITY of the
WESTERN CAPE

[1] Source: Republic of Namibia: Ministry of Agriculture, Water and Rural Development, Dept of Water Affairs, Hydrology Division.

Appendix D

Matlab script files used to minimize the negative log likelihood function with respect to unknown parameters of the Gumbel distribution, and for calculation of the observed information matrix.

```
function gnll=gevlikeG(params1,x)
n=numel(x);
glambda=params1(1);
gsigma=params1(2);
if glambda<0|gsigma<0
   gnll=1.0e20;
else
z=(x-glambda)./gsigma;
gnll=n*log(gsigma)+sum(exp(-z))+sum(z);
end

function paramEs1=gevmleG(x,params1)
[paramEs1,fval]=fminsearch(@gevlikeG,params1,[],x)

function guminfomatrix(x,paramEs1)
n=numel(x);
lambda=paramEs1(1,1);
sigma=paramEs1(1,2);
z=(x-lambda)./sigma;
z2=(x-lambda)./(sigma*sigma);
z1=(x-lambda)./(sigma*sigma*sigma);
gd2ld2lam=-sum(exp(-z))*(1./(sigma*sigma));
gd2ldlamsig=-sum(exp(-z).*z1-exp(-z)*(1./sigma*sigma))-(n./(sigma*sigma));
gd2ld2sig=(n./(sigma*sigma))-(sum((exp(-z)).*(z1.*z1)-2*exp(-z).*z1))-2*sum(z1);
guI=[gd2ld2lam gd2ldlamsig;gd2ldlamsig gd2ld2sig]
guinI=inv(guI)
```

UNIVERSITY of the WESTERN CAPE

Appendix E

Two Matlab script files for minimizing the negative log likelihood function with respect to the unknown parameters of the generalized extreme value (GEV) distribution. The third script file is used to calculate the observed information matrix.

```
function gnll=gevlike1(params,x)
n=numel(x);
glambda=params(1);
gsigma=params(2);
gk=params(3);
if gk <0|glambda<0|gsigma<0|max(x)>(gsigma./gk)+glambda
    gnll=1.0e20;
else
z=(x-glambda)./gsigma;
     z1=(1./gk)*log(1-gk*(z));
       gnll=n*log(gsigma)-(1-gk)*sum(z1)+sum(exp(z1));
end

function paramEs=gevmle(x,params)
[paramEs,fval]=fminsearch(@gevlike1,params,[],x)

function gevinfomatrix(x,paramEs)
gsigma=paramEs(1,2);
gk=paramEs(1,3);
y=psi(x);
p=((1-gk)^2)*gamma(1-2*gk);
q=(gamma(2-gk))*((psi(1-gk))-((1-gk)/gk));
y1=-psi(1);
n=numel(x);
m11=(n/(gsigma*gsigma))*p;
m22=(n/(gsigma*gsigma*gk*gk))*(1-2*(gamma(2-gk))+p);
m33=(n/(gk*gk))*((pi^2/6)+(1-y1-(1/gk))^2+(2*q/gk)+(p/(gk*gk)));
m12=(n/(gsigma*gsigma*gk))*(p-(gamma(2-gk)));
m13=(-n/(gsigma*gk))*(q+(p/gk));
m23=(n/(gsigma*gk*gk))*(1-y1-((1-(gamma(2-gk)))/gk)-q-(p/gk));
gI=[m11 m12 m13;m12 m22 m23;m13 m23 m33]
ginI=inv(gI)
Stderror=sqrt(diag(ginI))'
```

UNIVERSITY of the
WESTERN CAPE

Appendix F

Matlab script files for minimizing the negative Weibull log likelihood function and for calculating the observed information matrix.

```
function nll=weiblike2(pars,x)
n=numel(x);
lambda=pars(1);
sigma=pars(2);
beta=pars(3);

if sigma<0|beta<0|max(x)>lambda
   nll=1.0e20;
else
   z=(lambda-x);
         z1=(z./sigma).^beta;
  nll=-n*log(beta)+(n*beta*log(sigma))-((beta-1)*sum(log(z)))+sum(z1);

end

function pares=wmle(x,pars)
[pares,value]=fminsearch(@weiblike2,pars,[],x)

function weibinfomatrix(x,pares)
n=numel(x);
lambda=pares(1,1);
sigma=pares(1,2);
beta=pares(1,3);
z=lambda-x;
z1=z./sigma;
z2=z1.^beta;
z3=z1.^(beta-1);
z4=z./(sigma*sigma);
z5=z1.^(beta-2);
z6=z./(sigma*sigma*sigma);
dldlam=(beta-1)*sum(z.^-2)-sum(z3*beta./sigma);
dldbet=(n./beta)+n.*log(sigma)+sum(z)-sum(z2.*log(z1));
dldsig=-(n*beta./sigma)+sum((beta*z3.*z)./sigma*sigma);
d2ld2lam=((beta-1)*sum(z.^-2))+sum((z5.*(beta*beta-beta))./(sigma*sigma));
d2ld2bet=(n/(beta*beta))+sum(z2.*(log(z1).*log(z1)));
d2ld2sig=-((n*beta)./(sigma*sigma))+sum(((beta*beta-beta).*z5.*z4.*z4)+(2*beta.*z3.*z6));
d2ldlambet=-sum(1./z)+sum(((beta.*z3.*log(z1))./sigma)+(z3./sigma));
d2ldlamsig=-sum((((beta*beta-
beta).*z.*z5)./(sigma*sigma*sigma))+((beta.*z3)./(sigma*sigma)));
d2ldbetsig=(n./sigma)-sum((z4.*z3)+(beta.*z4.*z3.*log(z1)));
I=[d2ld2lam d2ldlamsig d2ldlambet;d2ldlamsig d2ld2sig d2ldbetsig;d2ldlambet d2ldbetsig
d2ld2bet]
I1=inv(I)
stderror=sqrt(diag(I1))'
```

UNIVERSITY of the
WESTERN CAPE