University of Khartoum
Faculty of Engineering & Architecture
Department of Civil Engineering

Thesis submitted for the partial Fulfillment of
M Sc. Degree in "Water Resources
Engineering"

Low Flow Analysis In The
River Nile

By:
Yathrep EL- Tigani Ahmed Mohamed
B sc. civil Engineering

Supervised by:
Dr. Kamal ElDeen Bashar

(2007)
Lowflow Analysis In The River Nile

Table of Contents

Table of Contents ............................................................... i
List of symbol ................................................................. iv
List of table ................................................................. v
List of figures ............................................................... vi
List of appendix (A) ......................................................... vii
List of appendi ............................................................. viii
Dedication ................................................................. ix
Acknowledgment ......................................................... x
Abstract (Eng) ............................................................. xi
Abstract (Arab) ............................................................. xii

CHAPTER ONE

Introduction:

1.1 General Introduction.................................................... 1
1.2 Area of the Study......................................................... 1
1.3 Statement of the Problem............................................. 3
1.4 Objective................................................................. 3
1.5 Layout of Research.................................................... 3

CHAPTER TWO

Literature Review:

2.1 Previous Studies......................................................... 6
2.2 Back Ground And Definition....................................... 8
2.2 Factors Affecting Lowflow .......................................... 9
2.3 Use of Annual Minimum Flow Series........................... 10
2.4 Low flow frequency analysis ....................................... 11
2.5 Analysis of Annual Minimum Flow Series.................... 13
2.6 Models Used In Hydrology........................................... 13
2.7 Difference Between Low Flow And Drought.................. 15
CHAPTER THREE

Statistical Concepts

3.1 Statistical Terminology.................................................. 16
3.2 Moments: Characteristic of a sample or Distribution Function… 17
   3.2.1. Mean................................................................. 17
   3.2.2. Variance............................................................ 18
   3.2.3 Skew................................................................. 19
3.3. Form of the Distribution.................................................. 21
3.4 Distribution Functions.................................................... 21
3.5 Random Variable (variate)................................................ 22
3.6 Distribution Used In This Study........................................ 22
   3.6.1. Normal Distribution.............................................. 23
   3.6.2. Exponential Distribution........................................ 24
   3.6.3. Person Type III Distribution.................................... 24
   3.6.4. 2 Parameter Lognormal Distribution.......................... 25
   3.6.5. 3 Parameter Lognormal Distribution.......................... 25
   3.6.5.1 Use Of Chows Table Of Frequency Factor For
   Lognormal distribution.................................................. 26
   3.6.6 Extreme Value type one Distribution............................ 27
3.7 Goodness of Fit Tests.................................................... 27
   3.7.1. Chi- Sq Test.......................................................... 27
   3.7.2. RMSE Test.......................................................... 28
   3.7.3 Visual inspective................................................... 28

CHAPTER FOUR

Data : Application & Results

4.1. Preface................................................................. 31
   4.2The Data.............................................................. 31
   4.3 Fitting Of The Distribution......................................... 32
   4.3.1 EDDIEM Station................................................ 32
CHAPTER FIVE:

Conclusion And Recommendations

Summary, Conclusion And Recommendations .......................... 43

References ................................................................. 45

Appendices (A) .......................................................... 47

Appendices (B) .......................................................... 68
Acknowledgement

It is difficult in a work like this to make suitable acknowledge-
ment to all the many people from whom I have learned. I hope silence
will not be taken to imply lack of gratitude.

I would like to express my appreciation to Dr. Bashar for his
enthusiasm and dedication to producing the best product possible. My
thanks are also due to Red Sea University, also acknowledge with thanks
for my family.

Especial note of appreciation goes to Dr. Omiyma Mohamed for
her valuable advice.

YATHREP
Dedication

To those who have paved the road for me and eliminated the darkness that overwhelmed my life to plain light.
Characterization of low stream flow is required for both project proponents and regulatory authorities, as water quantity and quality issues are frequently brought to public attention during times of low water availability.

This makes it worthy to study the low flows in the rivers of the Sudan and note their characteristics for better management of our resources.

This research presents the low flow characteristics of Nile River; low flow analysis techniques were applied to three Stations in the Nile namely Eddeim at Blue Nile, Malakal along the White Nile and Khartoum at the Main Nile. The annual minimum series were extracted for each station (42 years) and six statistical distributions were fitted to each set of data. The tested distributions were the: Normal, Two parameter log normal, Three parameter log normal, Extreme value type I, Person type III and Exponential distribution. Three methods of testing goodness of fit were used to judge the best distribution that suits the observed annual minimum data (RMSE, $\chi^2$ and Visual inspection).

The Normal and Log Normal distribution are commonly selected for low flow frequency analyses in the Nile River.
خلاصه

تعتبر خصائص ومميزات انخفاض المنسوب في النهر هامة لكلا من المشاريع المفترحة أو التي يجري تنظيمها على النهر، ويعزي ذلك إلى ان كمية ونوعية المياه عادة ما يتطرق إليها الاهتمام الشامل من قبل العامة ابان أزمن انخفاض المناسب.

لذا أصبح هذا الامر مستحقاً للدراسة وجدراً بها في مسائل انخفاض مياه الانهار في السودان ومعرفة خصائصها لجل الإدارة الأفضل للموارد المائية.

لعرض خصائص التدفق المنخفض في نهر النيل تم جمع ادبي مناسب سلسلة للنيل من ثلاث محطات مختارة (محطة الديم على النيل الأزرق، محطة ملكال على النيل الأبيض، محطة الخرطوم على النيل الرئيسي).42 سنة لكل محطة - تم تحليل البيانات إحصائياً باستخدام سنة أنوع من التوزيعات وهي: التوزيع الطبيعي- التوزيع الطبيعي اللوغوغرامي ذو المعاملان- والتوزيع الطبيعي اللوغوغرامي ذو الثلاث معاملات- وتوزيع القيمة القصوي النوع الأول- وتوزيع بيرسون النوع الثالث - والتوزيع الأسوي.

.After selecting a significant amount of data from the Nile, 42 years were collected for each station. Data analysis was carried out statistically using various statistical distributions: Normal distribution, Lognormal distribution with parameters, Lognormal distribution with three parameters, and the maximum value type first degree.

 выбрنا ثلاث محطات مختارة - محطة الديم على النيل الأزرق، محطة ملكال على النيل الأبيض، محطة الخرطوم على النيل الرئيسي - وتحليل البيانات إحصائياً باستخدام سنة أنوع من التوزيعات وهي: التوزيع الطبيعي- التوزيع الطبيعي اللوغوغرامي ذو المعاملان- والتوزيع الطبيعي اللوغوغرامي ذو الثلاث معاملات - والتوزيع القيمة القصوي النوع الأول- والتوزيع بيرسون النوع الثالث - والتوزيع الأسوي.

اختيار أفضل توزيع لكل محطة تم بناء على اختبارات جودة التفوق التي تم تطبيقها على البيانات. (اختبار RMSE، اختبار كاي²، الاختبار النظري). RMSE test, Chi² test, theoretical test).

خلصت الدراسة إن التوزيع الطبيعي والطبيعي اللوغوغرامي بشكل عام هما انصب توزيعين لتحليل البيانات في نهر النيل عند انخفاضه.
Chapter One

Introduction

1.1 General Introduction

Water resources studies are receiving considerable attention, nationally, regionally, and internationally through universities, research centers, United Nations Organisations and programmers in addition to conferences. This awareness is due to the fact that water becomes a vital element for the existence of almost every human activity.

Generally, human activities and poor water management led to decreasing water quantity, deteriorating water quality and low productivity per unit volume. Thus, due to scarcity in the fresh water as a result of pollution and increasing demand, the need for food and domestic water supply has increased.

No doubt," the adequacy of stream flow to supply requirements for disposal of liquid wastes, municipal or industrial supplies, supplemental irrigation and maintenance of suitable conditions for fish is commonly evaluated in terms of low flow characteristics.

Certain of these low flow characteristics also are useful as parameters in regional draft-stroge studies, as the basis for forecasting seasonal low flows, and as indiciators of the amount of ground water flow to the stream. In some states the legal index for pollution control is tied to a low-flow characteristic."(H-C-RIGGS, 1972)

1.2 Area of the Study

Sudan is the largest country in Africa with an area of 250 million hectares. Its current population is around 32 millions (Sudan population in 2002). More than half of the population live on just 15% of the land which is the area
neighbouring the Nile. Most of the rest live in towns and around water points away from the Nile. The country is dominated by gently slopping plain with the exception of Jebel Marra, the red sea hills, Nuba Mountains and Imatong hills.

Sudan has a unique position in the Nile system. It embraces 60% of the Nile basin within its borders and is a source, a path and a drainage basin of the Nile.

The Nile River receives its water from two major sources: the Equatorial lakes plateau with its year round rains and the Ethiopian plateau with its summer rains. Its worth noting that the flow contribution of the main tributaries of the Nile River from the Ethiopian highlands is 86% (Blue Nile 59%, Baro – Akobo 14% and Tekezze – Atbara 13%) and that from Equatorial lakes is 14%.

During flooding (July to October) over 90% of the Nile River flow is from Ethiopia and less than 10% from the Equatorial lakes.

The lowflow contribution of the Equatorial lakes basin to the Nile River is attributed to the fact that a great amount of water is lost by evaporation in the swamps while the Ethiopian plateau with steep slopes drains effectively into the Nile River. (Eldaw, 2003).
Fig (1.1): Area of the study
1.3 Statement Of The Problem

Low flows are the key factors in operation of reservoirs, in demand supply management, and in sustaining the environmental flows to conserve the environment.

Recently, the world is struck by several cycles of droughts which led to a substantial decrease in low flows.

This makes it worthy to study the low flows in the rivers of the Sudan and note their characteristics for better management of our resources.

1.4 Objective

The main objective of this study is to analyze the low flows in several rivers of the Sudan to identify their low flow characteristics for better management and operation of reservoir and water resources as general.

The specific objectives include:

- Extraction of annual minimum flows
- Fitting of Several distributions to the extracted data
- Realization of the Q-T relationship for low flows in the selected rivers.

1.5 Lay Out Of The Thesis

- Chapter one gives an introduction to the study where the problem is stated, the study area is outlined, and the objectives were set.
- Chapter two (Literature Review) gives previous studies.
- Chapter three (Theory) provides background information on the definition and characteristics of low flows, factors affecting it, use of annual minimum flow series and different between low flows and drought. At this chapter also information about models used in
hydrology. This chapter also gives (Statistical Terminology) the objective of this part is to define some of the more important statistical terminology and moments characteristics. Type of distribution used at this study and the goodness of fit tests.

- Chapter Four (Data, application and results) describe analysis of the data which have been collected from three gauging stations from Nile and tributaries on the Sudan (Blue Nile, White Nile and main Nile). It should be noted that all calculations have been made using EXCEL work sheet.

- The results of this research and recommendations for further work were determined at Chapter Five.
2.1 Previous Studies

Low-flow statistics indicate the probable availability of water in streams during times when conflicts between water supply and demand are most likely to arise. Because of this, low-flow statistics are needed by Federal, State, regional, and local agencies for water-use planning, management, and regulatory activities. These activities include
(1) developing environmentally sound river-basin management plans,
(2) siting and permitting new water withdrawals, interbasin transfers, and effluent discharges,
(3) determining minimum stream flow thresholds for maintenance of aquatic biota, and
(4) land-use planning and regulation.

Low-flow statistics are also needed by commercial, industrial, and hydroelectric facilities to determine availability of water for water supply, waste discharge, and power generation. Low-flow statistics can be calculated from stream flow data collected at locations where the U.S. Geological Survey (USGS) operates data-collection stations, but it is not possible to operate stations at every site where the statistics are needed. Because of this, methods are needed for estimating low-flow statistics for streams for which no data are available.

In 1988, the USGS began the first of three studies to develop and evaluate methods for estimating low-flow statistics for ungaged Massachusetts streams and to provide estimates of the statistics for selected locations on ungaged streams. These studies were done in cooperation with the Massachusetts Department of Environmental Management, Office of Water Resources (MOWR) and are referred to as the Basin Yield studies. The MOWR uses the stream flow statistics to develop water-resources management plans for each of the 27 major river basins in Massachusetts and provides the
stream flow statistics to other State and local agencies to support their decision making processes.

Five other reports have been published as a result of the Basin Yield studies (Ries, 1994a, 1994b, 1997, 1999, 2000). The first three reports describe studies done to develop regression equations for use in estimating low-flow statistics for ungaged sites. The fourth report describes and provides data for a network of 148 low-flow partial-record (LFPR) stations that was established in 1988 at the beginning of the first Basin Yield study and continued through 1996, during the third Basin Yield study. The fifth report describes a World Wide Web application that enables users to select sites of interest on streams and then to obtain estimates of stream flow statistics and basin characteristics for the sites.

In the Sudan there is a not study about low excepted drought studies. "Drought is a cause of great human sufferings including failure and pasture, widespread loss of livestock, famines, etc. as such it is important driving forces of environmental refugees.

Sudan is one of the African tropical countries that suffered extreme the year 1984 which had a severe impact on its human, animal and represent the duration-dependant termination ate of the data set.

The reliability analysis in terms of hazard rate and failure function are used to represent the drought duration function. Parameters are estimated using least-squares method.

The historical data of Nile annual stream flow at Eddeim station is analyzed by this methodology results obtained showed that one-year drought of Blue Nile has a return period of years and nine-years drought has a return period of 370 years - "Ibrahim & Seifeldin- 1994".
CHAPTER THREE

Theory

3.1 Low Flow Analysis

3.1.1 Background And Definition

In lowflow hydrology two questions were successively asked about a particular river site:

(a) Does the river supply a particular water demand at all times?

(b) If not, how much water must be stored in order to meet any deficiency which may arise?

In (a) "demand" covers a variety of meanings as discussed below. Also the term "at all times" can be replaced by a probabilistic statement such as "90% of the time".

The techniques of annual minimum flow analyses and annual drought volume analyses are useful when the demand is a small proportion of the total annual river flow.

In answering the question raised in (b), the assumption that demand remains constant at all times throughout the year greatly simplifies the analysis required. The assumption allows a certain technique, drought volume analysis, to be understood more easily. However, a method of analysis is required which takes into account a varying demand also because in times of extremely lowflows and water shortage, steps are usually taken to reduce (public/industrial) demand for water. This means that water extraction from a river or reservoir usually decreases in practice at times of low natural flow.

The term demand flow is used here to express total water requirement at a particular point on a river.

It includes both direct water requirements and residual water requirements or compensation water.
A. Direct water requirements

May be divided into categories:

(i) Water supply for domestic, trade, agricultural and municipal uses.

(ii) Water required for dilution of sewage/wastewater in rivers lakes, or estuaries.

(iii) Provision of water for hydroelectric power generators.

(iv) Provision of water for amenity, navigation and fisheries.

B. Residual water requirements:

In addition to direct requirements, sufficient water must be left in the stream or discharged from the new reservoir to allow other intakes downstream to continue to operate successfully. There must also be an allowance made for the rights of riparian owners and for migrating fish.

This is called compensation water. The amount needed is related to the state of the river and the effluents that are discharged into it. It should be large enough to maintain the river in a satisfactory state of "health".

(C. cunnane -1981)

3.1.2 Factors Affecting Low Flows

The low flow of streams is affected by many factors, such as underlying geology, impoundment regulations, stream diversions, basin topography, and short and long-term weather and climatic changes. The size of the drainage area is also a factor that affects the magnitude of low flow; generally, drainage area and low flow are directly proportional. Land-use changes can affect low flow.

If large areas of a basin are covered by impervious surfaces, such as parking lots, the increase in storm runoff and the reduced infiltration of precipitation will decrease the base flow of a watershed.
**Geology:**

The underlying geology of a basin affects low flow because at low flow (or base flow) most, if not all, of the stream flow is from ground water discharge. The more permeable or more highly fractured the underlying bedrock, the more water that can be stored and subsequently discharged to the stream.

Ground water flow paths seldom mimic local topography, and ground water and surface water divides may not coincide. Also, streams crossing carbonate valleys may lose water to sink holes or may gain water from springs. Therefore, estimating low flow statistics in streams underlain by carbonate rock where no stream flow-measurement station is present has questionable validity.

**Impoundments And Diversions**

Dams and reservoirs have a major effect on low flow of streams. Reservoirs are used for water supply and to augment low flow, which increases the magnitude of low flow. (Curtis-1998)

**3.1.3 Use Of Annual Minimum Flow Series**

On the assumption that a record of flows at the sites in question is available, the minimum flow in each year is noted. The resulting series $q_1, q_2 \ldots q_n$ is then assumed to be a random sample of size $N$ from a population in which the $q$ values have a distribution. An estimate of such a distribution may be obtained from the sample.
Figure (2-1) shows return period, $T$, in this context is now taken as the reciprocal of the non-exceedance probability. The value $q_1$ corresponding to the design probability, of failure $1/T$ is then obtained from the distribution. If the value of $q_T$ is large in comparison to $Q_D$, the demand flow, then the river is considered to be able to supply the demand satisfactorily. On the other hand if the $q_t$ is less than, or of the same order of magnitude as, $Q_D$ then the river alone without some form of regulation could not be considered satisfactory for supplying the demand.

It should be pointed out that $q_T$ for values $T=5, 10, 20$ years do not differ greatly from one another and are quite small in comparison to the mean lowflow of the river. (C. Cunnane - 1981).

### 3.1.4 Low-flow Frequency analysis:

Partially treated waste water is common processes improve the overall quality of the total flow. During periods of low flow, the volume of waste water may be too large to be safely discharged without reducing the quality of the water below established water – quality standards. When evaluating sites for suitability for a manufacturing or commercial business, it is important to assess the probability that the stream will almost always have sufficient flow to meet
the need for discharging wastewater. Such probabilities are estimated using a low – flow frequency analysis at the site.

While instantaneous maximum discharge is used for flood frequency analysis, low- flow frequency analysis usually specifies flow duration (e.g, 7 days). The instantaneous discharge is used with high flows because damage often occurs even if the site is inundated only for a very short period of time. For lowflows, however, high pollution concentrations over very short periods of time may not damage the aquatic life of the stream. Thus, the duration, such as 7 days or 1 month, is specified in establishing the policy.

One difference between low-flow and flood frequency analyses is that the data for low-flow analysis consist of the annual events that have the lowest average flow of the required duration D during each water year of record. Thus the records of flow for each water year are evaluated to find the period of D consecutive days during which the average flow was the lowest. These annual values are used as the sample data. The record of n yeas is then evaluated using frequency analysis.

Once the data record has been collected, that used for flood- frequency analysis is quite similar to that used for flow- frequency analysis. The major differences are listed here:

1- Instead of using the exceedence probability scale, the non- exceedence scale (i.e., the scale at the bottom of the probability paper) is used to obtain probabilities. The non-exceedence scale is important because the T- yr event is the value that will not be exceeded.

2- The data are ranked from low to high, with the smallest sample magnitude associated with a weibull probability of 1/(n+1) and the largest magnitude associated with probability of n/ (n+1). Any other plotting position formula could be used in place of the weibull. However, plotting position probabilities are non-exceedence probabilities.

3- The mathematical model for predicting magnitudes remains: \( Y = Y + z_s y \), but since non-exceedence is of interest, the T-yr values are from the lower tail of the
distribution; thus, values obtained from a normal distribution table should be from the lower tail (i.e., negative values of Z). (Richard 2005).

3.1.5 Analysis Of Annual Minimum Flow Series

The annual minimum flow series is formed by selecting the lowest flow occurring in each year of record. The set of observed annual minima at any gauging station is assumed to be a random statistical sample from the population of all possible annual minima at the site. An assumption is made about the form of the population distribution function and the validity of the assumption is tested. If it is not reasonable, an alternative distribution is proposed and example is obtained. The annual minimum flow of required return period is then estimated form the data with the help of this distribution.

3.1.6 Models Used In Hydrological Studies

In the mathematical sense the word model describes a system of assumptions, equations and procedures intended to describe the performance of a prototype system. Since the advent of the digital computer the term hydrologic model has come to mean a relatively complex mathematical description of the hydrologic cycle designed for solution on a computer (Linsley-1988).

Hydrological models generally use a very simplified picture of the process being modeled because an experience has often showed this to be adequate. It is a matter of convention to describe as forecasting models those which answer questions such as “what will the runoff be tomorrow given today's runoff, rainfall and evaporation?” , while the description of prediction models is reserved for those which answer the question “how often will a flow q be exceeded in this river during the next 50 years, with probability 0.1?” This convention is not universal, even among hydrologist but it is nevertheless useful. However, in control and systems engineering the word predictions nearly always used for what we would call forecasting). Table 2.1 shows classification of hydrologic models by Richard (2005).
### Table 2.1: Classification of hydrologic models

<table>
<thead>
<tr>
<th>I. Peak discharge models</th>
</tr>
</thead>
<tbody>
<tr>
<td>(A) Calibrated models</td>
</tr>
<tr>
<td>1- Single-return-period equations</td>
</tr>
<tr>
<td>2- Index flood</td>
</tr>
<tr>
<td>3- Moment estimation</td>
</tr>
<tr>
<td>(B) Uncalibrated model</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>II. Single-event hydrograph models</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Design storm</td>
</tr>
<tr>
<td>B. Actual record</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>III. Watershed (continuous) multiple-event models</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Actual record</td>
</tr>
<tr>
<td>B. Synthetic record</td>
</tr>
</tbody>
</table>

| IV. Joint probability models |
3.1.7 Difference Between Lowflow And Drought

Lowflow is the "flow of water in a stream during prolonged dry weather" according to the World Meteorological Organization. Many states use design flow statistics such as the 7 Q 10 (The lowest 7-day average flow that occurs on average once every 10 years) to define low flow for the purpose of setting permit discharge limits.

A drought is a more general phenomenon than lowflow and can be characterized by more than low stream flows. Drought can be classified as a meteorological, atmospheric, agricultural, hydrological, and water management. Typically, a drought is defined in terms of water availability for various designated uses. But "a hydrological drought" is typically described by a reduction in lake storage, a decrease of stream flow discharge and a lowering of groundwater levels over large areas, over one year or several consecutive years.

When drought and the lowflow were defined some misunderstanding appear between these two phenomenons:

A lowflow is a seasonal phenomenon (e.g. the "dry season") and is an important component of the flow regime in any river or stream. By contrast, a drought is a natural event that results from an extended period of below average precipitation. While droughts include lowflows, a continuous seasonal lowflow event is not necessarily a drought.
3.2 Statistical Terminology

Each year, considerable effort and resources are used on the collection of hydrologic and meteorological data. These data represent observations on a population. The set of observations recorded in the past represents a sample, not the population. It is important to understand that the sample is a random collection or subset of the population. As an example, we might setup an instrument to record the lowest flow level during the year at a particular location along a stream reach. The lowest observed flow for that year is an observation on a random variable because we can be assured that the annual lowflow will be different each year. If we obtain the measurement each year, then after n years we will have a sample of size n. The objective of the data analysis will be to use the sample to identify and draw inference about the population. After the population has been estimated, the assumed population will be the basis for designs. The population would consist of all values that could have been observed in the past and will occur in the future. Unfortunately we can not make measurements after the event has passed, nor can we measure even that have not yet occurred. Thus we can not know the population, and we have to use the sample to draw inferences about what will happen in the future.

Before providing a more formal discussion of these statistical terms, it is most important to point out the basis for much of statistical analysis. Statistical inference concerned with methods that permit an investigator to make generalizations about populations using information obtained from random samples. That is, samples are a subset of the population.

If knowledge of the population represents complete information, statistical analyses of sample data represent decision making under conditions of incomplete information (i.e., under conditions of uncertainty). Statistical methods are methods for collecting, organizing, summarizing, and analyzing quantitative information. (Richard, 2005).
3.2.1 Moments: Characteristic of a Sample or Distribution Function

Whether summarizing a data set or attempting to find the population, one must characterize the sample. The moments are useful description of data; for example, the mean, which is a moment, is an important characteristic of a set of observations on a random variable, such as rainfall volumes or concentrations of a water pollutant. A moment can be referenced to any point on the measurement axis; however, the origin (i.e., zero point) and the mean are the two most common reference points.

Although most data analyses use only two moments, in some statistical studies it is important to examine three moments:

1. The mean is the first moment of values measured about the origin.
2. The variance is the second moment of values measured about the mean.
3. The skew is the third moment of values measured about the mean.

3. 2.1.1 Mean

The mean is the first moment measured about the origin; it is also the average of all observations on a random variable. It is important to note that the population is most often denoted as $\mu$, while the sample mean is denoted by $\bar{x}$. For a continuous random variable, it is computed as:

$$ (x \text{ or } \mu) = \int_{-\infty}^{\infty} x f(x) \, dx $$

For a discrete random variable, the mean is given by:

$$ (x \text{ or } \mu) = \sum_{i=1}^{n} x_i f(x_i) $$
If each observation is given equal weight, then \( f(x_i) \) equals \( \frac{1}{n} \) and the mean discrete random variable is given by:

\[
( \bar{x} \text{ or } \mu ) = \frac{1}{n} \sum_{i=1}^{n} x_i
\]

Figure (3.1) shows the effect of mean as a location parameter on the distribution.

**Fig(3.1): Comparison of distribution with different mean.**

### 3.2.1.2 Variance

The variance is the second moment about the mean. The variances of the population and sample are denoted by \( S^2 \), respectively. The units of the variance are the square of the units of the random variable.
The variance of a sample can be computed by:

\[ S^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} x_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} x_i \right)^2 \right] \]

Figure (3.2) shows the effect of variance as a scale parameter on the distribution

\[ \text{Fig (3.2): Comparison distributions with the same mean but different variances.} \]

**Standard Deviation**

By definition, the standard deviation is the square root of the variance. It has the same units as both the random variable and the mean: therefore, it is a useful descriptor of the dispersion or spread of either a sample of data or a distribution function. The standard deviation of the population is denoted by \( \sigma \), while the sample value is denoted by \( S \).

**3.2.1.3 Skewness**

The skewness is the third moment measured about the mean. Unfortunately, the notation for skewness is not uniform from one user or text to another. The sample skewness can be denoted by \( g \), while \( \gamma \) can be used to indicate the
skewness of the population. Mathematically, it is given as follows for a continuous random variable:

\[(g \text{ or } \gamma) = \int_{-\infty}^{\infty} (x - \mu)^3 f(x) \, dx\]

For a discrete random variable, the skewness can be computed by:

\[(g \text{ or } \gamma) = \sum_{i=1}^{n} (x_i - \mu)^3 f_i\]

It has units of the cube of the random variable; thus if the random variable has units of pounds, the skewness has units of pounds^3.

The skewness is a measure of symmetry. A symmetric distribution will have a skewness of zero, while a non symmetric distribution will have a positive or negative skewness depending on the location of the tail of the distribution. If the more extreme tail of the distribution is to the right, the skewness is positive: the skewness is negative when this more extreme tail is to the left of the mean. This is illustrated in Fig. (3.3).

**Fig. (3.3): Comparison of the skewness of distributions**

(a) \(g = 0\)  -  (b) \(g > 0\)  -  (c) \(g < 0\)
3.2.2 Form of distribution

Either of the graphs f(x) or F(x) defines a distribution completely but the amount for details they contain may not always be useful. Usually the 1st three moments contain most of the useful information in practical cases. The practical range of variate values is:

\[ \mu - 45 \text{ to } \mu + 45 \text{ in symmetric distribution} \]
\[ \mu - 25 \text{ to } \mu + 65 \text{ in positive skewed distributions} \]
\[ \mu - 56 \text{ to } \mu + 25 \text{ in negative skewed distributions.} \]

Asymmetry is measured by the third moments, \( \mu_3 \)

- If \( f(x) \) is symmetric about \( X = \mu \) then \( M_3 = 0 \)
- If \( f(x) \) has larger tail to the right \( \mu_3 > 0 \) and the distribution is positive skewed.
- If \( f(x) \) has longer tail to the left \( \mu_3 < 0 \) and the distribution is negative skewed. (Bashar, UCWR)

3.2.3 Distributions

A distribution is an attribute of a statistical population. If each element of a population has a value X then the distribution describes the constitution of the population as seen through its X values. The distribution tells:

- About the location of the data on the x-axis, whether the data are bunched together or spread and whether it is symmetrically disposed on the x-axis or not. These are described by the mean, standard deviation and skewness.
- The relative frequency or proportion of various x values in the population same as what histogram do for a sample
• Probabilities because relative frequencies are probabilities. Pr(X<\(x\)) is the probability that X value on an element drawn randomly from a population would be less than a particular value \(x\).

3.2.4 Random Variable (Variate)

Any random variable which has an associated probability distribution is called random variable (r.v) or variate and may be continuous or discrete.

The complement of \(F(X)\) is called exceedance probability of \(X\), \(1 - F(X)\). The reciprocal of the exceedance probability is the return period \(T\).

\[
\frac{1}{1-F(X)} = T
\]

Or

\[
F(X_T) = 1 - \frac{1}{T}
\]

Where \(X_T\) has a return period \(T\).

In repeated random trials from the population, the \(X_T\) value is exceeded on average once in every \(T\) trials. The exceedance does not occur in a regular cyclic pattern.

3.2.5 Distributions Used In This Study

The distributions used in the study are:

• Normal Distribution.

• Exponential Distribution.

• Person type 111 Distribution (P111).

• 2 Parameter log normal Distribution (2LN).

• 3 Parameter log normal Distribution (3LN).
• Extreme Value Type1 (EV1).

### 3.2.5.1 Normal Distribution

It has a probability density function pdf of the form

\[
f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2}
\]

Where \( \mu \) and \( \sigma \) (\( \sigma > 0 \)) are parameters of location and scale.

f(x) is symmetric about \( X = \mu \) and is of bell shape.

![Diagram](image)

**Fig (3.4): The effect of location and scale parameter of distribution.**

Assigning different values for \( \mu \) and \( \sigma \) a family of different normal distributions can be obtained and figure (3.4) shows that.
3.2.5.2 Exponential Distribution

This is a two parameter distribution. Its manipulation depends on exponential and logarithmic tables. If has the form.

\[ f(x) = \frac{1}{\beta} e^{-\left(\frac{x-x_0}{\beta}\right)} \]

The cumulative form is

\[ F(x) = 1 - e^{-\left(\frac{x-x_0}{\beta}\right)} \]

3.2.5.3 The Person type 3 distribution

It is a distribution with three parameters and probability density function given by

\[ f(x) = \frac{(x-x_0)^{\gamma-1} e^{-\frac{x-x_0}{\beta}}}{\beta^\gamma \Gamma(\gamma)} \quad \text{and} \quad F(x) = \int_{x_0}^{x} \frac{(x-x_0)^{\gamma-1} e^{-\frac{x-x_0}{\beta}}}{\beta^\gamma \Gamma(\gamma)} dx \]

Notes:

1. When \( \gamma = 1 \) the Person Type 3 distribution is identical with the two parameter exponential distribution

2. The notation used for the exponential, gamma and Person Type 3 distributions and the use of \((x_0, \beta, \gamma)\) as parameters mean that the reduced or standardized variate has its distribution expressed in its simplest form. The parameters \(x_0, \beta\) and \(\gamma\) fall into the general categories of location, scale and shape parameters respectively. But the mean, standard deviation and coefficient of skewness fall into these categories also.

\[ \gamma = 4g^2 \]
\[ \beta = \sigma / \sqrt{\gamma} = \sigma g/2 \]

\[ x_0 = \mu - \beta \gamma = \mu - 2\sigma / g \]

### 3.2.5.4 Two Parameter Log Normal Distribution

The variate \( x \) is positive and the bound \( x_0 \) does not appear and hence the variate \( z = \ln(x) \) is normally distributed with mean \( \mu_z \) and variance \( \sigma_z^2 \). The pdf is

\[ \phi(z) = \frac{1}{\sigma_z \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{z - \mu_z}{\sigma_z} \right)^2} \quad \text{and} \quad f(x) = \phi(x) \left| \frac{dz}{dx} \right| \]

**Mean**

\[ \text{Mean} = E(x) = e^{\mu_z + \frac{1}{2} \sigma_z^2} \]

**Variance**

\[ \text{Variance} = E \left( x - E(x) \right)^2 = \left[ e^{\mu_z + \frac{1}{2} \sigma_z^2} \left( e^{\sigma_z^2} - 1 \right) \right]^2 \]

\[ CV = \left( e^{\sigma_z^2} - 1 \right)^{1/2} \Rightarrow \sigma_z^2 = \ln(1 + CV^2) \]

\[ g = 3CV + CV^3 \]

For practical values of \( \sigma_z^2 \) in the range \( 0.1 < \sigma_z^2 < 0.6 \) the relation is almost linear.

\[ g = 0.52 + 4.85 \, \sigma_z^2 \]

which is correct to within 2% in the stated range.

### 3.2.5.5 Three Parameter Lognormal Distribution

It differs from the two parameter lognormal distribution by introduction of lower bound \( x_0 \) so \( z = \ln(x - x_0) \) is normally distributed with mean \( \mu_z \) and variance \( \sigma_z^2 \). Thus it is a three parameters distribution; \( x_0 \) being location parameter while \( \mu_z \) and \( \sigma_z^2 \) control the scale and shape respectively. The pdf is.
\[ f(x) = \frac{1}{(x-x_0)\sigma_x \sqrt{2\pi} e^{-\frac{1}{2}\left(\frac{1}{\sigma_x} \log(\frac{x-x_0}{\mu_x})\right)^2}} \]

The moments of \( x \) can be obtained from the corresponding moments in the two parameter distribution because the two variates differ only in the location parameter \( x_0 \). Therefore, writing two parameter varaite temporarily as \( \xi \) the three parameter variate \( x \) is

\[ x = x_0 + 1.\xi \]

Then:

\[
\text{mean} = E(x) = x_0 + E(\xi) = x_0 + e^{\mu_x} + \frac{1}{2}\sigma_x^2 \\
\text{Variance} = \text{var}(x) = 1^2 \text{var}(\xi) = \left[e^{\mu_x} + \frac{1}{2}\sigma_x^2 \left(e^{\sigma_x^2} - 1\right)\right] \frac{1}{2}
\]

### 3.2.5.5.1 Use of chow’s table of frequency factor for lognormal distribution

Chow’s table give \( K_T \) for the expression

\[ xT = \mu - \sigma KT \]

Where \( K_T \) is a function of the return period \( T \) and skewness

\[ g = CS \]

or equivalently CV since

\[ CS = CV^3 + 3CV. \]
3.2.5.6 Extreme Value Type I Distribution

This is a two parameter distribution. It is also referred to as double exponential and/or Gumball distribution. Its form arises from consideration of the statistical properties of sample extreme values. It was first introduced in hydrology by Gumball. It has the form:

\[ f \left( x \right) = \frac{1}{\alpha} e^{-\left( \frac{x - \mu}{\alpha} \right)} - e^{-\frac{(x - \mu)}{\alpha}} \]

And

\[ F \left( x \right) = e^{-\frac{(x - \mu)}{\alpha}} \]

3.2.6 Goodness Of Fit Tests

The results obtained in samples do not always agree exactly with the theoretical results expected according to the rules of probability.

So goodness of fit test is needed. There are many methods to do that, at this thesis we selected three methods according to lecture notes on statistical hydrology (Gamal, 2004):

i) Chi-Sq Test

ii) "RMSE" Test

iii) Visual inspection.

3.2.6.1 Chi-square Goodness Of Fit Test:

One of the most commonly used tests for goodness of fit of empirical data to specified theoretical frequency distributions is the chi-square test. This test makes a comparison between the actual number of observations and the expected number of observations (expected according to the distribution under test) that fall in the class intervals. The expected numbers are calculated by
multiplying the expected relative frequency by the total number of observations. The test statistic is calculated from the relationship.

\[ \chi^2_c = \sum_{i=1}^{K} \left( \frac{(O_i - E_i)^2}{E_i} \right) \]

Where \( k \) is the number of class intervals, \( O_i \) is the observed and \( E_i \) the expected (according to the distribution being tested) number of observations in the \( i^{th} \) class interval. The distribution of \( \chi^2_c \) is a chi-square distribution with \( k-p-1 \) degrees of freedom where \( p \) is the number of parameters estimated from the data. The hypothesis that the data are from the specified distribution is rejected if

\[ \chi^2_c > \chi^2_{i-a, k-p-1} \]

3.2.6.2 Least Squares Test (Root Mean Square Error)

This method is used to compare the fit of different distributions to a data sample is to compute the

\[ \text{RMSE}_j = \left[ \frac{\sum_{i=1}^{n} (X_{i,j} - y_{i,j})^2}{n - mj} \right]^{\frac{1}{2}} \]

Where; \( x_i, i = 1,2,\ldots n \) are the recorded events \( y_i, i = 1,2,\ldots n \) are the event magnitudes computed from the \( j^{th} \) probability distribution at probabilities computed from the stored ranks. \( m_j \) is the number of parameters estimated from the \( j^{th} \) distribution.

3.2.6.3 Visual Inspection

The first step in examining the relationship between variables is to perform a graphical analysis. Visual inspection of the data can provide the following information:
1. Identify the degree of common variation, which is an indication of the
degree to which the two variables are related.

2. Identify the range and distribution of the sample data points.

3. Identify the presence of extreme events.

4. Identify the form of the relationship between the two variables.

5. Identify the type of the relationship.

Each of these factors is of importance in the statistical analysis of sample
data and decision making (Richard-2005).

Frequency curves arising in practice take on certain characteristic shapes, as
shown in Fig (3-5) (Murray-second edition).
Fig (3.5): Frequency curves
CHAPTER FOUR

Data, Application And Results

4.1 Preface

This chapter proceeds to analyze the data for three gauging stations namely, Blue Nile at Eddeim, White Nile at Malakal and the River Nile at Khartoum.

The data collected for each station is processed and prepared for analysis. Different distributions were fitted to each location. The Q-T relationships for the return periods 2, 5, 10, 15, 20, 100, and 1000 were obtained. $\chi^2$ Test, the RMSE and visual inspection were used to judge the best fit.

It must be noted that all the calculations were done using EXCEL worksheet.

4.2 The Data

The flow data for the three stations was collected from the Ministry of Irrigation and Water Resources, Khartoum.

The annual minimum flow in each year is noted (in cumecs) to be used for the analysis. Table (4.1) gives a summary of the annual minimum flow in the three stations and Table (4.5) gives summary of the results of $\chi^2$ Test and the results of RMSE Test are in Table (4.6).
Table (4.1): Summary of the annual minimum flow (Mm$^3$/d) in the three stations:

<table>
<thead>
<tr>
<th>Station</th>
<th>No. of years</th>
<th>Starting years</th>
<th>Mean</th>
<th>Std.</th>
<th>Skewness</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eddeim</td>
<td>42</td>
<td>1964</td>
<td>5.73</td>
<td>1.66</td>
<td>0.25</td>
<td>2.88</td>
<td>9.84</td>
</tr>
<tr>
<td>Malakal</td>
<td>42</td>
<td>1965</td>
<td>10.69</td>
<td>0.32</td>
<td>0.34</td>
<td>9.78</td>
<td>11.8</td>
</tr>
<tr>
<td>Khartoum</td>
<td>42</td>
<td>1965</td>
<td>11.59</td>
<td>4.77</td>
<td>-0.18</td>
<td>0.01</td>
<td>23.06</td>
</tr>
</tbody>
</table>

4.3 Fitting of the distributions

The observed annual minimum data extracted in section 4.2 above for the three selected stations were used to fit six distributions.

Mainly the method of moments is used in the fitting process. Judgment of the best fit is done on the basis of three methods namely the RMSE, $\chi^2$ and the visual inspection.

4.3.1 Eddeim Station

This station is the 1$^{st}$ gauging stations on the Blue Nile inside the border of Sudan.

The six distributions were fitted to the data of this station from (1964-2005) it must be noted that this station was stopped at 97,98 and 99. The fitted distributions are used to estimate the lowflows quantities for return periods of 2, 5, 10, 15, 20, 25, 100, 1000 year.

Figure (4.1) shows the observed data and Figure (4.2) shows a plot of observed and estimated lowflows, while Table (4.2) shows the station information.
The best fit test revealed that the best distribution is Two Parameters Log Normal (2LN), while Normal distributions (NOR) ranked the 2nd.

Both distributions fit data better than others.

Other distributions such as Person Type Three (P III), Exponential (Exp) and Extreme Value Type One (EVI) gave satisfactory results at relatively low return periods. Three Parameters Log Normal (3LN) is rejected at Elddeim stations.
Table (4.2): Station Information

- River: Blue Nile
- Station: Eddeim
- Unit Series: Mm$^3$/d
- Period: 1964-2005
- Time Step: daily

FIG(4.1): Observed Annual Minimum Flow At Eddeim Station
FIG(4.2): Eddeim Station-Discharge For Different Return Periods

Q (M m$^3$/d)

Year

FIG(4.2): Eddeim Station-Discharge For Different Return Periods
4.3.2 Malkal Station:

This station is on the White Nile. It has 42 years of data (1965-2005). The six distributions were fitted to data of this station. The fitted distribution is used to estimate the lowflows quantities for the return periods of 2, 5, 10, 15, 20, 25, 100, 1000 year. Figure (4.1) shows the observed data and Figure (4.4) shows a plotting of observed and estimated lowflows, while Table (4.3) shows the station information.

The best fit test revealed that the best distributions to fit data at this station are Normal and Two Parameters Log Normal (2LN) both fit data better than others. Other distributions such as Person Type Three (P III) and Extreme Value Type One (EVI) distributions have good results, also Exponential (Exp) distribution gave satisfactory results at relatively low return periods.

At this station Three Parameters Log Normal (3LN) rejected. The best distribution was selected on the basis of the Chi-Square Test ($x^2$), the Root Mean Square Error (RMSE) and Visual Inspection.
Table (4.3): Station Information

- River: White Nile
- Station: Malakal
- Unit Series: Mm³/d
- Period: 1965-2005
- Time Step: daily

FIG(4.3): Malakal Station - Observed Annual Minimum Flow
FIG (4.4): Malakal Station - Discharge For Different Return Periods
4.3.3 Khartoum Station:

This station is on the main Nile. It has 42 years of data (1965-2005). The six distributions were fitted to the data of this station. The fitted distributions are used to estimate the lowflows quantities for the return period of 2, 5, 10, 15, 20, 25, 100, 1000 year.

Figure (4.5) shows the observed data and Figure (4.6) shows a plot of observed and estimated lowflows, while Table (4.4) shows the station information.

The best fit test revealed that the Three Parameters Log Normal (3LN) is the best distribution to fit the data at this station.

Other distributions such as, Normal, Person III and Two Parameters Log Normal (2LN) distribution, gave satisfactory results at relatively low return periods.

But at this station Exponential (Exp) and Extreme Value Type 1 (EV1) distribution are rejected.
Table (4.4): Station Information

- River: Main Nile
- Station: Khartoum
- Unit Series: Mm$^3$/d
- Period: 1965-2005
- Time Step: daily

FIG(4.5): Khartoum Station - Observed Annual Minimum Flow
FIG(4.6): Khartoum Station-Dicharge For Different Return Periods

(\text{Mm}^3/\text{d})

(year)

EXP
NORMAL
EV1
Observed
2LN
3LN
PERSON

FIG(4.6): Khartoum Station-Dicharge For Different Return Periods
Table (4.5): Chi-Square Test Results

<table>
<thead>
<tr>
<th>STATION</th>
<th>EXP</th>
<th>NOR</th>
<th>EV1</th>
<th>2LN</th>
<th>PERSON111</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eddeim</td>
<td>1.74</td>
<td>0.91</td>
<td>11.849</td>
<td>0.4115</td>
<td>0.984328</td>
</tr>
<tr>
<td>Malakal</td>
<td>0.029</td>
<td>0.0255</td>
<td>0.229</td>
<td>0.0577</td>
<td>0.02465</td>
</tr>
<tr>
<td>$\chi^2$ Tablet</td>
<td>18.49</td>
<td>18.49</td>
<td>18.49</td>
<td>18.49</td>
<td>17.71</td>
</tr>
</tbody>
</table>

Table (4.6): RMSE Test Results

<table>
<thead>
<tr>
<th>STATION</th>
<th>EV1</th>
<th>EXP</th>
<th>P111</th>
<th>NOR</th>
<th>IOG NOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eddeim</td>
<td>1.75</td>
<td>0.96</td>
<td>3.18</td>
<td>0.33</td>
<td>0.27</td>
</tr>
<tr>
<td>Malakal</td>
<td>6.98</td>
<td>5.39</td>
<td>8.59</td>
<td>5.433</td>
<td>5.469</td>
</tr>
<tr>
<td>Khartoum</td>
<td>85.35</td>
<td>58.81</td>
<td>4.1869</td>
<td>61.44</td>
<td>29.78</td>
</tr>
</tbody>
</table>
CHAPTER FIVE
Conclusions And Recommendations

Lowflow analysis techniques were applied to three stations in the Nile namely Eddeim along the Blue Nile, Malakal along the White Nile and Khartoum at the Main Nile. The annual minimum series were extracted for each station (42 year) and six statistical distributions were fitted to each set of data. The distributions include Normal, Two Parameters Log Normal, Three Parameters Log Normal, Person Type Three and Exponential distribution.

Three methods of testing goodness of fit were used to judge the best distribution that suits the observed annual minimum data. The three methods are RMSE, Chi-Square $\chi^2$ and Visual Inspection.

Frequency analysis conducted on the three stations mentioned above revealed that, the annual minimum flow at Eddeim can be best fitted by the Two Parameters Log Normal (2LN) and Normal Distribution.

The goodness of fit indicators were persistently in favour of the selected distribution.

For Khartoum station Three Parameters Log Normal (3LN) is the best distribution to fit the data.

For Malakal station the annual minimum flow can be best fitted by the Normal and Two Parameters Log Normal (2LN) distribution.
**Recommendations for further work:**

- Further research is required to study Lowflow at seasonal rivers in Sudan because this study tackled only the Nile River. (White Nile, Blue Nile and Main Nile).

- Regional analysis should be done to determine Lowflow for sites that have no gauging station.

- More research is needed to determine characteristics of Lowflow by using more than “42” years record and different type of distribution.

- The study used the method of moments only; it will be preferable to use other methods of parameter estimation.
References


2. Bashar, K.E "Lecture Notes on statistical Hydrology"; UCWR/OIU.


13. Riggs- 1972 "Low flow investigations" - Book 4-Hydrologic Analysis and in interpretation U.S Government printing office Washington-