POTENTIAL HYDROELECTRIC ENERGY ESTIMATION FROM LIMITED DATA FOR RUN-OF-RIVER HYDROELECTRIC PLANTS

by

BRENT LANCE BURTON
B.A.Sc., The University of British Columbia, 1994

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
Department of Civil Engineering

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
June 1996
© BRENT LANCE BURTON, 1996
In presenting this thesis in partial fulfillment of the requirements for an advanced
degree at the University of British Columbia, I agree that the Library shall make it freely
available for reference and study. I further agree that permission for extensive copying of
this thesis for scholarly purposes may be granted by the head of my department or by his
or her representatives. It is understood that copying or publication of this thesis for
financial gain shall not be allowed without my written permission.

Civil Engineering

The University of British Columbia

2075 Wesbrook Place

Vancouver, B.C.

V6T 1W5

Date: June 11, 1996
Abstract

The estimation of the energy that will be produced by a proposed hydroelectric project, while obviously an essential component of any project feasibility analysis, is typically one of the more imprecise details associated with the investigation. The interpolated historical flow data used for any analysis is usually approximate, yet precise calculations of estimated energy are often conducted using this data. The objective of this thesis was to develop a computer program that directly takes into account the judgment of an experienced engineer in the estimation of potential energy output from a run-of-river hydroelectric plant when there is little or no flow data at the site in question.

The program requires the user to provide estimated curves of the annual energy, expressed as a percentage of the theoretical maximum energy, versus design flow expressed as a percentage of the mean annual flow; the user is also required to enter low, probable and high estimates of the mean annual flow. The program will create two additional estimates: a low-probable and a probable-high. The program will then look up the corresponding energy for each value of the mean annual flow from the graph provided by the user; it will also calculate and look up the corresponding standard deviation. Then, for a variety of design flows, the program will utilize a standard five by five normalized probability matrix derived from the normal distribution to calculate the mean value of the percentage of the theoretical maximum energy and the standard deviation of this value; this will be accomplished by multiplying the matrix of five energies and five standard
deviations by the probability matrix. The range of design flows and the step between design flows is chosen by the user.

The results of the program are compared with accurate results from stations with known hydrological data.
Table of Contents

Abstract ........................................................................................................................................ iii
List of Tables ................................................................................................................................ vii
List of Figures ........................................................................................................................... viii
Acknowledgements .................................................................................................................... ix
1. Introduction to Research ........................................................................................................ 1
   1.1 Reasons for Research ........................................................................................................ 1
   1.2 Objective of Thesis .......................................................................................................... 2
   1.3 Value of Classical and Proposed Method ........................................................................ 4
2. Types of Hydroelectric Plants Considered for Research ....................................................... 5
   2.1 Categories of Hydroelectric Plants .................................................................................. 5
      2.1.1 Pumped Storage Plants ............................................................................................. 5
      2.1.2 Storage Types Plants ............................................................................................... 5
      2.1.3 Run-Of-River Plants ............................................................................................... 6
   2.2 Type of Plant Considered for Research ............................................................................ 7
3. General Arrangement of a Run-Of-River Hydroelectric Plant ............................................... 9
   3.1 Main Components of Run-Of-River Hydroelectric Plant .............................................. 9
   3.2 Components of Hydroelectric Plant Considered for Research .................................. 10
4. Statistical Requirements of Hydrological Data .................................................................... 11
   4.1 Statistical Parameters of Hydrological Data ................................................................. 11
   4.2 Modelling of Hydrological Process .............................................................................. 11
      4.2.1 Stochastic versus Probabilistic Approaches ........................................................... 12
   4.3 Probability Distributions ............................................................................................... 13
   4.4 Statistical Criteria of Hydrological Data ....................................................................... 13
      4.4.1 Independence ........................................................................................................... 14
      4.4.2 Randomness ............................................................................................................ 14
      4.4.3 Homogeneity ........................................................................................................... 15
4.4.4 Stability ................................................................. 15
4.4.5 Type of Duration Series ........................................ 16

4.5 Approximation of a Continuous Time Series by a Discrete
Time Series ................................................................. 16

4.6 Observation Errors ...................................................... 17

5. Statistical Approaches to Analysis of Hydrological Data ......... 19
5.1 Single Station versus Regional Analysis ......................... 19
5.1.1 Single Station Analysis ............................................ 19
5.1.2 Regional Frequency Analysis ................................. 20
5.1.2.1 Physical and Climatic Characteristics ..................... 20
5.1.2.2 Index Flood Method ........................................... 21
5.1.2.3 Regression Method ............................................ 22
5.1.2.4 Combination of Site and Regional Estimates ............ 22

6. Summary of Classical Approach ...................................... 24

7. Discussion of Probabilistic Calculation Procedures .............. 27
7.1 Introduction ................................................................ 27
7.2 Choice of Probability Distribution ............................... 27
7.2.2 Discretization of Distribution .................................. 28
7.3 Required Data ............................................................ 30
7.3.1 Mean Annual Flows ................................................ 31
7.3.2 Graph of Annual Energy Output in Terms of Average
Flow Available .................................................................. 32
7.4 Calculation of Mean Energy and Standard Deviation ......... 33
7.4 Future Use of Baye’s Theorem to Update Probability Matrix . 34

8. Discussion of Program Output .......................................... 36

9. Comparison of Traditional and Probabilistic Method .......... 37
9.1 Comparison One .......................................................... 37
9.2 Comparison Two ......................................................... 39

10. Conclusions of Research ............................................... 41
10.1 Comparison of Classical and Probabilistic Approach ...... 41
List of Tables

Table 1: Probability Matrix ................................................................. 50
Table 2: Comparison One - Using Known Station Data to Predict
Unknown Station Data ............................................................... 50
Table 3: Comparison One - Mean Annual Flows ........................................ 51
Table 4: Comparison Two - Using Known Station Data to Predict
Unknown Station Data ............................................................... 51
Table 5: Comparison Two - Mean Annual Flows .......................................... 52
List of Figures

Figure 1: Classification of Hydrologic Process ............................................. 54
Figure 2: Discretization of Normal Distribution ............................................ 54
Figure 3: Pembeton Creek Near Pemberton
Comparison of Flow Duration Curves Obtained From Hourly, Daily and Monthly Flow Data ................................................................. 55
Figure 4: Example Graph of Energy Output in Terms of Average Flow Available ................................................................. 56
Figure 5: Graph of Energy Output in Terms of Average Flow Available for Station 08GA054 ................................................................. 57
Figure 6: Mean Energy and Standard Deviations for Station 08GA054 .......... 58
Figure 7: Comparison of Probabilistic and Explicit Approach for Estimation of Station 08GA064 ................................................................. 59
Figure 8: Energy and Standard Deviation Calculation by Probabilistic Method for Estimate of Station 8HB075 ................................................................. 60
Figure 9: Comparison of Probabilistic and Explicit Approach for Estimation of Station 8HB075 ................................................................. 61
Acknowledgment

I would like to thank Dr. S.O. (Denis) Russell for his guidance and advice throughout this research. His invaluable assistance with both the computing and research aspects of this project is appreciated.

I would also like to thank my parents and friends for their support and encouragement.
Chapter One

Introduction to Research

1.1. Reasons for Research

A feasibility analysis of a proposed hydroelectric project requires an estimate of the long term energy that will be produced. An accurate estimate normally requires complete historical flow records in order to identify long term variations. Flow records may or may not be available for a nearby location, but are rarely available at the location under consideration.

There are several approaches that may be considered when there is little or no flow data at the site in question. One method commonly used is to construct a synthetic record by interpolating between nearby hydrologic stations where flow records are available. Interpolation using a ratio of drainage areas is ordinarily required to obtain flow estimates at the proposed site, which will lead to uncertainties that are difficult to quantify. This construction of a synthetic record may also be accomplished by a computerized hydrological model such as the UBC Watershed Model; this model considers information from both historical flow data and basin characteristics.

However, the integrity of the flow data itself is often questionable; the data may be available for too short a time period to provide a reasonable estimate of long-term flow patterns, more data may be available for certain times of the year than for other times and the accuracy of the flow records themselves may be debatable.

In spite of the potential inaccuracies associated with historical flow data, an estimate of the energy produced is often calculated very precisely from the available data.
It is common to include explicit consideration of energy losses due to bends, valves and intakes as well as relative energy losses that may result from slight adjustments in penstock material. The wisdom of performing such precise calculations may be of doubt when the interpolated flow data itself may provide only a rough approximation of actual hydrological conditions.

An alternative approach is regional analysis; this requires using information from a nearby “similar” site and interpolating this information for use at the site in question. This information may be an estimate of potential energy at a nearby site or it may be flow parameters at a nearby site.

In this thesis, a regional analysis method is used to estimate the potential energy output for a run of river hydroelectric plant when there is little or no flow data at the site in question. This method requires the user to provide estimates of mean annual flows at the site without adequate data, as well as curves of energy as a percentage of theoretical maximum versus design flow as a percentage of mean annual flow. This required information is provided by intelligent interpolation of relevant flow parameters from nearby sites. The method is similar to a method developed by Russel (1982) for flood estimation; it has been adapted for use in prediction of long-term average annual energy at the site of interest.

1.2. Objective of Thesis

The objective of this thesis was to develop a computer program in Microsoft Visual Basic to simplify the task of probabilistically estimating the energy produced from a
hydroelectric project. The flow data are normally highly variable and stochastic in nature and thus are suitable for a probabilistic approach. (Louie, page 1) The user provides estimates of the low, probable and high curves of annual energy as a percentage of theoretical maximum energy versus design flow as a percentage of mean annual flow; the user also provides low, probable and high estimates of the mean annual flow. The curves of annual energy as a percentage of theoretical maximum energy are obtained from calculations performed on the familiar flow duration curve, which is a plot of flow versus fraction of time indicated flow is exceeded; the curves of annual energy are obtained by calculating the area underneath the flow duration curve for a variety of design flows. The curves are extrapolated from a site with known data to a site with unknown data. The details of the required information is detailed more fully in Chapter Seven.

The program uses a standard five by five probability matrix to perform the calculations. This matrix is used in conjunction with the given data to estimate mean energies and standard deviations: The probability of any particular combination of mean and standard deviation is obtained by multiplying the individual probabilities and normalizing to make the sum of all the probabilities add up to 1. (Russell, page 67)

There are rarely enough complete and precise historical flow records to perform a satisfactory frequency analysis; further, it is difficult to incorporate the engineer's judgment when a strict statistical procedure is followed. A “frequent situation is the inability of a practicing engineer to incorporate a good subjective understanding into a statistical process.” (Sparks, page 5) Often there is a surprising amount of additional information available to supplement the statistical analysis, and this information should be included in any exhaustive analysis.
The concept that is being applied here to estimate long-term energy from a proposed hydroelectric plant is similar to a concept used by Russell (Flood Probability Estimation, Jan., '82) to estimate peak flows. The method was adjusted to reflect that we are concerned here with long-term averages, not extreme values. An experienced engineer can use his or her judgment and incorporate all known information in determining both appropriate mean flows and graphs of energy output in terms of average flow available.

1.3 Value of Classical and Proposed Method

This probabilistic method could be used in conjunction with a more classical analysis which explicitly considers the given flow data to be an accurate representation of the true flow conditions. An experienced and wise engineer could examine the results of both analyses and make an appropriate judgment based upon consideration of all available information. The probabilistic approach, however, may prove particularly useful because it automatically provides values of the standard deviation; this can assist the engineer in conveniently determining the variability of the results.
2.1. Categories of Hydroelectric Plants

Hydroelectric plants are typically classified into three categories: run-of-river, storage and pumped storage. All three have their own advantages and disadvantages and are useful in different situations.

2.1.1. Pumped Storage Plants

A pumped storage plant uses all available flows during periods of peak load, but pumps water from the tailwater pool to the headwater pool during off-peak periods. It is common to power the pumps with energy from a run-of-river plant, where the energy would be wasted during off peak periods if it were not used to power the pumps. Pumped storage plants are economically efficient because they convert off peak power of low value to peak power of high value. Reversible pump-turbines are available which can efficiently function as either a pump or a turbine; these machines can make a pumped storage project feasible by eliminating the need for extra pumping equipment.

2.1.2. Storage Type Plants

A storage-type plant has a reservoir which allows for stabilization of energy output; some storage of water from the wet season allows energy to be produced during the dry season. Storage-type plants tend to be larger and produce more energy than either of the
other two types, although there are some exceptions. Such plants are becoming less common locally because of the potentially severe environmental implications of impounding large volumes of water.

2.1.3. Run-Of-River Plants

Run-of-river plants can use water only as it is provided by the flow of the river, although some plants do have a small storage capability, called pondage. This pondage allows for the storage of water during off-peak periods so it may be used during peak periods. Pondage can significantly increase the economic efficiency of a plant; the amount of power produced during periods of more expensive electrical rates is increased by reducing the power produced during periods of cheaper electrical rates.

Run-of-river plants are normally suitable only for streams which have a sustained flow during the dry season or where other reservoirs upstream provide the necessary storage. (Linsey, page 472) However, run-of-river plants may be suitable for streams which do not have a sustained dry season flow if the plant feeds into a grid that has storage plants; other storage plants may produce power when the run-of-river plants are not producing power. This situation is common in British Columbia, where storage plants are often available to supplement the power produced by run-of-river plants during periods of low natural river flow.

Run-of-river plants are often a suitable alternative in many locations in British Columbia, where it is common to have a sustained flow throughout the year. However, there are many locations in the world where the river runs dry during certain times of year; these locations would normally require either pumped storage plants or storage
plants, unless there is some special situation where the power from the hydroelectric plant is not needed at all times of the year.

Run-of-river plants are becoming increasingly common, both locally and globally. Since they involve no storage capacity (or very little storage capacity), there is little change in the natural flow of the stream. This is extremely important locally in British Columbia where safe conditions for salmon are a major concern; salmon are extremely susceptible to extreme variations in natural flow. The preservation of natural flow conditions is also important in general; there is increasing pressure on engineers to minimize the environmental impacts of their structures, and run-of-river plants can help meet this goal.

2.2. Type of Plant Considered for Research

This thesis considers only run-of-river plants without pondage, partially due to the ease of performing hydraulic calculations, but also because of their increasing significance. Many power producers are considering small hydroelectric plants, partially due to environmental concerns associated with larger plants; small plants are virtually all run-of-river plants. Further, storage is difficult to provide when the local terrain is steep, as is the case in many local areas where run-of-river plants are being considered. Run-of-river plants are also commonly used when there are other large storage projects that can produce power when the run-of-river plant is not producing.

Storage and pumped-storage plants introduce significant complications to the calculation process, since the power generated is no longer solely a function of natural
processes. Future research may consider the application of a similar probabilistic method to other types of hydroelectric plants, or to run-of-river plants with pondage.
Chapter Three

General Arrangement of a Run-Of-River Hydroelectric Project

3.1. Main Components of Run-Of-River Hydroelectric Plant

The main components of a run-of-river hydroelectric project include a headpond, tailpond, penstock, turbine system and electrical machines. The gross head is determined by the difference between the headwater and tailwater level; the effective head is determined by the gross head minus the sum of any friction losses in the penstock and form losses at specific locations such as bends, valves and entrances. For a run-of-river plant without pondage, the flow through the plant will depend directly on the natural flow of the river; for plants with pondage, the natural flow of the river may be supplemented by flow from storage during peak periods and the natural flow of the river may be reduced during off peak periods as flow is routed towards storage.

The energy that can be produced by the plant depends on the type, number and efficiency of the turbines; the efficiency of the turbines depends on the specific flow versus efficiency curve for the particular type of turbine. The electrical machines convert the mechanical energy of the turbines into useful electrical energy. The hydraulic engineer is not usually required to choose the electrical equipment, and this information is not pertinent to the thesis as we are concerned here with calculation of potential energy without consideration of losses.
3.2. Components of Hydroelectric Plant Considered for Research

The computer program is limited to calculation of potential energy from run-of-river plants without pondage and without consideration of losses. The judgment of the engineer is required to estimate appropriate friction and form losses that will result. Preliminary investigations have determined that it is often much more important for the engineer to choose appropriate values of flow; rough estimates of friction and form losses are often sufficient when consideration is made of the approximate nature of the original flow data. The friction losses through the penstock and the form losses at various locations throughout the conveyance system are thus not relevant for this thesis.

The headwater and tailwater levels may vary with flow; however, for these purposes the gross head is assumed constant. This assumption is made partially for reasons of simplicity in computer programming, but also because it is frequently a very reasonable assumption. The assumption of constant gross head is frequently made by practicing engineers, particularly for preliminary investigations. Future research may consider the effect of a varying tailwater level or headwater level.
Chapter Four

Statistical Requirements of Hydrological Data

4.1. Statistical Parameters of Hydrological Data

In any approach to estimating potential hydroelectric energy, the hydrological flow data is the most important consideration. There are three major statistical characteristics which can describe the flow data: central tendency, variability and skewness. The first, central tendency, is measured for our purposes by the familiar arithmetic mean average; the second, variability, is measured by the standard deviation; the final parameter, skewness, is not used for the purposes of this thesis, as a skewness of zero is assumed. The reason for this assumption of zero skewness is provided in Section 4.3.

4.2. Modelling of Hydrologic Process

"Hydrological models are mathematical formulations to simulate natural hydrologic phenomena which are considered as processes or as systems." (Chow, page 8-9) There are two main approaches to modelling: chance-independent methods or chance-dependent methods. Chance-independent methods are a deterministic process; such a process "ignores the chance of occurrence of the variables involved and the model is considered to follow a definite law of certainty but not any law of probability" (Chow, page 8-9). Chance dependent methods include stochastic and probabilistic approaches; here the chance of occurrence of the variables is considered and the model follows a law of probability.
4.2.1. Stochastic versus Probabilistic Approaches

While the terms “stochastic” and “probabilistic” are often used interchangeably, they do mean different things; probabilistic processes are time independent while stochastic processes are time dependent. Stochastic processes may be further classified into pure random processes or non-pure-random processes. In a pure-random process, the data is a random sequence, whereas in the non-pure-random process the data follows a non-random sequence. Both pure-random and non-pure-random processes can be further divided into stationary processes and non-stationary processes; the former is time-independent while the latter is time dependent. A non-stationary process occurs when either the flow characteristics of the basin are altered or the climatic processes are altered. The flow characteristics of the river would be affected, for instance, if a dam with a reservoir was placed upstream; the climatic process could be altered by a process such as global warming. A summary of the main classifications of mathematical simulations of hydrologic processes is shown in Figure 1.

Therefore, in reality, “all hydrologic processes are more or less stochastic,” (Chow, page 8-9), but an assumption of time independence can greatly simplify the analysis; a true stochastic approach can quickly become extremely complex. A stochastic model may need to consider details such as moving averages, sum of harmonics, autoregression and correlograms. For the purposes of this thesis, a probabilistic model was decided to be a reasonable representation of the true stochastic process.
4.3. Probability Distribution

The normal distribution, with its corresponding skewness of zero, is used in this computer program. The properties of the normal distribution are well known and well understood among most practicing engineers; it is symmetrical, bell-shaped and continuous. All types of averages, including the mean, mode and median are identical; the area under the curve is equal to one. Approximately sixty-eight percent of the observations are within one standard deviation of the mean; ninety-five percent are within two standard deviations of the mean and, for most practical purposes, all of the observations are within three standard deviations of the mean.

There are various standard checks for normality; the most obvious is to plot the observations and see if they exhibit the bell shape of a normal distribution, however this can be inexact if the observations are few in number. A good approach is given in Robert Hogg's standard engineering statistics book; (pages 130-133) it involves a normal probability plot where deviations from a linear tendency are evidence that the underlying data is not normal.

4.4. Statistical Criteria of Hydrological Data

While the long-term hydrological data is often automatically considered representative of the site of observations, it is prudent to ensure that the data meets all required statistical tests. These criteria are described by Sparks (pages 1-2) and elaborated on here with regard to the specific case of using hydrological data to estimate
long-term hydroelectric energy; they are: independence, randomness, homogeneity, stability and the type of data series.

4.4.1. Independence

The first consideration, independence, requires that the mean annual flows do not depend on previous events; the probability of a mean annual flow in any year is independent of the mean annual flow in any other year. This condition will usually be satisfied in the hydrological analysis, and is thus usually a safe assumption.

4.4.2. Randomness

The second criteria, randomness, requires that the sample of observations is representative of the entire population, or the actual long term flow at the site. This may be of particular concern because it is not uncommon for more data to be available at certain times of the year than at others; it is particularly common for results to be available in warm months, but unavailable in cooler months; this is often a result of ice in the river causing either inaccurate flow readings or equipment breakdowns. The actual effect of such an unrepresentative sample will vary from location to location, but the common effect in British Columbia will be to overestimate total potential hydroelectric energy. Cooler months tend to be months of low flow, so using only the available hydrological data will provide an overestimate of actual flow conditions at the site.

The randomness of the hydrological data will be jeopardized if the results are skewed. Skewed data can result from a temporary distortion of results, such as ice blockage in the river. If this or a similar condition is applicable, it is essential for an
engineer to attempt to adjust the data to provide more reasonable flow values. There are
different approaches, and the simplest is to use reasoned judgment to adjust the flow data;
it is likely, however, that the engineer may have to resort to complicated methods of data
correction, such as serial correlation analysis for persistence, moving averages for trend
and Fourier or harmonic analysis for periodicity. (Chow, page 8-18)

4.4.3. Homogeneity

Homogeneity requires that the sample data comes from the same population, or universe. This should not be a problem in the cases we are considering, as we are only concerned here with flows. “Homogeneity is of critical importance in the design of experiments and data collection programs where we can continuously confuse two phenomena. (i.e. the “hockey stick” effect of snowmelt and rainfall in peak flows in hydrology)” (Sparks, page 2)

4.4.4. Stability

The hydrological data must be stable to be directly used. For the purposes of hydroelectric potential energy estimation, the data should be both first and second order stationary; this means that both the long-term mean and standard deviation should not vary over time. This long term variation would occur if some long term effect, such as global warming, was systematically affecting long-term results; for the purposes of normal hydrological analysis, it is usual to assume that no such long-term effects occur. Future hydrological research, however, could well indicate that it is necessary to adjust historical flows to reflect variation in the long-term mean and standard deviation.
4.4.5. Type of Duration Series

Finally, the data should use a complete duration series; this consists of all of the available data on flows over some recorded time period. A partial duration series includes only information which exceeds a base value. An extreme value series includes the largest or smallest events in a series of time periods of equal length. Thus, unlike the partial duration series, the extreme value series will ignore everything but the largest event in one year, even though some of these ignored events may be larger than maximum events in other years. For our purposes, we will use a complete duration series as we are concerned with mean annual flows, and not with extreme flows.

4.5. Approximation of a Continuous Time Series by a Discrete Time Series

Hydrologic data is, of course, a continuous time series; hourly, daily, monthly and annual discharges represent a simplification into a discrete time series. A preliminary investigation was made of the comparison of hourly, daily, monthly and yearly data when constructing the graph of energy output in terms of average flow available; this graph is determined by calculating the area underneath the flow duration curve. The resulting increase in accuracy when using flows of a shortened discretization was examined; while it was evident that daily data provided significant information gains over both monthly and annual data, there was no evidence to support the use of hourly flow data as opposed to daily flow data; thus the discrete time series proved adequate for our purposes in comparison with a more continuous time series. For practical purposes, the increase in
information from hourly flow data was negligible; furthermore since accurate hourly data is frequently difficult to obtain, it was decided to limit the research to daily flow data.

Figure 3 is a graph of the comparison of flow duration curves obtained from hourly, daily and monthly flow data for Pemberton Creek near Pemberton (Station 08MG025) using three years of data; this graph is typical and shows that the differences between the hourly and daily flow curves are negligible for practical purposes, while monthly flow data probably fails to provide adequate accuracy.

However, this may not always be a reasonable assumption. In some regions, extreme fluctuations of flow throughout the day are common, necessitating a shorter period of time series discretization.

4.6. Observation Errors

The hydrological flows as published are not necessarily a true indication of actual flows at the time of measurement; instrumental and human errors are relatively common. These errors may be accidental or systematic. Accidental errors do not usually present a problem because these errors tend to be randomly distributed. Systematic errors are more of a concern because they are not random and can render the hydrological data less useful. These errors may result from consistent over or under-measurement of flow caused by human or equipment error; they may also result from errors that occur periodically under certain conditions. However, the hydrological engineer using flow data has little control over systematic errors; while it is possible for a wise engineer to identify and correct periodic systematic errors, it is not usually possible to identify
consistent systematic errors. For practical purposes, including ours, it must be assumed that no consistent systematic errors occur.
Chapter Five

Statistical Approaches to Analysis of Hydrological Data

5.1. Single Station versus Regional Analysis

While any flow can be analyzed, most studies have been concerned with peak floods. Some of these approaches are considered here; this section briefly explores the two main statistical methods of analysis of hydrological data contained in the literature: single and regional frequency analysis. A complete description of the statistical tests and methods described is contained in most standard literature on the subject; “Regional Analysis of Flood Frequencies” by Gary D Tasker is particularly good. The purpose of the inclusion of this information in this thesis is to identify a classical approach to combining information from different sites, even though most practicing engineers tend to ignore the regional database and give the single station data an inordinate weight.

5.1.1. Single Station Analysis

Single station analysis is appropriate if an adequate record of suitable length is available for the specific location of interest. As mentioned, however, this scenario is rare. As the period of record at the station becomes shorter, it becomes increasingly likely that the sample represented by the available flow data will not be representative of the actual flow values. The record is often too short to indicate the frequency and severity of severe droughts or floods; severe droughts can quickly render a hydroelectric project uneconomical.
5.1.2. Regional Frequency Analysis

Regional frequency analysis is appropriate when there is hydrological data available at a few other "similar" sites. The engineer should take care to ensure that the data reasonably represents the entire drainage basin and reflects various conditions within the basin.

There are two closely related approaches to regional analysis: index flood method and regression method. Both of these two main approaches to regional analysis involve first defining the appropriate regional boundaries and analyzing the data for acceptable station records; any stations which are obvious "outliers," will be discarded, as their inclusion could skew results. The second step involves determination of a relationship between the flow data and the physical and/or climatic characteristics of the drainage basin.

5.1.2.1. Physical and Climatic Characteristics

The characteristics considered must be independent of each other. The most important physical characteristic is usually basin area, which determines the volume of precipitation received by a watershed. Other common physical characteristics which may be considered are latitude, longitude, basin surface storage, drainage density, main channel slope and main channel length; other characteristics may be appropriate as well, as these criteria are site-specific. The climatic data is usually limited to precipitation information.
5.1.2.2. Index Flood Method

The index flood method tries to determine a dimensionless regional frequency curve and a corresponding regression equation; the equation will relate the mean annual flood flow to physical or climatic factors in the region. The combination of the curve and equation can then be used to determine the flow for varying recurrence intervals of interest. It is typical for the equation to depend solely on basin area, although other characteristics may be important at certain sites.

The application of this method requires a hydrologically homogeneous region, meaning that the flood response should be similar throughout the region. There are various homogeneity tests available in the literature; a commonly used test has been developed by W.B. Langbein. This test requires all stations to lie within the ninety-five percent control curves of a graph of recurrence interval in years versus effective length of record in years.

The requirement of hydrologic homogeneity leads into the main disadvantage of this method; there are often significant differences in flow response between different sites in the same region. However, the main advantage is the ease of calculation; it is only necessary to have flow data and, more than likely, the basin area. If there are no flow records at all, the index flood method can still be used as long as the mean annual flood is already known or can be estimated.
5.1.2.3. Regression Method

A pure regression analysis will consider the frequency distribution of one variable, like mean annual flood flow, when another variable, such as basin area, is fixed at certain levels. (Chow, 8-46) The regression method may be divided into the method of direct regression for quantiles and method of regression for distribution parameters. In the former method, a regression of flood characteristics on basin characteristics is performed; in the latter method, we perform regression of distribution parameters on basin characteristics. The regression analysis is typically performed by a computer program, which will include all independent variables which meet the F-statistic ninety-five percent confidence level; variables which do not meet this condition are removed. The F-statistic is a standard statistical test contained in most standard textbooks on the subject.

There are many regression procedures, such as development of all possible equations, forward selection, backward elimination, stagewise regression or stepwise regression; stepwise regression is the most common procedure used in practical computer programs. The regression equation will apply to all sites within the region.

5.1.2.4. Combination of Site and Regional Estimates

The best approach is typically to use engineering judgment to combine single station and regional analysis. It is common to place an excessive weight on the actual data from a single station analysis, even if this record is short, and to assign too small a weight to the regional database. An experienced engineer can determine an appropriate combination of the two results.
The two common ways to combine the information are numerically or analytically. The former approach is valid when the data set can be approximated by a normal-based distribution. It emphasizes the use of regional data when a single station record is short; short data records are extremely unreliable, as it is difficult to tell if flow variations are due to the existence of long term trends or due to chance. Numerical analysis utilizes Baye’s theorem; Baye’s theorem applies to mutually exclusive events whose union is the space where one of the events must occur.
Chapter Six

Summary of Classical Approach to Estimation of Energy

The estimation of energy at the site of a proposed hydroelectric project is simplified if there is accurate flow data available at this site in the form of a series of flows. For each record of flow, it is possible to estimate the energy that would have been available at the site; upon appropriate consideration of losses, an estimate may be made of the power that would have been produced if a hydroelectric plant had been present. If flow data is available for a long period in the form of a series of flows, a reasonably accurate estimate may be made of long-term energy potential, assuming that historical flow trends are an indication of future flow trends. This assumption will be true as long as there have not been any major changes to the flow pattern of the river, such as by man-made alteration of natural flow patterns.

However, there are rarely flow data present directly at the site of a proposed project and it is often necessary to estimate flow data at the proposed site. As mentioned, the traditional method first involves estimating the complete series of daily flows for the location under consideration and then analyzing these flows in detail; this is accomplished by interpolating the flows or by modeling the flows using a program such as the UBC Watershed Model.

Precise calculations are often performed on this relatively imprecise data, including explicit consideration of losses due to valves and bends; the procedure for the purposes of
this examination involved no consideration of losses. Consideration of losses, however, could easily be added; the energy available may be determined from as follows:

\[
\text{Energy Available in kilowatts} = 9.81 \times Q \times H \times \eta
\]

where \(Q\) is the flow in cubic metres per second, \(H\) is the difference in metres in headwater and tailwater elevation plus any head losses due to major and minor friction losses and \(\eta\) is the efficiency of the machines; this efficiency is calculated by multiplying the individual efficiencies of the turbines, generators and transformers. Alternatively, one could assume no head losses, as is the case in this thesis, and then make allowances for these after estimating the “potential” energy available.

Only the ratio of the energy for a particular design flow to the theoretical maximum energy is required for our purposes; the theoretical maximum energy is thus calculated by assuming that all the flows are used in their entirety to create power and then summing the total energy created. This theoretical maximum energy will be equivalent to the total area underneath the familiar flow duration curve, where the flow duration curve is the graph of flow versus fraction of time the indicated flow is exceeded for a particular location on the river. In theory, it is not possible to obtain more energy than the total area under the flow duration curve.

The energy for a particular design flow is determined by assuming that any flow greater than the design flow is “lost” for the purposes of energy production; the maximum energy that can be produced on any day is the energy that would be produced by the design flow. This procedure is then repeated for all known flows and the energies
for each flow are summed; this result is divided by the theoretical maximum energy, described above, to get the relevant values to be compared with the probabilistic method. Thus, the energy as a fraction of the theoretical maximum energy for any particular value of design flow is equivalent to the area underneath the flow duration curve which is less than the design flow divided by the total area underneath the flow duration curve.

As described in Section 4.5, it is usual to approximate the actual continuous record of flows by a daily discretization; the increase in information obtained from the use of hourly flows was shown to be inconsequential for our purposes. This may not be the case, however, in situations where the flow of the river tends to fluctuate highly throughout the day.

This approach was used for our purposes where adequate flow records were known to exist, so that the regional estimate may be interpolated.
Chapter Seven

Discussion of Probabilistic Calculation Procedure

7.1 Introduction

This method involves first estimating the mean annual flow at the site in question, then estimating the amount of energy available as a fraction of the theoretical maximum energy. The curves of annual energy as a percentage of theoretical maximum energy are obtained from calculations performed on the familiar flow duration curve, which is a plot of river flow versus fraction of time that the indicated flow is exceeded; the curves of annual energy are obtained by calculating the area underneath the flow duration curve for a variety of design flows. This information is then combined to ultimately create an estimate of long-term energy production. The estimates of flow parameters at the site in question are determined by intelligent interpolation of data from nearby sites; a portion of the computer program will assist the user in construction of the energy curve as a percentage of maximum energy versus design flow as a percentage of mean annual flow. The energy curve is described in Chapter 7.3.

7.2 Choice of Probability Distribution

There are several common probability distributions, and all have their advantages and disadvantages. It is important to recognize, however, that no distribution will exactly match hydrologic data; they are only mathematical approximations of real conditions. The distributions most commonly used for floods are the Log Pearson Type III, Gumbel,
lognormal and normal. The Log Pearson is commonly used by government agencies in North America and is a compromise between the Gumbel and lognormal distributions. The Gumbel distribution is widely used in Europe; the lognormal distribution is often used because it can often give reasonable results and is only a slight variation of the well recognized normal distribution.

However, the Gumbel, Log Pearson and lognormal are more suitable for extreme events, such as floods; since we are concerned only with long term average energy here, we use the normal distribution as described earlier. The normal distribution is often used because it is convenient and readily understood by practicing engineers; it is specified by two parameters, a mean and a standard deviation. Further, for this particular case the normal distribution does seem to be the most appropriate distribution to use: the mean flow during each year is an average of the 365 daily flows during that year. By the Central Limit Theorem, one would thus expect the series of mean annual flows to follow a normal distribution.

Future research may consider the effect of different distributions, although for these practical and theoretical reasons, it is assumed that the normal distribution will provide the best results.

7.2.1. Discretization of Distribution

A compound probability distribution "has been described as one where the parameters are themselves random variables, but it is more convenient to consider it as a weighted mixture of individual distributions." (Russell, page 64) In this case, we are using a compound probability distribution consisting of twenty-five unique combinations
of the standard deviation and mean of the mean energy as a percentage of the total
theoretical energy; each of the twenty-five individual numbers in the matrix corresponds
to a probability of that particular combination of the mean and standard deviation being
correct. The relevant standard deviation is the sample standard deviation, because we are
only considering a sample of possible data.

The standard deviation is calculated from the curves of energy as a fraction of
maximum energy versus design flow as a percentage of mean annual flow. For each point
on each of the three curves provided by the user, the standard deviation is calculated as
follows:

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

where \( n \) is the number of data points and \( x_i \) is appropriate value of energy as a
fraction of maximum energy. This will be the standard deviation corresponding to a
particular value of energy as a fraction of maximum energy and design flow as a fraction
of mean annual flow; this is the "regular" standard deviation.

The "regular" standard deviation considers only variation in the fraction of annual
energy as a percentage of the theoretical maximum. The "corrected" standard deviation
considers that the ratio of design flow to mean annual flow has its own source of
variation; thus the corrected standard deviation considers variation in both variables. For
example, if \( y = f(x) \), then the corrected variance will be:

\[
\sigma^2 = \sigma_e^2 + \left( \frac{\delta y}{\delta x} \cdot \sigma_x \right)^2
\]

The identical procedure is followed here where Energy As a Fraction of Maximum
Theoretical Energy = f(Design Flow As a Fraction of Mean Annual Flow).
The well known probability density function of the normal distribution is:

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

If we set the mean equal to zero and the variance equal to one and also set

\[Z = \frac{x-\mu}{\sigma},\]

then the corresponding distribution function is:

\[
f(z) = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{z^2}{2} \right)
\]

By using standard statistical tables of the values of this function, it is a simple task to discretize the distribution and calculate appropriate probabilities. The probabilities for any row or column are, respectively, 0.1255, 0.2321, 0.2848, 0.2321 and 0.1255; these correspond to areas under the normal curve between 0 and 20%, 20% and 40%, 40% and 60%, 60% and 80% and 80% and 100%.

The appropriate discretization of the normal probability curve is shown in Figure 2; the probability matrix is shown in Table 1.

7.3. Required Data

The program requires the user to provide three estimates of mean annual flows at the site in question: a low, probable and high flow corresponding to ninety percent, fifty percent and ten percent levels of exceedence. The program will produce two additional estimates which are the average of the lowest and middle value and the middle and highest value. The user is also required to provide a graph of energy output in terms of average flow available; this is interpolated from a nearby site with known data. For a variety of design flows, the program will look up the corresponding value of energy
output in terms of average flow available; it will also calculate the standard deviation of this value as described earlier. The program will then use a standard five by five probability matrix derived from the normal distribution to calculate the mean value of the percentage of maximum theoretical energy and the corresponding standard deviation.

7.3.1. Mean Annual Flows

The low, probable and high estimates of the mean annual flow are provided by the user; these are obtained by using engineering judgment to interpolate data from gauged basins. The low, probable and high mean annual flows provided by the user should correspond to, respectively, flows which have a ten, fifty and ninety percent probability of exceedence; the low, probable and high curves of energy versus flow should correspond to those expected for the corresponding three values of mean annual flow. Mattison and Russell have shown that it may be appropriate to ask the user to indicate five, fifty and ninety-five percent probability of exceedence, yet use the ten, fifty and ninety percent probabilities for actual calculation. This is due to the recurring tendency of people, including experienced engineers, to slightly underestimate the ninety percent level of exceedence and overestimate the five percent level of exceedence. However, for the purposes of this computer program, the user is explicitly asked to enter the ten, fifty and ninety percent probabilities of exceedence.

The program will interpolate to create estimates between low and probable values and probable and high values. The program will automatically calculate the standard
deviation of the energy curve from the data provided; these standard deviations will be required for the program to look up the relevant standard deviations.

7.3.2. Graph of Annual Energy Output in Terms of Average Flow Available

The high, probable and low curves for annual energy as a percentage of theoretical maximum energy versus design flow as a percentage of mean annual flow are required; an example set of curves is shown in Figure 4. The curves for annual energy as a percentage of the energy from the theoretical maximum may be determined for each basin from calculations on the flow duration curve for that basin; the annual energy curves for each basin are then intelligently combined to create a single curve representing the engineer’s estimate of conditions at the ungauged site in question. An example of this is shown in Comparison Two of Chapter Nine; involves using three months of winter data from one station and three months of autumn data from another station. The mean flow values for each station are independently adjusted to try to reflect estimated flow conditions for the entire year and two separate annual energy curves are obtained. Since these curves are approximately equidistant from the proposed site, the curves of annual energy as a percentage of mean annual flow are averaged.

By calculating the total area underneath the flow duration curve, the theoretical maximum energy can be determined. The area underneath the curve can then be calculated for a variety of design flows; when this area is divided by the theoretical maximum energy, this gives the ordinate of the graph of energy output in terms of
average flow available. When the design flows are divided by high, probable and low estimates of the mean flow, the corresponding abscissa values of the graph of energy output in terms of average flow available are provided.

A separate part of the program is provided to assist the user in creating the required graph of energy output in terms of average flow available. If the user provides historical flow information in the form of a Water Survey of Canada card file (a file with the extension *.crd) or a UBC Watershed Model file (a file with the extension *.cal), the program will strip the information to create a simple file with an INF extension. This file will contain a year, month, day and historical flow for each date of record. Using this simple historical data text file, the program will create a flow duration curve and calculate low, probable and high values of the mean annual flow corresponding to exceedence levels of ninety, fifty and ten percent. From these mean estimates and the flow duration curve, the curves of energy output in terms of average flow will be calculated.

In a typical situation, the engineer would use his or her judgment to intelligently combine the curves from two or more hydrological stations; however, investigation has indicated that the curves tend to very similar for stations within the same hydrological region and only minor adjustments, if any adjustments at all, will usually be required to obtain useful results.

7.4. Calculation of Mean Energy and Standard Deviation

As discussed, the program provides a mean energy as well as two values of the standard deviation; one standard deviation considers only variation in values of the energy as a percentage of the maximum energy, while the other also considers that the ratio of
design flow to mean annual flow has its own source of variation. Once the program has
created five estimates of the mean energy, then the total mean energy as a fraction of the
maximum energy is calculated as follows:

\[
\text{total mean energy} = \sum_{i=1}^{5} (\text{energy}_i \times \text{probability of mean}_i)
\]

where \(\text{energy}_i\) represents the five values (low, low-probable, probable, probable-
high and high) and \(\text{probability of mean}_i\) refers to the corresponding values of the
probability matrix.

Once the program has created five estimates of the mean energy, the standard
deviation of the energy as a percentage of maximum energy is calculated as follows:

\[
(totol \sigma)^2 = \sum_{i=1}^{5} (\sigma_i^2 \times \text{probability of } \sigma_i)
\]

where \(\sigma_i\) refers to the five values of standard deviation (low, low-probable,
probable, probable-high and high) and \(\text{probability of } \sigma_i\) refers to the corresponding
values of the probability matrix. The total standard deviation is calculated in this manner
for both the regular and corrected standard deviation.

7.5. Future Use of Baye's Theorem to Update Probability Matrix

Baye's theorem has been shown to be suitable for hydrological data (Russell) such as
mean annual flows. This theorem requires discretization of a continuous distribution of
mean annual flows. The current program does not update the probability matrix with
additional information, such as actual mean annual flows. The output of this program,
thus, may be affected by an inaccurate initial estimate of low, probable and high mean
annual flows. It is important that the engineer reflect his or her judgment in the creation
of an appropriate graph of energy output in terms of average flow available, as well as
with provision of accurate values of mean annual flows. Future research could explore
the effect of updating the probability matrix with additional data.
Chapter Eight

Discussion of Program Output

The mean energy as calculated from the probability matrix is provided. In addition, both the regular and the corrected standard deviations of the energy fraction are provided. The former standard deviation considers only variation in the fraction of annual energy as a percentage of the theoretical maximum. The latter considers that the ratio of design flow to mean annual flow has its own source of variation; thus the corrected standard deviation considers variation in both variables. The calculation procedure for both the mean energy and the standard deviations is discussed in Section 7.3.

At high values of annual energy as a percentage of the energy from the theoretical maximum, where the curves tend to flatten out, the two standard deviations will tend to be closer together; at low values, the two standard deviations can be different, although either standard deviation would tend to be adequate for most practical purposes. Thus the difference between the two standard deviations may be significant, but are probably negligible.

The standard deviations calculated are relevant for individual years; to obtain the standard deviation of the long term average, we must divide the standard deviations provided by the program by the square root of the relevant number of years under consideration.
Chapter Nine

Comparison of Traditional and Probabilistic Method

9.1. Comparison One

The first comparison involves Station 08GA054 on Mamquam River above Mashiter Creek and Station 08GA064 on Stawamus River Below Ray Creek; the objective was to use the Mamquam data to predict the known Stawamus River data. This comparison was used to test the effect of adjusting only the mean annual flows for the prediction; it was assumed that the graph of energy output in terms of average flow available is the same as the two stations are quite close. This would apply to a case where there was only one hydrological station in the region of the proposed hydroelectric facility; if the station and the proposed site are close enough, it may be reasonable to assume that the curves would be very similar and that it is only necessary to proportionally adjust the mean annual flows. Some records of flow were purged from each file to ensure that each file contained flow records for identical dates. The relevant information including precise locations, drainage areas and dates of record is shown in Table 2.

The first consideration was to use the program to assist in the creation of curves of annual energy as a percentage of energy from the theoretical maximum for the Manquam River data. The curves are shown in Figure 5; the low, probable and high mean annual flows were, respectively, 8.21, 20.28 and 50.08 cubic metres per second. The engineer must now use this information to intelligently adjust the mean annual flows; as mentioned,
it was decided not to adjust the data curve and therefore to assume that only the mean annual flows needed to be adjusted at the new location. Assuming that the engineer knows the drainage area of the Station 08GA064 reasonably precisely, the low, probable and high mean flow estimates may be calculated; these values may be compared with the actual mean values as shown in Table 3. In this situation, it is assumed that the engineer decides that the low and probable mean flow estimates are adequate, but that the high mean flow may be too low; his judgment and experience with the area indicates that a more reasonable 10% exceedence level may be 9 m³/s. This would be an appropriate procedure where the engineer felt that the distribution of flows at this location may be skewed, while it is not skewed at the location where the data is collected. In spite of these differences in distribution, it may be felt that the source of the data can still provide a reasonable approximation of conditions at the proposed site.

Thus the graphs of energy output in terms of average flow available are used without adjustment, while the mean flows the engineers uses are, respectively, 0.983, 2.428 and 9 m³/s. The resulting output graph of mean energy as a fraction of theoretical maximum versus design flow with the corresponding standard deviations is shown in Figure 6. It may be seen that the differences between the regular and corrected standard deviations are probably negligible for practical purposes. A comparison of the mean energy with the explicit method is shown in Figure 7. It is evident that the probabilistic and explicit approach both yield similar results, particularly at high and low design flows; there are slight differences in the medium range of design flows (2-7 m³/s), but these differences are minor. Both methods do provide a high estimate of the energy as a fraction of maximum theoretical energy for any particular design flow when compared
with the actual results. This should not be considered a constant feature of either method, as there is no reason to suggest that either method should favour over-estimation of energy as opposed to under-estimation of energy.

### 9.2. Comparison Two

This comparison was used to test the influence of extremely limited data on results; it involved using 3 months of winter data from Station 08HB011 (Tsolum River near Courtenay) and 3 months of autumn data from Station 08HB006 (Puntledge River at Courtenay) to predict the energy from a proposed hydroelectric plant at Station hb075 (Dove Creek near the mouth). The relevant station data is shown in Table 4.

In this case, the engineer decides to somewhat arbitrarily adjust the data to reflect two important details: the winter flows for Station 08HB011 will be lower than at other times of the year, resulting in low values of the mean annual flows and; the autumn flows for Station 08HB006 will be higher than at other times of the year, resulting in high values of the mean annual flows. The engineer decides, somewhat arbitrarily, to make the following adjustments to the data: for Station 08HB006, the engineer decides that the high mean flow may be accurate, but that the probable and low flow should be reduced by twenty and forty percent, respectively; for Station 08HB011 the engineer decides that the low mean flow will be accurate, but that the probable and high flows should be increased by twenty and forty percent respectively. A summary of the mean flow calculations for this comparison is shown in Table 5. It is important to stress that these calculations reflect the engineer’s judgment only; different values of the mean flows could have been justified as well.
The stations are both approximately equidistant from the proposed site, so the engineer now decides that it may be prudent to average the adjusted mean flows to obtain an estimate of the mean annual flow at the site; this results in a low, probable and high mean annual flow of, respectively, 0.5, 1.394 and 3.82 m³/s. He or she also decides to average the curves of annual energy as a percentage of mean annual flow.

The graph of results with the corresponding standard deviations is shown in Figure 8; a comparison with the explicit method may be seen in Figure 9. Figure 8 shows that the differences between the standard deviations may be more significant in this case as compared with Comparison 1; however, for practical purposes, the differences between the two standard deviations are probably negligible. Figure 9 indicates that the differences between the probabilistic and explicit approach are significant; however, there is probably little evidence to suggest one method over the other. The differences between the two methods are primarily due to the extremely poor data available. Both methods provide only a rough approximation of actual conditions, so this particular case does provide some evidence to question the wisdom of performing precise calculations on transposed data in cases of extremely limited data.
10.1. Comparison of Classical and Probabilistic Approach

The estimation of the energy produced by a hydroelectric plant is typically one of the more imprecise details associated with a feasibility analysis; despite the approximations involved in transposing known hydrological data to the site under consideration, precise calculations are often performed. Furthermore, it is difficult in a traditional analysis to incorporate the engineer's judgment in a structured fashion. This thesis proposed a probabilistic method of estimating annual energy as a percentage of the theoretical maximum energy; it requires both low, probable and high graphs of energy output in terms of average flow available, as well as estimates of the mean annual flow. The engineer may effectively incorporate his or her judgment when providing the data.

The results of various program runs have indicated that the additional effort associated with a traditional explicit analysis may not be justified; a simple procedure where an engineer approximates the graph of energy output in terms of average flow available as well as the mean flows may be just as accurate when consideration is given to the imprecise nature of the original data. Experimentation has indicated that the graphs of energy output in terms of average flow available tend to be very similar for hydrological stations in the same region. It may often be necessary for the engineer to adjust only the values of low, probable and high mean annual flows at the proposed site.
to obtain reasonable results; the curves may often remain unadjusted, or be subject to minor adjustments only.

A brief discussion of statistically sound regional analysis methods was provided to show how a precise hydrological analysis would be carried out. However, practicing engineers often feel uncomfortable relying heavily on statistical methods; their tendency is to give undue consideration to a station data at a nearby site, however short the period of record, while ignoring the large amounts of valuable information which may be contained in the regional database. It is thought that the computer program may be useful to practicing engineers because the statistics and probabilities are largely hidden from the user; the engineer is only required to enter estimates of mean annual flow as well as a graph of energy output in terms of average flow available. Most engineers will feel comfortable with adjusting a graph of high, probable and low curves.

While the results of the classical and probabilistic methods may be similar, it is suggested that an effective procedure may be to consider both. Hydroelectric plants are obviously extremely expensive projects and a careful estimation of available energy as a percentage of the maximum is important when choosing a design flow. The information produced by this program provides a convenient method to assist the engineer in deciding on an appropriate design flow. It is often useful for an engineer to have a statistical indication of the variability of the results; this variability is conveniently provided by the program. In any case, it is always prudent for an engineer to consider the results of a second procedure independent from the first; if the results are similar, it provides stronger evidence that the results are correct.
10.2. Computer Pitfalls

Various pitfalls of the usage of computer programs by engineers were discussed by Sparks. (pages 43-44) I have elaborated on several of his comments, and have included some of my own.

10.2.1. Computer Output

Computer programs are increasingly used by civil engineers to perform typical engineering tasks. They are especially valuable because they aid in “automating” the engineering process: instead of performing calculations independently for each new project, a good computer program will consider aspects that are common to many projects and allow the engineer to spend more time on matters of judgment rather than matters of calculation. The rapid increase in the use of computer programs has enabled engineering technicians to perform much of the calculation details that used to be performed by engineers.

The danger, of course, arises when the output of a computer program is automatically considered to be correct without further verification. A computer program will necessarily be based on a number of assumptions; it is imperative that the user verify that these assumptions hold for the particular case in question. While computers are useful for performing redundant calculations, they are no substitute for an engineer’s reasoned judgment.
10.2.2. Commercially Available Programs versus User Specific Programs

There are a number of commercially available computer programs available to perform a variety of engineering tasks, including potential hydroelectric energy estimation. My experience with commercial programs has indicated that, while typically the easiest to learn and use, they frequently do not meet the exact needs of a specific user or include all of the functions that a particular user would like. If sufficient time is available, it is often far better if an engineer can write a computer program tailored to his or her specific needs and/or the needs of the employer.

However, "civil engineers are not computer scientists, yet we need to see reasonably formatted output." (Sparks, page 43) It has been tempting in the past for an engineer writing a computer program to spend an inordinate amount of time both making the program excessively user-friendly and making the output professional; while the program should be reasonably easy to use and the output should be legible, the creation of professional-looking programs has traditionally been the domain of computer scientists. However, with the proliferation of object-oriented languages, it is rapidly becoming easier for those without a programming background to create professional programs. This program was written in Microsoft Visual Basic 3, an object-oriented version of the traditional BASIC language; the language proved easy to learn and enabled the author to create a reasonably professional program without spending excessive amounts of time on matters of "housekeeping." As the use of object oriented languages widens, it will likely become increasingly popular for engineers to write their own user-friendly programs to address their particular needs.
10.3. Future Research

Suggestions are made for future directions of research in the use of probabilistic approaches to estimation of long-term hydroelectric energy production; these suggestions consider hydraulics, probability distributions and the updating of the probability matrix with additional information.

10.3.1. Hydraulic Considerations

The present thesis considers only run-of-river plants; as mentioned this was partially for ease of calculation, but also because the use of such plants is rapidly increasingly both locally and globally. The program does not consider pondage; future research may consider this as well as consider storage plants and pumped storage plants. Headwater and tailwater levels may vary with flow, although it is quite common for the levels to remain reasonably constant. Future research may consider the effect of a varying headwater or tailwater level.

10.3.2. Probability Distribution

The present thesis considered only the application of the normal distribution. This distribution was chosen primarily because it is convenient to use and easily understood by practicing engineers. While preliminary indications suggested that the normal distribution would be appropriate for the purposes of this program, it may prove useful for future research to consider the effect of using different probability distributions, such as the Gumbel, Log Pearson and log-normal. However, these three distributions do tend to be
more useful for extreme events, such as floods; since we are concerned with long-term average energy here, it is theorized that the normal distribution provides the best approximation. The mean flow during each year is an average of the 365 daily flows during that year. By the Central Limit Theorem we would expect that the series of mean annual flows would follow a normal distribution.

10.3.3. Updating of Probability Matrix

This program does not update the probability matrix based on additional information entered by the user, such as actual mean annual flows. Bayesian statistics is useful for such updating; "Bayesian statistics offers a framework for combining different types of information and making the best use of what is available." (Russell, page 82) Thus the present program may prove excessively sensitive to an inaccurate initial estimate by the user; the program could be adjusted to consider the effect of additional information and use Bayesian probabilities to update the probability matrix as appropriate. At present, therefore, it is essential that an experienced engineer determine acceptable low, probable and high estimates of the mean annual flows and use the graph of energy output in terms of average flow available to reflect his or her judgment and experience with the area.
Bibliography


Probability Matrix

<table>
<thead>
<tr>
<th>Standard Deviations</th>
<th>Means</th>
<th>Marginal Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>L</td>
<td>(L+P)/2</td>
</tr>
<tr>
<td>L</td>
<td>0.0158</td>
<td>0.0291</td>
</tr>
<tr>
<td>(L + P)/2</td>
<td>0.0291</td>
<td>0.0539</td>
</tr>
<tr>
<td>P</td>
<td>0.0357</td>
<td>0.0661</td>
</tr>
<tr>
<td>(P + H)/2</td>
<td>0.0291</td>
<td>0.0539</td>
</tr>
<tr>
<td>H</td>
<td>0.0158</td>
<td>0.0291</td>
</tr>
<tr>
<td>Marginal Probability</td>
<td>0.1255</td>
<td>0.2321</td>
</tr>
</tbody>
</table>

### TABLE 1

**Comparison One - Using Known Station Data to Predict Unknown Station Data**

<table>
<thead>
<tr>
<th>Known Station Data</th>
<th>Unknown Station Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Station No.</strong></td>
<td><strong>Station Name</strong></td>
</tr>
<tr>
<td>08ga054</td>
<td>Manquam River above Mashiter Creek</td>
</tr>
<tr>
<td><strong>Longitude</strong></td>
<td><strong>Latitude</strong></td>
</tr>
<tr>
<td>123° 06' 20&quot; West</td>
<td>49° 43' 44&quot; North</td>
</tr>
<tr>
<td><strong>Drainage Area</strong></td>
<td></td>
</tr>
<tr>
<td>334 km²</td>
<td>40 km²</td>
</tr>
<tr>
<td><strong>Dates of Record</strong></td>
<td>For both stations, the records used were from 04/20/72 to 12/31/86, with the exception of 1981 data. Between this two dates, some records were purged to ensure that both sets of data contained records for the same days.</td>
</tr>
</tbody>
</table>

**TABLE 2**
### Comparison One - Mean Annual Flows

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Station 8GA064 - Mean Flows Estimated From Data</td>
<td>Station 8GA064 - Values Estimated From Column A*</td>
<td>Station 8GA064 - Actual Mean Flow Values</td>
<td>Station 8GA064 - Values Estimated By Engineer</td>
</tr>
<tr>
<td>Low Annual Mean</td>
<td>8.21</td>
<td>0.983</td>
<td>0.638</td>
<td>0.983</td>
</tr>
<tr>
<td>Probable Annual Mean</td>
<td>20.28</td>
<td>2.43</td>
<td>2.31</td>
<td>2.43</td>
</tr>
<tr>
<td>High Annual Mean</td>
<td>50.08</td>
<td>6</td>
<td>8.35</td>
<td>9</td>
</tr>
</tbody>
</table>

*estimated by multiplying by the ratio of relevant drainage areas

**TABLE 3**

### Comparison Two - Using Known Station Data to Predict Unknown Station Data

<table>
<thead>
<tr>
<th>Station Number</th>
<th>Known Station Data 1</th>
<th>Known Station Data 2</th>
<th>Unknown Station Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Station Name</td>
<td>Tsolum River near Courtenay</td>
<td>Puntledge River at Courtenay</td>
<td>Dove Creek near the mouth</td>
</tr>
<tr>
<td>Longitude</td>
<td>125° 00' 41&quot; West</td>
<td>125° 01' 57&quot; West</td>
<td>125° 05' 00&quot; West</td>
</tr>
<tr>
<td>Latitude</td>
<td>49° 42' 26&quot; North</td>
<td>49° 41' 17&quot; North</td>
<td>49° 44' 13&quot; North</td>
</tr>
<tr>
<td>Drainage Area</td>
<td>258 km²</td>
<td>583 km²</td>
<td>583 km²</td>
</tr>
</tbody>
</table>

Dates of Record: Known Station Data 1 runs from 07/01/71 to 09/30/71; Known Station Data 2 runs from 10/01/77 to 12/31/77.

**TABLE 4**
## Comparison Two - Mean Annual Flows

<table>
<thead>
<tr>
<th>Station 8HB006</th>
<th>Actual</th>
<th>Estimates of 8HB075</th>
<th>Correction Factor</th>
<th>Corrected Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>21.14</td>
<td>1.49</td>
<td>-40%</td>
<td>0.89</td>
</tr>
<tr>
<td>Probable</td>
<td>46.26</td>
<td>3.25</td>
<td>-20%</td>
<td>2.6</td>
</tr>
<tr>
<td>High</td>
<td>101.2</td>
<td>7.12</td>
<td>0</td>
<td>7.12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Station 8HB011</th>
<th>Actual</th>
<th>Estimates of 8HB075</th>
<th>Correction Factor</th>
<th>Corrected Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>0.438</td>
<td>0.06954</td>
<td>0</td>
<td>0.07</td>
</tr>
<tr>
<td>Probable</td>
<td>1.194</td>
<td>0.1897</td>
<td>+20%</td>
<td>0.23</td>
</tr>
<tr>
<td>High</td>
<td>3.255</td>
<td>0.5174</td>
<td>+40%</td>
<td>0.72</td>
</tr>
</tbody>
</table>

**TABLE 5**
FIGURES
Classification of Hydrologic Process

- **Hydrologic process**
  - Deterministic process (Chance-independent)
  - Stochastic or probabilistic process (Chance-dependent)

- **Probabilistic process** (Time-independent or sequence ignored)
  - Stochastic process (Time-dependent or sequence considered)

- Pure-random process
- Non-pure random process

- Stationary process (Time-independent distribution)
- Nonstationary process (Time-dependent distribution)

from Chow (page 8-10)

FIGURE 1

Discretization of Normal Distribution

Low | Low-Probable | Probable | Probable-High | High

from Sparks (page 40)

FIGURE 2
Pemberton Creek near Pemberton - Station 08MG025
Comparison of Flow Duration Curves Obtained From Hourly, Daily
and Monthly Flow Data

FIGURE 3
Example Graph of Energy Output in Terms of Average Flow Available

![Diagram](image)

**Figure 4**
Graph of Energy Output in Terms of Average Flow Available - Station 08GA054

FIGURE 5
Comparison of Probabilistic and Explicit Approach

Estimation of Station 8G064

FIGURE 7

Design Flow in cubic metres per second

Probabilistic

Actual

Explicit
Energy and Standard Deviation Calculated By Probabilistic Method

Estimate of Station 8HB075

![Graph showing energy and standard deviation](image)

- Mean Energy
- Corrected Standard Deviation
- Regular Standard Deviation

**Design Flow in cubic metres per second**

**FIGURE 8**
Comparison of Probabilistic and Explicit Approach

Figure 9

Probabilistic

Explicit

Actual

Design Flow in cubic metres per second

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1
20
15
10
5
0

0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0