C-1

UNITED STATES
DEPARTMENT OF THE INTERIOR
Geological Survey

ECONOMIC WORTH OF HYDROLOGIC DATA IN PROJECT DESIGN: AN APPLICATION TO REGIONAL ENERGY DEVELOPMENT

by

Michael R. Karlinger

Open-File ~~Report~~ 76-316

Reston, Virginia

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

SYMBOL	MEANING
A	Input-output transaction matrix
a_o	Proportionality constant between capacity and water supply
a_{ij}	Technical coefficient of I-O matrix
$\hat{a}_{ij}$	Element of I-O matrix
B	Regional Benefit function
C	Cost of data collection
crf, η	Capital recovery factor
C_v	Coefficient of Variation
d	Design Discharge
d*	Optimal d
FD	Final demands
f(.)	Probability distribution of X given θ
g(.)	Marginal density function of X
h(.)	Joint probability density function of X and θ
I	Identity matrix
K	Capital Cost
K_j	j^{th} flow parameter's equivalent year information
L	Length of project life
N_B	Number of gaged basins

LIST OF SYMBOLS

SYMBOL	MEANING
N_Y	Average record length
NW	Net worth of information
n	Length of time for data collection
n*	Optimal length of data collection
OM	Operation and maintenance costs
PVNB	Present value of net benefits
P(.)	Prior distribution function on θ
r	Interest rate
TGO	Total gross outputs of the economy
$\hat{V}$	Transpose of regional income vector
VI	Value of information
w	Short run weighting factor
XB	Expected benefits from the project
XXB	Expected-expected benefits from the project
X'	Flow value during project life
X	Sample information flow value
X_t	Streamflow in time t
Y	True equivalent year value
$\hat{Y}$	Apparent equivalent year value
Y_t	Plant output in time t
Z	Plant capacity

LIST OF SYMBOLS

SYMBOL	MEANING
β_o	Proportionality constant between regional benefits and water supply
δ	Present value factor
γ	Regression model error
μ	Mean of streamflow distribution
$\psi(.)$	Conditional probability density of X' on θ
$\phi(.)$	Joint probability density function of X and θ
ρ_c	Cross correlation between gaging stations
ρ_s	Serial correlation of stations
ρ_{ij}	Household coefficient from $[I-A]^{-1}$ matrix
σ	Standard deviation of streamflow distribution
τ	Benefit time reference
θ	Streamflow parameter set

Conversion Table

Multiply	By	To obtain
Cubic feet	.028	Cubic metres
Square miles	2.59×10^{6}	Square metres
Barrels (42 gal)	.159	Cubic metres
Acre-ft	1219.68	Cubic metres
Cubic feet per second	.02832	Cubic metres per second

ECONOMIC WORTH OF HYDROLOGIC DATA
IN PROJECT DESIGN: AN APPLICATION TO REGIONAL ENERGY DEVELOPMENT

ABSTRACT

The linkage between the benefits to a regional economy from a water-dependent industry and the statistical uncertainty of the water supply that is needed in the production process is analyzed using a Bayesian scheme of expected-expected benefits of hydrologic data. In this analysis, expected benefits are calculated using anticipated streamflow data and these benefits are averaged over all possible data values. The increase in these average expected benefits is subsequently compared to the costs of obtaining the data and benefits foregone by delaying the project to collect the data. Thus, the value of collecting the data may be determined.

A Leontief type production function is assumed for an oil shale retorting plant. The economic impact of hydrologic uncertainty is investigated by examining potential changes in plant output along with changes in regional incomes which are calculated by means of a regional input-output model. Alternatives for reducing the uncertainty included establishing water discharge measuring stations, transferring streamflow information from one site to another via regionalization, and a combination of both. Given a specific alternative, the optimum time period for data collection was found. Regionalization considered statistical properties of network designs based on simulated regression analysis.

CHAPTER I

INTRODUCTION

The incremental value of information as measured in an economic study is the expected improvement in the net benefits of a project as a result of gathering additional data or transferring information from one site to another. The difference between these added net benefits and the cost of obtaining the data is the net worth of the data. The aim of this study is to determine the marginal value of hydrologic data for the optimal initiation of a water intensive industrial process. That is, in light of the statistical nature of hydrologic events, at what point do the costs of obtaining more data overcome the improved net benefits from a revised project?

The benefits of a water-resource project are a function of random streamflow values, and as such are subject to risk and uncertainty: uncertainty arising from the lack of knowledge in parameter definition of the streamflow probability distribution and risk resulting from imperfect information concerning the outcome from a known distribution. If more data were available, an improved project could be designed to be better accommodated to the randomness of the streamflow input, and hopefully having an improved flow of benefits from the project. However, in practice, neither the extra data nor the effect that they might have on the benefits derived is known at the time the design decision is required, and the value of gathering more data must be analyzed probabilistically.

Plant size and the production function (process), the basis of the project's direct benefit function, are assumed designed on a prespecified water availability. This quantity is often based on a concept of firm yield; an amount of water that must be guaranteed a certain percentage of time in order for output to reach the desired level. The yield used in the design of the project can be derived using a probability density function characterizing the streamflows. The dependence of the output from a given plant on the input of the streamflow, therefore, results in the production process explicitly being a function of the actual streamflow values and implicitly a function of the parameters of the probability function of the streamflow.

Bayesian analysis techniques are used to manage the problems of risk and uncertainty in the benefit stream. In such an analysis, probabilities of obtaining a particular sample value of streamflow from a given probability distribution, based on assumed parameters of the distribution, are weighted with (prior) probabilities of those assumed parameter values. This weighting process is then normalized to redefine a new (posterior) probability function, which measures the probability of knowing the true parameter values given the additional sample information.

These posterior probabilities are used to weight the benefit function and the result is summed over all possible parameter and streamflow values. The resulting increase in the expected value of benefits is compared with costs of obtaining data to determine the payoff from better definition of the parameters; that is, from gathering additional data.

This report is essentially the same as the author's thesis for the degree of Master of Arts in Economics, which was submitted to the University of Colorado.

Selected Previous Studies

Several studies relating to the worth of data have been performed in the past, each looking at the problem in a different manner. Dawdy and others (1970) have approached the problem via simulation procedures whereby the uncertainty portion of the problem was managed by using the statistics of a 500-year simulated data record to define an optimum design. This total record was subdivided into smaller records of various lengths, and designs were based on these truncated records and compared to the optimum. Moss (1970) used information content of data based on the variance of streamflow to determine the optimum length of a streamflow record and the supplemental frequency of measurement for purposes of building a water resources project. The differences in the benefits derived from a project could be matched against the differences in the records' lengths, and consequently costs, to determine if the benefits outweighed the costs of data collection.

Davis and Dvoranchik (1971) analyzed in a decision theoretic framework this problem of investment justification and presented an example of bridge design with regard to optimal piling depth in a scour prone creek. A linkage of decision theory with simulation modeling and mathematical programming was developed by Maddock (1973) in studying the worth of data in ground water systems. His results were incorporated in a ranking of the variables in terms of their error contribution and the effect of data collection priorities on the management of objectives. And while not directly dealing with the value of specific data collection operations, a report by Howe and Cochrane (1975) illustrated a method for evaluating improvements in weather forecasting accuracy. The study presented here is based on the ideas embodied in these and other previous works pertaining to the value of information about an economic-hydrologic system.

CHAPTER II

BENEFIT ANALYSIS

The benefits examined in this study are based on decisions made under the existence of uncertainty in the supply of inputs. Turnovsky (1972) studied the problem of optimum investment in uncertain supply systems by assuming a defined utility function for the inputs and arriving at a step-by-step procedure for obtaining an expected demand curve for the system outputs.

The approach used here in analyzing the social benefits obtained from a particular use of a water-resource project is to define a production process and transform it via a regional model into a regional benefit function.

It is assumed that from engineering design considerations, the output of a particular plant or project is proportional to the available water supply. That is, a long run production function exists such that

$$Z = a_0 d \tag{1}$$

where a_0 is an industry determined proportionality constant, d is the design water-supply system discharge upon which the plant is sized, and Z is the plant capacity.

Regional benefits, B, from any given plant size are necessarily dependent not only on Z, but also on the particular streamflow realization of time period t, X_t, whose distribution should determine d. The regional benefits are assumed to accrue linearly with X_t up to the flow value of d, where they reach a maximum and remain constant for all X_t greater than d (Figure 1). This form of the benefit function implies that any spilled or excess water has no beneficial use.

$$B(Z(d),X_t) = \beta_0 X_t \quad X_t < d \tag{2}$$

$$= \beta_0 d \quad X_t \geq d \tag{3}$$

The proportionality factor in the benefit function, β_0, is itself dependent on the proportionality constant, a_0, of the production function. Because of the expectation operation, d becomes a function of the probability density assumed for X_t. A maximization over d of the expected net benefits yields an optimal design value d*.

The remaining step in the benefit analysis is to determine β_0. It is necessary to express the short run output of the production process as a general function of X_t and d. If Y_t is the short run output in period t, then the following relationships can be used

$$Y_t = \begin{cases} a_0 d & X_t \geq d \\ a_0 X_t & X_t < d \end{cases} \tag{4}$$

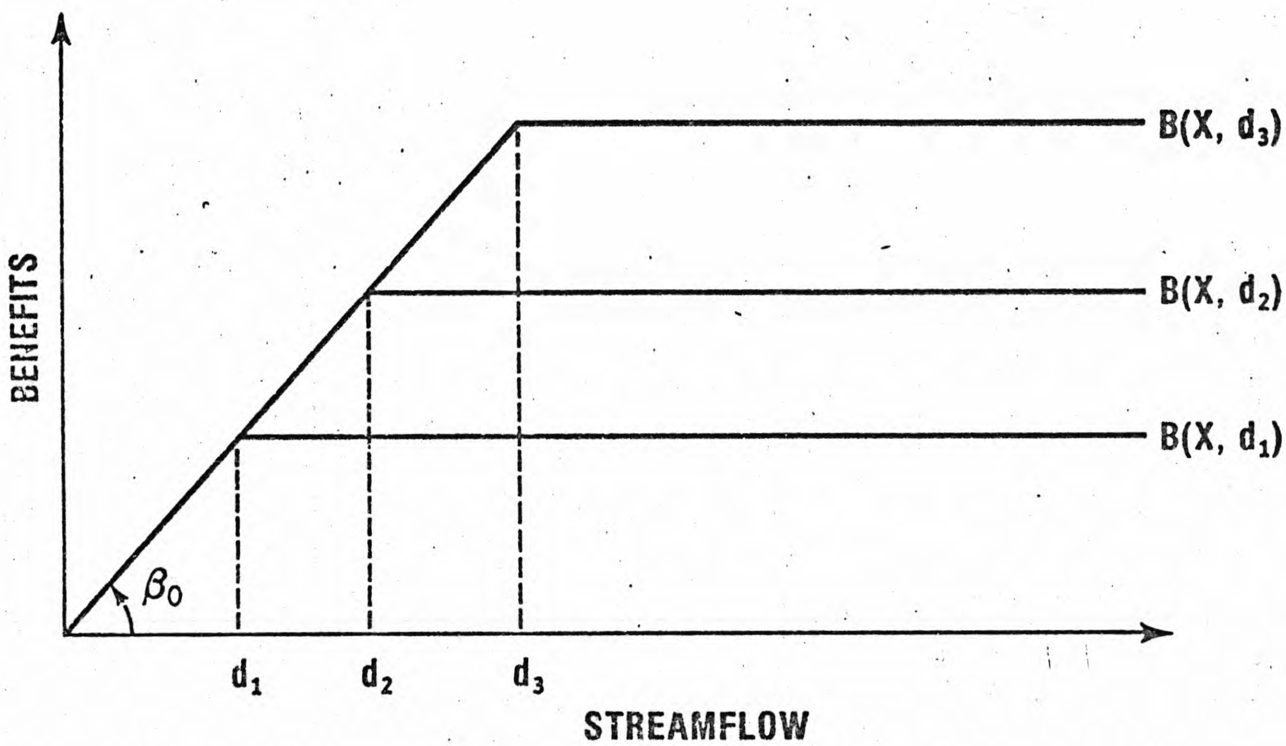

Figure 1. General Form of the Benefit Function.

The production output, Y_t, is entered into an input-output model by utilizing the linear properties of matrix algebra. The benefits can then be obtained through an interpretation of the total gross output (TGO) vector.

The input-output model referred to above is a transactions matrix of coefficients showing the interactions of the various sectors of an economy. These interactions also include the intra-relationships of all sectors. The basic form of the matrix is as follows:

$$
\begin{array}{cc}
 & \text{Sectors } j \\
\text{Sectors } i & \begin{bmatrix} \hat{a}_{11} & \hat{a}_{12} & \cdots & \hat{a}_{1n} \\ \hat{a}_{21} & & & \\ \vdots & \ddots & & \\ \vdots & & \ddots & \\ \hat{a}_{n1} & & & \end{bmatrix}
\end{array}
$$

where the $\hat{a}_{ij}$ describe the input from sector i into sector j. Division of the $\hat{a}$ coefficients by the total outlays of their respective sectors furnishes technical coefficients (a_{ij}), from which direct and indirect coefficients of production will be obtained. These coefficients show the quantities that a given sector must produce in order to satisfy any given set of final demands of all the sectors. In matrix notation

$$[I - A] \ [TGO] = [FD]$$

$$\text{or} \tag{5}$$

$$[TGO] = [I - \bar{A}]^{-1} \ [FD]$$

where [A] is the matrix of $[a_{ij}]$ factors, [I] is the identity matrix and $[I-A]^{-1}$ is the matrix of direct and indirect coefficients. [TGO] is a vector showing the production of each sector to meet the final demands, [FD], of the economy. A change in total output might be taken as an indicator of the health of an economy. The change in the TGO vector resulting from changes in final demands might thus be used as a surrogate measure of benefits. More specific assumptions will be discussed later.

Initial conditions for the input-output model will be defined as that state of the relevant economy where the industry in question does not yet exist and final demands (FD) in all other sectors have a base value of zero. The change in the final demand vector as a result of constructing the project is, therefore, a vector with zero entries for all sectors except for the newly defined sector. The A matrix is augmented by a row and column for this new industry, and the new $[I-A]^{-1}$ is computed. It is assumed that all of the new industry's output is delivered to final demands, for example, exported from the region.

$$\Delta FD_t = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ a_0 X_t \\ 0 \\ 0 \\ \cdot \\ \cdot \end{bmatrix} \text{ for } X_t < d \qquad \Delta FD_t = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ a_0 d \\ 0 \\ 0 \\ \cdot \\ \cdot \end{bmatrix} \text{ for } X_t \geq d$$

If the household sector is endogenous to the input-output model, that is, it is included in the transactions matrix, the change in household TGO approximates the change in regional values added, a reasonable measure of regional income before transfer payments. Let ρ_{ij} be that entry of the $[I-A]^{-1}$ matrix indicating the direct and indirect requirements on the household sector from the establishment of the project sector. The change in total gross output of the household sector is thus deferred as regional benefits and is given by

$$B = \beta_0 Y_t = \begin{cases} \rho_{ij}\ a_0\ X_t \text{ for } X_t < d \\ \rho_{ij}\ a_0 d \text{ for } X_t \geq d \end{cases} \tag{6}$$

If the household sector is not endogenous to the structure of the I-O model, then regional benefits can be defined slightly differently as

$$B = \hat{V}\ \Delta TGO_t \tag{7}$$

where $\hat{V}$ is the transpose of the vector of regional income or value added coefficients from each of the other sectors. In this case the benefit function is defined as

$$B' = \hat{V}\,\Delta TGO = \hat{V}\,[I-\bar{A}]^{-1}\,\Delta FD \tag{8}$$

where

$$\Delta FD = \begin{cases} [0,\ 0,\ .\ .\ .\ a_0X_t\ .\ .\ .] \text{ for } X_t < d \\ [0,\ 0,\ .\ .\ .\ a_0d\ .\ .\ .] \text{ for } X_t \geq d \end{cases} \tag{9}$$

The above procedure for using direct and indirect coefficients of the I-O matrix implies a total response of the various sectors to any variation or disturbance in the short run operations of the new plant. In reality, this response may be damped because of the assumption in the other sectors that any shortfall in production is only temporary, and there is no need to gear down services to the extent that the inverse matrix indicates.

A damping factor may be incorporated in the TGO calculation by introducing a diagonalized matrix containing weights suggesting how each of the other sectors might react to a reduction in the new sector's output due to a shortfall of the design water yield. These weights have a lower limit of zero and an upper limit of unity, where the weights are designated w_i.

Equilibrium:

$$\Delta TGO = [I-A]^{-1} \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ a_0 d \\ 0 \\ 0 \\ \cdot \\ \cdot \end{bmatrix} \qquad \text{for } X < d \tag{10}$$

Short term variations:

$$\Delta TGO = \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix} [I-A]^{-1} \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ a_0 d \\ 0 \\ 0 \\ \cdot \\ \cdot \end{bmatrix} \text{for } X \geq d \tag{11}$$

This procedure for mitigating the effects of downward production swings in the new sector is one method of accounting for the shortrun fluctuations. However, if no damping factors are used, the impact of variations in production of the plant considered illustrates the equilibrium condition of the plant operations.

The input-output analysis above is developed to establish the coefficient of the unit direct and indirect benefit in the lower range of the benefit function below d_i in Figure 1. The random nature of the benefits must still be dealt with probabilistically in order to fix optimal values of d_i for further analyzing the value of information.

CHAPTER III

VALUE OF SAMPLE INFORMATION

Water-resource project designs may be based on estimates of the statistical parameters of streamflow; the project's purpose defines which of these parameters are the most important. Methods available for obtaining an improved measure of these estimates include collection of at-site data and transferring information from other sites by statistical methods. The question is, how much data is enough to optimally design the project in terms of expected net benefits?

In the preceding economic model the streamflow experienced in a period during the life of the project determines the applicable segment of the benefit function for that period. The parameters of the cumulative distribution function of streamflow determine the frequency with which the two segments of the benefit function are applicable. If the estimated parameters lead to overdesign, the sloped portion of the benefit function is used more often, while project underdesign utilizes the horizontal section more often. Therefore, any hydrologic information which will narrow the possibilities of overdesign or underdesign will have value.

The objective is to calculate the expected net benefits of the project at various times in the future, and determine when the increase in expected benefits equals the marginal costs of collecting data and delaying construction to these future dates. Postponing development into the future implies more information and a higher probability of appropriate project design, but the added benefits will be received later by the amount of delay needed to collect the data. This increase of net benefits is diminished, therefore, by the effect of the discount factor. Deferment of the project will involve costs of the continued operation of gage(s) which are also subject to the discount rate. The point when the (timewise) marginal costs equal or overcome the (timewise) increase in benefits establishes the optimal period for the project initiation.

Although several good expositions of Bayesian schemes can be found in Pratt and others (1965), Raiffa (1970), and Ferguson (1967), this analysis of the value of additional data will begin with a description of "predictive" probability density functions presented in Zellner (1971). The problem is one of updating prior probabilities of states of nature (parameters) by anticipating certain sample information.

If X' is the set of yet unobserved flow values, X, the set of sample information, and θ a vector of the flow distribution's defining parameters, mean, μ, and standard deviation, σ, a predictive probability density function of project flows can be obtained from the following:

$$\phi(X',\theta|X) = \psi(X'|\theta)\ P(\theta|X) \tag{12}$$

ϕ is the joint probability density function for X' and the flow parameters, ψ is the conditional probability density function for X' given the parameters θ, and P is the posterior probability of θ conditioned on the sample information.

If the parameter values of θ are integrated out of the ϕ function, the predictive probability density function is obtained:

$$\phi(X'|X) = \int_\theta \psi(X'|\theta)\ P(\theta|X)\ d\theta \tag{13}$$

This formulation, the conditional probability of unobserved data given sample data, is essentially an averaging of the sample data, ψ function, with the posterior density P acting as a weighting function.

To arrive at the posterior density P, a prior distribution, $p(\theta)$, of the true flow parameters must be defined. This will either be subjective as a non-data-based prior or will use some previous information to arrive at a data-based prior. With X assumed to follow some distribution $f(X|\theta)$, sometimes called a likelihood function, the joint density of X and θ, $h(X,\theta)$, is given by

$$h(X,\theta) = f(X|\theta)\ p(\theta) \tag{14}$$

The marginal density function of X, $g(X)$, is obtained by integrating with respect to θ

$$g(X) = \int_\theta h(X,\theta) d\theta = \int_\theta f(X|\theta)\ p(\theta)\ d\theta \tag{15}$$

With these last two results and the definition of conditional probability, the posterior distribution, $P(\theta|X)$, will be given by

$$P(\theta|X) = \frac{h(X,\theta)}{g(X)} \tag{16}$$

The preposterior analysis outlined by Zellner (1971) and discussed above can be thought of in more explicit terms as follows. The time at which the decision is made to undertake a project is defined as $t = 0$. The span of time between $t = 0$ and the final design and construction is the period in which data are to be taken and thereby defines the sampling time (n years) over which X is considered (X_i, $i = 1,n$). For simplicity in explanation, it is assumed that sampling is carried through the construction period. Computationally, it would not be difficult to define a time interval during construction in which no data were taken. This would simply define a different reference point from which data costs and benefits would be discounted.

Once the project is built, the streamflow X_{n+j}, $j = 1,L$, where L is the project life, begin to generate the benefit stream; these streamflows constitute the set X'. Note that data collected previous to $t = 0$ are used only to define prior distributions and to estimate the form of distribution which the X follows. These data are not used directly in the weighting and averaging process in the preposterior analysis. With the above explanation of X and X', the predictive probability of X' can be rewritten

$$\phi(X_{n+j},j=1,L|X_i,i=1,n) = \int_\theta \Psi(X_{n+j},j=1,L|\theta,X_i,i=1,n) \cdot P(\theta|X_i,i=1,n)\, d\theta \tag{17}$$

The posterior probability $P(\theta|X_i, i = 1,n)$ can be derived from a set of n independent random variables each having $f(X|\theta)$ as its density function. If the X's are mean annual streamflow values, they can be assumed and shown to exhibit inconsequential serial correlation. The likelihood function in this case would be $f(X_1 \ldots X_n|\theta)$, and the assumption of independence allows this joint distribution to be written as a product of univariate distributions: $\prod_{i=1}^{n} f(X_i|\theta)$.

This expression can be substituted for $f(X|\theta)$ in (14), resulting in $h(X,\theta) = \prod_{i=1}^{n} f(X_i|\theta)p(\theta)$.

The predictive probability function (equation 13) can be coupled with a benefit function and expectations can be taken over values of X' and X. This is done for various values of n to determine the expected effect of n on the benefit function, that is, the value of increasing n.

The benefit function, $B(X',d)$, is dependent on two types of variables, elements of the set X' effecting short run fluctuations in benefits, and a design discharge d reflecting a long run objective where d is now a function of the data X_i. After expectations of the benefit function are taken with respect to X', a maximization of the net benefit function over d should yield a d^* or optimal discharge on which the project design is based on a given X in the i^{th} year, that is X_{i_o}. This masimization cannot be performed until costs of construction, operation, and maintenance (OM) are considered, and this will be illustrated in the next section.

Therefore expected total benefits with respect to an X_{i_0} are given by

$$XB(d,X_{i_0}) = \int_0^\infty B(d,X')\ \phi(X'|\,X_{i_0})\ dX' \qquad (18)$$

which is a weighting scheme using functions as shown in Figure 2. Averaging equation 18 over all X_{i_0} will give an expected total benefit function

$$XXB(d) = \int_0^\infty XB(d,X)\ g(X)\ dX \qquad (19)$$

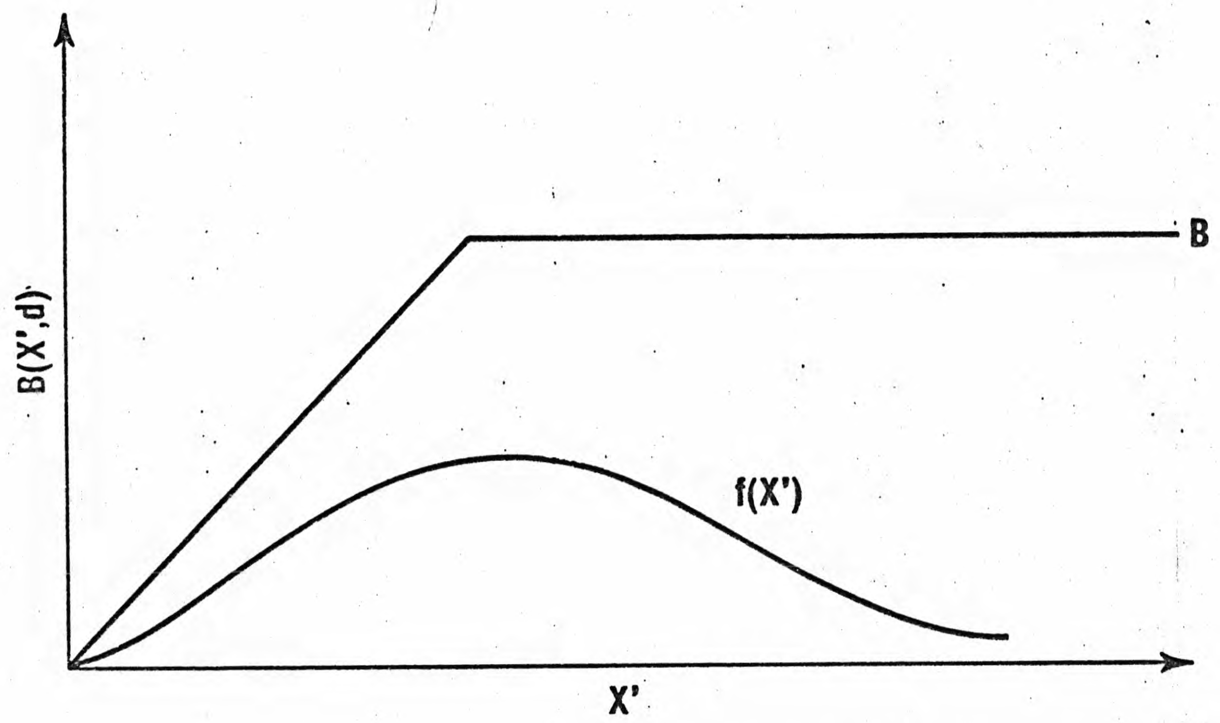

Figure 2. Benefit Function with Underlying Distribution as Weighting Function.

where again $XB(d,X)$ and $g(X)$ are dependent on n.

With numerical results for $XXB(d)$ over various n values, considered together with costs of construction, OM, and data collection, a value of information can be calculated for each added year of data. Specifically, the marginal value of the $n+1^{st}$ year of data collection (VI_{n+1})

$$VI_{n+1} = XXB_{n+1}(d) - XXB_n(d) \tag{20}$$

Although the VI is assumed to be at least non-negative, the value may not be large enough to warrant any change in the decision regarding the project. With this consideration in mind, the added information can be said to have value only if a resulting change in the decision process can be expected.

CHAPTER IV

NET BENEFITS VERSUS COSTS OF DATA

The expectation of expected benefits (which will be referred to subsequently only as benefits) as a function of d can now be combined with the costs of construction and maintenance of different sized water systems to determine the optimal length for additional data gathering. If K is the capital invested and OM is the annual operation and maintenance costs of the system, the discounted net benefit function (PVNB) takes the following form:

$$PVNB(d^*,(n)) = \delta^n \int_0^\infty \{\max_d [XB(X,d) - K(d)crf(L) - OM(d)]g(X)\}dX \tag{21}$$

which is assumed to be a concave function in the range of d values considered (crf is the capital recovery factor and is equal to $(r(1+r)^L) / ((1+r)^L-1)$ and δ is equal to $\frac{1}{1+r}$ where r is the interest rate).

In order to determine the marginal values of delaying construction of a project, two relative time periods should be defined. τ is an arbitrary future point in time denoting the initial realization of benefits, and is the reference for the net benefits assessed during the project life (Figure 3). Real time is designated t, and $t(0)$ refers to the time of project conception. Therefore $t(\tau) - t(0)$ is the length of time the project is postponed for data collection and defines n. To compare results on a common base, all net benefits are discounted back to $t(0)$ by multiplying them by the factor δ^n.

If C is defined to be the cost of collecting data for one year at an already established site, the present value of data collection costs (PVC) for n years equal $\eta(n)C$ where $\eta(n) = \dfrac{(1+r)^n-1}{r(1+r)^n}$. Therefore the net worth (NW) of the information collected over the n years would be

$$NW(n) = PVNB - \eta(n)C \tag{22}$$

For a continuous case, the maximum value of $NW(n)$ could be evaluated by differentiating the expression with respect to n and setting the result equal to zero, which gives

$$\frac{d}{dn}(PVNB) = \frac{d}{dn}\eta(n)C \tag{23}$$

The value of n where the marginal discounted net benefits equal the marginal discounted costs of data collection defines the optimal length of time for data gathering, n^*, as implied in Figure 4 for a well behaved function.

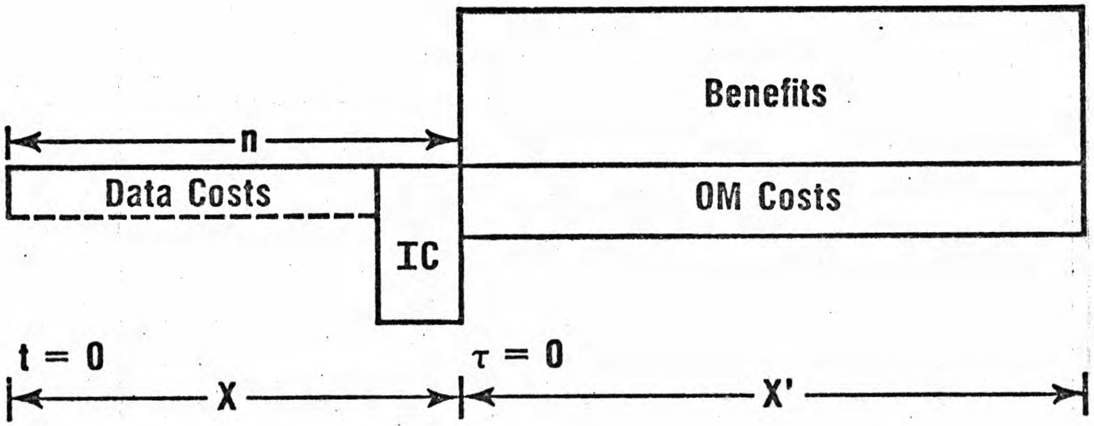

Figure 3. Schematic of Benefit and Cost Stream of the Assumed Project.

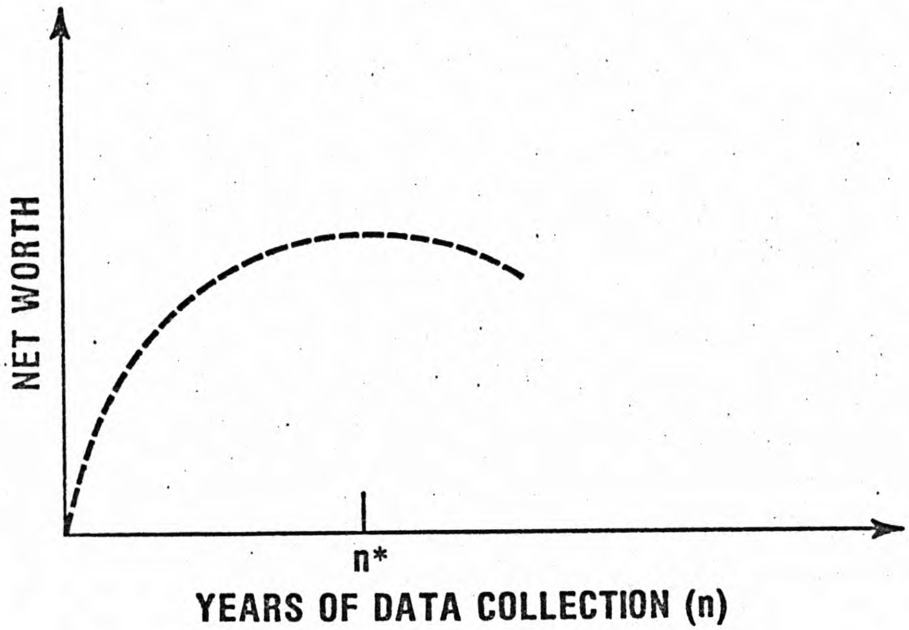

Figure 4. Net Worth as a Function of n Years of Streamflow Data.

With an existing gage at the particular project site, data collection costs are simply the costs of manpower, travel, and data reduction. In this case, the collection costs take the simplest form as described above.

If the project site is not a gaged site, regionalization techniques for transferring hydrologic data can be used to obtain streamflow information. Specifically, a regional regression analysis is performed for the data transfer and the concept of equivalent years of record as described by Hardison (1969) can be used in the decision process. If such is the case, extra equivalent years of data can be obtained by various changes in the gaged network makeup: one can increase collection at existing sites; more sites can be added; or a combination of both. These various options are developed in Moss and Karlinger (1974) and will be discussed in the next section.

There is also a third option for the data collection. A gage can be established at the site, with a combination of regionalization values and the at-site data used to obtain estimated values of the streamflow parameters.

In the case of no at-site data, the data collection cost function requires new consideration, because capital costs are incurred in establishing any new stations. Since the stations will be established at $t = 0$, the time reference point, no discounting adjustment of the capital costs is made. If the number of stations remains fixed, the operation and maintenance costs of continued gaging will be applied to all of the stations. With the simplifying assumption of constant costs for all stations, the OM discounting factor, $\eta(n)$, is simply multiplied by the number of gages. If new stations are added, a combination of capital and OM costs must be considered.

The number of alternatives for obtaining the additional information for an ungaged site are, of course, limited by constraints of the data collection agency. These can be in terms of limited funds or in terms of manpower limitations. The assumptions made here will be that the minimum cost combination, noting constraints, for obtaining a specified increase in record will be the cost incorporated into calculations of the net worth of data.

A decision on a network configuration based on the cost-benefit tradeoff should not be a static one, but rather a reevaluation of the network should be made after gathering some additional data. This additional data will reflect any improvement in the time sampling error, and a reestimation of any statistical parameters can be made. Hence the additional information could be used in an absolute sense to evaluate the present strategy, whereas three years previously it was used in an expected or weighted sense. The reanalysis may indicate a change in strategy and corrections can thus be made. This updating of network information is an iterative process but should be convergent to an optimal configuration.

CHAPTER V

DESCRIPTION OF HYDROLOGIC NETWORK DESIGN BY REGRESSION ANALYSIS SIMULATIONS

This section will briefly review a methodology which defines alternatives for obtaining additional information on streamflow parameters via a regionalization technique.

Benson and Matalas (1967) suggested regionalization of hydrologic parameters as a means of transferring information from gaged sites to ungaged sites. This regionalization mechanism has taken the form of logarithmic regression analysis at gaged-site flow parameters, such as means and standard deviations, versus basin parameters such as drainage area, forest cover and soil moisture properties. The standard error of the regression line can be computed and through a logarithmic transformation, a value of equivalent years of record can be calculated (Hardison, 1969). This equivalent years of record concept is defined as the expected length of time an ungaged site would have to be gaged before the statistical parameter computed from the at-site record would be as accurate as that obtained from the respective regression lines.

One regression line for a basin constitutes a single realization of a random variable and therefore the equivalent year value derived from the regression's standard error is also only a single realization of a random variable. The gaged sites used in the regression might have been located in other basins, thereby incurring a space sampling error, and the record length at the given sites can be longer than that used in the regression, thereby reducing the time sampling error.

Moss and Karlinger (1974) have outlined a procedure whereby the statistical randomness of the regression lines could be analyzed via Monte Carlo techniques. The resulting distribution of apparent equivalent year values $(\hat{Y})$ for mean flow (μ) and standard deviation of mean flow (σ), can be used within a Bayesian framework to obtain a probability distribution of equivalent year values of the same true parameters of the streamflows (Y).

The statistical properties of the streamflow in the present network are based on an assumed known regression model of true flow parameters. A synthetic streamflow generator is used to develop sample flow statistics based on the true parameters. These sample statistics are regressed on the same basin parameter as the true flow statistics. The sample flow regression yields an apparent equivalent year statistic, while a transformation of this standard error relative to the true model yields the true statistic of equivalent years. By repetitions of the regression-analysis simulation, likelihood functions of the apparent equivalent years are developed and conditional probabilities of the true equivalent years with given input parameters are obtained.

The initiating parameters needed are: (1) the cross correlation between streamflows within the basin ρ_c, (2) the serial correlation of the streamflows, ρ_s, (3) the coefficient of variation of the streamflow C_v, and (4) the error inherent in the regression model, γ. In addition to these parameters, the standard error of an actual sample regression performed over the network, the number of stations, N_B, in the basin, and the number of years of record, N_Y, for each station are also required. Prior probabilities of ρ_c, ρ_s, C_v, and γ are used in conjunction with the likelihood functions of the apparent equivalent years conditioned on values of these parameters; Bayes formula is subsequently applied to develop posterior probabilities for the parameters, i.e.,

$$P(\phi|\hat{Y}) = \frac{P(\hat{Y}|\phi)P_r(\phi)}{\sum_{\phi} P(\hat{Y}|\phi)P_r(\phi)}$$

where ϕ is the parameter set $\{\rho_c, \rho_s, C_v, \gamma\}$ and P_r designates prior probability.

The model error, γ, is the most difficult parameter to be assigned a prior probability, since there is no analytical way of measuring it. Therefore, the likelihood functions of the sample equivalent years for the regression performed over the existing network are used to define the range of model errors to be considered. This is accomplished by determining the largest model error over all sets for the given $\hat{Y}$ that have a nonzero likelihood function, $P(\hat{Y}|\phi)$. This value of model error becomes the upper limit of model errors used in the analysis. All of the model errors between zero and this upper bound are assigned an equal prior probability of occurrence and are entered into subsequent calculations.

The conditioning of the true equivalent year values on the various ϕ sets can now be released with the posterior probabilities. There will be a regression and analysis for each streamflow parameter, μ and σ, and, therefore, the conditioning will be released for true equivalent year values of each parameter separately. The expected value of each parameter j will be

$$E(Y_j) = \sum_{\phi}(Y_j|\phi)P(\phi|\hat{Y}_j)$$

These same posterior distributions, $P(\phi|\hat{Y}_j)$, can be used with the distributions of true equivalent years for other N_B and N_Y combinations, yielding expected true values for these points. A plot of N_B versus N_Y with true equivalent record lengths contoured between the defined points yields a production function for information. The contours are convex to the origin and increase in value with increasing N_B and N_Y. Therefore the N_B and N_Y values on a given contour define the possible combinations of attaining the contour value from an initial state, subject to any given constraints (Figure 5). The minimum cost combination from among these is then used to establish the cost functions of obtaining the additional years of information.

There will now be a set of contours for equivalent year values of each parameter, μ and σ. Because the contour set from each of these parameters will probably have different convexities, the sets will intersect and define a new joint production function for information. The shape of the new contours will depend on the amount of information needed from each parameter. For instance, if it is decided that the information needed for σ is K_2 and that needed for μ is K_1, then the desired network alternatives would be defined by the K_1 contour for the mean intersecting the K_2 contour for the standard deviation (Figure 6). Note that for simplicity the information content for each parameter along the newly defined contour is assumed to equal those contour values from which the new contour was drawn.

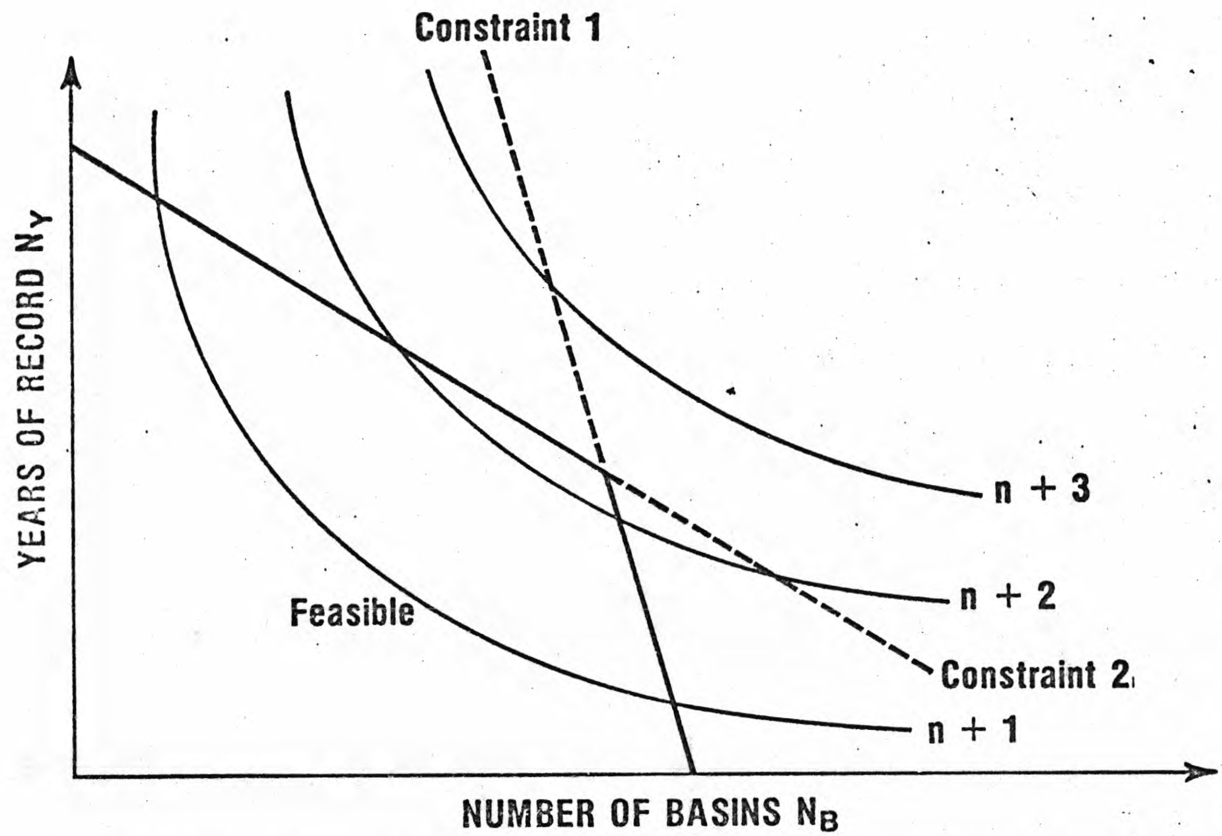

Figure 5. Typical Information Production Function (for equivalent years) and Feasible Solution Space.

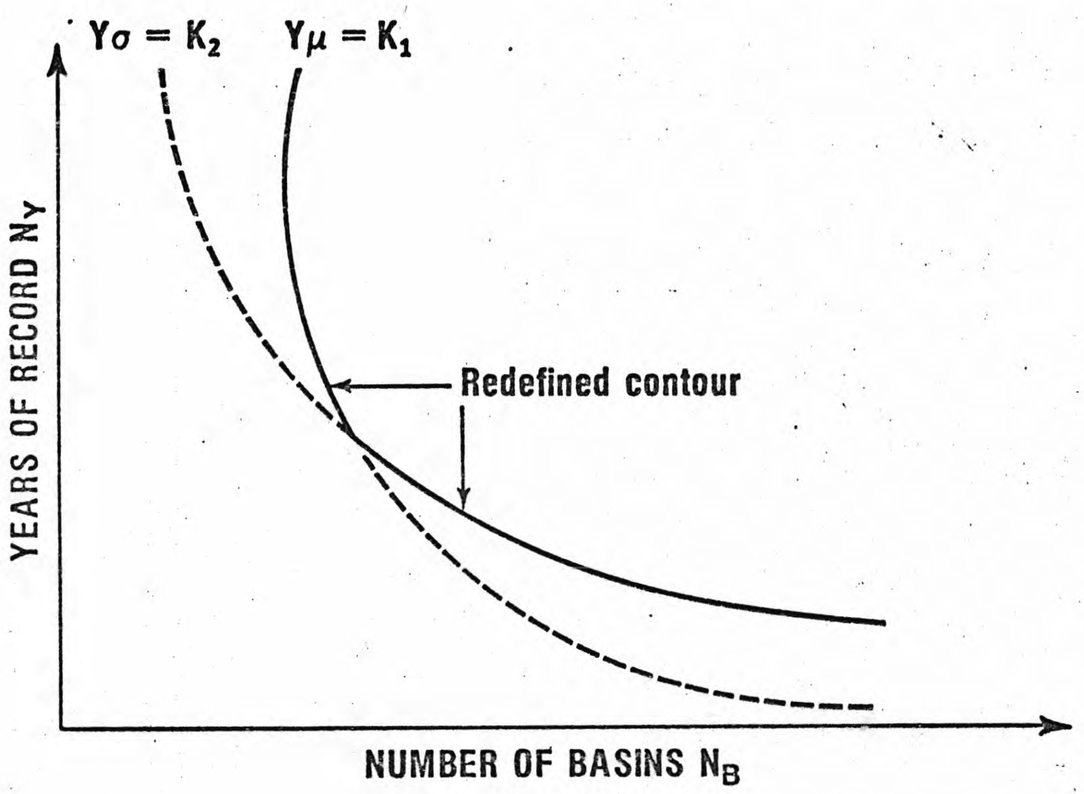

Figure 6. Redefinition of Information Production Function as a Result of Different Parameter Isopleth Intersection.

Determination of the proportionality factor of K_2/K_1 is an important part in a complete economic study of the worth of data. However, not only would the initial determination of the factor be a difficult task, but the proportionality probably will vary in the analysis as more data are used.

CHAPTER VI

EXAMPLE APPLICATION IN THE GREEN RIVER BASIN

The ongoing interest in the development of alternative energy sources is the basis for the particular example presented in this section and is that of the value of hydrologic data in the establishment of an oil-shale plant. The main consideration in the study is the statistical relationship between regional economic changes as a result of building the plant and the hydrologic information needed in the proper plant design.

The particular region chosen for the study is the Green River sub-basin of the Colorado River basin. The Green River sub-basin comprises the northern most part of the Colorado River basin encompassing the northwestern corner of Colorado, the southwestern portion of Wyoming, and the northeastern section of Utah. These three sections result in a total area of 46,000 square miles.

The political units within the sub-basin include three counties in Colorado, five in Utah, and four in Wyoming. There are no large cities in the region and the total population of the sub-basin is approximately 100,000 people, giving a small average population density of about 2.2 persons per square mile.

The principal areas of economic activity in the sub-basin are oil and gas extraction, agriculture, and mining - these three accounting for 45 per cent of the economy of the region (Udis 1967). Prior to the recognition of the energy crisis, a sharp decline in the mining sector had been noted. However, with the energy situation as it exists today, not only will such things as coal mining and gas extraction increase, but the oil-shale development is expected to strongly influence the economy of the area. Before proceeding with the analysis at a hypotehtical plant site, it is necessary to state two important assumptions on the economic side of the study. Although the econometric input-output model deals with the economy of the entire sub-basin, it is assumed that the transaction matrix and total gross output vector are representative of any smaller unit within the sub-basin. Secondly, the benefit study is assumed to be analyzed from a regional accounting stance. This is important because costs are not considered in terms of unemployment or lost revenue to other areas as a result of people and businesses moving into the region from the outlying areas. These costs would have to be accounted for if a study from a natioanl viewpoint were considered.

For purposes of illustrating the methodology presented in the previous chapters, a hypothetical oil-shale development is to be located on Lost Creek near Buford, Colo. It is assumed that the only limitation of the plant in terms of barrels of shale oil produced per day (BPD) is the availability of process water. This water will be withdrawn from Lost Creek via a diversion structure. The plant may use as much water in its production process as is furnished from the regional government. The design problem is to size the water system such that the resulting plant has diminished underutilization and underdesign over its life.

With a production function of the form of equation 1 (p. 5), a value of a_0 may be estimated to be 0.0426 (10^5 barrels/cfs-day), where Z is in 10^5 BPD and d is in ft^3/s (cubic feet per second). This value of a_0 was computed from production coefficients in the range of those described by the National Petroleum Council (1971). The relationship between water input and yearly plant output, Y_t (in 10^5 barrles/year) is given as

$$Y_t = \begin{matrix} 15.50\ X_t & X_t \leq d \\ 15.50\ d & X_t \geq d \end{matrix}$$

The coefficient of the above equation has units of 10^5 barrels/cfs-year. The updating included the addition of an oil-shale sector, with price levels adjusted to 1970 dollars. This modification together with the endogeneity of the household sector allows the coefficient of the benefit function to be computed directly from the $[I-A]^{-1}$ entry, which denotes the direct and indirect requirements on the household sector from the oil-shale sector. This value was calculated to be 0.11 (Udis and others, 1973) and this indicates that \$0.11 on each barrel of oil produced is added directly and indirectly to the economy of the region in terms of household income. Equation 7 can therefore be used to calculate β_0.

$$\beta_0 = \rho_{ij}\ 15.50 \times 10^5$$

$$= 0.11 \times 15.50 \times 10^5 + \$1.7 \times 10^5/\mathrm{ft}^3/\mathrm{s}\text{-year}$$

The benefit function in dollars per year to be used in subsequent calculations of the worth of hydrologic data is

$$B(X_t, d) = 1.7 \times 10^5\ X_t \qquad X_t < d$$

$$= 1.7 \times 10^5\ d \qquad X_t \geq d$$

Hydrologic Evaluation

Two regression models were used for the estimation of the hydrologic parameters of mean annual flow ($\hat{\mu}$) and the standard deviation of mean annual flow ($\hat{\sigma}$) for the potential site on Lost Creek. The regression model used for estimating the mean streamflow was a channel geometry model of channel width and depth versus flow as described by Hedman and others (1972). The regression model in esimating the standard deviation was a basin parameter model of drainage area and precipitation versus the standard deviation of flows. A portion of the gaged stations used in the latter regression is a subset of the stations used by Livingston (1970). The stations used were concentrated in the Colorado and Wyoming portions of the Green River sub-basin.

The regression estimates of $\hat{\mu}$ and $\hat{\sigma}$ were subsequently used in estimating the means of the prior distributions of mean and standard deviation, that is, for the θ set in equation 14. A bivariate normal distribution in log space was used for $p(\theta)$. The parameter values of the prior distribution were estimated to be 3.08 natural log units (referred to as only log hereafter) or 21.8 ft^3/s for the mean of the mean flow and 1.86 log units or 6.45 ft^3/s for the mean of the standard deviation of the flows.

The determination of an estimate of the standard deviation of the prior distribution entailed a design analysis of the network models for both μ and σ as described in Chapter V. The standard errors about the sample regression lines were translated into standard errors about the true regression models. These translations result in a distribution of true standard errors from which an expected value of true standard error can be calculated. These expected values are then used as estimates for the standard error (deviation) parameters in the prior distribution functions. The estimation procedure yielded a standard error of the mean of 1.19 log units and a standard error of the standard deviation of 1.34 log units.

The expected value of sample information can now be analyzed with the above definition of the prior probability distribution, together with the assumptions of log normality and independence for streamflow distribution.

Numerical Results

The bivariate normal prior distribution, $p(\theta)$, for the logs of mean flow and logs of standard deviation of flow, that is, $\{\theta\} = \{\log \mu_X, \log \sigma_X\}$ is of the form (Fisz, 1963)

$$p(\log \mu_X, \log \sigma_X = \frac{1}{2\pi\sigma_W\sigma_Z(1-R^2)^{\frac{1}{2}}} \quad x$$

$$\exp \quad \frac{1}{2(1-R)^2} \quad \frac{(W-\mu_W)^2}{\sigma_W^2} + \frac{(Z-\mu_Z)^2}{\sigma_Z^2} - \frac{2R(W-\mu_W)(Z-\mu_Z)}{\sigma_W\sigma_Z} \, dWdZ \qquad (24)$$

where $W = \log \mu_X$, $Z = \log \mu_X$, and R is the correlation coefficient between W and Z. The derivation of R is given on page 60.

The random variates from this distribution, however, cannot be used directly in the calculation of expected benefits because the log of the mean flows ($\log \mu_X$) does not necessarily equal the mean of the logs (and standard deviation of the logs) of flows ($\mu_{\log X}$), which are the distribution parameters of interest. To ensure consistency in the following calculations, a transformation is used to calculate the mean and standard deviation of the logs from the log of the means and log of standard deviation. The transformation can be found, for example, in Aitchison and Brown (1957);

$$\mu_{\log X} = \log \mu_X - \tfrac{1}{2} \log \left(1 + \left(\frac{\sigma_X}{\mu_X}\right)^2\right) \tag{25}$$

$$\sigma_{\log X} = \left(\log \left(1 + \left(\frac{\sigma_X}{\mu_X}\right)^2\right)\right)^{\frac{1}{2}} \tag{26}$$

It is these parameters upon which the logs of the flows experienced during the project life and during the years of additional data collection are conditioned.

The remaining item of a cost curve can be developed assuming a linear function defined by passing a line through the origin and a cost point for a prototype 100,000 BPD oil-shale plant (National Petroleum Council 1971). The particular extrapolated functions for capital and OM costs are illustrated in Figure 7. Converting the capital costs

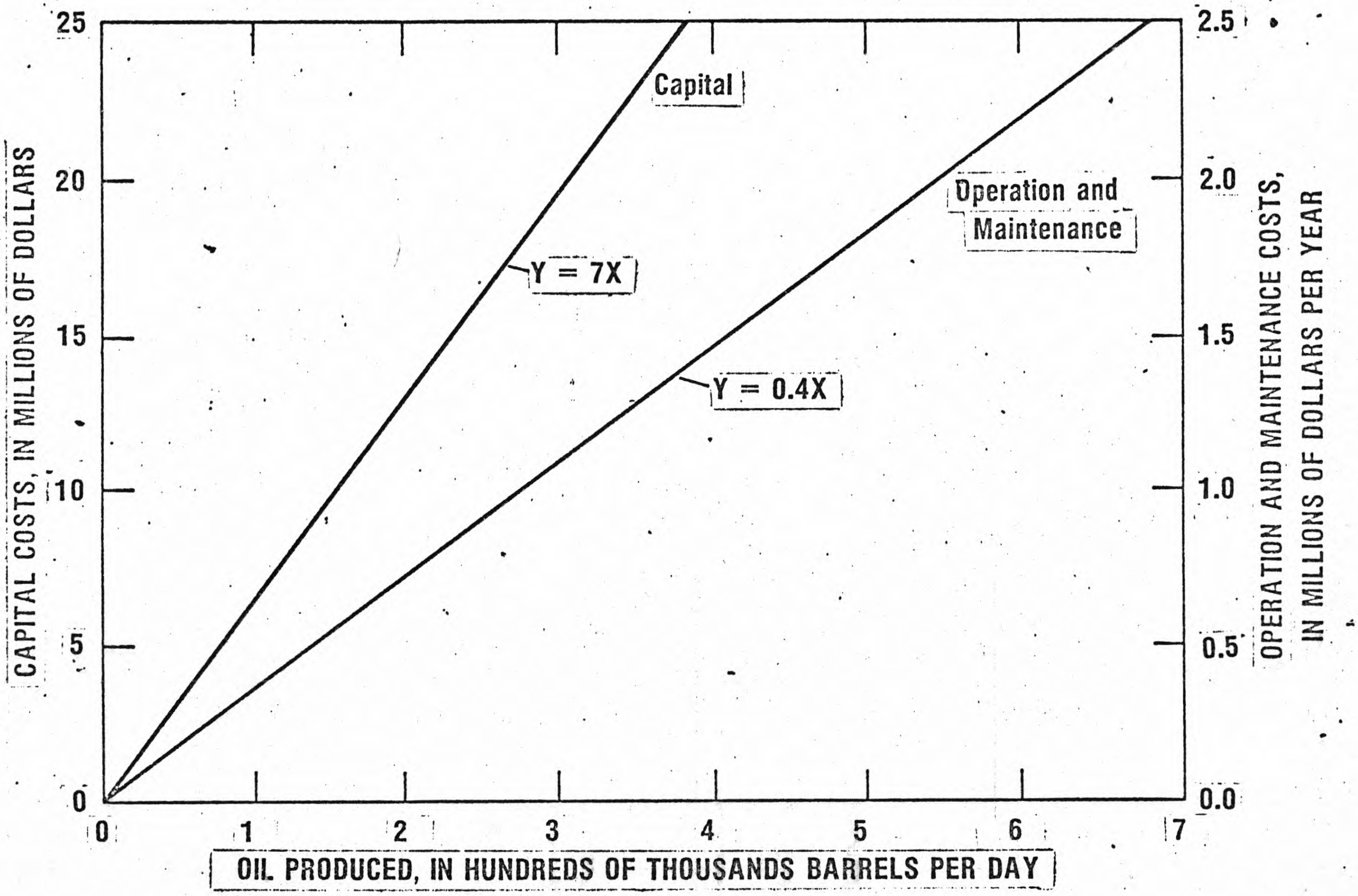

Figure 7. Cost Curves for an Oil Shale Water System.

into a yearly annual cost $(r = 0.12,\ L = 20)$ via the capital recovery factor and combining it with the OM cost yields a total yearly cost of \$60,500 per ft^3/s. An expression for expected net benefits can be obtained.

$$XB(d) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [1.7 \times 10^5 \int_0^d X' \cdot f(\log X' | \mu_{\log X'}, \sigma_{\log X'}) dX'$$

$$+ 1.7 \times 10^5 \int_d^{\infty} d \cdot f(\log X' | \mu_{\log X'}, \log X') \, d X'] p(\log \mu_X, \log \sigma_X)$$

$$d \log \mu_X \, d \log \sigma_X \tag{27}$$

and

$$XNB(d^*) = \int_X \max_d \, [XB(d) - 60,500 \, d] g(X) dX \tag{28}$$

The expected net benefits after n years of data collecting are

$$XNB(d^*) = \int_{X_n} \ldots \int_{X_i} \max \, [XB(d) - 60,500 \, d] g(X_i)$$

$$\ldots g(X_n) \, dX_i \ldots dX_n \tag{29}$$

where the $g(X_i)$ are defined in equation 15. Figure 8 illustrates the tightening of the distribution of the expected net benefits as a result of going from 1 to 2 years of additional data collection. The graphs depict the distribution of expected

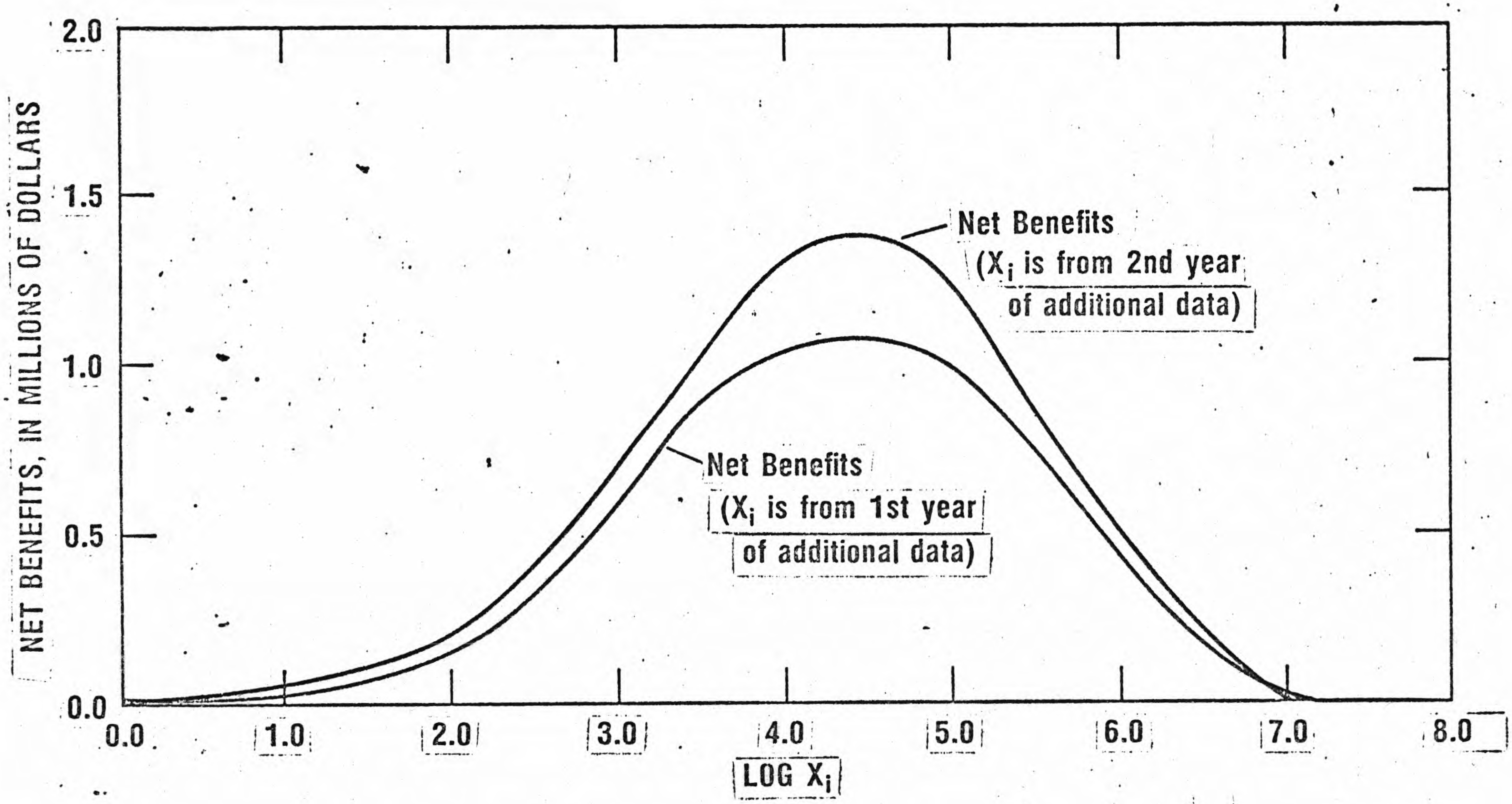

Figure 8. Benefit Distributions as a Function of Data.

net benefits as a function of the logs of possible values of streamflow X_i of the n^{th} year of data collection. (The taller curve has the effects of the first year of data implicitly contained in it.) Table 1 and Figure 9 show the expected net benefits after 1 and 2 years of additional data collection and the corresponding discounted net benefits. In this figure the costs of data collection are not considered. Moss (1970) has shown that the expected benefit function may be hyperbolic and approaching an asymptote by $\frac{1}{n}$ for each additional year of data collection. With this consideration in mind the projection of the expected benefits in Figure 9 and the corresponding discounting, the optimal length of data collection for this project appears to be two years. This projection is made because the costs of computer time in establishing the value of benefits after the second year of data collections are much greater than the value of obtaining the point, which would only confirm an obvious conclusion.

Alternatives for Obtaining Data

Mean Annual Streamflow

Figure 10 is the production function for information with regard to the true equivalent length of record for mean annual streamflow as described in Chapter V. The existing network from which the mean flow regression line was obtained is seen

TABLE 1

EXPECTED NET BENEFITS

(r = 12 percent)

n	Absolute	Discounted by δ^n
0	1,380,462	1,380,462
1	3,314,778	2,950,152
2	4,047,931	3,226,985

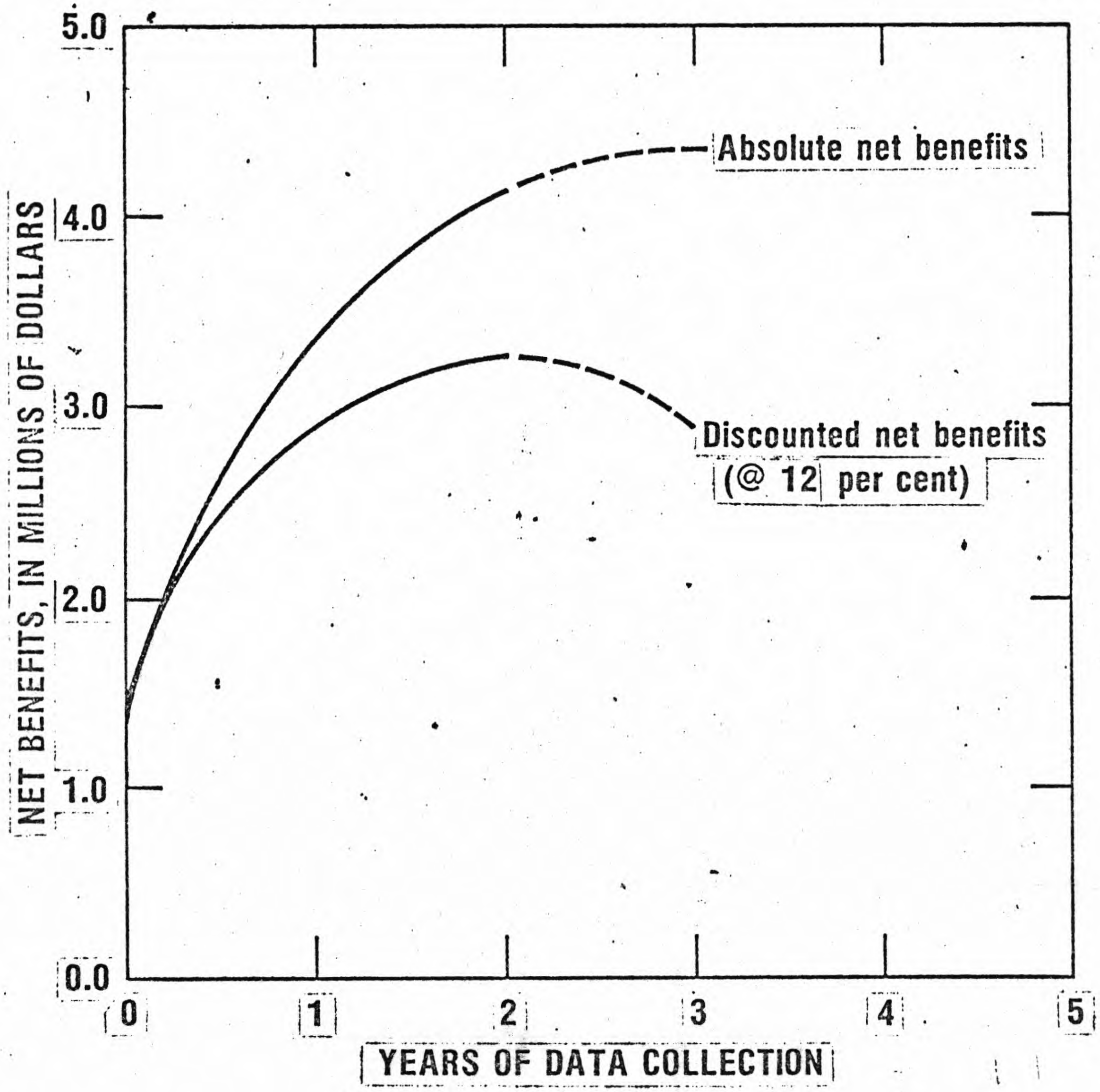

Figure 9. Net Benefits as a Function of n.

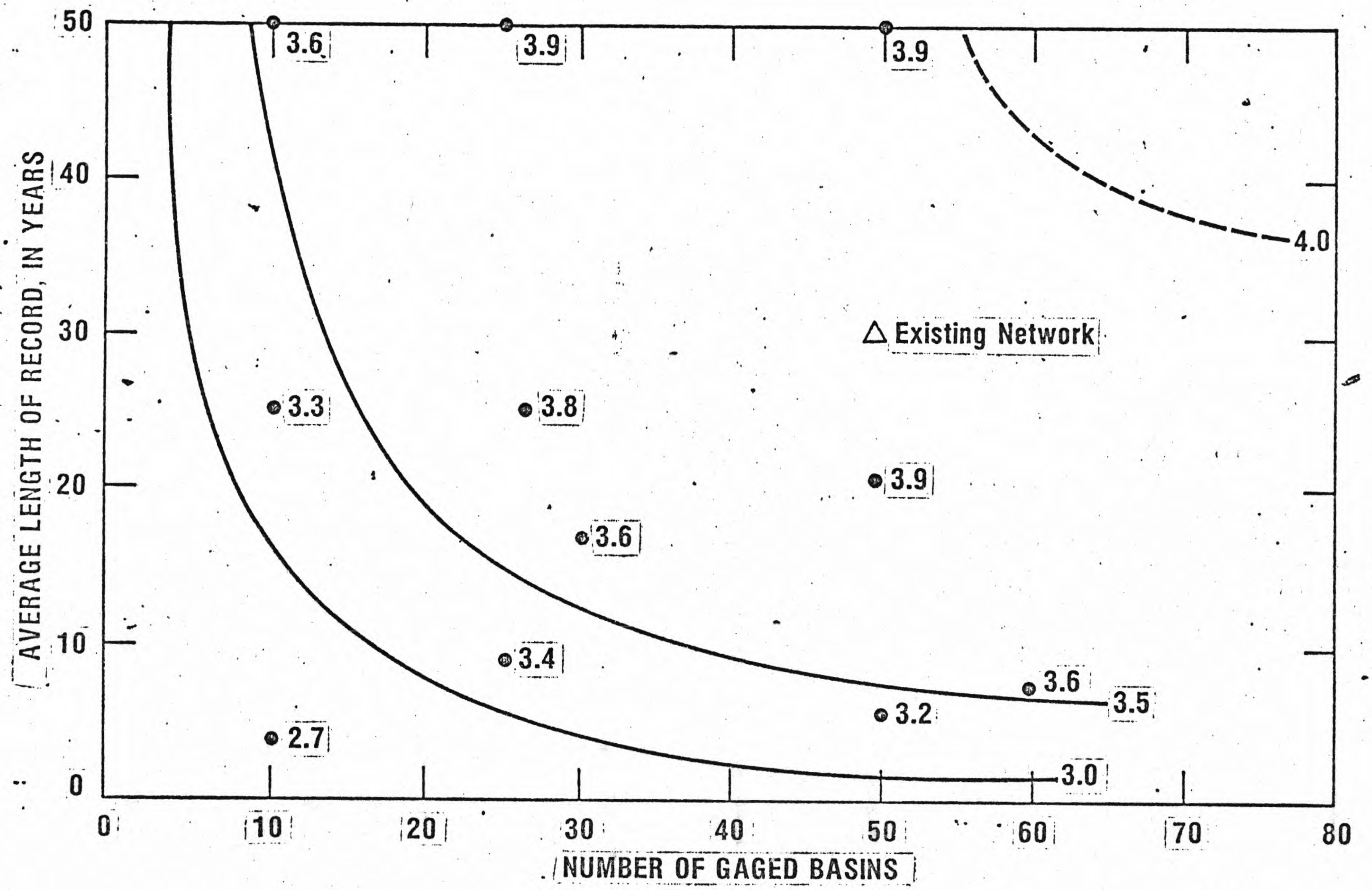

Figure 10. Production Function for Equivalent Record Information (Means).

to consist of 50 gaged sites each with an average length of record of 30 years. The true accuracy of the regression in terms of equivalent record length is somewhere between 3 and 4 years of record as is shown with the two isopleths. The 4-year isopleth is dashed to indicate that its location cannot be precisely located because of the flatness of the production function for greater numbers of gages and length of record.

To increase the present state of knowledge to an accuracy of one more additional year of equivalent record would require establishing at least one additional station. This can be easily seen because moving strictly vertically (increasing time sampling only) from the present network would not accomplish any increase in equivalent years, that is, the 4-year isoline would never be intersected. However, adding only one station is only a necessary but not sufficient condition to be met to increase the accuracy because the final network must be on the concave side of the 4-year isopleth, which requires several additional stations. Since this is a logical concession, and if the need for information lies at only the Lost Creek site, it seems reasonable and more economical to establish an at-site station. With this at-site station, direct estimates for the mean streamflow can be obtained, and these estimates can be pooled with those from the regression model:

$$\overline{X}_C = \frac{\overline{X}_R N_R + \overline{X}_G N_G}{N_R + N_G} \qquad (30)$$

where N_R and N_G are the regionalized and gaged lengths of record, respectively, $\overline{X}_R$ and $\overline{X}_G$ the respective individual estimates, and $\overline{X}_c$ the pooled estimate.

Therefore, the costs incurred to lengthen the period of record for estimating mean streamflow with the above option include the capital cost of establishing one station, the discounted operation and maintenance costs for the specified period, and the benefits foregone due to the time delay.

Standard Deviation of Annual Streamflow

The isopleths of actual equivalent year values of the standard deviation in Figure 11 indicate an existing accuracy of about ten years. There are more options for increasing the information content of the regression here than in the analysis of the mean, but the flattening of the contours immediately beyond the present network limit the amount of network improvement by additional stations. This consideration leaves three options for network modification to reach any desired information contour of equivalent years. The first option is to simply continue gaging at the existing sites, which would result in a vertical expansion path from the present network. An expansion path to the right and up, a second choice, would indicate building more gages and extending the record length at all gages. The final network modification choice, an expansion path to the left and up, results in discontinuing certain gages but extending the record at the remaining stations.

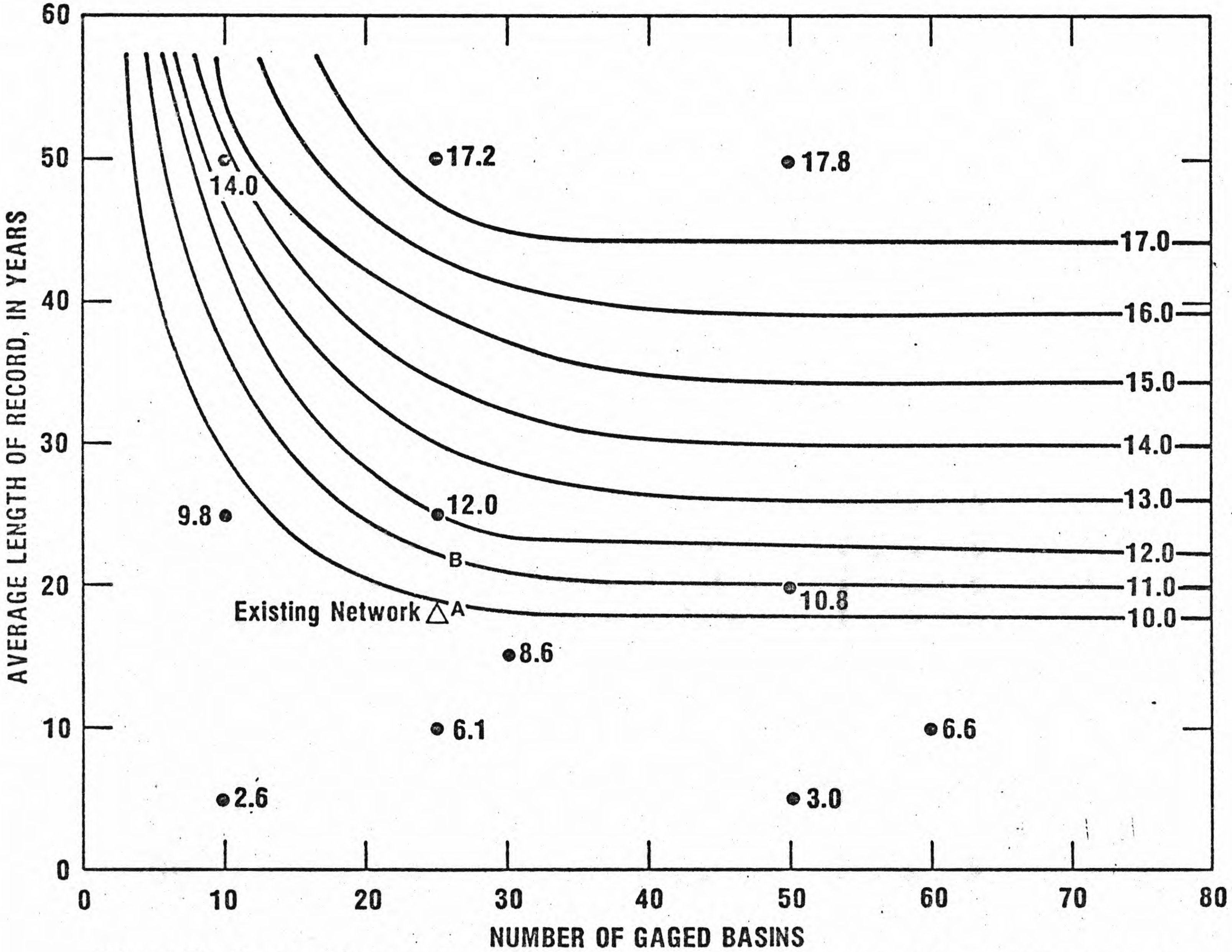

Figure 11. Production Function for Equivalent Record Information (Standard Deviations)

The ratio of capital cost to operation and maintenance cost will be a deciding factor as to which option is used for modifying the network, along with the effect of differing time frames in discounting when the trade-off of additional equivalent years versus calendar years to attain them is considered (Figure 11). If operation and maintenance costs are very small compared to capital cost, then it could be more economical to extend the record for the existing stations and perhaps discontinue some. If the OM costs are large, constructing new stations may be the least expensive option, because extending the record for 25 stations would become costly.

It has been estimated that the cost of installation in 1974 of a conventional float gage is $12,500 with an OM cost of $2,000 per year. (These figures are gross costs averaged over various sections of the country.) With these figures in mind, again the question must be asked if regional authorities want a uniform information increase for all ungaged sites or if the current project site is the only site where information is needed.

If the answer to the second question is yes, then the station cost figures would indicate that an at-site station would again be the most economical solution. It would roughly cost $85,000 to continue gaging the existing network of 25 stations for 2 calendar years to obtain one additional equivalent year (see Figure 11). Establishing an at-site station and gaging for two years would cost around $15,000. These 2 years of at-site record can, as in the case of the mean, be combined with the regionalized record length to acquire a pooled record length.[2] If it is the case that the entire network is gaged for two years to obtain an additional year of equivalent record, the benefits of the data will be discounted by two years thereby making the PVNB a smaller value. Another point to consider is that if the 25 stations are to be gaged regardless of a new project, the cost of data collection would be negligible and the only cost would be in terms of benefits foregone. With these considerations on the cost of data collection, the costs of establishing an at-site gage will do little to alter the optimum length of time for data collection, since these costs are small compared to the net benefits.

If it is the case that the decision-maker wishes to increase the record length by means of expanding the existing network, Figures 10 and 11 define the outcomes of possible alternatives for doing so, and although little can be done to increase the value for mean flow, several possibilities exist for lengthening the record for standard deviations. Again, if the existing network is modified, the least expensive choice is to gage 25 stations for an additional 2 years (points A to B in Figure 11)

---

[2]This can be shown to be a reasonable conclusion if the at-site data were taken long enough to get an estimate of the actual sample standard deviation, which is 2 years for an unbiased estimate.

which will cost approximately $85,000 (discounted). This type of adjustment will soon have an effect on the optimal length of record, which would begin to decrease because the cost of data collection becomes an important consideration in the decision process. This effect of the network expansion costs of decreasing the optimal length of record will be amplified by the cost of benefits foregone because the trade-off of calendar years for equivalent years from regionalization of the hydrologic data is greater than one for one.

TABLE 2

DISCOUNTED COSTS OF DATA COLLECTION

(r = 12 percent)

n	Costs of Data Collection (dollars)
0	0
1	13,780
2	15,360

CHAPTER VII

DISCUSSION AND CONCLUSIONS

The foregoing analysis is a direct application of a Bayesian preposterior statistical method to the problem of the worth of data. The objective of the study was to incorporate the uncertainty of hydrologic data with its economic consequences to a regional economy via this Bayesian scheme. The purpose was not to design an oil-shale plant or devise operating rules of a water-resources project and, therefore, the simplifying assumptions concerning the production function, water-system input, and cost analysis seem justified. The important point is that if such assumptions are not acceptable for a particular application, a detailed analysis can be done on each and the output from these analyses can be as easily assimilated by the Bayesian scheme as were the more simplifying assumptions.

An obvious extension to the study would be that of sensitivity analysis of the exogenous parameters. Parametric analysis of the $[I-A]^{-1}$ coefficient and costing sensitivity could be performed. A particularly sensitive parameter to analyze would be the interest rate.

Perhaps the biggest shortcoming in the methodology is the cost incurred in running the computer program. The cost itself follows a geometric progression in terms of added years of data, and this fact is the reason for projecting the third year of data instead of computing it for this particular study. Had the oil-shale plant been an actual project, the cost needed to confirm the third-year point might well have been justifiable. Other solutions for different parameters or projects may not be obvious after computer runs for 2 or 3 years of additional data collections, but this apparent shortcoming is indeed only that--apparent. As was stated in Chapter IV, the analysis of the worth of data should not end in a static conclusion. If the results say that 3 years of data collection are not yet optimal, the regional authorities need not spend vast amounts of computer funds to find the optimum. The procedure can be done in less expensive iterative steps; that is, the analysis can be redone after 2 or 3 years of actual data collection which has already been justified by the initial analysis. The 2 or 3 years of lead time given by the preposterior analysis should be sufficient for proper reevaluation of the data collection program with respect to its impact on the benefits from the proposed project. Eventually with the reevaluation and updating of parameters during these iterations, an eventual convergent solution could be arrived at. Therefore, the 3- or 4-year time frame for restaging analyses using this methodology limits the scheme to a short-term planning tool, and precludes its use as a long-term forecasting mechanism.

Selected References

Aitchison, J., and Brown, J. A. C, 1957, The Log-Normal Distribution: London, England, Cambridge Univ. Press, 176 p.

Benson, M. A., and Matalas, N. C.,1967, Symthetic Hydrology Based on Regional Statistical Parameters: Water Resources Research, 3(4), p. 931-935.

Davis, D. R., and Dvoranchik, W.M., 1971, Evaluation of the Worth of Data: Water Resources Bull., 7(4), p. 700-707.

Dawdy, D. R., Kubik, H. E., and Close, E. R., 1970, Value of Streamflow Data for Project Design--A Pilot Study: Water Resources Research, 6(4), p. 1045-1050.

Ferguson, Thomas S., 1967, Mathematical Statistics - A Decision Theoretic Approach: New York, London, Academic Press, 396 p.

Fisz, Marek, 1963, Probability Theory and Mathematical Statistics: New York, John Wiley and Sons, 677 p.

Hardison, C. H., 1969, Accuracy of Streamflow Characteristics in Geological Survey Research: U.S. Geol. Survey Prof. Paper 650-D, p. D210-D214.

Hedman, E. R., Moore, D. O., and Livingston, R. K., 1972, Selected Streamflow Characteristics as Related to Channel Geometry of Perennial Streams in Colorado: U.S. Geol. open-file report, 14 p.

Howe, C. W., and Cochrane, H. C., 1976, A Decision Model for Adjusting to Natural Hazard Events with Application to Urban Snow Storms: Review of Economics and Statistics, 58(1), p. 52-58.

Livingston, Russell K., 1970, Evaluation of the Streamflow Data Program in Colorado: U.S. Geol. Survey open-file report, 72 p.

Maddock, Thomas, III, 1973, Management Model as a Tool for Studying the Worth of Data: Water Resources Research 9(2), p. 995-1005.

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, 6(4), p. 1051-1061.

Moss, M. E. and Karlinger, M. R., 1974, Surface Water Network Design by Regression Analysis Simulation: Water Resources Research 10(3), p. 427-433.

National Petroleum Council, 1971, U.S. Energy Outlook-An Interim Report: An Initial Appraisal by the Oil Shale Task Group from 1971 to 1985, 122 p.

Pratt, V. W., Raiffa, H., and Schlaifer, R., 1965, Introduction to Statistical Decision Theory: New York, McGraw-Hill.

Raiffa, Howard, 1970, Decision Analysis-Introductory Lectures on Choices Under Uncertainty: Reading, Mass., Addison-Wesley, 309 p.

Turnovsky, Stephen J., Optimum Government Investment in Supply Systems Yielding Uncertain Output in Models for Managing Regional Water Quality: edited by Dorfman, R., Jacoby, H. D., and Thomas, H. A., Jr., Cambridge, Mass., Harvard Univ. press, 453 p.

Udis, B., 1967, An Analysis of the Economy of the Green River Sub-basin of the Colorado River Drainage Basin in 1960 with Emphasis on Heavy Water-Using Industries, Univ. of Colorado, 394 p.

Udis, B., Howe, C. W., and Kreider, J. F., 1973, The Inter-Relationship of Economic Development and Environmental Quality in the Upper Colorado River Basin: An Interindustry Analysis: Univ. of Colorado, a report to U.S. Dept. of Commerce, 642 p.

Zellner, Arnold, 1971, An Introduction to Bayesian Inference in Econometrics: New York, John Wiley and Sons, 430 p.

Derivation of $\rho_{\log \mu_X, \log \sigma_X}$.

The correlation coefficient between the log of the mean streamflows $(\log \mu_X)$ and the log of the standard deviation of the mean streamflow $(\log \sigma_X)$ needed in equation 23 is obtained using an estimated distribution of the coefficient of variation (C_v). The derivation is as follows:

$$C_v = \frac{\sigma_X}{\mu_X}$$

$$\log C_v = \log \sigma_X - \log \mu_X$$

$$\mathrm{Var}(\log C_v) = \mathrm{Var}(\log \sigma_X) + \mathrm{Var}(\log \mu_X) - 2\,\mathrm{Cov}(\log \sigma_X, \log \mu_X)$$

Solving for the covariance results in

$$\mathrm{Cov}(\log \sigma_X, \log \mu_X) = \frac{\mathrm{Var}(\log \sigma_X)+\mathrm{Var}(\log \mu_X)-\mathrm{Var}(\log C_v)}{2}$$

Dividing the covariance by the square root of the product of the variances yields:

$$\mathrm{Cor}(\log \sigma_X, \log \mu_X) = \rho_{\log \mu_X, \log \sigma_X}$$

$$= \frac{\mathrm{Var}(\log \sigma_X)+\mathrm{Var}(\log \mu_X)-\mathrm{Var}(\log C_v)}{2[\mathrm{Var}(\log \sigma_X)\ \mathrm{Var}(\log \mu_X)]^{\frac{1}{2}}}$$

Table A-1 depicts the estimated distribution of C_v. Because the prior probability distribution of the parameters $\log \mu_X$ and $\log \sigma_X$ are estimated using the regional regression analyses, the intervals of the C_v distribution are based upon those values of C_v chosen in the ϕ sets described in Chapter V. These population values of C_v on which priors are placed are 0.3, 0.5, 0.8, 1.0 and 1.5. The frequency counts of recorded values of C_v averaged over the network are distributed between these population values and thus define Table A-1. These values can be inserted into the above equation and the following result is obtained:

$$\frac{1.8 + 1.4 - .14}{2\ (1.34 + 1.2)} = \frac{3.06}{3.20} = 0.95$$

TABLE A-1

DISTRIBUTION OF COEFFICIENT OF VARIATION

C_v	$P(C_v)$	$\text{Log } C_v$
0.2	.857	- 1.61
.525	.107	- 0.64
.775	.036	- 0.26

$\mu_{\text{Log } C_v} = -1.457$ $\qquad$ $\sigma^2_{\text{Log } C_v} = .142$