ECONOMIC VALUE OF SEDIMENT DISCHARGE DATA

by

SVEN JACOBI and EVERETT V. RICHARDSON

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
Fort Collins, Colorado

67

ECONOMIC VALUE OF SEDIMENT DISCHARGE DATA

by
Sven Jacobi*

and

Everett V. Richardson**

HYDROLOGY PAPERS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO

April 1974 No. 67

*Post-Doctoral Research Associate, Department of Civil Engineering, Colorado State University, Fort Collins, Colorado.
**Professor of Civil Engineering, Colorado State University, Fort Collins, Colorado.

TABLE OF CONTENTS

Chapter		Page
	ACKNOWLEDGMENTS	iv
	ABSTRACT	iv
I	INTRODUCTION	1
	1.1 Background	1
	1.2 Purpose and Scope of Study	1
II	THE STATISTICAL DECISION APPROACH	3
III	WORTH OF ADDITIONAL DATA	5
	3.1 Use of all Existing Primary Data	5
	3.2 Alternatives of Project Postponement to Obtain Additional Data	5
	3.3 Use of a Regression Model	7
IV	CASE STUDY IN THE RIO GRANDE BASIN	10
	4.1 Problem Description	10
	4.2 The Goal Function	10
	4.3 Application of the Method	12
	4.3.1 Analysis of the Observed Sample	13
	4.3.2 EVEOL, Expected Bayesian Risk, and Optimum Record Length	14
	4.3.3 Use of Secondary Data and Equivalent Length of Record	18
	4.4 Results of Case Study	24
V	EFFECTS OF ECONOMIC UNCERTAINTY	25
	5.1 General Remarks	25
	5.2 The Unit Cost Density Function, f(K)	26
	5.3 The Discount Rate Density Function, f(r)	27
	5.4 Applications	28
VI	SUMMARY AND CONCLUSIONS	30
	6.1 Summary	30
	6.2 Conclusions	30
	6.3 Recommendations for Further Studies	31
	APPENDIX A. FACTS ABOUT THE COCHITI LAKE PROJECT IN NEW MEXICO	32
	A.1 Comparison of Sediment Storage Design	32
	A.2 Project Benefits	32
	APPENDIX B. THE "MINIMAL INFORMATION" DENSITY FUNCTION	34
	APPENDIX C. LIST OF SYMBOLS	35
	REFERENCES	37

ACKNOWLEDGMENTS

The material in this paper is based on the Ph.D. dissertation submitted by S. Jacobi to Colorado State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The research leading to the dissertation was financially sponsored by the Colorado State University Experiment Station. This financial support is gratefully acknowledged by both authors. Thanks are due to Dr. Carl F. Nordin, Jr. and other personnel of the U. S. Geological Survey for their valuable review and criticism of the manuscript of this report. The research leading to this paper has been conducted within the Civil Engineering Department, Engineering Research Center, Colorado State University. Appreciation is extended to the U. S. Geological Survey, Fort Collins, Colorado, and the U. S. Army Corps of Engineers, Albuquerque District, New Mexico, for their willingness and helpfullness in making the appropriate data available.

ABSTRACT

The design engineer is always faced with the problem of taking a particular action and making decisions under uncertainty. This report explains and applies a statistical decision approach which produces an expected minimum cost decision, together with a measure in monetary terms of the value of the given data sample used for design purpose. This latter concept is introduced as decrease in expected opportunity loss (EOL).

In the study it is shown how information is increased in the decision process by incorporating more data in the sample, either (1) through the use of more existing data, (2) by a postponement of the project to collect additional data, or (3) by the use of a regression model. The investigation defines the term "expected economic optimum record length" and the regression related term "equivalent length of secondary data," found in the framework of the probabilistic method.

The hydrologic parameters treated as uncertain are the mean and variance of an annual sediment load series. The investigation also deals with the question of economic uncertainty; for example, the consequences in the decision analysis of not knowing the exact value of a discount rate and/or a unit cost figure.

The theory and procedures are applied in a case study in the Rio Grande Basin, with the considered design alternative being the storage allocated for sediment deposition in a reservoir. It is shown how the use of extra data reduces the uncertainty in connection with a particular design alternative; however, the more information included, the less incremental value is gained from these additional data. The decrease in EOL using secondary data via a regression model is not as pronounced as in the case when the given data sample is augmented by means of extra primary data, due to the prediction error inherent in the regression model. This information transfer depends on the value of the cross-correlation coefficient.

In the sediment storage case an economic analysis shows that if the observed data sample is less that 12 years, further data collection is recommended. The gain in information expressed in monetary terms as a result of a more economically efficient design will in that case offset the cost of data collection plus the loss of benefits which occurs with a delay in the construction of the project. This recommendation is made assuming that a five-year sample already is available; the economic record length is dependent on the existing data at the time of decision.

The study points out that uncertainty related to economic parameters should have as equal a role within the decision process as uncertainty in the hydrologic parameters. Certain types of a beta distribution are found appropriate to relate to uncertainty in the discount rate, and a normal distribution is found to be applicable for a unit cost figure, like cost of sediment removal per ton or per acre-feet.

The decision concerning the allocation of sediment storage is made based on an economic efficiency criterion; objectives like environmental quality and social benefits are disregarded in this analysis.

CHAPTER I
INTRODUCTION

1.1 Background

It has always been the case that engineers in the design phase of a project have been forced to make decisions based on the data that are at hand or might be readily obtained. The decision-maker continually asks, "Is there enough information in the data I possess, or should I collect more data in order to reduce the uncertainty of the decision?"

Inherent in every decision process is an amount of uncertainty about the "true" state of nature (the population characteristics). If every population parameter in the decision process were known, there would be no need for decision-makers under the uncertainties, though the basic risk would remain as an input to decisions.

The value of additional data has sometimes been evaluated by examining long historic records to determine when parameter estimates "settle down." Dawdy et al. (1970) use a rather quantitative approach which involves generating synthetic traces based on the historic record, and evaluating the increase in benefit of a designed project with an increase in record length. This type of analysis assumes that the parameters used are known with certainty; i.e., the value of additional data is only known for the assumed true parameters.

Tschannerl (1971) uses the concept of opportunity loss to evaluate the worth of additional data, and in the data generation he also assumes the true values of the parameters being represented by estimates from historic records. Opportunity loss is defined as money "lost" associated with not making the best choice of action possible in light of the "true" state of nature.

Several methods are available to the engineer to make a decision under uncertainty. The most commonly used is the well known Minimax Principle. One form of this states that the investigator must choose the decision minimizing the highest cost which the state of nature can inflict. A severe objection against this method is that is does not introduce a probability for the different states of nature to occur. It is concerned solely with their consequences in relation to the various possible decisions, and takes no account of the greater or lesser likelihood (i.e. the uncertainty) of the events that might happen. Therefore, such course of action might often cause excessive conservatism in the design alternative.

The Bayesian decision approach is a method for choosing and evaluating design alternatives for a project, when the "true" state of nature or other factors are not known. The effect of uncertainties is taken into consideration through the use of probability density functions. This type of decision theory focuses on the decision to be made and not on the hydrologic parameters as an end result; from that point of view the statistical decision analysis is appealing to the design engineer. Also, such analysis makes it possible to estimate the dollar value of the uncertainties considered in the problem. The Bayesian decision is used throughout this study.

The variance of a parameter estimate or of a goal function is often used as a measure of uncertainty. There are objections to this use of variance as a measure of risk; first, it may be an oversimplification and second, it is not brought together with the economics of the project. Use of variance implies an MSE (mean square error) type of loss function, which means that such a risk evaluation is nonobjective in an economic sense.

Statistical decision theory has been developed over the past two decades to help in making decisions with uncertain information. The term Bayesian decision theory is frequently met in the literature, due to the fact that the more than two hundred years old Bayes Theorem often is introduced in such type of analysis.

Statistical decision theory primarily has been used in connection with business and industrial decisions. Very little work involving Bayesian decision theory has been done in the field of hydrology with the exception of the comprehensive study by Davis (1971). He concludes that decision theory is a rational method for making decisions necessary for the design of hydrologic projects. The decision takes into account the economics of the project, the risks involved, and the uncertainty in some of the parameters used. Decision theory is not developed thoroughly enough to be routinely used as a tool by the project engineer, and research is needed in the computational aspects of applied decision theory, for instance, by new types of case studies. The few studies published in that field, with the exception of Gates (1972) and Davis et al. (1972b), all use peak streamflow data to make decisions about flood protection levels.

1.2 Purpose and Scope of Study

This study considers river sediment load data in connection with allocation and design of storage for sediment in a reservoir. River sediment load is defined as the amount of solid inorganic material being transported by the water either in suspension or along the bed. Sediment trapping in reservoirs is one method used to cope with siltation problems. Economic losses as a result of sedimentation are considerable. As early as in the late 1940's Brown (1948) estimated for the United States that the value of damage and storage lost in reservoirs used for power, water supply, irrigation, and flood control amounted to $50 million annually. This figure is indeed comparable with losses due to annual flood damages in the States. These latter economic losses are reported by the Task Committee for Preparation of Manual on Sedimentation (1969, p. 193) to be approximately $200 million per year. Furthermore, it should be mentioned that suspended load is an important factor to take into consideration during the design of nearly all types of hydraulic structures in natural waterways.

Sediment load data is of vital importance in order to make judgement about future sedimentation conditions; but extensive collection of sediment load data is a costly operation. This naturally forces the engineer to face the problem of answering questions like: Is it at all worthwhile to continue sampling? What is the value in monetary terms of a given sample? What is the economical optimum record length when sediment data are used to design a specific project? Questions like these are treated in this investigation.

Bayesian decision analysis is a way to evaluate the economic gain obtained by extending the data record either by postponing a proposed project and collecting additional data or by the use of generated data through a regression model. In this study a primary data set is defined as the data series which is used directly for design purposes and whose population characteristics are treated as uncertain. A secondary data set is related in some manner to the primary data, and therefore, it can be used to extend the primary data set. The logarithms of annual river flows and annual sediment loads can be assumed linearly related, and this feature is used to find the worth of augmenting a primary set of data by a secondary set, taking into account the uncertainty embedded in the regression model. This type of analysis actually makes it possible to redefine the classical concept of the equivalent length of a set of secondary data, as being the length of a set of primary data which provides us with the same economic gain (increased information content) as the use of the secondary data set does. As a sidelight a comparison of such type of equivalent length with the concept introduced by Roche (1963) is done, because the equivalent length is affected by the degree of cross-correlation.

Tschannerl (1970) studied the problem of using secondary data (tree rings and mud varves) to extend a primary set (streamflow data). Different approaches are used in his study in order to find an optimal estimate of a particular streamflow parameter, for instance, minimization of expected opportunity loss and method of least squares. Comparison is made utilizing the historical record alone and the historical record plus data obtained through the application of a regression model. An optimal estimate of the parameter might thereby be found for use in a generating model for streamflows. Again it should be noted that the data simulation which is an important basis in Tschannerl's investigations is carried out with a particular prescribed value for the population parameters.

In Chapter II, Bayesian decision theory is outlined in the context of sediment load data series. Chapter III discusses the value of extending a given data base with additional data. Information is thereby gained concerning the unknown parameters of the time series - but for a certain price. Concepts like the expected value of the expected opportunity loss, called EVEOL, and the economic optimal record length are explained and investigated in Chapter III. Furthermore, the regression model in connection with the statistical decision theory is taken up for consideration. Chapter IV treats in detail the goal function, presents an extensive application of the theory, and discusses the different results. The U. S. Army Corps of Engineers Cochiti Lake project on the Rio Grande in New Mexico is used as a case study to test the feasibility of using the methods developed.

Besides uncertainty in the hydrologic parameters (the mean and variance of the sediment load series), uncertainties in the decision process might also arise because of the estimation of the cost figures used in the goal function. Hydrologic decision problems up to now have treated economic factors as being fixed or at least "stepwise" fixed in a sensitivity type of analysis. In Chapter V a procedure is outlined for treating those factors - unit cost factor and discount rate - as variables, and for adopting "subjective" uncertainty distributions for them. The procedure is applied to enable us to find the consequences in the decision process of not knowing the exact future value of money (rate of inflation, technological advances, etc.), taking into account the unpredictable fluctuations of the future interest and discount rates.

In Chapter VI, conclusions from the worth of data study are drawn, and a summary of the investigation is presented.

The U. S. Army Corps of Engineers design of the sediment storage part of Cochiti Lake on the Rio Grande in Appendix A is taken up for comparison with the design obtained in the present study. However, it should be kept in mind that decision making in water resources planning is often a more complex process than this investigation might indicate where only part of an economic efficiency criterion is applied. According to the Water Resources Council (1973) the following main objectives should be considered by federal, state, and local governments for planning the use of the nations water resources and related land:

1) Economic efficiency,

2) Quality of environment, and

3) Social benefits.

In a realistic decision process with regard to allocation of reservoir storage for different purposes, all three objectives have to be considered. In the presented study only a portion of the first objective about an economically efficient project has been used as the determining factor in the different decisions reported in the following. Omitted is, for example, the economic consequences of taking the flood control purpose of the reservoir into consideration.

CHAPTER II
THE STATISTICAL DECISION APPROACH

The statistical - or Bayesian - decision approach is simply a procedure for applying logical thinking and cannot be called a strict method. Howard (1966) did formalize the thought processes to reach a decision. The following decision analysis procedure, outlined by Davis (1971) is in idea and principle Howard's but to a certain extent changed to fit the case study problem treated in this paper. An explanation of the different concepts introduced in the following outline is given in the further text.

A. Define the decision to be made and identify the alternatives.

B. Form the goal function

   1. Select the variables which describe the "state of nature" (arguments in goal function)
   2. Establish possible time preference (discount or interest rate)

C. Derive stochastic properties of state variables (probability density functions)

D. Select best alternative

   1. Calculate the expected value of the goal function for each alternative
   2. Choose alternative which minimizes the expected value of the goal function (Bayesian Risk)

E. Evaluate uncertainties and find the worth of additional data

   1. Determine expected opportunity loss, EOL (due to uncertainty)
   2. Determine the reduction in EOL either by
      (a) including more existing primary data
      (b) collecting more data, or
      (c) transmitting information from a secondary set of data by a regression model.

   EOL results with additional data included in the sample is found by going back and performing the calculations outlined in step C through E.1. The reduction in EOL is the difference between the EOL values obtained before and after the addition of extra data.

Increased information concerning the unknown population parameters is measured as this reduction in expected opportunity loss. The net worth of additional information is defined in monetary terms as the positive decrease in EOL minus the cost of obtaining these extra data. Raiffa and Schlaifer (1961) treat some theoretical aspects of additional information in a comprehensive way. The following presents an explanation of the above outline applied to river sediment load data.

According to Nordin and Sabol (1973) a two-parameter lognormal distribution is very often adaptable to describe the variable of annual sediment loads. If the random variables are the natural logarithms, log to base e, of the annual sediment data, a Gaussian distribution can be employed on the transformed data. This feature will be used throughout this study. The true population parameters, mean μ_t and variance σ_t^2, of the normal distribution of logarithms constitute the state parameters, which describe the "state of nature" and are not known with certainty. The assumed population parameters, μ_a and σ_a^2, can be considered to be a set of random variables having a joint probability distribution. To simplify the use of symbols in the following text, μ_a and σ_a^2 will be designated from now on as μ and σ^2, by understanding that they are not the population constants but the random variables of assumed parameters. The alternatives considered in the decision process are the possible design sizes of the sediment storage part of a, reservoir, large enough to store the deposited material accumulated over a prescribed lifetime of the project.

Raiffa and Schlaifer (1961, p. 300) and Benjamin and Cornell (1970, p. 628) derive the joint distribution of the random variables μ and σ^2. It is shown that the distribution, given a set of sample statistics, takes the form of a so-called normal chi-square density function,

$$f(\mu,\sigma^2|n,\bar{x},s^2) =$$

$$\sqrt{\frac{n}{2\pi\sigma^2}} \cdot e^{-\frac{n(\mu-\bar{x})^2}{2\sigma^2}} \cdot \frac{(\frac{n}{2})^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \cdot \frac{1}{\sigma^2} \cdot \left(\frac{s^2}{\sigma^2}\right)^{\frac{n-1}{2}} \cdot e^{-\frac{ns^2}{2\sigma^2}}. \tag{2.1}$$

Equation 2.1 is derived assuming that the underlying series is independent, in this case the annual sediment loads. The $f(\mu,\sigma^2|n,\bar{x},s^2)$ distribution is in principle an aposteriori distribution, because a sample (length n) is given and incorporated in the functional form appearing as the sample mean $\bar{x}$ and sample vairance s^2. The aposteriori distribution is found as a product of a sample likelihood function and an apriori distribution of μ and σ^2. The source of data for this conjugate prior is the longest available data sample. The posterior has the same form as the prior, because the normal chi-square distribution belongs to a group called natural conjugate density functions. This characteristic makes significant computational savings, mainly because the functional form of the prior and posterior is preserved and the parameters are related to simple statistics of a sample. All that must be done to reach an aposteriori distribution from an apriori is to revise the sample statistics $\bar{x}$ and s^2 in Eq. 2.1. For further explanation of such conjugate relationship, consult Benjamin and Cornell (1970, pp. 625-631). This is an important feature in the present study, because the results of the following calculations (Bayesian Risk, EOL, etc.) are functions of the particular sample of data used in the analysis. Furthermore, this dependency on the given sample allows us to find the relationship between Bayesian Risk, EOL, etc., and the sample length, as will be shown in the next chapter.

For Convenience, the statistics $\bar{x}$ and s^2 are omitted in the notation - when not subject to misunderstanding - and the joint probability density function is simply called $f(\mu,\sigma^2)$.

The goal function, $G(Q_s^{alt}|\mu,\sigma^2)$, is in this study a so-called penalty function. It indicates the excess cost that has to be paid because of either a realized overdesign or underdesign of the sediment storage part of a reservoir. The explicit functional form is defined in Section 4.2. Taking the expectation of this goal function with respect to the probability distribution $f(\mu,\sigma^2)$ yields the Bayesian Risk. The decision is made by choosing the alternative Q_s^* that minimizes the Bayesian Risk:

$$R(Q_s^*) = \underset{Q_s^{alt}}{\text{Min.}} \iint G(Q_s^{alt}|\mu,\sigma^2) \cdot f(\mu,\sigma^2) \cdot d\mu \cdot d\sigma^2 \tag{2.2}$$

Q_s^* is often called the Bayesian solution.

The concept of opportunity loss is introduced to represent a measure of the value of perfect information on the population parameters. If the true values (μ_t,σ_t^2) of the state variables were known, this information would yield Q_s^t, the alternative that gives the minimum variable cost:

$$G(Q_s^t|\mu_t,\sigma_t^2) = \underset{Q_s^{alt}}{\text{Min.}} \left[G(Q_s^{alt}|\mu_t,\sigma_t^2)\right] \tag{2.3}$$

Having used Q_s^* instead of Q_s^t, an opportunity loss has resulted because the economic optimum design alternative for the Bayesian Risk will differ from the optimal alternative if the parameters were known with certainty.

The suffered opportunity loss (OL) is

$$OL(OL(Q_s^*|\mu_t,\sigma_t^2) = G(Q_s^*|\mu_t,\sigma_t^2) - G(Q_s^t|\mu_t,\sigma_t^2), \tag{2.4}$$

which represents the extra costs which have to be paid because our decision was made on the basis of imperfect knowledge about the state variables. Obviously the "true" values of the state variables are never known, but the probability density function $f(\mu,\sigma^2)$ makes it possible to calculate an expected opportunity loss (EOL):

$$EOL = \iint \left\{G(Q_s^*|\mu,\sigma^2) - G(Q_s'|\mu,\sigma^2)\right\} \cdot f(\mu,\sigma^2) \cdot d\mu \cdot d\sigma^2, \tag{2.5}$$

where Q_s' indicates the design alternative, which minimizes the goal function for each particular set of the parameters μ and σ^2, as they show up in the integration. The integration is over all "possible" values of the state variables. The EOL represents the expected value in monetary terms which we are willing to pay to obtain perfect information, and may be used to judge the effect of uncertainty about μ_t and σ_t^2 with respect to a specific project. It should be kept in mind that the Bayesian Risk and EOL values are functions of a set of sample statistics as they appear in the density function $f(\mu,\sigma^2)$.

In the outline of the decision analysis procedure, step A through step E point 1 have been covered. Explanation concerning the last and very important part of the analysis, which treats the worth of additional data is given in the next chapter.

WORTH OF ADDITIONAL DATA

The current chapter with Sections 3.1 through 3.3 corresponds to the outline of the decision analysis step E point 2(a), (b), and (c), respectively. Different types of additional data might be used to gain inforamtion about the state variables. The decrease in EOL is a measure of the reduction in uncertainty. The most common ways to get more information are described in the following sections.

3.1 Use of all Existing Primary Data

It is not surprising that the design engineer should use the longest available and reliable record of the primary data; after all, that is his only real guide to judge the "true" state of nature. The present section is included merely to support that fact. In this case the primary data consist of a sample of annual sediment loads. Many investigations have found quantitatively how the marginal worth of data decreases with the increasing number of data points in the sample, see as examples Dawdy et al. (1970), Davis (1971), and Tschannerl (1970). Their results for streamflow data show that this decrease is quite rapid as the sample size approaches 40 to 50 years, after which the EOL is rather insensitive to the incorporation of additional data. Herfindahl (1969) makes a comparison with other areas and states that the added information is similar to the economy-related concept "law of diminishing returns," such that the more information one includes, the less incremental value is gained from this additional information.

Because sediment load data rarely exceed 25 to 30 years, and most often is of a length as short as 5 to 10 years, it is of interest to establish the marginal worth of that type of data as a function of the number of historic observations. In that way it is possible to evaluate the importance of using all the available data and not just parts of the observed record.

3.2 Alternatives of Project Postponement to Obtain Additional Data

The most common practice among design engineers has been to use available data for design purpose. However, they have often failed to take into consideration the possibility that further data collection might result in an economically more efficient design. It is always desirable to improve the decision-making process, and one way to achieve that goal is to obtain the financially optimal amount of data on which to base decisions.

In the paper by Moss (1972), the concepts of expected optimum record length are discussed. Contained therein is a graph which shows the general trend of the marginal worth of additional data that are collected as a function of record length, and a curve which defines the total marginal cost of obtaining those data versus record length. The intersection of these two curves defines the optimal record length.

Figure 3.1 shows that data should be collected as long as their marginal worth exceeds its expected marginal cost. These expenditures must include the costs of operating the data-collection facilities and the cost of delaying the design which shows up as benefits foregone.

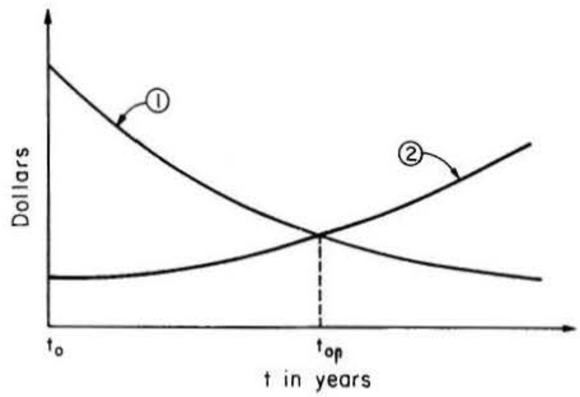

Fig. 3.1 Marginal worth and cost versus record length: (1) marginal worth of additional data; and (2) marginal cost of collecting data and of benefits foregone. t_o = present time, and t_{op} = optimal length of additional data collection.

Needless to say, no one can predict the actual sediment load for future years. Davis (1971) was the first who took up the problem in the hydrological field of including future data with the underlying population characteristics not known. He introduces the concepts of expected expected opportunity loss (EEOL) as an average of the EOL's taken over all possible values for the next datum. However, in the present paper this concept will be referred to as the expected value of the expected opportunity loss, EVEOL. The classical Bayes' Theorem enables him to get a revised (aposteriori) distribution of the state variables given a new data point, Q:

$$f(\mu,\sigma^2|Q) = \frac{f(\mu,\sigma^2) \cdot N(Q|\mu,\sigma^2)}{\iint f(\mu,\sigma^2) \cdot N(Q|\mu,\sigma^2) \cdot d\mu d\sigma^2}, \quad (3.1)$$

where $f(\mu,\sigma^2)$ is a normal-gamma distribution, which depends on a set of sample statistics $\bar{x}$ and s^2. $N(Q|\mu,\sigma^2)$ is the likelihood of the data point Q, given a set of values of μ and σ^2, which in his case were assumed as coming from a normal distribution after a logarithmic transformation of the data is performed. The predictive distribution of Q can now be found as

$$g(Q) = \iint N(Q|\mu,\sigma^2) \cdot f(\mu,\sigma^2) \cdot d\mu d\sigma^2. \quad (3.2)$$

It is seen that this expression is identical to the denominator in Bayes' Theorem, Eq. 3.1. By means of the predictive distribution Davis' expected value of the expected opportunity loss may be calculated,

$$\text{EVEOL}(1) = \int_{Q_1} \text{EOL}(Q_1) \cdot g(Q_1) \cdot dQ_1, \quad (3.3)$$

where the subscript of Q indicates the subsequent period of the additional observation being considered (in this case the first period). It is seen that for all possible values of the next period, the EOL has been weighted by the probability of obtaining that particular figure.

This is a statistically elegant method. An objection to Davis' presentation is that he talks about the integration even though the functional form of EOL is not defined. Furthermore, the method often requires very complex mathematics - or extensive computer use for numerical integrations - in order to achieve an answer. However, in Davis' study the calculations to find g(Q) were simplified considerably because the distributions involved belong to a certain type called natural conjugate distributions. As explained earlier in Chapter II, this characteristic implies that the distribution has the property that parameters have apriori and aposteriori distributions belonging to the same family. Using this particular distribution feature Bayes' Theorem yields directly an explicit expression for g(Q):

$$g(Q_1) = \left[\frac{n}{\pi(n+1)}\right]^{\frac{1}{2}} \cdot \frac{\left[n \cdot s^2\right]^{\frac{n-1}{2}}}{\left[(n+1)\cdot s_r^2\right]^{n/2}} \cdot \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}, \qquad (3.4)$$

where n is the number of original data points, s^2 is the sample variance, and s_r^2 is the revised sample variance with the "new data" included. With the conjugate relationship and considering only one additional year of data, as Davis did, the procedure seems usable with a reasonable consumption of computer time.

The procedure introduced by Davis can now be extended in order to evaluate the expected worth of more than one year of additional data. By including two new observations, the expected EOL might be calculated as

$$EVEOL(2) = \int_{Q_1}\int_{Q_2} EOL\ (Q_1,Q_2)\cdot g(Q_1)\cdot g(Q_2)\cdot dQ_1\cdot dQ_2, \qquad (3.5)$$

in case the observations from year to year are assumed independent of each other. Using the same scheme, N years of future observations will have the following expected EOL associated with them:

$$EVEOL(N) = \int_{Q_1}\int_{Q_2}\cdots\int_{Q_N} EOL(Q_1,Q_2,\cdots,Q_N)\cdot g(Q_1) \cdot g(Q_2)\cdots g(Q_N)\cdot dQ_1\cdot dQ_2\cdots dQ_N, \qquad (3.6)$$

which theoretically is a satisfactory expression. However, as the formula demonstrates, multiple, N, integrations have to be carried out over an expected EOL value. If it is assumed in the numerical integration that it takes m calculations of EOL to get one year's EVEOL, actually m^N different EOL computations must be made in order to find EVEOL(N). Even with a very powerful numerical integration technique it is virtually impossible to find EVEOL(N) by strictly employing Eq. 3.6, because of the exponential increase of EOL computations with respect to the number of future years. The investigator is therefore forced to approach the problem by means of other methods.

Moss and Dawdy (1973) recommended in their paper a combination of a data simulation method and the statistical decision approach. By combining the two methodologies a technique has been achieved which eliminates some of the shortcomings because none of them are common to both methods. The combination approach can be employed in many different ways depending on the investigator's particular case. The Monte Carlo approach has the serious restriction that the hydrologic parameters must be known or assumed prior to the analysis. This deficiency can be avoided as will be explained in the following.

For a particular set of assigned values of the population parameters μ and σ^2 it is possible to generate synthetic annual sediment loads S_i using the model recommended by Matalas (1967):

$$S_{i+1} = \mu + \rho(1)\cdot\left[S_i - \mu\right] + \sqrt{1-\rho(1)^2}\cdot\sigma\cdot\varepsilon_{i+1}, \qquad (3.7)$$

which is the general equation for simulation of data which are normally distributed and possess autocorrelative dependency represented by the lag-one serial correlation coefficient $\rho(1)$. S_i are annual sediment loads in logarithmic form. This model is valid under the condition of stationarity, i.e., the distribution of S_i is identical to S_{i+k} for all integer values of k. ε_{i+1} is a random normal component with zero mean and unit variance, and independent of S_i. According to Nordin and Sabol (1973) annual sediment load series generally possess a very small degree of autocorrelation (an assumption used in the derivation of Eq. 2.1). Nevertheless, this small dependency is taken into consideration here in order to make the generated data sample as realistic as possible.

Through the application of Eq. 3.7 a sequence of n_2 future events can be generated. This sample of data pooled with the observed sample of length n_1 yields a new sample mean and variance, which are used to develop a revised distribution of μ and σ^2 according to the comments related to Eq. 2.1. Bayesian Risk and EOL calculations can now be carried out in the regular manner as described in Chapter II. What has been obtained is a value of EOL given the synthetic record, or $(EOL_{n_1+n_2}|\mu,\sigma^2)$. By repeating the above procedure a sufficient number of times for many different values of μ and σ^2 (covering the "possible range" of these parameters in accordance with the apriori distribution) a set of EOL's can be defined. The average over this set is an estimate of the expected value of the expected opportunity loss (EVEOL). The average value is found as a weighted average of the EOL's using the original distribution of μ and σ^2, $f(\mu,\sigma^2)$ derived from the observed sample, to weight the different synthesized values. It is seen that this method is a blend of a Monte Carlo simulation technique and an expected-value criterion. That means,

$$EVEOL_{n_1+n_2} = \sum_{i=1}^{n_\mu} \sum_{j=1}^{n_{\sigma^2}} (EOL_{n_1+n_2}|\mu_i,\sigma_j^2) \cdot f(\mu_i,\sigma_j^2) \cdot \Delta_i \cdot \Delta_j, \quad (3.8)$$

where n_μ and n_{σ^2} are the number of intervals into which the ranges of the mean and variance respectively are divided, and Δ_i and Δ_j are respectively the sizes of the i-th interval of the mean and the j-th interval of the variance. Following the same scheme expected Bayesian Risk (EBR) can be defined as

$$EBR_{n_1+n_2} = \sum_{i=1}^{n_\mu} \sum_{j=1}^{n_{\sigma^2}} (EBR_{n_1+n_2}|\mu_i,\sigma_j^2) \cdot f(\mu_i,\sigma_j^2) \cdot \Delta_i \cdot \Delta_j. \quad (3.9)$$

The above analysis can now be repeated for each consecutive year of data or groups of data added beyond the available observed sediment load series (for example $n_2=1$, $n_2=5$, $n_2=10$, $n_2=25$, etc.). The worth-of-data curve is hereby defined, and the optimum record length can be found if the cost of collecting the data and the cost of project postponement can be specified, as will be shown in Section 4.3.2.

The merit of the described combination approach is that it keeps the mathematical considerations and computational efforts down to a reasonable level, and at the same time takes advantage of the statistical validity of the Bayesian method.

3.3 Use of a Regression Model

A third way to gain information about the distribution of the annual sediment load is to augment this primary data with a set of secondary data, which are related to each other. In hydrology one of the most commonly used relationship between two types of data is a regression model. In this section linear regression for a short and a long sequence of hydrologic events is used to increase the information of the short sequence.

Some users of regression models directly extend the primary set of data without taking into account the uncertainty inherent in the prediction given by the regression equation. Gray and Davis (1972) outline a procedure in the framework of Bayesian decision analysis which copes with the problem of introducing regression error. Given a secondary data point x_i they obtain the regression estimate of y:

$$\hat{y} = b_o + b_1x_1 + b_2x_2 \cdots + b_nx_n. \quad (3.10)$$

From the classical regression analysis it is known that the difference between the actual value of y and the predicted value $\hat{y}$ follows a Student-t distribution, for the statistic written in the following form:

$$\frac{y - \hat{y}}{s\sqrt{1+x_o' \bar{c} \bar{x}_o}}$$

In this expression s is the square root of the residual variance, $\bar{x}_o$ is a vector of secondary observations x_i, and $\bar{c}$ is a constant matrix whose elements are functions of the data used to establish the regression equation. The above expression is true under the assumption that the variable $(y-\hat{y})$ is distributed normally with mean zero and estimated variance s^2. By means of this distribution Gray and Davis (1972) form a posterior distribution of the state parameters θ (μ and σ^2 in the present study), considering all possible values of y in proportion to their likelihood:

$$F(\theta|\bar{x}_o) = \frac{1}{k} \int F(\theta) \cdot f(y|\theta) \cdot t(y|\bar{x}_o) \cdot dy, \quad (3.11)$$

where k is a standardization constant. There are objections to this procedure: (a) no new information is added to the primary set of data, only uncertainty stemming from the regression model is considered, (b) the numerical computations involved are very extensive, even with the use of only one year of secondary data, and (c) the proper use of the t variable is for one prediction of y. For several predictions, which is the case in the integration over y, the regression coefficients and s should be reestimated each time a prediction on y is made (see Mood, 1950, p. 299).

In the present study the ideas introduced by Matalas and Jacobs (1964) are applied. Using their results makes it possible to add information to the uncertainty distribution of the state parameters in question, and at the same time take the regression error into account.

It is recalled from Chapter II that in order to make a decision, Bayesian Risk calculations must be made. In those computations an important feature is the probability distribution of the state parameters given a set of sample statistics:

$$f(\mu,\sigma^2|n_1, \bar{S}_1, s^2_{S_1}).$$

After the augmentation of the primary set of data, S_i, the distribution can be revised by the use of estimates of the mean and variance for the lengthened series.

As pointed out by Nordin and Sabol (1973), the logarithm of annual sediment load in a river is linearly related to the logarithm of annual water discharge. Furthermore, they show how these two transformed series both follow a normal distribution. Often a short series of sediment load data and a longer series of water discharge data are available with n_1 concurrent years:

Logarithms of annual sediment load:

$$S_1, S_2, \cdots, S_{n_1}$$

Logarithms of annual water discharge:

$$W_1, W_2, \cdots, W_{n_1}, W_{n_1+1}, \cdots W_{n_1+n_2}.$$

It should be noted that the emphasis in the following is only on the logarithmic transformed series.

The correlative interdependency between the two series is expressed in a linear regression equation:

$$E[S_i | W_i] = \alpha + \beta \cdot W_i, \qquad (3.12)$$

where α and β are the population values of the intercept and the slope of the regression line, $E[\cdot]$ denotes the expected value of S_i for a particular value of W_i.

S_i can be considered normally distributed around the regression line with constant variance independent of W_i; the regression system of S_i on W_i is thus homoscedastic. As reported by Matalas and Jacobs (1964) this variance around the regression line is equal to $(1-\rho^2) \cdot \sigma_S^2$, where ρ denotes the true correlation coefficient, and σ_S^2 is the population variance of the sediment load series. Considering this term for the residual variance, it is possible to express S_i as:

$$S_i = \alpha + \beta \cdot W_i + \sqrt{1-\rho^2} \cdot \sigma_S \cdot \varepsilon_i, \qquad (3.13)$$

where ε_i is a random normal variable with zero mean and unit variance, and the term $\sqrt{1-\rho^2} \cdot \sigma_S \cdot \varepsilon_i$ represents the prediction error, or noise as it is often called in the literature.

In the following text subscripts "1" and "2" indicate that an estimate is based on the sample period n_1 and n_2, respectively; subscript "1+2" is used when estimates are based on the total period (n_1+n_2).

Using the method of least-squares on the n_1 concurrent data, estimates of the regression coefficients and the correlation coefficient can be found as,

$$\hat{\beta} = \frac{\sum_1^{n_1} (S_i-\overline{S}_1) \cdot (W_i-\overline{W}_1)}{\sum_1^{n_1} (W_i-\overline{W}_1)^2}, \qquad (3.14)$$

$$\hat{\alpha} = \overline{S}_1 - \hat{\beta} \cdot \overline{W}_1, \qquad (3.15)$$

and

$$\hat{\rho} = \frac{\sum_1^{n_1} (S_i-\overline{S}_1) \cdot (W_i-\overline{W}_1)}{\left[\sum_1^{n_1} (S_i-\overline{S}_1)^2 \cdot \sum_1^{n_1} (W_i-\overline{W}_1)^2\right]^{1/2}}. \qquad (3.16)$$

If the population parameters are replaced by their sample estimates, Eq. 3.13 may be expressed in the following way:

$$S_i = \hat{\alpha} + \hat{\beta} \cdot W_i + C \cdot \sqrt{1-\hat{\rho}^2} \cdot s_{S_1} \cdot \varepsilon_i, \qquad (3.17)$$

where C is a constant which is defined below.

For a particular value of the secondary data W_i and the use of a normal random number generator, Eq. 3.17 yields an estimate of S_i outside the concurrent period. These estimates are now pooled with the observed data to form the augmented series,

$$S_1, S_2, \cdots, S_{n_1}, S_{n_1+1}, \cdots, S_{n_1+n_2}.$$

For this lengthened sequence of the sediment load, the estimate for the mean can be calculated as

$$\overline{S}_{1+2} = \frac{n_1 \cdot \overline{S}_1 + n_2 \cdot \overline{S}_2}{n_1 + n_2}$$

$$= \frac{n_1 \overline{S}_1}{n_1+n_2} + \frac{n_2}{n_1+n_2} \cdot (\hat{\alpha}+\hat{\beta}\cdot\overline{W}_2+C\cdot\sqrt{1-\hat{\rho}^2} \cdot s_{S_1} \cdot \overline{\varepsilon}_2). \qquad (3.18)$$

Substituting $\hat{\alpha} = \overline{S}_1 - \hat{\beta} \cdot \overline{W}_1$ yields

$$\overline{S}_{1+2} = \overline{S}_1 + \frac{n_2}{n_1+n_2} \cdot \hat{\beta} \cdot (\overline{W}_2-\overline{W}_1) + \frac{n_2}{n_1+n_2} \cdot C \cdot \sqrt{1-\hat{\rho}^2} \cdot s_{S_1} \cdot \overline{\varepsilon}_2, \qquad (3.19)$$

where $\overline{\varepsilon}_2 = \frac{1}{n_2} \sum_1^{n_2} \varepsilon_i$. Because $E(\overline{\varepsilon}_2)$ asymptotically will approach zero, Eq. 3.19 in the limit (for $n_2 \to \infty$) simplifies to

$$\overline{S}_{1+2} = \overline{S}_1 + \frac{n_2}{n_1+n_2} \cdot \hat{\beta} \cdot (\overline{W}_2-\overline{W}_1). \qquad (3.20)$$

Matalas and Jacobs (1964) have derived the following approximate estimate for the variance of the lengthened series:

$$s_{S_{1+2}}^2 = \frac{1}{n_1+n_2-1} \cdot \left[(n_1-1) \cdot s_{S_1}^2 + (n_2-1) \cdot \hat{\beta}^2 \cdot s_{W_2}^2 + \frac{n_1 \cdot n_2}{n_1+n_2} \cdot \hat{\beta}^2 \cdot (\overline{W}_2-\overline{W}_1)^2 + (n_2-1) \cdot C^2 \cdot (1-\hat{\rho}^2) \cdot s_{S_1}^2\right]. \qquad (3.21)$$

They found C^2 to be equal to $\frac{n_2(n_1-4)(n_1-1)}{(n_2-1)(n_1-3)(n_1-2)}$, which implies that $s_{S_{1+2}}^2$ in the limit $(n_2 \to \infty)$ will be an unbiased estimator of σ_S^2.

The addition of regression noise is reflected by the last term in brackets in Eq. 3.21. Nonstationarity in the secondary series also has a strong effect as it shows up in the term $(\overline{W}_2-\overline{W}_1)$. More about the interplay between these different effects will be explained in Section 4.3.3.

The new sample mean and variance for the augmented sediment load series enable us to get a revised normal-chi-square distribution for the state parameters,

$$f(\mu,\sigma^2|n_1+n_2,\ \overline{S}_{1+2},\ s^2_{S_{1+2}}).$$

A condition for this being true is that the extended S_i series still follows a normal distribution. Going back to Eq. 3.17 it is seen to be the case, because S_i is assumed to be a linear function of the normal variables W_i and ε_i. However, it should be noted that this statement is only valid when the normality of the long W_i series is preserved and not distorted by some type of nonhomogeneities.

With a revised probability function as indicated in the above term Bayesian Risk and EOL computations can be pursued. A measure of the value of using a regression model to extend a primary set of data - the annual sediment load - has thereby been obtained. It should be noted that the uncertainty inherent in the prediction given by the regression equation has been accounted for by this method, if the regression model is, in fact, correct. The worth of the secondary data set is equal to the difference between the EOL values before and after the incorporation of those data in the statistical decision analysis.

CHAPTER IV
CASE STUDY IN THE RIO GRANDE BASIN

4.1 Problem Description

One of the best known sediment-oriented problem areas in the world is the Rio Grande river basin, especially in the New Mexico region. The consumptive use of water for irrigation and the heavy sediment contribution from tributaries have resulted in considerable and harmful aggradation of major reaches along the rivers in the system. In that connection Woodson and Martin (1963) reported that during a 20-year period about 17,500,000 cubic yards of sediment was deposited on the channel and floodplain of the Rio Grande between Cochiti and San Antonio, respectively about 50 miles above and 100 miles below Albuquerque. The U. S. Army Corps of Engineers was authorized to investigate and control the situation, and it resulted in a project including four major reservoirs to store sediment; Cochiti on the Rio Grande, Abiquiu on the Rio Chama, Jemez on the Rio Jemez, and Galisteo on the Rio Galisteo. It is estimated that these reservoirs will reduce the sediment load in the Rio Grande near Albuquerque by 70 percent after 20 years. With the exception of Cochiti Dam, all of the reservoirs are at present completed and in operation.

In order to design such a sediment control program, the required information is usually obtained from published records of these data. But what is the value of such a set of data? How representative are they? It is necessary to look at a specific project to determine and answer that type of question. It was decided to concentrate on the data for the yet uncompleted Cochiti Dam project where the Corps of Engineers could make realistic cost figures available, and the U. S. Geological Survey's sediment and water discharge sampling station upstream of Cochiti on the Rio Grande could provide a long record of water discharge and a fairly long record of concurrent discharge and sediment load data.

In order to investigate the effect of the correlative dependency between water discharge and sediment load another gaging site was selected in the Rio Grande Basin. The U. S. Geological Survey sampling station at the Pecos River near Artesia, New Mexico, was found appropriate for that purpose. The reason for this selection was the fact that it was desirable to investigate data from rivers draining the same type of climatic and geological regions, i.e. with equivalent natural features. Furthermore, the water discharge and sediment load records for both stations had the same length and were gaged over the same time period. Another reason, why those two stations were suitable for comparison, was that the two sets of concurrent data of water discharge and sediment load showed a pronounced difference in cross-correlation coefficients. More about this matter is explained in Section 4.3.3.

Because there is no "true" or exact knowledge of the annual sediment load series (the characteristics of the population are never known), an overdesign or underdesign of the long-term reservoir storage for sediment deposition might be the case. The lack of information about the "true" population characteristics is especially distinct dealing with sediment load series, because they normally have been gaged for a much shorter period of time compared to other hydrologic variables, such as precipitation and runoff.

An important feature in applying the statistical decision analysis is the use of a goal or objective function, often called also the benefit or cost function, dependent on the study case. The problem is to determine the design alternative (the size of the sediment storage), which minimizes the cost due to either building the reservoir too large or constructing it too small. An overdesign results in an increased cost measured in dollars per acre-foot of excess storage; an underdesign causes either a need for a removal of sediment from the reservoir or a loss of reservoir storage allocated for other purposes (flood control, water supply, etc.). It is the trade-off between these two types of costs which are the basis for the economic minimization problem.

4.2 The Goal Function

The goal function in this study is an "opportunity cost" function, often called in the literature, a penalty function. It indicates the excess cost that has to be paid because of either an overdesign or an underdesign during the lifetime of the reservoir. If the future amount of deposited sediment exactly matches the alternative chosen for design, this cost will be zero. However, the more the realized sediment load diverges from the design value the greater is the opportunity cost.

The variable which will be of concern here is the mean sediment load over a time period equal to design lifetime. If the probability of outcomes of this variable can be described by means of a density function, it is feasible to calculate the expected value of the future cost. The integration will then be over all "possible" values of the mean sediment load during the design lifetime as shown in Fig. 4.1.

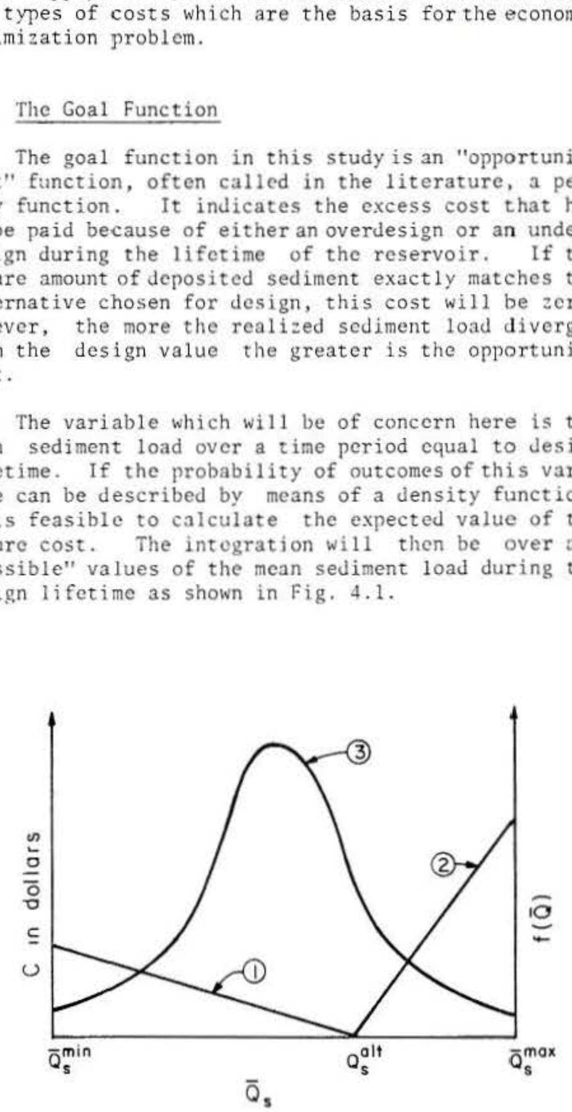

Fig. 4.1 Penalty and density functions:
(1) Overdesign cost (C);
(2) Underdesign cost (C);
(3) Probability density curve of mean annual sediment inflow into the reservoir;
Q_s^{alt}=Decision variable.

In Fig. 4.1, curve 1 represents costs for overdesign of sediment storage. Curve 2 represents costs for underdesign; these costs would show up either as (a) removal of sediment, (b) additional construction, or (c) loss in expected benefits (say in decrease of flood damage). Curve 3 represents the probability density curve which the future mean sediment load $\overline{Q}_s$ follows. The estimate of this density curve and the definition of $\overline{Q}_s^{min}$ and $\overline{Q}_s^{max}$ are given in the following text. The density function of $\overline{Q}_s$ is designated $\phi(\overline{Q}_s|\mu,\sigma^2)$, which indicates the obvious fact that the sample mean of the annual sediment load is a function of the state parameters μ and σ^2 (the population parameters of the assumed lognormal distribution). In Aitchison and Brown (1957) the explicit relationship between the mean and variance of a series in logarithmic transformation and the mean δ and variance η^2 of the original series, is given as

$$\delta = e^{\mu+\sigma^2/2} \tag{4.1}$$

and

$$\eta^2 = e^{2\mu+\sigma^2}\,(e^{\sigma^2}-1). \tag{4.2}$$

The random variable is the mean sediment load in N years, where N is the design lifetime of the reservoir. Because N is large (say about 50), the Central Limit Theorem justifies the use of a normal probability distribution for the average 50-year sediment load inflow, or

$$N(\overline{Q}_s|\delta,\frac{\eta^2}{N}) = \sqrt{\frac{N}{2\pi\eta^2}}\cdot e^{\frac{-N(\overline{Q}_s-\delta)^2}{2\eta^2}} . \tag{4.3}$$

With the relationships of Eqs. 4.1 and 4.2, it is now possible to express the desired probability density function as,

$$\phi(\overline{Q}_s|\mu,\sigma^2) = N(\overline{Q}_s|\delta,\frac{\eta^2}{N}). \tag{4.4}$$

The goal function is dependent on the alternative chosen for design and on the values of the state parameters in question. The functional form is:

$$G(Q_s^{alt}|\mu,\sigma^2) = \int_{\overline{Q}_s^{min}}^{\overline{Q}_s^{max}} \text{Cost }(\overline{Q}_s,Q_s^{alt})\cdot\phi(\overline{Q}_s|\mu,\sigma^2)\cdot d\overline{Q}_s, \tag{4.5}$$

where, $\phi(\overline{Q}_s|\mu,\sigma^2)$ is defined above by Eqs. 4.3 and 4.4,

$$\overline{Q}_s^{min} = \delta - 3.0\cdot\frac{\eta}{\sqrt{N}}, \tag{4.6}$$

$$\overline{Q}_s^{max} = \delta + 3.0\cdot\frac{\eta}{\sqrt{N}}, \tag{4.7}$$

δ and η are determined using Eqs. 4.1 and 4.2, and

N = design lifetime of reservoir (50 years in this study).

The limits of the integration are chosen on the basis that over 99 percent of the possible data points are included in this range.

The cost function, Cost $(\overline{Q}_s,Q_s^{alt})$, has to be divided into two cases in the integration (see Fig. 4.1), dependent upon whether the future mean load is less than or greater than the alternative chosen.

Case 1: Overdesign, $\overline{Q}_s < Q_s^{alt}$:

$$\text{Cost }(\overline{Q}_s,Q_s^{alt}) = P\cdot F\cdot N\cdot(Q_s^{alt}-\overline{Q}_s)\cdot K_1\cdot(1+i)^n, \tag{4.8}$$

where

P = proportionality factor between total sediment load and suspended load (P = 1.2, see Section 4.3);

F = factor which converts the sediment load data from tons per year to acre-feet per year; F = 0.00061, equivalent to a unit weight of the deposited sediment of 75 pounds per ft^3, which is the average density used by the U. S. Army Corps of Engineers in their designs;

N = design life of reservoir (50 years);

$\overline{Q}_s$ = mean annual sediment load (in tons per year);

K_1 = unit cost of construction (in dollar per acre-feet); Through written communication with U.S. Army Corps of Engineers, Albuquerque District, the figure K_1 = \$150 per acre-feet was used;

i = the interest rate of borrowed money to finance construction costs;

n = number of years between the commencement of the loan and the start of reservoir operation.

Case 2: Underdesign, $\overline{Q}_s > Q_s^{alt}$:

If an underestimation of the sediment load is realized in the future, one approach would be to remove the excess deposited sediment a certain number of times during the "life" of the reservoir. According to the Task Committee for Preparation of Manual on Sedimentation (1969, p. 195) this approach will be adopted more and more as a general practice whereas the traditional use of additional storage as a means of solving sediment problems will not be feasible in the future. This is caused not only by growing construction costs, but more importantly by the fact that sites for low-cost dams are disappearing.

Let us assume constant time intervals between two successive removals, say each M years. Let N be a multiple of M. For every time money is spent in the future, these costs have to be discounted back to get the equivalent-time cost figure. If the interest rate for discounting is r, we then have:

$$\text{Cost }(\overline{Q}_s, Q_s^{alt}) = P\cdot F\cdot M\cdot(\overline{Q}_s - Q_s^{alt})\cdot K_2 \cdot\left[\frac{1}{(1+r)^M} + \frac{1}{(1+r)^{2M}} + \cdots + \frac{1}{(1+r)^{N-M}} + \frac{1}{(1+r)^N}\right], \tag{4.9}$$

where K_2 = unit cost factor for removal of sediment (in dollars per acre-feet); estimated to be \$1700 per acre-feet. (See Chapter 5.2 for explanation of this figure).

The other terms in Eq. 4.9 are explained above. During every M-year period there will be a minor loss of storage; for simplification this cost is not included in the analysis as Eq. 4.9 indicates. In the computations throughout this chapter the interest rate i and the rate r used for discounting have simply been selected to take the same value of six percent.

4.3 Application of the Method

The procedures described in Chapter III to achieve sediment storage design criteria, to find worth of sediment load data, optimum record length and so forth, are now applied in an actual case study. The river subject for investigation is the Rio Grande upstream of the Cochiti Dam site in New Mexico. The U. S. Geological Survey station (number 8.3130) at Otowi Bridge near San Ildefonso has been recording suspended sediment load for more than 20 consecutive years. The sediment load data are compiled from the U. S. Geological Survey annual reports on quality of surface waters in the United States.

Table 4.1 presents 20 years of observed sediment load used, together with the concurrent water discharge data. Figures 4.2 and 4.3 show a cumulative plot of the annual sediment load and annual runoff, respectively. It is seen that the two single mass curves are

Table 4.1 Annual Sediment Load and Water Discharge as Recorded at Otowi Bridge, Rio Grande, New Mexico. Data taken from U. S. Geological Survey, Water-Supply Papers, Part 8, Western Gulf of Mexico Basins.

Year	Annual Sediment Load in Tons	Annual Water Discharge in Acre-feet
1948	4,306,000	1,359,500
1949	3,681,000	1,301,300
1950	1,733,000	662,300
1951	900,700	394,600
1952	4,475,400	1,375,700
1953	732,000	547,700
1954	1,329,500	749,900
1955	2,430,700	431,200
1956	714,300	376,400
1957	455,700	1,295,300
1958	7,562,200	1,522,800
1959	1,424,500	508,900
1960	2,074,300	819,700
1961	1,971,900	674,400
1962	3,253,000	1,037,700
1963	862,100	559,400
1964	946,600	382,900
1965	3,377,900	1,175,900
1966	2,255,600	942,000
1967	2,651,000	578,500
Annual Means	2,367,000	835,000

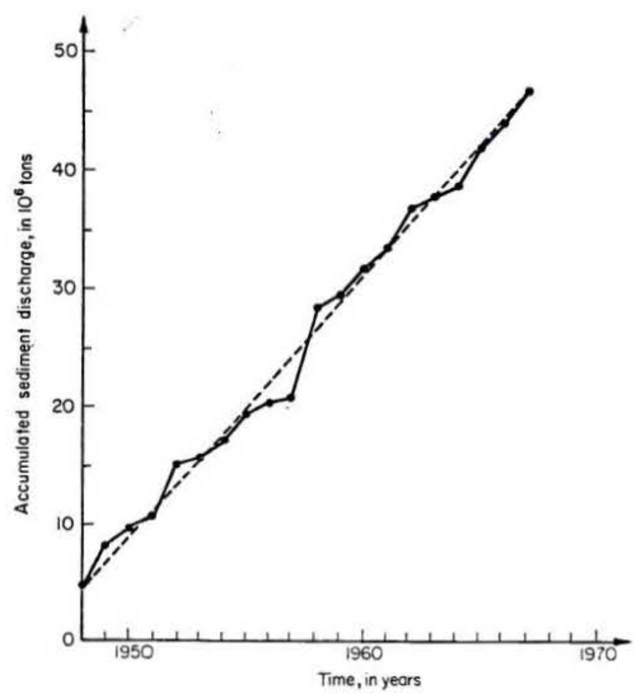

Fig. 4.2 Mass curve for annual sediment load, Rio Grande at Otowi Bridge, N.M.

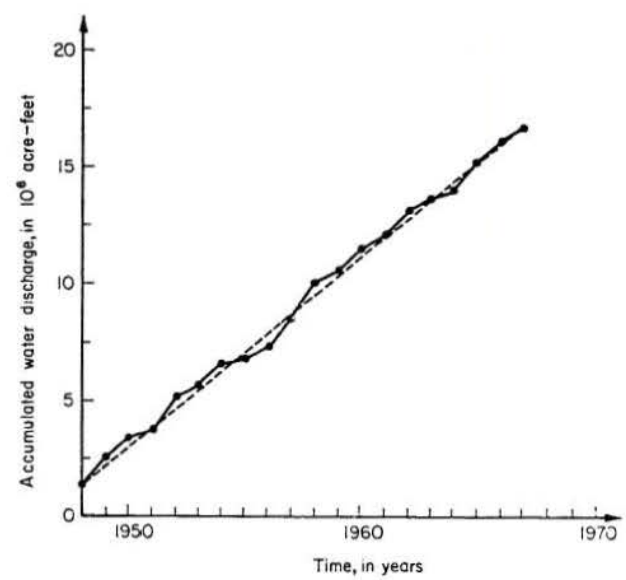

Fig. 4.3 Mass curve for annual water discharge, Rio Grande at Otowi Bridge, N.M.

approximately linear without any marked change in slopes. The graphs indicate in a qualitative way that nonstationarity cannot be detected in these observed annual data. In the sampling period apparently no large-scale factors have been introduced into the drainage basin that would significantly affect the annual sediment yield conditions at the Cochiti Lake site. The probability density function of the state parameters of Chapter II is based on the assumption that the logarithms of the annual sediment load follow a normal distribution. To test this assumption a plot of the frequency curve is given on logarithmic-probability graph paper, as shown in Fig. 4.4. The Kolmogorov-Smirnov Δ-statistic is used to test the goodness of the graphical fit of a straight line.

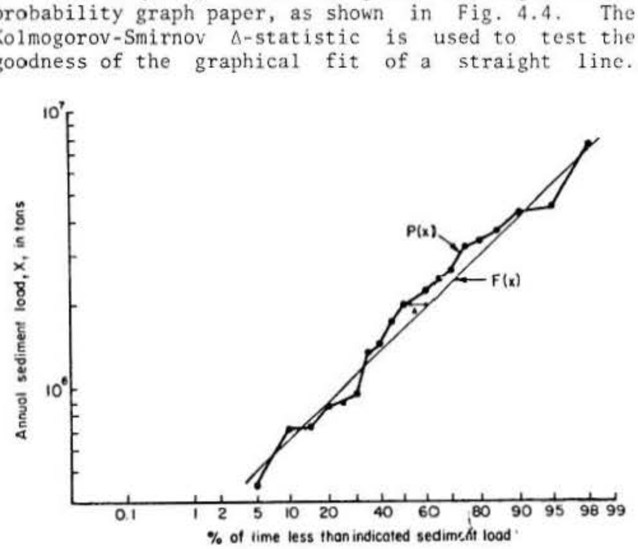

Fig. 4.4 Kolmogorov-Smirnov test for normal distribution of logarithms of annual sediment load.

$$\Delta = \max|F(x) - P(x)| = 0.10 \quad (4.10)$$

From a table over critical values Δ_o of the Kolmogorov-Smirnov statistic (see for instance Yevjevich, 1972, p. 229) one reads for $n = 20$ and $\alpha = 0.05$:

$$\Delta_o = 0.29 \quad (4.11)$$

Because $\Delta = 0.10 < \Delta_o = 0.29$, the fit of Fig. 4.4 is acceptable by the Kolmogorov-Smirnov test on a five percent significance level, which further implies that the Gaussian assumption after a logarithmic transformation of the data should be accepted. Therefore, the input data in form of natural logarithms of the observed annual sediment load will be used in all computations.

The drainage area above the gaging site is about 14,300 square miles in which the Rio Grande runs mainly in deep canyons or relatively narrow alluvial valleys. According to a report by the U. S. Corps of Engineers (1971), the Colorado portion of the Rio Grande watershed is mostly mountainous and contributes relatively little sediment to the Rio Grande in proportion to the area and volume of runoff. In the upper New Mexico part of the watershed the sediment load is increased heavily, mainly due to tributary contributions, especially by the Rio Chama which flows through highly erosive materials. The Abiquiu Dam was placed in operation across the Rio Chama 30 miles upstream of the junction with the Rio Grande in 1963. However, Figs. 4.2 and 4.3 indicate that the dam does not yet significantly affect the sediment load of the Rio Grande. This implies further that a degradation of the 30 miles reach along the Rio Chama has resulted since 1963.

Nordin and Beverage (1965) report on investigations of many aspects of the sediment transport in the Middle Rio Grande in New Mexico. This report describes among other things hydraulic data, observed and computed sediment concentrations, and size distributions of bed-material samples for nearly 300 observations. The ratio between the total amount of sediment load and the suspended load is introduced in the goal function in Section 4.2. This ratio reflects a complex interdependence between many different types of factors such as hydrodynamic variables and geophysical features of the basin. In this study the value of the ratio is estimated to be 1.2. This figure can be derived from Table 1, of the report by Nordin and Beverage (1965).

4.3.1 Analysis of the Observed Sample

Design alternatives, expected cost as related to the chosen alternative (Bayesian Risk), and EOL were computed on the CDC 6400 computer at Colorado State University applying Eqs. 2.1 through 2.5. The computer program is given and explained in the study by Jacobi (1974, Appendix C).

Table 4.2 shows the results for the use of various periods of annual sediment data, and Figs. 4.5 and 4.6 present some of these results in graphical form, with

Table 4.2 Worth of Data and Sediment Storage Results. The Rio Grande, 20 Years of Observed Data, and Design Lifetime of 50 Years

Time Period		Design Alternative Acre-feet	Bayesian Risk 10^6 \$	EOL 10^6 \$
5-year period	1948-52	178,000	28.83	13.84
	1953-57	60,000	8.22	4.70
	1958-62	153,000	27.95	14.33
	1963-67	105,000	13.30	6.96
10-year period	1948-57	89,000	11.82	5.96
	1958-67	110,000	8.65	4.92
20-year period	1948-67	90,000	5.97	2.69
Average for different periods of record	5 years	125,000	19.56	9.94
	10 years	100,000	10.23	5.44
	20 years	90,000	5.97	2.69

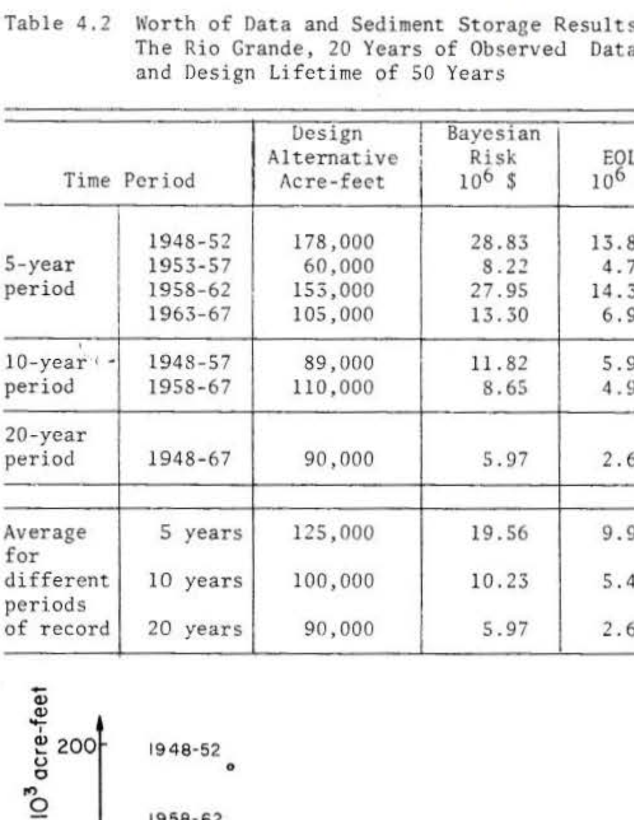

Fig. 4.5 Bayesian Risk design for various periods of data.

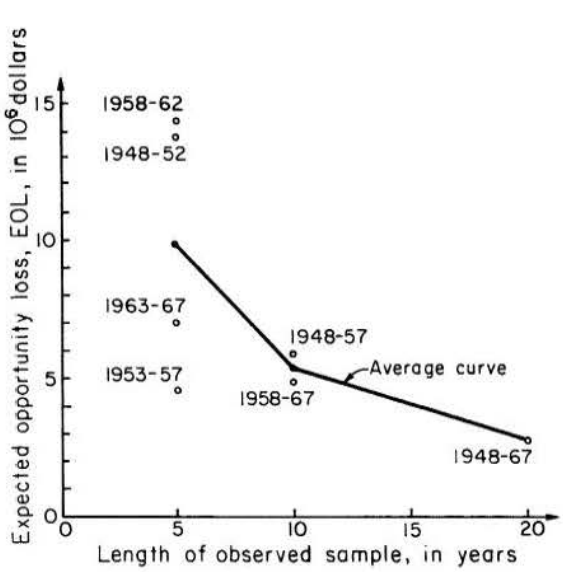

Fig. 4.6 Expected opportunity loss for various periods of data

Fig. 4.5 pertaining to the design and Fig. 4.6 to the worth of data. The goal function involves the cost figures, design lifetime, etc., which were obtained through personal communication with the U. S. Army Corps of Engineers, Albuquerque, New Mexico District. Thus, the minimum Bayesian Risk design achieved in this study is comparable with the storage the Corps of Engineers has allocated for the sediment deposition in the design of Cochiti Reservoir. This storage is 110,000 acre-feet for a 50 years design lifetime (U. S. Corps of Engineers, 1971, p. 16). In Appendix A are the two types of design criteria and results further taken up for comparison and discussion.

Figure 4.5 shows the variation in storage design using the four five-year periods and the two 10-year periods. The design recommended using the 1948-52 sample is nearly three times greater than the design based on the 1953-57 sample. For the two 10-year samples the variation is only a little greater than 1.2. This sharp drop in the spread of design results indicates how uncertain it is to base a design decision on a short (5-year) record. The variation in the spread might also be due somewhat to the fact that there is a difference in sample sizes because the investigation is over four samples from the five-year periods and only two samples representing the 10-year periods. The average design curve in Fig. 4.5 shows a downward trend from 125,000 acre-feet to 90,000 acre-feet. This is explained by the probability density function containing more uncertainty when based on a five-year sample than on a 20-year sample. With a higher degree of uncertainty there will be a tendency in the decision process to favor an overdesign rather than underdesign, because the cost of overdesign is smaller than that of building the storage capacity too small.

Figure 4.6 shows as anticipated a decrease in the expected opportunity loss going from five-year sample, to 10-year sample, and ending up with the observed record of 20 years length. EOL is the reduction expected in the Bayesian Risk due to a better information. It is seen that the average EOL curve still has a marked negative slope after the incorporation of all 20 data points; that is a longer observed record should decrease further the uncertainty about the state parameters. The slope of the EOL curve determines the relative importance of the gain in information by including additional data points. Since the computed EOL figures are nothing but an estimate of an unknown "population" EOL curve, fluctuations around a smooth decreasing curve should be expected. In other words, the sampling variability from the expected curve is anticipated, a fact which the sample points in Fig. 4.6 indicate.

4.3.2 EVEOL, Expected Bayesian Risk, and Optimum Record Length

In Section 3.2 a procedure was outlined to achieve the expected Bayesian Risk and EVEOL through simulation. In the generation procedure the first-order autocorrelation coefficient $\rho(1)$ for the sediment load data is used in Eq. 3.7 and is estimated from the given 20 years sample to be $\hat{\rho}(1) = -0.19$. In this case a negative $\hat{\rho}(1)$ might be explained by a type of "cleaning-out" action where in one year if a large sediment yield has occurred the next year there will be a deficiency of material readily available to be transported by the water. The values associated with the extended samples are calculated by incorporating the information with that of the observed sample. This information is reflected in the apriori probability density function used to find the expected Bayesian Risk and EVEOL. Table 4.3 shows results for the extension of the observed data by the application of the expected EOL concept as defined in this study. Computations are

Table 4.3 Expected Bayesian Risk and EVEOL Results. The Rio Grande, Design Lifetime 50 Years.

	Sample Size	Design Alternatives Acre-feet	Bayesian Risk 10^6 \$	EVEOL 10^6 \$	EOL(5)-EVEOL(n) 10^6 \$
Observed data points used	5	125,000	19.56	9.94	----
	10	100,000	10.23	5.44	4.50
	20	90,000	5.97	2.69	7.25
Generated records included	30	90,000	4.27	1.62	8.32
	40	90,000	4.19	1.35	8.59
	50	90,000	3.77	1.07	8.87
	60	90,000	3.71	0.93	9.01

carried out for an ascending number of additional years up to a total record length of 60 years. It should be noted that no new information has been introduced by this procedure and therefore, as seen in Table 4.3, no change in storage design is realized using the generated data compared to the result obtained by utilizing the entire observed record. The last column in Table 4.3 is included for use later in this Section. In Fig. 4.7 the EOL curve is plotted as a combination of the values obtained from the given 20 years sample and the expected EOL values from 20 to 60 years. In Fig. 4.7 it is seen that the curve appears to have a shape similar to a decreasing power function. It is further noted from the graph that practically speaking there is no decrease in EOL after incorporation of roughly 30 additional data points. Beyond a sample size of approximately 50, relatively there is no noticeable gain in information, which is in agreement with the findings reported in Section 3.1.

Records of suspended sediment load samples are generally short compared to other related hydrological records. Most likely a design using sediment data must be based on only five to ten years of observed data or less. Only 10 stations out of the 900 U.S. Geological Survey stations which take sediment measurements have a record length of 25 years or more as of 1972. This

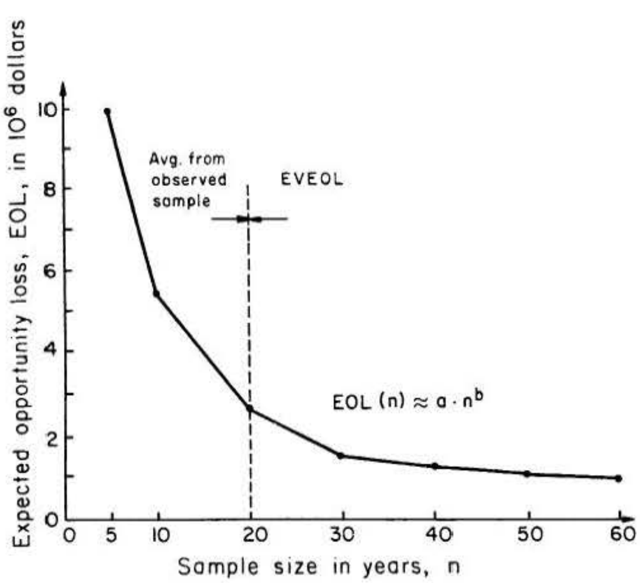

Fig. 4.7 The relationship of EOL(n) to sample size, n.

points out the importance of using every bit of data available in order to extract as much information as possible about the characteristics of the sediment load series.

The above mentioned feature of the EOL curve can be presented quantitatively in the following way. Figure 4.8 shows a plot on log-log scale paper of the EOL values versus the number of years in the underlying sample. A straight line is seen to fit well the points,

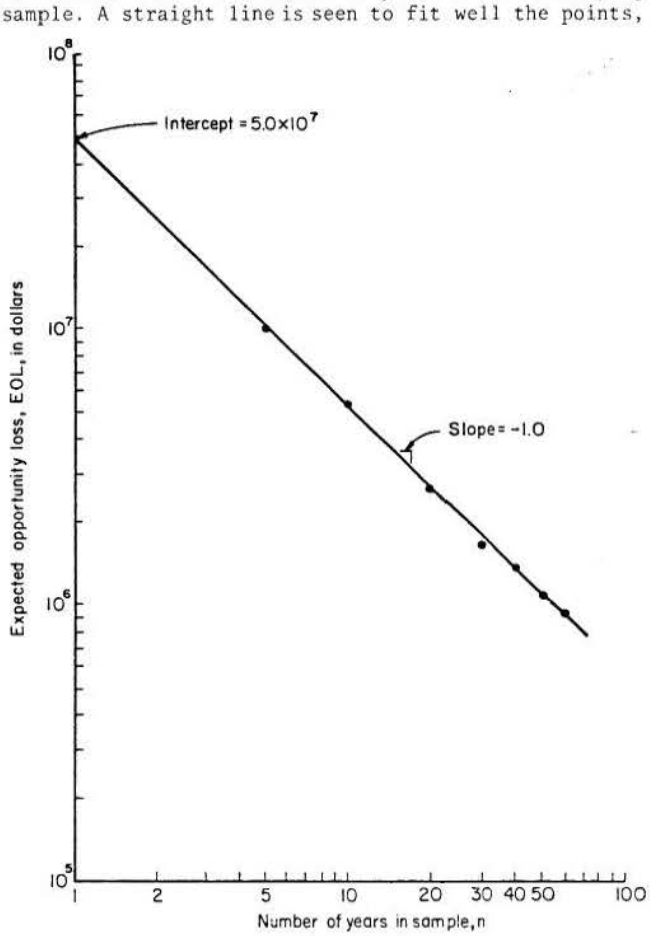

Fig. 4.8 Straight line fit to log-log plot of EOL versus n.

therefore the expected opportunity loss as a function of sample length, EOL(n), is assumed to be a power function: $EOL(n) = a_1 \cdot n^{b_1}$, where the constants are found as the intercept and slope of the straight line in Fig. 4.8:

$$a_1 = 5.0 \cdot 10^7, \tag{4.12}$$

$$b_1 = -1.0, \tag{4.13}$$

which implies

$$EOL(n) = \frac{50 \cdot 10^6}{n}, \tag{4.14}$$

with EOL in dollars and n in years. In this particular case the EOL-curve is seen to be an equilateral hyperbola since $b_1 = -1.0$. With the explicit functional relationship established in Eq. 4.14, it is now possible to compute in quantitative terms the expected marginal worth of one increment of data (one additional data point). This is obtained as the absolute value of the first derivative of the EOL-function with respect to n.

$$\left|\frac{d\{EOL(n)\}}{dn}\right| = \left|10^6 \cdot 50 \cdot (-1.0) \cdot n^{-2.0}\right| = \frac{50 \cdot 10^6}{n^2} \tag{4.15}$$

For different number of years in the sample the following values are obtained:

n	5 yr.	20 yr.	50 yr.
$\left\|\frac{d\{EOL(n)\}}{dn}\right\|$	2,000,000 $	125,000 $	20,000 $

It is of interest to note that the worth of one extra data point added to a sample, already consisting of 50 points is one percent of the worth of an extra point when added to a five-year sample. From a statistical point of view--and without any cost considerations taken into account--it might be concluded that sampling should continue as long as possible since a decrease of EOL is continuously realized. However, it should be kept in mind, as the above table shows, that the value of one additional data point is highly dependent on the length of the available sample.

From an economical viewpoint unlimited data collection is naturally never acceptable when the data are used for design purpose in a proposed project. The question therefore arises: What is the expected economical optimum record length of the sample to be used in the decision making process? The decrease in expected opportunity loss (EOL) with the incorporation of more data represents in monetary terms the reduction of uncertainty about the state parameters in the decision process. Furthermore, a smaller EOL figure implies a smaller anticipated cost because of realizing either an overdesign or underdesign of the sediment storage. The decrease in EOL with the addition of more data to the sample can therefore be considered as actual benefits caused by those extra data points. But there are two costs of getting the additional data: (1) the cost of continued data collection, and (2) the cost of delaying the construction of the project which shows up as missing benefits. With the benefits and costs defined an economic analysis can produce the optimum record length.

Assume that the given data base consits of five years of observed sediment loads. How many additional years of data - if any - would be economically worthwhile to include in the given data base with respect to the Cochiti sediment storage project?

In order to make such an economic analysis the total benefit and cost curves have to be found explicitly. The estimated values for the benefit curve are computed and given in Table 4.3, where the decrease in expected opportunity loss is calculated going from five years of data to any additional number of years. The analytical expression for the benefit function is established because the functional form of EOL as a function of the number of years, n, already is found by Eq. 4.14.

That means, the total benefit function in this analysis is:

$$B(n) = EOL(5) - EVEOL(n), \qquad (4.16)$$

or substituting from Eq. 4.14 yields,

$$B(n) = (1 - \frac{5}{n}) \cdot 10^7, \qquad (4.17)$$

where, n is in years (≥ 5), and B(n) is expressed in dollars.

The functional form of the total cost curve is a sum of two terms: the annual expenditures of operating the sediment load sampling facilities, c_1, and the benefits foregone, c_2, of not having the project in operation. A detailed explanation of how an estimate of the benefits foregone is obtained is given in Appendix A. In this case it is the annual net benefit - based on 1973 figures - provided by the trapping of sediment in Cochiti Lake.

The annual cost of data collection and project postponement (benefits foregone) changes with the number of years of delay as an interplay between these factors: (1) an increasing trend because of continued development of the middle Rio Grande flood plain, and (2) the discount factor used to bring the future costs back to present time (including the inflation allowance). For the benefits foregone part of the total cost function, it is often important to take a third factor into account. It reflects the fact that benefits from a project normally improve with a larger data sample used in the design of the project. In the present analysis this increasing tendency can be neglected. The reason is that the sediment trapping efficiency (and therefore the benefits) is not affected to any noticeable degree of either an underdesign or an overdesign of the sediment storage. After all, the storage allocated for sediment constitutes only 15 percent of the total capacity of the reservoir.

If the two described trends are represented by the annual percentages r_1 and r_2, respectively, the functional form of the total cost curve is,

$$C(n) = c_1 \cdot \left[1 + \frac{1}{(1+r_2)} + \frac{1}{(1+r_2)^2} + \cdots + \frac{1}{(1+r_2)^{n'-1}}\right]$$

$$+ c_2 \cdot \left[1 + \frac{(1+r_1)}{(1+r_2)} + \frac{(1+r_1)^2}{(1+r_2)^2} + \cdots + \frac{(1+r_1)^{n'-1}}{(1+r_2)^{n'-1}}\right] \qquad (4.18)$$

where,

- $n' = n - 5$, since $n \geq 5$; n' and n are expressed in years.
- c_1 = \$12,000. c_1 is the average annual cost of having a fully equipped sediment sampling station in operation.
- c_2 = \$275,000. This estimated value of c_2 is found in Appendix A through written communication with the U. S. Army Corps of Engineers, Albuquerque District, N. M.
- r_1 = four percent. This figure is achieved assuming that the increase in benefits foregone is proportional to the population growth in the region. U. S. Census Reports show approximately a four percent annual increase in the population of the Rio Grande Valley taken as an average of rural, farm, and urban areas.
- r_2 = two percent. Discount rate including inflation allowance; obtained as a combination of the same unadjusted discount rate of six percent as used in the goal function (Section 4.2) and an assumed inflation rate of four percent.

Using the standard formula for the sum of a geometric progression series, Eq. 4.18 reduces to,

$$C(n) = c_1 \cdot \frac{\left(\frac{1}{1+r_2}\right)^{n'} - 1}{\left(\frac{1}{1+r_2}\right) - 1} + c_2 \cdot \frac{\left(\frac{1+r_1}{1+r_2}\right)^{n'} - 1}{\left(\frac{1+r_1}{1+r_2}\right) - 1} \qquad (4.19)$$

For the given values of r_1 and r_2, and when (n-5) at the same time is substituted for n', Eq. 4.19 can be written as,

$$C(n) = \frac{c_1}{0.02} \cdot (1 - 0.98^{n-5}) + \frac{c_2}{0.02} \cdot (1.02^{n-5} - 1). \qquad (4.20)$$

Inserting the cost factors yields the final form of the total cost function as,

$$C(n) = 0.6 \cdot 10^6 \cdot (1 - 0.98^{n-5}) + 13.75 \cdot 10^6 \cdot (1.02^{n-5} - 1), \qquad (4.21)$$

with n expressed in years (≥ 5), and C(n) in dollars.

In Fig. 4.9 the total benefit function, Eq. 4.17, and the total cost function, Eq. 4.21, are plotted versus the number of data points in the sample.

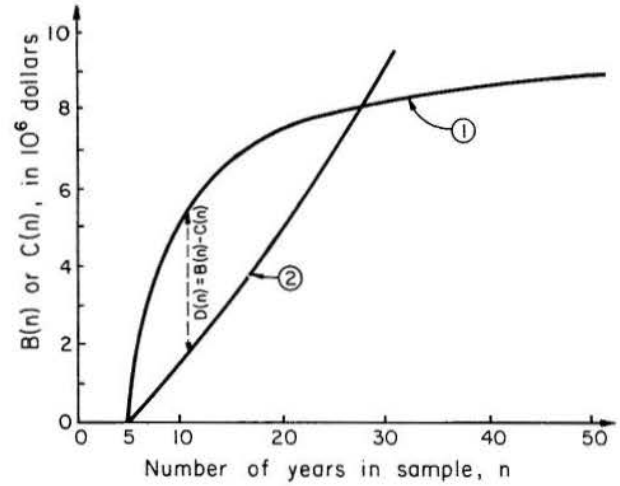

Fig. 4.9 Total benefit and total cost curves plotted against length of data sample:
(1) Total benefits, Eq. 4.17
(2) Total costs, Eq. 4.21

The economic optimal point is defined as that point where the difference between total benefits and costs is as large as possible. That means, the function $D(n) = B(n) - C(n)$ is subject to maximization with respect to n. D(n) is drawn in Fig. 4.9. In order to find the point the derivative $\frac{dD(n)}{dn}$ is set equal to zero, or

$$\frac{dB(n)}{dn} = \frac{dC(n)}{dn} \qquad (4.22)$$

Substituting Eqs. 4.17 and 4.21 for B(n) and C(n) in Eq. 4.22, and taking the derivatives yields,

$$\frac{5.0 \cdot 10^7}{n^2} = -0.6 \cdot 10^6 \cdot 0.98^{n-5} \cdot \ln(0.98)$$
$$+ 13.73 \cdot 10^6 \cdot 1.02^{n-5} \cdot \ln(1.02) \ . \qquad (4.23)$$

Rearranging the terms results in,

$$2.72 \cdot 1.02^{n-5} + 0.12 \cdot 0.98^{n-5} - 500/n^2 = 0. \qquad (4.24)$$

Eq. 4.24 cannot be solved explicitly. By the "trial and error" method the equation is found to be satisfied for n = 12.4. Left side of Eq. 4.24 is equal to -0.24 for n = 12, and equal to 0.33 for n = 13. Therefore, the conclusion can be drawn that the economic optimal length of the sample used in the decision process is approximately 12 years.

In economic terminology that point is reached where the marginal benefit curve intersects the marginal cost curve as explained in Section 3.2. Figure 4.10 shows those curves and gives the graphical solution to the problem.

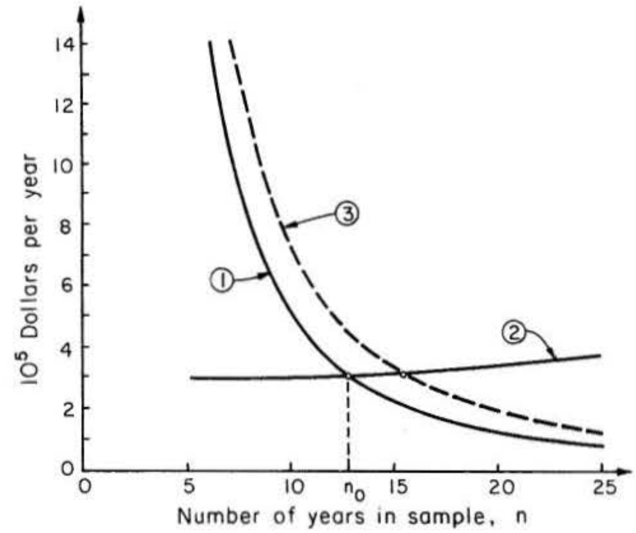

Fig. 4.10 Marginal benefits and marginal costs curves plotted against length of data sample:
(1) Marginal benefits 50-year design lifetime;
(2) Marginal costs:
(3) Marginal benefits 100-year design lifetime; and
n_o = optimal record length for 50-year design lifetime.

The number of data points were assumed to be five. That means, from an economic point of view, it would be worthwhile to delay the construction of the reservoir for seven years and thereby collect seven more data points to form the desired 12 years sediment load sample. Said in another way, if more than 12 years of annual sediment load data are available, which is the actual case for the Rio Grande at Otowi, no postponement of construction is recommended. The design should be calculated and decided with the given information inherent in the already available sample.

This analysis shows how the design engineer, after a political decision has been made to build a project, might be forced to suggest a postponement of the construction for economic reasons. However, if the political decision-makers want the project immediately, it is possibly because they consider its intangible benefits (for instance, regional development, recreation, environmental quality) as more important than the extra costs created by an economic suboptimal design.

The expected optimal record length is a result of a complex interrelationship between statistical, economic, and hydrologic characteristics. As an example, a sensitivity analysis was performed with respect to one of the factors involved, namely the design lifetime of the project. Up to this point the entire decision process and the investigation of data value have been carried out with a lifetime of the sediment reservoir of 50 years. The analysis is now repeated with a 100-year lifetime of the deposition storage. The numerical results of the computer calculations of design alternative and EOL are shown in Table 4.4.

Table 4.4 100-Year Design and EOL Results. The Rio Grande.

Sample Size	Design Alternative Acre-feet	EOL 10^6 \$	EOL(5)-EOL(n) 10^6 \$
5	180,000	13.90	-----
10	172,000	7.40	6.50
20	175,000	4.01	9.89
30	175,000	2.85	11.05
50	175,000	1.70	12.20

The total benefit curve, similar to Eq. 4.17 is in this case found to be,

$$B(n) = 1.39 \cdot 10^7 - 6.0 \cdot 10^7 \cdot n^{-0.85} \qquad (4.25)$$

with n expressed in years (≥ 5), and B(n) in dollars. Accordingly the marginal benefit function is,

$$\frac{dB(n)}{dn} = \frac{5.1 \cdot 10^7}{n^{1.85}} \qquad (4.26)$$

Eq. 4.26 is plotted in Fig. 4.10, and shown as a dashed curve. The marginal cost curve in this graph is valid also in the case where a 100-year lifetime design is considered. The annual benefits foregone and the cost of data collection are both independent of the physical life of the reservoir according to the definition of the cost function in this study.

The location of the two intersection points in Fig. 4.10 indicates that the economical optimum length of the sample used in the design phase can be considered rather insensitive to the design lifetime, at least when dealing with the long-term projects like reservoirs. Mathematically the optimum record length in this case is found to be n = 15.0. Only an increase of three years from 12 to 15 years of recommended data is needed when the lifetime is extended from 50 to 100 years. By comparing the two marginal benefit curves in Fig. 4.10 it is found - which also should be expected - that additional data have a highest value when used for design of the most expensive of the projects, in this case the 100-year sediment storage reservoir. However, it should also be noted that this difference in the worth of data depends on the length of record as a relative decrease with an increase in the sample used for design purposes.

As recalled from Section 4.2, the design lifetime appears in the goal function as a proportionality factor in the formula for the future costs. Nevertheless, a comparison of the chosen 50 and 100 years design alternatives, respectively given in Table 4.3 and 4.4, demonstrates that a relative smaller design is suggested for the long lifetime case. This can be explained by the fact that the future cost of underdesign receives less and less weight as the reservoir lifetime increases, due to the discounting of future sediment removal costs. This tendency towards a more conservative design in connection with the shorter project lifetimes is more noticeable when the sample used for design is short. It was pointed out earlier (Section 4.3.1) how the degree of uncertainty in the decision process affects the choice of alternatives. If, at the same time, it is kept in mind that the cost of underdesign now plays a smaller role, the observed characteristic concerning the changes in design alternatives with respect to sample length and lifetimes can be explained.

4.3.3 Use of Secondary Data and Equivalent Length of Record

In the investigations and results reported, the emphasis has been on the observed annual sediment load data. The information used to find the expected optimum length was extracted from the given sediment sample. Section 3.3 describes a method to take the advantage of another information source through a linear regression model and to use it in the statistical decision approach. As the application of this method will show, the investigator must be very cautious when he uses long-term hydrologic samples for information augmentation purposes. It is sometimes the case that hydrologic data when gaged for more than 50 to 60 years indicate nonstationarities in the time series. That means that in a regression process there are two types of uncertainties with overlapping effects, namely a nonstationarity in the long-term water discharge series, and an uncertainty embedded in the regression model; the latter described by the cross-correlation coefficient. Concerning the long-term sample, the design-engineer has to decide in one way or another how much, if at all, to adjust that set of data before it is used for augmentation. This in itself is a complex question which involves considerations like (a) wet versus dry spells in the climate, (b) lower groundwater table (increased rate of infiltration), (c) irrigation development, (d) construction of upstream reservoirs, etc. The points under (a) and (b) cannot be classified as permanent changes in the hydrologic system, while (c) and (d) might be so.

For comparison to the gaging station on the Rio Grande, data were used from the U.S. Geological Survey discharge and sediment measurement station on the Pecos River near Artesia, New Mexico, Station no. 8-3965. The Pecos River is a tributary to the Rio Grande with the confluence at the Texas-Mexico border. The drainage area at Artesia is approximately the same as the area above the Otowi Bridge station on the Rio Grande. The upper part of the Pecos River watershed has natural features similar to the Rio Grande area; in the lower parts of the basin the Pecos River runs out from the mountains and onto the plains. A sediment storage construction similar to the Cochiti Lake project is assumed to be designed on the Pecos River. That means the goal function is the same as for the Rio Grande with preservation of unit cost figures, interest rates, design lifetime of 50 years, etc. Both gaging stations provide more than 20 years of suspended sediment load data and more than 60 years of water discharge data, with both series rated as reliable. Figures 4.11 and 4.12 present time series plots of the annual water discharge for the Rio Grande and the Pecos River, respectively.

These two graphs show a typical runoff characteristic in the Southwest of the United States, namely a declining trend in annual streamflow has been recorded in the last decades. This is due partly to the increasing depletion of water for supply and irrigation purposes, and partly to a change to somewhat dryer weather in the last decades. This nonstationary trend can be simplified with a step-function with a single negative jump in the mean, located in the middle of the sampling time period as indicated in Figs. 4.11 and 4.12. In that way, it is possible to obtain an idea about the percentage decrease in the mean annual streamflow. A similar analysis is done for the precipitation data in the region. Table 4.5 shows the results for the Rio Grande station at Otowi, and at two locations along the Pecos River upstream and downstream of Alamogordo Reservoir. The precipitation column in Table 4.5 shows that the long term climatic fluctuation

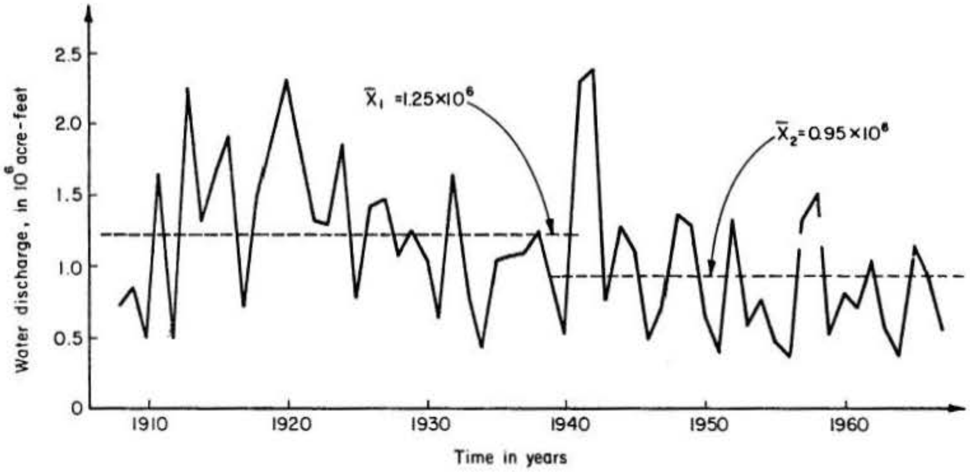

Fig. 4.11 Time series of annual water discharge of the Rio Grande at Otowi Bridge, N.M.

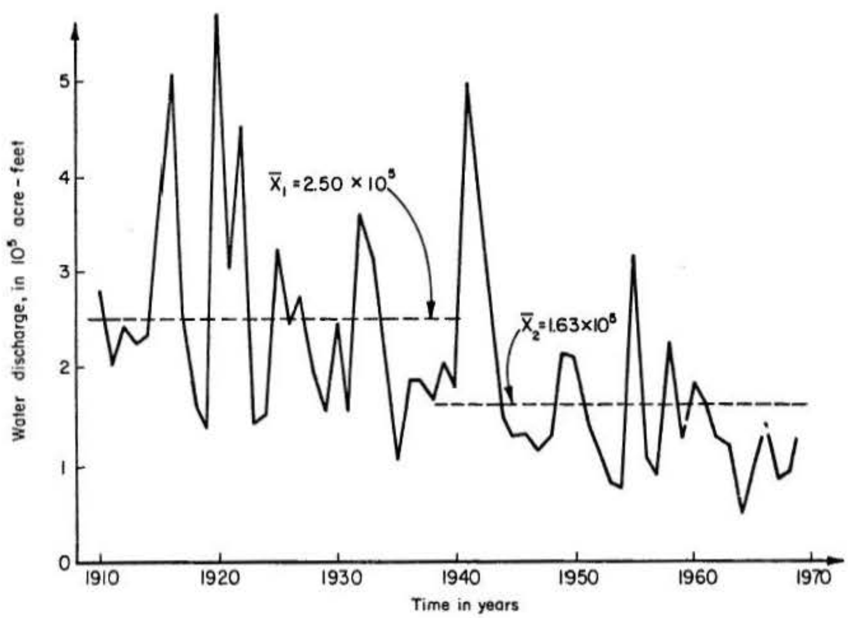

Fig. 4.12 Time series of annual water discharge of the Pecos River near Artesia, N.M.

Table 4.5 Decline in Annual Averages from Period 1910-1940 to Period 1940-1970

Location	Water Discharge*	Precipitation**
Rio Grande at Otowi	24%	11%
Pecos River at Artesia (downstream of Alamogordo Reservoir)	37%	15%
Pecos River at Pecos (upstream of Alamogordo Reservoir)	20%	13%

* Data from U. S. Geological Survey Water-Supply Papers.
** Data from the National Weather Service.

appears to be a dry spell in the last half of the investigated time period. In addition to natural fluctuations, the Pecos River watershed also has been subject to significant man-made changes, in the form of the Alamogordo irrigation reservoir which was completed in the late 1930's. This explains the introduction of the negative jump in the annual means around 1940 in order to compare the man-made nonstationarity with the natural fluctuations.

With a certain negative adjustment in the annual discharges recorded before 1940 in the Pecos River at Artesia, this data series can be made comparable with the data series from the Rio Grande. The purpose of the adjustment is to obtain relatively the same decreasing trend, which, in both series, would result mainly from natural changes in the Rio Grande basin. But how much should the first half of the Pecos River data be diminished? Following the Bayesian approach the value of such an adjustment parameter (called AP) must be considered uncertain. However, the information reported in Table 4.5 yields some helpful background in the assessment of certain values of AP. Because of the 37 percent change in the mean discharge in the Pecos River at Artesia in relation to the changes encountered at the two other locations in the basin, an AP in the neighborhood of minus 17 percent seems reasonable. The jump in the mean annual precipitation appears to be somewhat higher in the Pecos River area compared to the Rio Grande region. This is the reason for weighing the small values of the adjustment parameter heavier than the higher values in Table 4.6, which gives a "subjective" discrete probability function for the possible values of AP.

Table 4.6 Discrete Probability Function for the Value of the Adjustment Parameter (AP)

i	AP_i	$p(AP_i)$
1	-25%	0.05
2	-21%	0.10
3	-17%	0.40
4	-13%	0.30
5	- 9%	0.15
		Σ = 1.00

It is recalled that the design alternatives and EOL are functions of a given sample, in this case, the augmented series. Because the augmented record is dependent on the adjustment parameter used to reduce the measured water discharges, EOL can be written as

$$EOL(n_1+n_2,\ \overline{S}_{1+2},\ s^2_{S_{1+2}}\,|AP),$$

with the symbols of Section 3.3 used.

In the spirit of Bayesian approach, the EOL for a given augmented data sample would then be,

$$EOL(n_1+n_2,\ \overline{S}_{1+2},\ s^2_{S_{1+2}}) = \sum_{i=1}^{5} EOL\ (n_1+n_2,\ \overline{S}_{1+2},\ s^2_{S_{1+2}}\,|AP_i)\cdot p(AP_i). \qquad (4.27)$$

The same type of equation is also valid for finding the appropriate design alternative. It is seen how the consequences of not knowing the exact value of the adjustment parameter in the decision process is incorporated by the application of this expected-value criterion.

A computer program was developed to calculate design alternative and EOL in accordance with the method and formulas described in Section 3.3. The input data now consist of the sediment series and the adjusted longer water discharge sample. In order to investigate the sensitivity on the value of AP and to apply Eq. 4.27, one set of test runs was performed for a total record length of 60 years. The results are given in Table 4.7 with the values of AP_i and $p(AP_i)$ taken from Table 4.6.

Table 4.7 Expected-value Criterion to find EOL and Design Alternative for Different Adjustment Parameters. The Pecos River at Artesia

i	$(EOL\|AP_i)$ in \$	(Design Alt $\|AP_i$) in Acre-feet	$EOL\cdot p(AP_i)$	Design $\cdot p(AP_i)$
1	319,275	34,800	15,960	1,740
2	302,611	35,300	30,260	3,500
3	322,723	36,000	129,090	14,400
4	342,998	38,000	102,900	11,500
5	354,858	39,000	53,230	5,860
			Σ = \$331,440	37,000 acre-feet

Columns 2 and 3 in Table 4.7 show that the value of the adjustment parameter is not a crucial factor with regard to the EOL results and design alternatives. However, it should be noted that the design alternative increases almost 15 percent going from the minimum case (AP = -25 percent) to maximum design (AP = -9 percent). As a result of this analysis, and mainly in order to keep the computational effort down to a reasonable level, it was decided that the computations throughout this section be made with one fixed value of the adjustment parameter. Consideration of the summation results in Table 4.7 justifies the adoption of a 15 percent reduction factor in order to decrease the water discharge data recorded before 1940 in the Pecos River at Artesia.

The Alamogordo Reservoir is built for the purpose of irrigation. The resulting depletion of water from the river causes the change in the annual mean discharge, which is accounted for above. Another common result from a reservoir is a reduction in the variance of the water discharge series downstream of the dam. As seen from Eq. 3.21, the variance of the long-term sample appears in this formula. Therefore, the question arises as to what effect a nonstationarity in the variance might have in the decision process. Because the design in this study is concerned with sediment accumulation over a long time span, the variance should not be a determining factor compared to the mean of the series. To illustrate this fact, the consequence in the decision analysis of introducing a reduction of the variance in the streamflow data is investigated. In order to make a reasonable adjustment, a comparison between sample standard deviations for the period 1910-39 and 1940-69 is carried out as shown in Table 4.8. When the Pecos River and the Rio Grande data are compared, it seems that the Alamogordo Reservoir causes roughly a 15 percent reduction in the standard deviation of the annual discharge series. According to this difference in the percentage decline, the standard deviation was lowered for the period before the reservoir was put into operation, as shown in the last column in Table 4.8. Two decision calculations using

Table 4.8 Comparison of Standard Deviations for Different Time Periods. Rio Grande and Pecos River Water Discharge Data Used.

s in Acre-feet	Rio Grande Data	Pecos River Data	Adjusted std. dev. Pecos Riv.
1910-39	$5.27 \cdot 10^5$	$1.18 \cdot 10^5$	$1.00 \cdot 10^5$
1940-69	$5.17 \cdot 10^5$	$0.98 \cdot 10^5$	$0.98 \cdot 10^5$
% Decline	2%	17%	2%

the long-term Pecos River streamflow sample through the regression analysis are presented: the first without any adjustment of recorded data, the second with the reduction incorporated. Results are given in Table 4.9. It should be emphasized that the standard deviation is reduced so that the sample mean remains constant.

Table 4.9 Design Alternatives and EOL for Varying Values of the Sample Variance. Pecos River at Artesia. 20 Years of Sediment Data Plus 40 Years of Streamflow Data

	Design Alternative in Acre-Feet	EOL in $
Data as recorded	42,100	403,922
Adjustment made	41,200	406,184

Table 4.9 shows how little the decision process is affected by a nonstationarity in the variance of the long-term water discharge series. Therefore, no adjustment of this sample statistic is considered necessary in the present analysis.

With the above mentioned type of adjustment considerations of man-made and other changes in the hydrologic system, the regression technique can provide us with valuable information about the short sediment load series in the framework of statistical decision theory.

In the following, identical computer runs are presented for both the Rio Grande and the Pecos River. Annual water discharge data (secondary data) are utilized to extend the 20-year long sediment load sample with 10, 25, and finally 40 extra data points. The decision analysis is then performed with these augmented primary data samples. In order to compare the results, it was necessary to find the expected EOL curve for the Pecos River data by using primary data, as already done for the Rio Grande in Section 4.3.2.

The cross-correlation coefficient between the logarithm of the 20 years of concurrent sediment load and discharge data was estimated to be:

Rio Grande at Otowi Bridge: $\hat{\rho} = 0.62$ (4.28)

Pecos River at Artesia: $\hat{\rho} = 0.92$ (4.29)

These two values of the correlation coefficient more or less give the lower and upper limits for the range in which the value of ρ normally is found. Nordin and Sabol (1973) report in their investigation of 24 rivers of the United States that the correlation coefficients are found in the range between 0.56 and 0.96, with an average value of 0.79.

In the framework of Fisher's well-known information concept, criteria for information transfer via regression were found in the early 1960's. Criteria for obtaining more reliable estimates of the mean as well as the variance were reported with the value of the cross-correlation coefficient being the determining parameter. Matalas and Langbein (1962) found that the condition for the cross-correlation to provide additional information of the mean is $\rho^2 > \frac{1}{N-2}$ for a random series. In this case we have 20 concurrent data (i.e. N = 20) and the inequality implies, $\rho > 0.24$. Matalas and Jacobs (1964, Table 2) give the critical minimum values of ρ for various values of sample lengths in order to obtain an improved estimator of the variance. In the present study, where the noise term is included in Eq. 3.21, the condition $\rho > 0.52$ has to be satisfied. Therefore, additional information about both the mean and variance of the primary series can be extracted form the secondary data via cross-correlation according to these tests.

Everytime cross-correlation is applied, the investigator has to see whether spurious correlation is present or not. If the scatter graph shows a clustering of points in two or more groups, a seemingly large correlation may result, due merely to the heterogeneity in data. Such correlations are called spurious. The scattered plot in Fig. 4.13 of sediment load versus water discharge data in the Rio Grande case does not indicate any marked spurious correlation in the data.

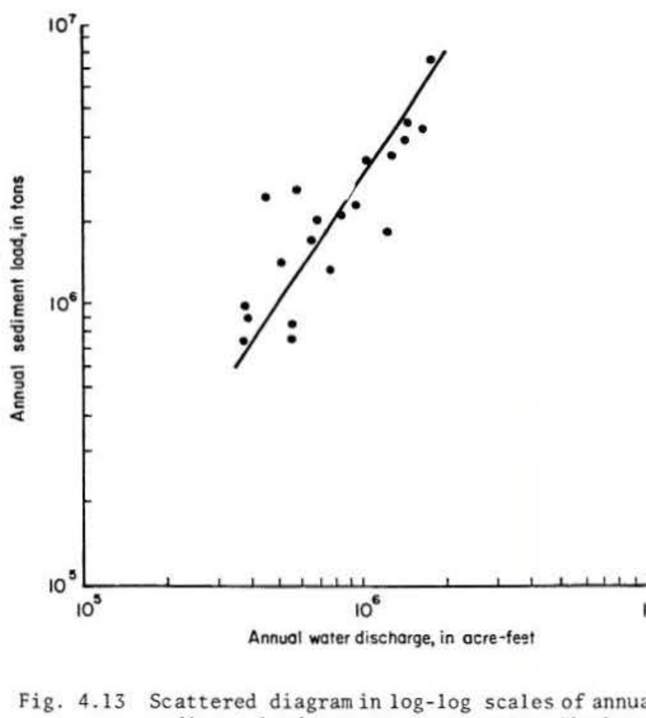

Fig. 4.13 Scattered diagram in log-log scales of annual sediment load versus annual water discharge for the Rio Grande data.

The results of the above mentioned computer runs are presented in the following text, with Table 4.10 and Fig. 4.14 related to the Rio Grande data, and Table 4.11 and Fig. 4.15 pertaining to the Pecos River data. All designs are for a 50-year reservoir lifetime.

Table 4.10 Primary Data and Regression Results. Rio Grande, New Mexico

	No. of Years in Sample	Design Alternative Acre-feet	EOL 10^6 \$
Primary data observed sample	20	90,000	2.69
Primary data generated values	30 45 60	90,000 90,000 90,000	1.62 1.14 0.93
Secondary data via regression $\hat{\rho}$ = 0.62	30 45 60	100,000 108,000 120,000	2.20 1.51 1.34

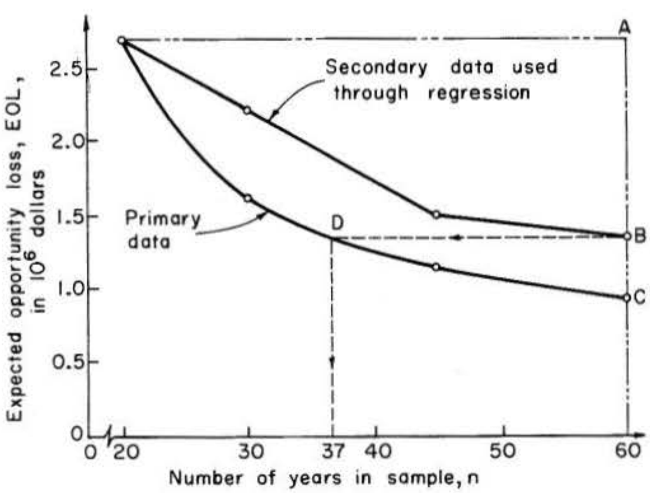

Fig. 4.14 EOL values using primary data and secondary data via regression. Rio Grande, $\hat{\rho}$ = 0.62.

Table 4.11 Primary Data and Regression Results. Pecos River, New Mexico

	No. of Years in Sample	Design Alternative Acre-feet	EOL 10^5 \$
Primary data observed sample	20	25,000	5.46
Primary data generated values	30 45 60	25,000 25,000 25,000	3.95 2.74 2.18
Secondary data via regression $\hat{\rho}$ = 0.92	30 45 60	31,000 34,000 37,000	5.25 3.71 2.78

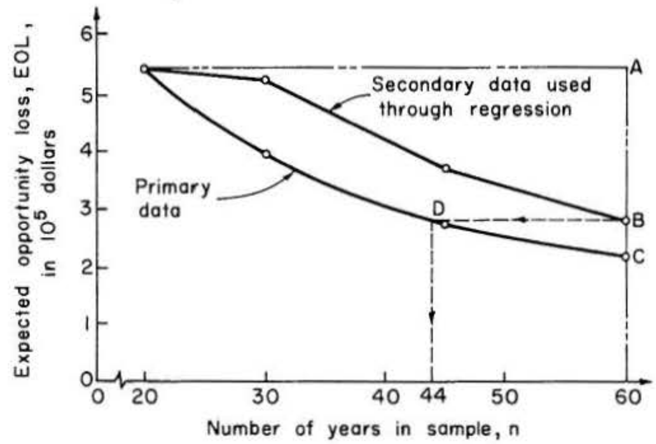

Fig. 4.15 EOL values using primary data and secondary data via regression. Pecos River $\hat{\rho}$ = 0.92

Tables 4.10 and 4.11 clearly show how the new information affects the decision process. The design alternative in the Rio Grande case has increased from 90,000 acre-feet to 120,000 acre-feet, approximately a 33 percent increase, in the Pecos River case the increase is nearly 50 percent. What causes this marked difference in the change of design? After all, the percentages of nonhomogeneities in water discharge series are the same. The difference is due to the fact that the Pecos River data set has a much higher cross-correlation coefficient between runoff and sediment load than the Rio Grande data. A high correlation coefficient further indicates a high regression coefficient because the relationship between the two coefficients is $\hat{\beta} = \hat{\rho} \cdot \frac{s_x}{s_y}$. This statement is correct when the ratio s_x over s_y is considered constant; s_x and s_y are the standard deviations for the independent and dependent variables, respectively. Recalling Eqs. 3.20 and 3.21, it is seen that the higher the regression coefficient $\hat{\beta}$, the more weight is put on nonstationarity in the secondary data series. This influences the mean of the augmented sediment load sample, and finally the chosen storage design.

Figures 4.14 and 4.15 show how a decrease in EOL is realized when more and more water discharge data are used to extend the given 20 years sample of sediment load. However, it is also noted that the decrease is not nearly as pronounced as it is for the curve based on the primary data alone. The secondary data curves are drawn as straight lines between the computed EOL points contrary to the smooth primary data curve. This is to indicate that the points found to be the means of secondary data can be considered one set of sample points of a "population" curve, while the primary data points estimate the "population" curve directly, because they are found by using an expected EOL procedure.

The reason for the higher EOL values using a secondary data set results from uncertainty introduced in the decision process as part of the regression model represented by the noise term described in Section 3.3. The EOL curves can now be employed to obtain a measure of how much this prediction error counts as a function of the correlation coefficient. According to the notation used in Figs. 4.14 and 4.15, the following can be stated. Extending the 20 years of observed sediment load data with 40 years of the same type of data results in an expected EOL decrease of size A-C. Using 40 years of water discharge data through regression yields a decrease of magnitude A-B. Because of the uncertainty associated with the regressed values they do not provide as much information as the same number of primary data points. A horizontal line through B intersects the primary EOL curve at a point D, where the same decrease in EOL has been realized using either of the additional data. In the Rio Grande case (Fig. 4.14), point D corresponds to $n = 37$ years.

What has been achieved through this analysis is a definition of the equivalent length of a secondary set of data when used for regression in the framework of statistical decision theory. For the Rio Grande case with $\hat{\rho} = 0.62$ the result is: 40 years of water discharge data are equivalent to 17 years of sediment load data. One obtains these values by subtracting the number of concurrent years (in this study 20) from the numbers obtained in Fig. 4.14. The same type of analysis (Fig. 4.15) for the Pecos River with $\hat{\rho} = 0.92$ leads to the result: 40 years of water discharge data are equivalent to 24 years of sediment load data. Because the analytical expression for the EOL function in the Rio Grande case is defined (see Eq. 4.14), the equivalent record length can also be found by solving the equation,

$$\frac{50 \cdot 10^6}{n} = 1.34 \cdot 10^6, \qquad (4.30)$$

where the right side of this equation is taken from Table 4.10. Equation 4.30 yields the solution $n = 37.3$, a result which is in agreement with the graphical solution in Fig. 4.14.

Expected opportunity loss (EOL) is the average loss to be anticipated form incorrect design and is comparable in principle to the variance of an estimate in classical statistics as used in Fisher's information concept. As reported by M. Roche (1963, p. 48), R. Véron was one of the first to use the Fisherian concept in order to define equivalent length of record when secondary data are used to augment a primary set of data.

R. Véron gives the ratio, R, of the variances for the mean of the given sample ($s^2_{\bar{x}_n}$) and for the mean of the lengthened sample after augmentation ($s^2_{\bar{x}_N}$) as,

$$R = \frac{s^2_{\bar{x}_N}}{s^2_{\bar{x}_n}} = 1 + (1 - \frac{n}{N}) \cdot [\frac{1 - (n-2) \cdot \hat{\rho}}{n-3}]. \qquad (4.31)$$

The two variances can be expressed by means of the variance for the individual members of the series, σ^2_x, in the following way, which yield a formula for the equivalent record length N_e:

$$\frac{\sigma^2_x}{N_e} = R \cdot \frac{\sigma^2_x}{n} \Rightarrow N_e = \frac{n}{R}. \qquad (4.32)$$

This derivation is correct under the assumption of time independence in the series. Numerical examples:

(i) Rio Grande case with $n = 20$, $N = 60$, $\hat{\rho} = 0.62$ $\Rightarrow R = 0.60$. Equation 4.32 yields the equivalent length, $N_e = 33$ years.

(ii) Pecos River case with $n = 20$, $N = 60$, $\hat{\rho} = 0.92 \Rightarrow R = 0.38$. Equation 4.32 yields the equivalent length, $N_e = 52$ years.

Subtracting the 20 years of concurrent data from the two N_e values provide us with comparable figures between equivalent length using the EOL concept and Véron's variance method.

Table 4.12 sums up the results from the two ways of achieving a measure of equivalent length of record.

Table 4.12 Results Using Different Methods to Find Equivalent Record Length

	EOL Analysis		Véron's Variance Concept	
	$\hat{\rho} = 0.62$	$\hat{\rho} = 0.92$	$\hat{\rho} = 0.62$	$\hat{\rho} = 0.92$
Length of secondary data set, in years	40	40	40	40
Length of equivalent primary data set, years	17	24	13	32

The EOL method results in a considerable smaller variation of the equivalent length as a function of the correlation coefficient compared to the classical Fisherian concept. Many studies in the past of correlation, information content, effective length, etc., by Matalas and Langbein (1962) and Fiering (1963) and others, all use the same type of concepts as Véron introduced. The numbers presented in Table 4.12 indicate that the strength of the correlative dependence represented by the value of $\hat{\rho}$ is not as important as previous studies might suggest. However, it should be kept in mind in connection with the results reported in Table 4.12 that there is a basic difference between the two approaches for finding the equivalent record length. While Véron's variance concept operates solely in the "log space" of the data series, the EOL method uses an "economic space," i.e. as part of the latter method, the nontransformed form of the series is used. Nevertheless, for the design-engineer, the decrease of EOL as a measure of gained information makes sense, because worth of data can be assessed only when the purpose for which the data are to be used is defined. The statistical decision (EOL) approach for defining the equivalent length involves the incorporation of many different factors, such as, (1) length of primary and secondary samples, (2) cross-correlation coefficient, (3) hydrologic characteristics of the basin, and (4) economic factors. Because only factors (1) and (2) are used to define the equivalent length in the Véron's variance approach, it seems that the engineering related factors (3) and (4) tend to "smooth out" the effect caused by the statistical factors (1) and (2). That further indicates a dependence of the equivalent length on the particular type of project under investigation.

The above results can be interpreted in another way. The EOL method shows that for $\hat{\rho} = 0.62$, the 40 years of secondary data provide the decision process with as much information as 17 years of primary data. It can be asked, which value of $\hat{\rho}$ is required by using the Véron's method to obtain the same amount of information transfer. By setting $N_e = 20 + 17 = 37$ years in Eq. 4.32 and substituting for R in Eq. 4.31, this equation is solved for $\hat{\rho}$ with $n = 20$ and $N = 60$. The result is $\hat{\rho} = 0.71$. It is seen that the requirement on the value of $\hat{\rho}$ is stricter in the Fisherian concept than in the EOL concept.

It can be concluded from this analysis that the value of the cross-correlation coefficient when the regression analysis is used in a statistical decision process is not a critical factor concerning the amount of information transfer as defined by the equivalent length of record. Important, however, for the choice of design alternative is the question of nonstationarity in the long-term record used for augmentation purpose, as recalled from the adjustment considerations earlier in this Chapter. Questions like man-made versus natural changes in the hydrologic regime, permanent versus temporary trends and jumps in time series, etc., have to be considered carefully by the design-engineer before a secondary set of data is used in a regression procedure to increase information in comparison with the information contained in the primary data. The design is dependent on the value of cross-correlation coefficient and the amount and type of nonstationarity.

4.4 Results of Case Study

A recapitulation of investigations and results reported in Chapter IV is given in this section. The Rio Grande basin in the southwest United States is one of the most distinct sediment problem areas in the world. The study of data worth is carried out by using observed samples from two locations in the basin, with the watersheds upstream of the two gaging sites possessing similar natural features. Furthermore, streamflow data and sediment load data at the two stations have been gaged over the same time period.

The form of the goal function is a penalty function, which indicates the loss of money the designer might expect because of a realized overdesign or underdesign. The decision variable is the sediment storage part of a reservoir, intended to trap the transported solid material in the Rio Grande.

Section 4.3.1 shows an evident fact, namely that design sizes generally have a marked variability when extremely short samples are used to make decisions about design alternatives. The amount of uncertainty inherent in the decision process, represented by an EOL figure, reduces sharply in the case a five-year sample is compared to a 10-year sample. The decrease in EOL gets smaller and smaller with an increase of the length of record used. An explicit functional expression for the EOL curve is found in Section 4.3.2 (in this particular case a hyperbola). The worth of one extra point of data added to a given sample already consisting of 50 data, is only one percent of the worth when added to a five-year record. However, the EOL curve continuously decreases which is the same as saying that information is gained on the unknown parameters for every new additional data point. This very often implies that a fundamental difference of interest exists between the data gatherer and the engineer. While the data expert wishes to improve the reliability and accuracy of his product, which will require more years of data collection, the engineer generally wants to or has to get on with his work.

In order to satisfy both interest groups, the concept of expected economic optimal record length is introduced. It is found as the point in time where the marginal cost of data (cost of sampling plus cost of project postponement) is equal to the marginal benefits provided by additional data. For annual sediment load data in connection with the allocation of storage for sediment deposition in a reservoir, it is shown that the expected economic optimal sample has a length of 12 years, under the condition that five years of data already exist. That is, a project delay of seven years is recommended so that the remaining data can be obtained. It should be kept in mind that this record length is what the decision-maker anticipates to be economically optimal, given the information already available in the observed five-year sample. This implies further that there is no guarantee for the additional data collection to change the recommended economic optimal record length. The result of such an analysis is reasonably conditioned, on the available information at the time of decision. Also the sensitivity to the selected lifetime of the reservoir is investigated. The result indicates that the expected optimal record length only increases from 12 to 15 years while the physical lifetime increases from 50 to 100 years.

A regression model is included in the statistical decision analysis in Section 4.3.3. The purpose was to extract information from a long-term streamflow series and to transfer it to the short sediment load sample. Investigation of nonstationarity (type and amount) in the long secondary data sample is made in order to perform a reasonable adjustment of the data. In this case a 15 percent decrease in the mean of the first half of water discharge data in the Pecos River was justified, due to the completion of Alamogordo irrigation reservoir in the late 1930's. The consequences in the decision analysis of different values of cross-correlation coefficient, ρ, are studied. This led to a definition of the equivalent length of the secondary set of data. It is found that the EOL method results in a considerable smaller variation of equivalent length as a function of ρ compared to figures obtained (Table 4.12) when the classical Fisherian definition is used.

Up to this point the economic factors have been treated as constant. Therefore, in the following Chapter the statistical decision analysis is extended by incorporating the uncertainty in economical parameters.

CHAPTER V
EFFECTS OF ECONOMIC UNCERTAINTY

The purpose of this Chapter is to illustrate the need for incorporating economic uncertainty in the decision process of hydrologic related designs. The expected opportunity loss concept is used. It is shown how the mathematical considerations might be an aid in the development of "subjective" probability density functions of economic parameters.

5.1 General Remarks

Projects have many associated uncertainties. Such uncertainties are: construction may be more or less difficult than projected, the values of inputs and outputs may differ from expectations, just to name a few. Some risks result from engineering uncertainties, others from the stochastic nature of hydrologic variables, and many from economic and political factors on which the project success or failure so much depend.

In previous chapters only uncertainties in hydrologic parameters have been investigated, while other factors affecting the decision process have been regarded as constants. Sensitivity analysis, such as Moss (1970) and Davis et al. (1972a), indicates that the lack of perfect knowledge of economic parameters of a project may affect significantly the decision reached. This result is in good agreement with the findings by James, Bower, and Matalas (1969), who demonstrate that the relative importance of planning variables in their descending order of effect is: (1) the economic variable, (2) the political variable, and (3) the hydrologic variable. Because it is recognized that uncertainty in economic factors plays a significant role, the next step is to find the consequence of this uncertainty in the decision process and to compare it with the effects of uncertainties in hydrologic variables.

The expected opportunity loss (EOL) concept as outlined in Chapter II can be used to cover uncertainties in the economic as well as the hydrologic parameters. As demonstrated in the chapter treating the goal function, the economic parameters involved in the cost function, Cost $(\overline{Q}_s, Q_s^{alt})$, are K and r. The parameter K is an estimate of a unit cost figure (dollars per acre-feet), and might, as indicated by Kazonowski (1972), vary by 10 to 50 percent of the estimated mean within the United States. The parameter K is expressed as a current value; however, it is also used as a cost factor for future works (removal of sediment), which obviously increases the uncertainty in the evaluation of K. This additional uncertainty is caused by the fact that the value of money changes from one time period to another, either by inflation or deflation; the first is more common than the latter. Another reason is possible technological advances, which in the future might decrease the expected expenditure.

The other fiscal parameter introduced in the cost function is the rate, r, which appears in the discount factor $1/(1+r)^N$ The selection of particular value for this rate is a controversial subject among economists, who have even created different "schools of thought," see for example James and Lee (1971, pp. 126-131). Furthermore, the discount rate cannot be expected to stay constant in the coming years, especially when dealing with reservoirs of designed project lives between 50 and 100 years.

The most widely used approaches in economic analysis in treating uncertainty of this kind include: (1) applying preselected percentages to increase or reduce costs and benefits, (2) limiting the period of analysis (the time horizon), or (3) adding a risk increment to the discount rate, so the discount factor becomes

$$\frac{1}{(1+r+\Delta r)^N}, \tag{5.1}$$

where r represents the riskless discount rate and Δr represents an additional fraction to the discount rate (positive or negative), to account for what is likely to be the expected variability in the rate, r. However, these approaches require the use of a numerical factor, and the estimation of such a specific value often turns out to be more or less an arbitrary choice. A more satisfactory approach, instead of using single fixed values of the economic parameters, is to represent the potential values by probability distributions to be introduced in a probabilistic decision making process.

The concept of expected cost (and benefit) in economic analysis is used, but it should be emphasized that most often the expectations are found by means of probability functions describing the occurrence of physical events, say floods greater than a certain size. One of the better known earlier studies, which treats probability functions connected to economic measures, was done by Altouney (1963). Using past data of construction costs for more than one hundred water resources projects and population growth records, probability density functions for construction cost and benefit were developed. To accomplish the latter, Altouney assumed benefits to be proportional to the size of population. However, he did not show the use of these density functions, as should have been done in an expectation calculation. He makes an important conclusion, namely that the longer the project lifetime and the lower the discount rate, r, used for planning purposes, the larger should be the minimum acceptable estimated benefit-cost ratio.

Considering a benefit-cost ratio as a variable implies a very "bulky" economic parameter to work with. In the present study the uncertainty analysis is carried out directly with the factors which convey nonperfect knowledge into the present value of future costs (or benefits).

The economic uncertainty is taken into account and quantified by assigning μ and σ^2 (the hydrologic variables), and K and r (economic variables) as the state variables. The decision variable is the same as previously, the design size of the sediment storage part of a reservoir. Using the same notation as in Chapter II then the Bayesian Risk becomes,

$$R(Q_s^*) = \min_{Q_s^{alt}} \iiiint G(Q_s^{alt} | \mu, \sigma^2, K, r) \cdot f(\mu, \sigma^2) \cdot f(K) \cdot f(r) \cdot d\mu \cdot d\sigma^2 \cdot dK \cdot dr \tag{5.2}$$

with Q_s^* the alternative chosen as the Bayesian solution. The "loss" in monetary terms of not knowing both the exact fiscal figures and hydrologic parameters in the cost function is

$$EOL = \iiiint \{G(Q_s^* | \mu,\sigma^2,K,r) - G(Q_s' | \mu,\sigma^2,K,r)\} \cdot f(\mu,\sigma^2) \cdot f(K) \cdot f(r) \cdot d\mu \cdot d\sigma^2 \cdot dK \cdot dr. \tag{5.3}$$

$f(\mu,\sigma^2)$, $f(K)$, and $f(r)$ are the probability density functions of the mean and variance of the sediment load (in log transformed form), the unit cost factor K, and the discount rate r, respectively.

The problem now focuses on the selection of these probability functions which characterize the stochastic properties of the parameters. Probability functions can be determined either theoretically or experimentally. The theoretical approach uses the mathematical statistics in deriving the density functions given a certain state of nature. The joint probability function - discussed in Chapter II - for the mean and variance of the annual sediment load belongs to that category. The experimentally determined distributions can be subdivided into two groups. It is often the case that useful information about an uncertain parameter is contained in samples of past data, in the way that a "solid" frequency distribution of past parameter values is available. It would then be reasonable to assign a probability distribution which as closely as possible match the frequency distribution of the actual historical values. There exists a substantial amount of objectivity in such a selection of density function which in the literature often is called a "data-based" distribution. This method was used by Altouney (1963) in his derivation of the benefit-cost ratio density function as mentioned earlier.

In other cases the prior information may arise from sources other than currently available samples of past data - "nondata-based" distributions. This is a debatable and crucial point in the Bayesian decision theory, which often is a subject of relevant criticism from opponents of such type of probabilistic analysis. In these situations the distribution represents simply the investigator's personal view and belief, which also reflects his work experience and knowledge of the subject in question. Needless to say, one person's "nondata-based" distribution can greatly differ from that of another. Because of psychological difficulties involved in this assessment, it is usually recommended that the decision-maker does not specify his subjectively derived distribution in more details than by a few summary measures such as the mean, standard deviation, a few percentiles, or just the shape. For a thorough treatment of the controversy about objectively versus subjectively based probability functions a book by L. J. Savage (1972) is recommended, especially its Chapter 4. In the following determination of probability distributions $f(K)$ and $f(r)$, it is shown how the mathematical tools might be used in the process of estimating a distribution subjectively.

5.2 The Unit Cost Density Function, f(K)

As mentioned earlier K is a total unit cost figure which can be considered composed of subactivity costs like (1) equipment, (2) manpower, and (3) disposal sites; all expressed in dollars per acre-feet of removed sediment. Moder and Phillips (1970) describe the PERT statistical approach which has been widely used in managerial engineering to determine time schedules of technically oriented programs. However, the employment of that technique to find probability distributions of cost figures, which can be broken down into subcosts, has not been recognized among water-resource engineers and planners.

In this method, three estimates are made of each of the subactivity costs involved. A natural choice is to choose a most likely cost, M_1, and a range from a low estimate, A, to a high estimate, B, as shown in Fig. 5.1.

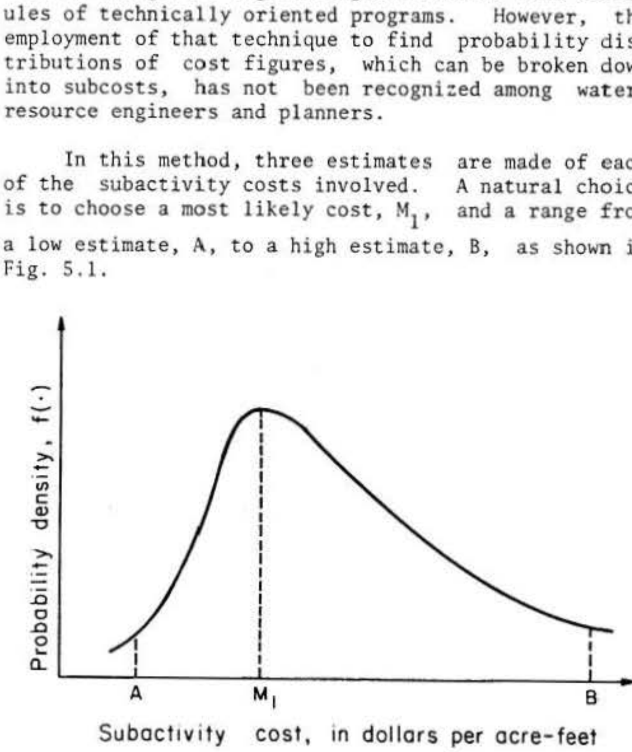

Fig. 5.1 Choice of cost estimates:
A = low cost estimate,
B = high cost estimate,
M_1 = most likely value.

If the limits A and B are assumed to be the 5 and 95 percentiles of the distribution, then it is common that the difference, (B-A), often varies from around 3.1 to approximately 3.3 (average 3.2) of the standard deviation. This is true for a wide variety of distributions, ranging from the exponential distribution to the normal distribution, including rectangular, triangular, and beta-type functions. The estimator of the standard deviation is robust to variations in the shape of the distribution of the cost. Therefore, without knowing the exact form of the distribution the standard deviation can roughly be estimated as

$$\hat{s}_i = \frac{(B-A)}{3.2}, \tag{5.4}$$

with subscript i indicating one of the subcosts considered.

A simple formula for estimating the mean, $\overline{C}_i$, has also been suggested (Moder and Phillips, 1970) as being the weighted average of the mode, M_1, and the midrange, $\frac{A+B}{2}$:

$$\overline{C}_i = \frac{1}{3}(2M_1 + \frac{A+B}{2}) = \frac{1}{6}(A + 4M_1 + B). \tag{5.5}$$

This formula for the mean is only valid with the assumption of some functional form for the unknown distribution, such as indicated in Fig. 5.1. It could for example be a gamma distribution, beta distribution or normal distribution; that is single peak distributions skewed or symmetrical.

Following this scheme, values for $\overline{C}_i$ and $\hat{s}_i$ can be obtained for each of the variable subcosts. Although it is realized that the well known Central Limit Theorem has its limitations, it is used under "very general conditions," as described by Benjamin and Cornell (1970, pp. 251-253), to presume that the total cost factor, K, follows a normal distribution. That means,

$$f(K) = N(\overline{K}, s_K^2). \tag{5.6}$$

If the total cost consists of n subcosts, the distribution characteristics can be found as,

$$\overline{K} = \overline{C}_1 + \overline{C}_2 + \overline{C}_3 + \cdots + \overline{C}_n \tag{5.7}$$

$$\hat{s}_K = \sqrt{\hat{s}_1^2 + \hat{s}_2^2 + \hat{s}_3^2 + \cdots + \hat{s}_n^2}, \tag{5.8}$$

Equation 5.8 is valid only when the subcosts are statistically independent, i.e., if they can be considered as uncorrelated variables. This is an assumption which might appear too strict because economic factors generally are dependent. This is a point the investigator has to keep in mind when such type of subcost method is used. However, in this particular case, the assumption is considered adoptable: prices of equipment for sediment removal, manpower, and disposal sites will find their own level without reference to each other, so that the assumption of independence has been met and the use of Eq. 5.8 can be justified.

By the described procedure, a probability distribution of K is obtained reflecting a quantitative measure of the cost uncertainty. In the case of sediment removal from a reservoir it was mentioned above that it would be relevant to consider only three subcosts, namely expenses for equipment and machinery, labor, and compensation for sites to be used for disposal of the removed sediment. The Task Committee for Preparation of Manual on Sedimentation (1969, Ch. IV) reports a study by Ferrell and Barr (1965), who give an estimation of the cost of excavating sediment and debris from reservoirs in the southern California region. They found a price of approximately \$1.00 per cu. yd. (\$1630 per acre-feet), which includes the disposal site costs. Using the Task Committee's manual it is possible to deduce illustrative figures for the subcosts. Those estimates are presented in Table 5.1 together with the probability distribution parameters calculated by Eqs. 5.4 and 5.5.

By applying Eqs. 5.7 and 5.8 with n = 3 we then get:

$$\overline{K} = 800 + 600 + 300 = 1700 \text{ \$ per acre-feet}, \tag{5.9}$$

and

$$s_K = \sqrt{125^2 + 375^2 + 250^2} = 470 \text{ \$ per acre-feet}. \tag{5.10}$$

Thus, one form of an uncertainty distribution for the total cost factor is then symbolized by

$$f(K) = N(1700, 470^2). \tag{5.11}$$

Table 5.1 Subcost Estimates for Sediment Removal. All Figures are in Dollars per Acre-feet. The Rio Grande Case Study

		Subcost Estimates		
5% - 95% Range:		Equipment (1)	Manpower (2)	Disposal Site (3)
Low Cost	A =	600	200	100
Most Likely	M_1 =	800	500	200
High Cost	B =	1000	1400	900
Distribution Parameters:		(1)	(2)	(3)
Mean	$\overline{C}_i$	800	600	300
Standard Dev.	$\hat{s}_i$	125	375	250

5.3 The Discount Rate Density Function, f(r)

The probability distribution for the interest rate also belongs to the "nondata-based" group. The possible values of future interest rates are strongly related to a nation's general financial and economic situation, which further is very dependent on political decisions concerning price levels, wages, and so forth. According to the fluctuations of interest rate in the past it seems reasonable to expect the future rates to be contained inside certain limits say two percent and ten percent when public (either state or federal) discount factors are considered.

Because of reasons stated above, the value of the future rates are difficult to predict. Therefore, the situation is that a distribution representing the state of "knowing little" is desired, under the assumption that the rates can be in a certain range from a to b. It seems agreeable to depict such vague or diffuse information about a parameter by taking f(r) constant, which implies,

$$f(r) = \frac{1}{b-a}, \text{ for } a \leq r \leq b. \tag{5.12}$$

The question about whether the choice of a rectangular function really represents "very little" information about the value of a parameter is discussed in Appendix B, where it is found that the rectangular probability function indeed is a so-called "minimal information" density function.

Because sometimes the investigator knows more about a particular parameter than "minimal information," he must weigh the chances of getting the values in the middle of the range to be higher than at the lower and upper portions of the range. A suitable model which covers all cases mentioned above is the beta distribution. It has the desirable properties of being contained inside a finite interval (a to b), and can be either rectangular, symmetrical or skewed. The beta distribution is therefore selected for the variable discount rate r as

$$f(r) = \frac{(l-1)!}{(k-1)! \cdot (l-k-1)!} \cdot \frac{1}{(b-a)^{l-1}} \cdot (r-a)^{k-1} \cdot (b-r)^{l-k-1}, \tag{5.13}$$

where $a \leq r \leq b$, and l and k are distribution parameters which have to be fixed according to the shape wanted. The most common cases the investigator will meet are,

(i) rectangular, $l=2$ and $k=1$,

$$f(r) = \frac{1}{b-a}, \qquad (5.14)$$

(ii) symmetrical with "flat" peak, $l=4$ and $k=2$,

$$f(r) = \frac{6}{(b-a)^3} \cdot (r-a) \cdot (b-r), \qquad (5.15)$$

(iii) symmetrical with "marked" peak (high kurtosis), $l=12$ and $k=6$,

$$f(r) = \frac{11}{5! \cdot 5!} \cdot \frac{1}{(b-a)^{11}} \cdot (r-a)^5 \cdot (b-r)^5, \qquad (5.16)$$

(iv) skewed right, $l=6$ and $k=2$,

$$f(r) = \frac{20}{(b-a)^5} \cdot (r-a) \cdot (b-r)^5. \qquad (5.17)$$

As the beta distribution is flexible and adjustable, a sensitivity analysis on the choice of the underlying distribution can be carried out. This analysis goes from a "complete" uncertainty in the discount rate over "some" information ("flat" peak case), and end up with essentially no uncertainty which means a rectangular distribution is used with an extremely narrow range for the value of r. The degree of sensitivity in the distribution selection is measured as the change in the expected opportunity loss (EOL).

5.4 Applications

With the probability density functions selected in the previous two sections, the statistical decision analysis can be pursued. The computer program was extended to cope with the required number of multiple integrations over both the hydrologic and economic parameters. A maximum of four uncertain parameters is considered in this investigation.

Table 5.2 presents results of the economic uncertainty analysis, carried out on the Rio Grande data. The hydrologic parameters, mean μ and variance σ^2, follow the data-based normal chi-square distribution of Eq. 2.1, where the sample statistics are determined from the observed 20 years of sediment load data. As described in Sections 5.2 and 5.3, the uncertainty distributions for the unit cost factor and discount rate are, respectively,

$$f(K) = \frac{1}{470\sqrt{2\pi}} \exp\left[-\frac{(K-1700)^2}{2 \cdot 470^2}\right] \qquad (5.18)$$

and

$$f(r) = \frac{1}{0.10-0.02}, \quad \text{for } 2\% < r < 10\% \qquad (5.19)$$

As Table 5.2 indicates, the uncertainty analysis is done with a varying number of uncertainty factors in the decision process. The different cases are arranged according to the increasing values of the EOL figures obtained. In cases where one or both of the hydrologic parameters are considered certain (constants), they are assumed equal to the sample mean and variance, respectively, computed from the longest available data series.

Table 5.2 Economic Uncertainty Analysis. The Rio Grande Data

Case	Uncertainty Parameters		Design Alternative Acre-Feet	EOL 10^6 \$
	Hydrologic	Economic		
1	----	r,K	91,000	0.31
2	- σ^2	---	89,000	0.90
3	- σ^2	r,K	93,000	1.36
4	μ -	---	90,000	1.72
5	μ -	r,K	92,000	2.22
6	μ,σ^2	---	90,000	2.69
7	μ,σ^2	- K	93,000	2.91
8	μ,σ^2	r -	98,000	3.42
9	μ,σ^2	r,K	99,000	3.50

The expected opportunity loss expresses the cost the designer may anticipate because of either an overdesign or underdesign. Therefore, it is a measure of the uncertainty inherent in the decision process in monetary terms in connection with a particular design alternative. The last column in Table 5.2 shows how the different types of parameters introduce uncertainty in relation to each other. For instance, case 1 and case 2 indicate that it is more important to have perfect knowledge about the variance of the sediment load series than about the two economic parameters. The consideration of the EOL values leads to the following ranking of relative importance of design variables (in descending order): (1) mean of annual sediment load, μ, (2) variance of sediment load series, σ^2, (3) discount rate, r, and (4) unit cost factor, K. This is an interesting result in the sense that it has reversed the order in the relative importance of parameters compared with the findings reported in Section 5.1. This result might be due to the fact that the study by James, Bower, and Matalas was carried out in the planning stage of a water resource project, while the present analysis is concerned with the design phase of a proposed project.

Furthermore, Table 5.2 shows a strong dependence of EOL on the type and number of uncertain parameters involved in the analysis. As expected, incorporating all four parameters (case 9) results in the highest EOL value, namely $3.50 \cdot 10^6$\$. This figure is more than 25 percent higher than the value shown in case 6, which is taken from the analysis presented in Chapter IV. This indicates further that taking the economic uncertainties into account affect the result concerning the optimum record length, as found in Section 4.3.2.

It is also interesting to note, as Table 5.2 shows, that the expected opportunity loss is superadditive. If the terminology EOL(p) is used, which means that this particular EOL value is found by treating p parameter(s) as uncertain, it is seen that,

$$EOL(\mu) + EOL(\sigma^2) + EOL(r,K) < EOL(\mu,\sigma^2) + EOL(r,K)$$
$$< EOL(\mu,\sigma^2,r,K) \quad .$$

Such superadditivity indicates an interaction between all four parameters as they appear interrelated in the goal function. Unfortunately, this property makes it impossible for the investigator to find the EOL values "piecewise," which otherwise could have resulted in considerable savings in computational efforts and computer time.

In order to find the sensitivity in the choice of the shape of probability distributions, another set of runs was performed with three types of the beta distribution function of the discount rate r:

(a) "complete" uncertainty - wide rectangular distribution,

$$f(r) = \frac{1}{0.10-0.02}, \text{ for } 2\% < r < 10\%, \qquad (5.20)$$

(b) "some" uncertainty-symmetric peaked distribution,

$$f(r) = \frac{6}{(0.08-0.04)^3} (r-0.04)\ (0.08-r), \text{ for } 4\% < r < 8\%, \qquad (5.21)$$

and

(c) nearly no uncertainty - narrow rectangular distribution,

$$f(r) = \frac{1}{0.0650-0.0550}, \text{ for } 5\tfrac{1}{2}\% < r < 6\tfrac{1}{2}\%. \qquad (5.22)$$

It should be mentioned that the unit cost factor K is kept constant in these runs. This means that the run (a) corresponds to case 8 of Table 5.2. Table 5.3 gives the results of this analysis.

Table 5.3 Sensitivity in the Shapes of Probability Distribution Functions for the Discount Rate

Run	Description of r	Design Alternative Acre-Feet	EOL 10^6 \$
(a)	"complete" uncertainty	98,000	3.42
(b)	"some" uncertainty	93,000	2.98
(c)	"little" uncertainty	91,000	2.81

As expected, we find these EOL values inside the limits determined by case 6 and 8 in Table 5.2, with run (c) - "little" uncertainty - nearly matching the case where both economic parameters are fixed.

Although the degree of information on economic parameters influences the uncertainty analysis in the decision process, the decision reached is rather insensitive. However, as seen in Table 5.3, a slightly increasing trend in the design alternative occurs as knowledge about the discount rate grows smaller. This is explained by the fact that the economic uncertainty tend to favor an overdesign, because a certain overdesign is less expensive than the same size of underdesign.

All the calculations reported in this Section are characterized by a common mean discount rate of six percent, the value used in all previous chapters. Table 5.2 shows that the discount rate plays a more significant role in the uncertainty analysis than the other economic factor, at least in this sediment storage case study. Consequently, a third series of computer runs was carried out to investigate the influence of a change in the mean discount rate, keeping the amount of uncertainty attached to that rate in accordance with Eq. 5.22, i.e., $r\% \pm \tfrac{1}{2}\%$. The runs are performed for both a 50-year and a 100-year design lifetime of the reservoir. The results are presented in Table 5.4.

Table 5.4 Sensitivity to Change in Mean Discount Rate

Design Lifetime	50 Years		100 Years	
Discount rate, r	$3\% \pm \tfrac{1}{2}\%$	$6\% \pm \tfrac{1}{2}\%$	$3\% \pm \tfrac{1}{2}\%$	$6\% \pm \tfrac{1}{2}\%$
Design alternatives in acre-feet	108,000	91,000	196,000	174,000
EOL in 10^6 \$	3.83	2.81	6.69	4.18

By considering the EOL values which represent the expected costs, it is seen that the future costs receive less weight with a higher discount rate and more weight with a lower. The relative difference in EOL is larger in the 100 years case compared to the shorter lifetime. The longer a proposed project lifetime, the more crucial is the selection of a proper discount rate in testing of economic feasibility. This is a manifest of the economical principle that high discount rates favor projects with little initial investment, while low discount rates favor capital intensive projects.

Perhaps of more interest for the design-engineer is the apparent jump in decision value with a change in the mean discount rate. In regard to the 50-year lifetime case, it is seen that the recommended sediment storage has increased from 91,000 acre-feet to 108,000 acre-feet, or nearly 20 percent, by decreasing the mean discount from a six percent rate to a three percent rate. Again, that is explained by the fact that future costs are more important when the discount rate is low, and therefore, an immediate cost at the time of construction in form of an overdesign is preferable.

To recapitulate, it can be stated, that lack of perfect information on economic parameters does have an important effect in a statistical decision analysis. For the type of goal function used in this study concerning design of sediment deposition storage, it is found that uncertainty in the selection of a value of the discount rate is more crucial than the uncertainty in the unit cost factor, like the price of removal of sediment per acre-feet. The choice of the shape of the probability density function (different types of beta distribution) and the choice of the mean of the discount rate result in marked changes in the uncertainty analysis represented by the EOL, which in turn, will effect the length of an economic optimum sample. Furthermore, the selection of the mean discount rate has a strong influence on the design decision reached.

Investigations presented in this Chapter point out the inseparable interrelationship between the parameters involved in a decision process. Economic uncertainty should be considered carefully on equal terms with uncertainty in the hydrologic parameters in order to achieve a realistic economic optimum design. The design alternative and worth of a particular data sample, as found by the statistical decision approach, show a high degree of dependence on both types of inadequate information.

CHAPTER VI
SUMMARY AND CONCLUSIONS

6.1 Summary

The design engineer is always faced with the problem of taking a particular action and making decision under uncertainty. This study uses a probabilistic method making it possible to choose a design alternative that minimizes the future costs of a project. The applied statistical decision approach produces an expected minimum cost decision, together with a measure in monetary terms of the value of the given data sample used for design purpose. This latter concept is introduced as the decrease in expected opportunity loss, (EOL).

Information can be increased in the decision process by incorporating more data in the sample, either (1) through the use of more existing data. (2) by a postponement of the project to collect additional data, or (3) by the use of a regression model with secondary data. Item (2) leads automatically to a definition of economic optimum record length. This is an important concept for the design engineer and data collection manager to keep in mind, since data cannot be treated as a "free resource."

In the process of finding the expected opportunity loss when future "unknown" data are incorporated in the observed sample, a special data generation technique was introduced. In brief, it synthesizes new realizations (traces, samples) for various randomly selected population parameters, with a weighted average procedure used to achieve an estimate of the expected value of the expected opportunity loss (EVEOL). Furthermore, definitions of the regression related term "equivalent length of secondary data," found in the framework of both the statistical decision theory and the classical Fisherian concept, are compared in this study.

Uncertainty in the decision process may stem from sources other than hydrology related uncertainties. The investigation also deals with the question of economic uncertainty. For example, consequences in the decision analysis of not knowing the exact value of a discount rate and/or a unit cost figure are found. For that purpose the expected opportunity loss concept is used. It is shown how mathematical considerations might be an aid in developing the subjective probability density functions of uncertain economic parameters.

The theory and procedures are applied in a case study of the Rio Grande Basin. Annual sediment load data are subject to investigation, with the considered design (decision variable) being the storage allocated for sediment deposition in the not yet completed Cochiti Lake project on the Rio Grande in New Mexico. Decisions concerning design alternatives are made on the basis of an economic efficiency criterion; planning objectives like quality of environment, social benefits, and similar are not taken into consideration. The goal function in this study is nonlinear and consists of a linear penalty function (for either a realized overdesign or underdesign), and a normal probability distribution, which the sample mean of annual sediment load follows, taken over a period equal to the design lifetime. The decision variable (or design alternative) is the sediment trapping part of a reservoir, and the state variables in the problem are the uncertain population parameters, the mean and the variance, which characterize the assumed lognormal probability distribution used in this study to describe the outcome of annual sediment loads.

6.2 Conclusions

One of the advantages of the statistical decision theory is the fact that it takes the uncertainty in the determining parameters and the economics of the project into account. It focuses on the engineering problem of having to decide, with the available information, which design alternative to recommend. The realistic Rio Grande case study shows how the statistical decision method provides a rational and practicable tool in a decision-making process.

Conclusions reached from this study are the following:

(1) The EOL curve follows a decreasing power function, when more and more data are included in the sample. The curve levels off fast and tends to be horizontal, especially after incorporating 40 to 50 data points into the sample. This indicates that the incremental worth of data in connection with such long samples is negligible from a practical point of view.

(2) From an economic point of view it might be worthwhile for the decision-maker to request a postponement of a proposed project in order to collect more data, particularly if this sample is small, say five to ten points. A longer sample increases the information about the uncertain parameters. The gain in information expressed in monetary terms as a result of a more economic and efficient design might offset the cost of data collection plus the loss of benefits which occur by a delay in the construction of a project. For annual sediment load data in the Rio Grande Cochiti Lake project case, it was found that it would be an economic optimum decision to delay the construction of the dam for seven years if a five-year sample was available at the time of design, in order to obtain seven more years of data. The ultimate goal should always be to obtain the expected economic optimum record length for design purpose. It should be noted that only one purpose of the Cochiti Lake has been considered; the consequence of taking benefits from flood protection into account in the economic analysis was disregarded.

(3) Secondary data can be incorporated via a regression model to extend the length of primary set of data. The decrease of uncertainty, i.e., the decrease in the expected opportunity loss, depends on many factors among which the most important are the value of the cross-correlation coefficient and any nonstationarity in the secondary series. The decrease in EOL using the secondary data is not as pronounced as in the case when the given sample is augmented by means of additional primary data, due to the prediction error inherent in the regression model.

(4) The equivalent record length of secondary set of data is less sensitive to the degree of the correlative dependency, represented by ρ, when defined by means of the EOL concept compared to the classical definition by using Fisher information concept. A smaller variation of equivalent length as a function of ρ is found.

(5) The long-term secondary data have to be examined carefully in regard to any nonstationarity before being used for the transfer of information to the primary series. Man-made versus natural changes in the hydrologic regime should be considered and adjusted for, if necessary. Such a consideration is a crucial point in regression analysis. The design alternative and EOL values obtained in this study emphasize this aspect.

(6) Uncertainty related to economic parameters should have as equal a role within the decision process as uncertainty in physical (hydrologic) parameters. A circumspect selection of a subjective probability density function for economic parameters makes it possible to incorporate that type of uncertainty in the statistical decision process. Certain types of a beta distribution were found appropriate to relate to uncertainty in the discount rate, and a normal distribution to be applicable for a unit cost figure, like the cost of sediment removal per acre-feet. It was found that hydrology related parameters introduce more uncertainty into the decision process than the economic variables. The mean value of the discount rate distribution has a strong effect on the chosen design alternative. A low discount rate, in the case of a sediment storage project, results in an increased reservoir space; the lower this rate is, the more an overdesign is favored compared to an underdesign.

(7) The design alternative found in the Cochiti Lake case study, using the sediment load data alone, is smaller than the design by the U.S. Corps of Engineers. The inclusion of uncertainty distributions through the application of the statistical decision approach seems to result in a more economical optimal design compared to the classical design methods. The possibility of a costly overdesign is reduced by the use of some degree of mathematical considerations and statistical sophistication in the design phase of a project.

6.3 Recommendations for Further Studies

During the conduct of this study the following items showed up as being worthwhile for further investigations:

(1) The underlying assumptions for the proper use of a normal chi-square distribution to describe the unknown mean and variance of the data series are (i) independence in time, and (ii) normality. Further study should be undertaken in order to develop a joint distribution where the autocorrelation is considered. As in the present investigation, a complete independent series often seems to be a rough simplification of reality. In order to meet the condition of normality, a logarithmic transformation of input data was necessary. The validity of treating the state parameters in a form of their logarithms, when used in the context of a design problem, might be questioned, although the proper conversion between "log space" and "economic space" are incorporated in the goal function. This controversy can be avoided, if it is possible to describe the stochastic properties of the state parameters by probability distributions not based only on normality. That means, that one should find an alternative to the normal-chi-square density function, which will describe the unknown state parameters in "real space." Attention should also be paid to the common case where the required probability distributions are completely unknown and have to be found through "experimentation" by means of special Monte Carlo simulation techniques. A step in extending the traditional synthesizing procedure in that sense is done in the present study.

(2) The worth of data was found for one particular use, i.e., only one objective was considered. The reason is that the data worth was obtained in connection with a design problem and not in the planning phase of a proposed project. The latter case very often forces the engineer to consider multiple uses of hydrologic data which will require economic studies of tradeoffs among various objectives. Under such, more general conditions, the application of the statistical decision approach as a tool for the engineer has to be investigated to determine whether it is tractable or not.

(3) The study may be extended to incorporate into the economic analysis, all types of benefits foregone from the reservoir. Taking the consequences of flood control into account will undoubtedly decrease the economic optimal record length.

(4) In connection with the use of secondary data, the question arose of how to cope with nonstationarity in a long term hydrologic series. More research should be carried out in order to make proper adjustment for such series. In those considerations, knowledge from disciplines like geology, climatology, population movement, and regional development, must be used by the hydrologist.

(5) The study points out the need for a thorough analysis of uncertainties in economic parameters. The value and the reliability of economic parameters have a noticeable impact on the project feasibility and on the selected design alternative. Economic uncertainties should not be ignored in any type of engineering project, and therefore, should be investigated with the same attention as given to uncertainties in hydrology related parameters.

APPENDIX A.

FACTS ABOUT THE COCHITI LAKE PROJECT IN NEW MEXICO

The information reported in this Appendix is mainly obtained through written communication with the U. S. Army Corps of Engineers, Albuquerque District in New Mexico.

Cochiti Lake was authorized by the 1960 Flood Control Act as a flood and sediment control dam on the main stem of the Rio Grande. The Cochiti Dam is located near the Pueblo Indian village of Cochiti about 50 miles upstream from Albuquerque. When completed the dam will be among the largest earthfill dams in the world, with its 5.4 miles long embankment rising 251 feet above the river.

A.1 Comparison of Sediment Storage Design

Storage allocations to spillway crest are 110,000 acre-feet for sediment reserve and 492,000 acre-feet for flood control. An additional 188,000 acre-feet of storage will be between spillway crest and the top of dam making a total of 790,000 acre-feet.

The Corps of Engineers design of 110,000 acre-feet for sediment storage is based on fairly old sediment flow records which indicate an average of 2,200 acre-feet of sediment flow annually into Cochiti Lake. The sediment storage is designed to hold the inflows for 50 years or 110,000 acre-feet. That means, the value of the sample mean is used to represent the future annual inflow without taking any probability concepts into considerations.

The statistical decision approach using primary data only results in a 50 years lifetime design of 90,000 acre-feet as found in Section 4.3.1. That is nearly 20 percent smaller than the design by the Corps of Engineers.

It is not the purpose of this Appendix to state what design is "right or wrong," because other considerations than strictly economic related factors go into the decision process as pointed out in Section 1.2. However, a few facts concerning this matter is appropriate to mention. The Corps of Engineers uses an average annual flow without relating the sediment load series to any probability distribution. The statistical decision approach accounts for that; in this case the annual sediment loads are found to follow a lognormal distribution. This feature gives one reason why the present study might result in a smaller storage allocation compared to the old design procedure. By taking the logarithm of a set a data more "weight" is put on the lower values in the sample in comparison with the higher numbers, and a smaller average value is the result. In light of the present study the Corps of Engineers seems to have made a conservative design based on the sediment load data. The "overdesign" for the sediment storage alone is 20,000 acre-feet, which in construction cost amounts to approximately $3,000,000. It should be emphasized that this investigation only covers the sediment storage part of the entire project, or 15 percent of the total capacity of the reservoir.

In Section 4.3.3 is the information increased about the unknown mean sediment load by the incorporation of a long water discharge series. Hereby the statistical decision approach ends up with a design alternative of 120,000 acre-feet. These extra considerations consequently result in a fairly good agreement between the Corps of Engineers design of 110,000 acre-feet and the design obtained in the present study. Therefore, in this particular case, it can be stated that the amount the Corps of Engineers allocated of extra storage as a type of safety margin, paradoxically enough, turned out to make their design alternative an economic optimal choice from a Bayesian Risk analysis point of view with a regression model included.

A.2 Project Benefits

The benefit foregone for the project is used (Section 4.3.2) in order to find an expected optimum record length. As recalled, the factor c_2 appears in the total cost function (Eq. 4.18). An estimate of c_2 can be determined in the following way.

Trapping sediment in Cochiti Lake is expected to provide the listed benefits:

(a) Reduce the cost of raising 170 miles of the present leveed system three feet to offset streambed aggradation and maintain the existing flood protection.

(b) Reduce the cost of extending drains to maintain their effectiveness with streambed aggradation and reduce the cost of repairing drains damaged by seepage from the river.

(c) Maintain the present cost of disposal of effluent by preventing the streambed aggradation that would block the outlets and increase the cost of operation and maintenance of sanitary and storm sewer outlets.

(d) Reduce the removing of sediment from irrigation canals and lessen crop losses caused by fine sediment deposited on the land.

No standard method exists to evaluate sediment benefits, but the Corps of Engineers did make estimates in order to determine project feasibility. Table A.1 presents those estimates of damages and benefits for the Cochiti Lake project, updated to the July 1973 price level.

Table A.1 Estimates of Annual Sediment Damages and Benefits in Dollars

Item	Average Annual Damages		Average Annual Benefits
	Existing Conditions	With Cochiti Lake	
Levee raising	101,000	0	101,000
Extending drains	25,000	0	25,000
Drain bank sloughing	25,000	0	25,000
Sediment in Irrigation water	177,000	78,000	99,000
Storm and sewer outlets	25,000	0	25,000
TOTAL:	353,000	78,000	275,000

The benefits foregone factor has hereby been found as c_2 = 275,000 dollars based on 1973 figures. It should be kept in mind that the above calculated benefits are provided only by the part of the reservoir which acts as sediment trapping pool (approximately 15 percent of the total storage) and considered not to have connection with the other purposes (and benefits) of the reservoir.

APPENDIX B.

THE "MINIMAL INFORMATION" DENSITY FUNCTION

The purpose of this Appendix is to find a probability density function, which represents the state of having minimum information about a parameter, r, where r in this case represents a discount rate. In order to cope with that problem, a measure of information has to be introduced. Shannon's information content, as described by Shannon (1948), was selected as being appropriate for that purpose:

$$I = \int_a^b f(r) \cdot \log \{f(r)\} \cdot dr \; . \tag{B.1}$$

The information content, I, in the probability density function f(r), is going to be minimized, subject to the constraint:

$$\int_a^b f(r)dr = 1 \; . \tag{B.2}$$

Minimum I is the same as maximum uncertainty associated with f(r). For convenience f(r) is called F in the following derivations. The Lagrangian expression is formed:

$$L = I + \lambda \cdot \left[\int_a^b F \cdot dr - 1 \right] , \tag{B.3}$$

where λ is a Lagrange multiplier. The differential of the term L with respect to F is equal to zero at a minimum point:

$$\int_a^b (1 + \log F) \cdot dr + \lambda \cdot \int_a^b dr = 0 \; . \tag{B.4}$$

Eq. B.4 yields

$$1 + \log F + \lambda = 0 \; . \tag{B.5}$$

The solution is,

$$F = e^{-(\lambda + 1)} \; . \tag{B.6}$$

The constraint has to be satisfied:

$$\int_a^b e^{-(\lambda + 1)} dr = 1 \; , \tag{B.7}$$

or

$$(b - a) \cdot e^{-(\lambda + 1)} = 1 \; . \tag{B.8}$$

Combining Eqs. B.6 and B.8 yields: $F = \frac{1}{b-a}$, which is identical to $f(r) = \frac{1}{b-a}$ for $a \leq r \leq b$. Thus, the rectangular probability function has been shown to be a "minimal information" density function, for the discount rate r, when it is contained inside a specified range.

APPENDIX C.

LIST OF SYMBOLS

Symbol	Definition
A	Low estimate of subactivity cost
AP	Adjustment parameter, used in the regression analysis
a	Lower limit for the range of the possible values of the discount rate
a_1	A constant which defines the EOL curve
B	High estimate of subactivity cost
B(n)	Total benefits in economic analysis as a function of sample size
b	Upper limit for the range of the possible values of the discount rate
b_1	A constant which defines the EOL curve
C	A constant which makes $s^2_{S_{1+2}}$ an unbiased estimator, used in Eq. 3.21
$\overline{C}_i$	Estimate of mean subactivity cost (i)
C(n)	Total costs in economic analysis as a function of sample size
Cost $(\overline{Q}_s \mid Q_s^{alt})$	Cost function, dependent on the annual mean of sediment load given a design alternative
$\overline{c}$	A constant matrix whose elements are functions of two sets of concurrent data used to establish a regression equation
c_1	Annual cost of having a sediment sampling station in operation
c_2	Benefits foregone parameter
D(n)	Defines difference between total benefits and costs as a function of sample size (net worth of data)
EOL(n)	Expected opportunity loss as a function of sample size
$(EBR_{n_1+n_2} \mid \mu_i,\sigma_j^2)$	Expected Bayesian Risk for a total record length of $n_1 + n_2$ years given a particular set of values of the population parameters
EVEOL(n)	Expected value of the expected opportunity loss as a function of sample size
F	A factor which converts sediment load data from tons per year to acre-feet per year
F(X)	Theoretical distribution used in Kolmogorov-Smirnov test
f(K)	Unit cost probability density function
f(r)	Discount rate probability density function
$f(\mu,\sigma^2)$	Density function for the state parameters
$G(Q_s^{alt} \mid \mu,\sigma^2)$	Goal function dependent on the design alternative and the state parameters
g(Q)	Predictive distribution of Q
I	Shannon's information content
i	Interest rate
$\overline{K}$	Estimated mean of unit cost factor
K_1	Unit cost of reservoir construction (dollars per acre-feet)
K_2	Unit cost for removal of sediment (dollars per acre-feet)
k	Shape parameter for the beta distribution
l	Shape parameter for the beta distribution
M	Number of years between two successive removals of sediment
M_1	Most likely value of subactivity cost
$\underset{Q_s^{alt}}{Min}[G(Q_s^{alt} \mid \mu_t,\sigma_t^2)]$	Minimum value of goal function for different design alternatives
N	Design lifetime of reservoir
N_e	Equivalent record length of secondary data
n	Size of observed data sample
n_1	Size of concurrent water discharge and sediment load data samples
n_2	Size of water discharge sample excluding the concurrent data
$OL(Q_s^* \mid \mu_t,\sigma_t^2)$	Opportunity loss due to the selection of the minimum Bayesian Risk solution as design alternative
p	Proportionality factor between total sediment load and suspended load

Symbol	Definition
$P(X)$	Sample distribution used in Kolmogorov-Smirnov test
$p(AP)$	Discrete probability function for the possible values of the adjustment parameter
Q	Data point in Davis' derivation (Eqs. 3.1 through 3.6)
$\overline{Q}_s$	Annual mean sediment load taken over a period equal to the design lifetime of the reservoir
Q_s^{alt}	Design alternative for the size of the sediment storage expressed in tons per year
Q_s^t	Design alternative which yields minimum value of the goal function when the true values of the state variables are known
Q_s^*	Minimum Bayesian Risk solution
R	Ratio of the variance for the mean of a given sample and for the mean of the lengthened sample after augmentation with secondary data
$R(Q_s^*)$	Minimum Bayesian Risk
r	Discount rate
S_i	Logarithm of annual sediment load used in regression analysis
$\overline{S}_{1+2}$	Sample mean of S_i after augmentation with water discharge data
s^2	Sample variance of sediment load data in logarithmic transformation
$\hat{s}_i$	Estimate of standard deviation of subactivity cost (i)
s_K^2	Estimated variance of unit cost factor
$s^2_{S_{1+2}}$	Sample variance of S_i after augmentation with water discharge data
W_i	Logarithm of annual water discharge used in regression analysis
$\overline{x}$	Sample mean of sediment load data in logarithmic transformation
$\overline{x}_o$	Vector of secondary observations

Symbol	Definition
y	Actual value of primary data
$\hat{y}$	Predicted value of y by means of a regression equation
α	True values of the intercept of the regression line
$\hat{\alpha}$	Estimate of α
β	True value of the regression coefficient
$\hat{\beta}$	Estimate of β
$\Gamma(n)$	Gamma function as a function of sample size
Δ	Kolmogorov-Smirnov statistic
δ	Population mean of annual sediment load (Eq. 4.1)
ε_i	Normal random component with zero mean and unit variance
η^2	Population variance of annual sediment load (Eq. 4.2)
θ	Represents the uncertain parameters in a Bayesian decision analysis
λ	Lagrange multiplier
μ	Assumed value of population mean of annual sediment load in logarithmic form
μ_t	True value of μ
ρ	Correlation coefficient
$\hat{\rho}$	Estimate of ρ
$\rho(1)$	First order autocorrelation coefficient
$\hat{\rho}(1)$	Estimate of $\rho(1)$
σ^2	Assumed value of population variance of annual sediment load in logarithmic form
σ_t^2	True value of σ^2
$\phi(\overline{Q}_s \mid \mu, \sigma^2)$	Probability density distribution for $\overline{Q}_s$ given a set of values of μ and σ^2

REFERENCES

Aitchison, J. and J. A. C. Brown, 1957, "The lognormal distribution," Cambridge University Press, England.

Altouney, E. G., 1963, "The role of uncertainties in the economic evaluation of water-resources projects." Institute in engineering economic systems. Stanford University, Report EEP-7.

Benjamin, J. R. and C. A. Cornell, 1970, "Probability, statistics and decision for civil engineers." McGraw Hill Book Company, Inc., New York.

Brown, C. B., 1948, "Perspective on sedimentation." Proceedings of Federal Inter-agency sedimentation conference, U. S. Bureau of Reclamation, p. 307.

Davis, D. R., 1971, "Decision making under uncertainty in systems hydrology." Tech. Report No. 2, Hydrology and water resources interdisciplinary program, University of Arizona, Tucson.

Davis, D. R., C. C. Kisiel, and L. Duckstein, 1972a, "Bayesian decision theory applied to design in hydrology." Water Resources Research, V. 8, No. 1, pp. 33-41.

Davis, D. R., C. C. Kisiel, and L. Duckstein, 1972b, "Optimum design of mountainous raingage networks using Bayesian decision theory." Presented at Int. symposium on distribution of precipitation in mountainous areas, Norway, July 31- Aug. 5, 1972.

Dawdy, D. R., H. E. Kubik, and E. R. Close, 1970, "Value of streamflow data for project design - A pilot study." Water Resources Research, V. 6, No. 4, pp. 1045-1050.

Ferrell, W. R. and W. R. Barr, 1965, "Criteria and methods for use of check dams in stablizing channel banks and beds." Proceedings from Federal Inter-agency sedimentation conference 1963. U. S. Dept. of Agriculture, Agr. Res. Service, Miscellaneous publication, No. 970, 1965, pp. 376-386.

Fiering, M. B., 1963, "Use of correlation to improve estimates of the mean and variance." U. S. Geological Survey Professional Paper, 434-C, Washington.

Gates, J. S., 1972, "Worth of data used in digital computer models of groundwater basins." Tech. Report No. 8, Hydrology and water resources systems and Industrial Engineering Departments, University of Arizona, Tucson.

Gray, H. and D. R. Davis, 1972, "Augmenting peak flow data by rainfall data in Bayesian decision analysis." Presented at the A.G.U. Nat. Fall Meeting, San Francisco, California, Dec. 7, 1972.

Herfindahl, O., 1969, "Natural resources information for economic development." John Hopkins Press, Baltimore.

Howard, R., 1966, "Decision analysis: applied decision theory." 4th Int. conference of operations research, John Wiley, pp. 55-71, New York.

Jacobi, S., 1974, "Economic worth of sediment load data in a statistical decision framework," Ph.D. dissertation , Colorado State University, Civil Engineering Department. Ft. Collins, Colorado.

James, I. C., B. T. Bower and N. C. Matalas, 1969, "Relative importance of variables in water resources planning." Water Resources Research, V. 5, No. 6, pp. 1165-1173.

James, L. D. and R. L. Lee, 1971, "Economics of water resources planning." McGraw-Hill Book Company, New York.

Kazanowski, A. D., 1972, "Treatment of some of the uncertainties encountered in the conduct of hydrologic cost-effectiveness evaluations." Proc. of the Int. Sym. on Uncertainties in hydrology and water resources systems, Tucson, Arizona, Dec. 11-14, 1972.

Matalas, N. C., 1967, "Mathematical assessment of synthetic hydrology," Water Resources Research, V. 3, No. 4, pp. 937-945.

Matalas, N. C. and W. B. Langbein, 1962, "The information content in the mean." Jour. Geophysical Research, V. 67, No. 9, pp. 3441-3448.

Matalas, N. C. and Barbara Jacobs, 1964, "A correlation procedure for augmenting hydrologic data." U. S. Geological Survey professional paper 434-E, Washington.

Moder, J. J. and G. R. Phillips, 1970, "Project management with CPM and PERT." Van Nostrand Reinhold Company, New York.

Mood, A, 1950, "Introduction to the Theory of Statistics." McGraw-Hill Book Company, New York.

Moss, M. E., 1970, "Optimum operating procedure for a river gaging station established to provide data for design of a water supply project." Water Resources Research, V. 6, No. 4, pp. 1051-1061.

Moss, M. E., 1972, "Expected optimum record length as a basis for hydrologic network design." Proc. of Int. Sym. on water resources planning, Mexico City, Mexico, Dec. 4-8, 1972.

Moss, M. E. and D. R. Dawdy, 1973, "The worth of data in hydrologic design." Journal of the Highway Research Board. In preparation.

Nordin, C. F. and J. P. Beverage, 1965, "Sediment Transport in the Rio Grande, New Mexico." U. S. Geological Survey professional paper 462-F, Washington.

Nordin, C. F. and G. V. Sabol, 1973 , "Estimating average sediment yields from annual streamflow and sediment records." Presented at Int. Sym. on river mechanics, Bangkok, Thailand, Jan. 1973.

Raiffa, H. and R. Schlaifer, 1961, "Applied statistical decision theory." MIT Press, Boston.

Roche, M., 1963, "Hydrologie de Surface." Gauthier-Villars Editor, Paris, France.

Savage, L. J., 1972, "The foundations of statistics." Dover Publication Inc., New York.

Shannon, C. E., 1948, "A mathematical theory of communications." Bell Systems Technical Journal, V. 27, pp. 379-423.

Task Committee for Preparation of Manual on Sedimentation, 1969, "Sedimentation engineering, Chapter IV: economic aspects of sedimentation." Journal of Hydraulics Division, ASCE HY1, 1969, pp. 191-207.

Tschannerl, G., 1970, "A decision framework for the efficient use of data sources." Center for population studies, Report No. 4, Harvard University, Cambridge, Mass.

Tschannerl, G., 1971, "Designing reservoirs with short streamflow records." Water Resources Research, V. 7, No. 4, pp. 827-833.

U. S. Army Corps of Engineers, Southwestern Division, Dallas, Texas. "Water Resources Development in New Mexico 1971."

U. S. Geological Survey, Water-Supply Papers, part 8, Western Gulf of Mexico basins.

Water Resources Council, 1973, "Principles and standards for planning water and related land resources." Federal register, Washington, D.C., V. 38, No. 174 .

Woodson, R. C. and J. T. Martin, 1963, "The Rio Grande comprehensive plan in New Mexico and its effects on the river regime through the Middle Valley." Proceedings of Federal Inter-agency Sedimentation conference.

Yevjevich, V., 1972, "Probability and Statistics in Hydrology," Water resources publications, Fort Collins, Colorado, U.S.A.

KEY WORDS: Statistical decision theory, sediment load, storage, economic analysis, worth of data, equivalent length, optimum record length.

ABSTRACT; Statistical decision approach produces a minimum cost decision with regard to a specific design, and yields a measure in monetary terms, called expected opportunity loss, EOL, of the uncertainty inherent in the decision process. The uncertainty stems from the fact that the "true" state of nature (the population parameters) is unknown. Additional information can be obtained either by collecting extra primary data or by incorporating a long-term secondary data set by means of a regression model. The value of additional information is measured by decrease in EOL. The investigation defines the term"expected economic optimum record secondary data," found in the framework of the probabilistic method. The hydrologic parameters treated as uncertain are the mean and variance of an annual sediment load

KEY WORDS: Statistical decision theory, sediment load, storage, economic analysis, worth of data, equivalent length, optimum record length.

ABSTRACT; Statistical decision approach produces a minimum cost decision with regard to a specific design, and yields a measure in monetary terms, called expected opportunity loss, EOL, of the uncertainty inherent in the decision process. The uncertainty stems from the fact that the "true" state of nature (the population parameters) is unknown. Additional information can be obtained either by collecting extra primary data or by incorporating a long-term secondary data set by means of a regression model. The value of additional information is measured by decrease in EOL. The investigation defines the term"expected economic optimum record secondary data," found in the framework of the probabilistic method. The hydrologic parameters treated as uncertain are the mean and variance of an annual sediment load

KEY WORDS: Statistical decision theory, sediment load, storage, economic analysis, worth of data, equivalent length, optimum record length.

ABSTRACT; Statistical decision approach produces a minimum cost decision with regard to a specific design, and yields a measure in monetary terms, called expected opportunity loss, EOL, of the uncertainty inherent in the decision process. The uncertainty stems from the fact that the "true" state of nature (the population parameters) is unknown. Additional information can be obtained either by collecting extra primary data or by incorporating a long-term secondary data set by means of a regression model. The value of additional information is measured by decrease in EOL. The investigation defines the term"expected economic optimum record secondary data," found in the framework of the probabilistic method. The hydrologic parameters treated as uncertain are the mean and variance of an annual sediment load

KEY WORDS: Statistical decision theory, sediment load, storage, economic analysis, worth of data, equivalent length, optimum record length.

ABSTRACT; Statistical decision approach produces a minimum cost decision with regard to a specific design, and yields a measure in monetary terms, called expected opportunity loss, EOL, of the uncertainty inherent in the decision process. The uncertainty stems from the fact that the "true" state of nature (the population parameters) is unknown. Additional information can be obtained either by collecting extra primary data or by incorporating a long-term secondary data set by means of a regression model. The value of additional information is measured by decrease in EOL. The investigation defines the term"expected economic optimum record secondary data," found in the framework of the probabilistic method. The hydrologic parameters treated as uncertain are the mean and variance of an annual sediment load

series. The study also deals with the question of economic uncertainty; for example, the consequences in the decision analysis of not knowing the exact value of a discount rate and/or a unit cost figure. The theory and procedures are applied in a case study in the Rio Grande Basin, with sediment deposition in a reservoir. The EOL curve is found to follow a decreasing power function as the size of the observed sample increases. The study points out that uncertainty related to economic parameters should have as equal a role within the decision process as uncertainty in the hydrologic parameters.

"Economic Value of Sediment Discharge Data," by Sven Jacobi and Everett V. Richardson.

HYDROLOGY PAPER NO. 67
COLORADO STATE UNIVERSITY

series. The study also deals with the question of economic uncertainty; for example, the consequences in the decision analysis of not knowing the exact value of a discount rate and/or a unit cost figure. The theory and procedures are applied in a case study in the Rio Grande Basin, with sediment deposition in a reservoir. The EOL curve is found to follow a decreasing power function as the size of the observed sample increases. The study points out that uncertainty related to economic parameters should have as equal a role within the decision process as uncertainty in the hydrologic parameters.

"Economic Value of Sediment Discharge Data," by Sven Jacobi and Everett V. Richardson.

HYDROLOGY PAPER NO. 67
COLORADO STATE UNIVERSITY

series. The study also deals with the question of economic uncertainty; for example, the consequences in the decision analysis of not knowing the exact value of a discount rate and/or a unit cost figure. The theory and procedures are applied in a case study in the Rio Grande Basin, with sediment deposition in a reservoir. The EOL curve is found to follow a decreasing power function as the size of the observed sample increases. The study points out that uncertainty related to economic parameters should have as equal a role within the decision process as uncertainty in the hydrologic parameters.

"Economic Value of Sediment Discharge Data," by Sven Jacobi and Everett V. Richardson.

HYDROLOGY PAPER NO. 67
COLORADO STATE UNIVERSITY

series. The study also deals with the question of economic uncertainty; for example, the consequences in the decision analysis of not knowing the exact value of a discount rate and/or a unit cost figure. The theory and procedures are applied in a case study in the Rio Grande Basin, with sediment deposition in a reservoir. The EOL curve is found to follow a decreasing power function as the size of the observed sample increases. The study points out that uncertainty related to economic parameters should have as equal a role within the decision process as uncertainty in the hydrologic parameters.

"Economic Value of Sediment Discharge Data," by Sven Jacobi and Everett V. Richardson.

HYDROLOGY PAPER NO. 67
COLORADO STATE UNIVERSITY