Nordic Hydrology, 7, 1976, 57-72

Economic Value of Low Flow Data

Jens Chr. Refsgaard and Eggert Hansen
Technical University of Denmark,
Copenhagen

By means of the Bayesian decision approach the economic value of low flow data is considered with special reference to the design of treatment plants for typical Danish conditions. For a specific case the worth of primary data in form of direct observations of low streamflows has been investigated in terms of an expected opportunity loss *(EOL)*. Indirect information concerning the low flow properties of interest may be obtained by means of a secondary data set in combination with a regression model. In this study is suggested a heuristic method for the economic evalution of a secondary set of data taking into account the uncertainty embedded in the regression model.

Introduction

Low streamflows are of great importance. One of the reasons being that very often the streams are utilized as recipients for effluents from waste water treatment plants. As a certain minimum of dilution is required the critical period of the year is the time when low flows occur. Consequently low streamflows are crucial in the design of treatment plants discharging their effluents into streams. Ideally in this design the magnitude as well as the duration of low flows should be considered. Very often in practice, however, only the magnitude of low flows is taken into account. In Denmark e.g. the median value of the minimum daily discharge during each year is taken as one of the basic design quantities for treatment plants. Thus it is accepted that cases of dilution less than a certain critical value are experienced on the average every second year.

For at particular stream under investigation primary data in terms of direct observations of streamflow are often available only for a short period of time on the basis of which it is impossible to get more than a rather crude estimate of the relevant design factors. Thus the engineer, who has to design the treatment plant, is faced with a general problem of taking a particular action and making decisions under uncertainty. In this situation the engineer evidently has to consider the possibility of improving the basis of his decision either by postponing the project to collect additional primary data or by transmitting further information about the relevant design factor from a secondary set of hydrological data e.g. by means of a mathematical model.

The Bayesian decision approach is a rational statistical method for choosing and evaluating design alternatives when the »true« state of nature (the design factors) is unknown. Furthermore such an analysis makes it possible to evaluate in monetary terms the worth of primary data as well as secondary data. The effect of uncertainty is taken into account through the use of probability density functions for the hydrological quantities treated as uncertain. This type of decision theory focuses on the decision to be made and not on the hydrological parameters as the final results.

Within the field of hydrology the Bayesian approach has only been applied recently, but the interest is increasing. Besides a comprehensive study by Davis (1971), Davis et al. (1972) applied the Bayesian approach to a flood levee design problem. Jacobi (1974, 1975) considered the problem of economic value of sediment load data and optimum record length, respectively. Wood et al. (1974) applied the Bayesian approach to extreme hydrologic events while Vicéns et al. (1974) applied the Bayesian approach to hydrological time series modelling.

Theoretical Background

The statistical decision approach

According to Jacobi (1974) the different steps in the Bayesian decision procedure can be outlined as follows:

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

B) Form the goal function including the selection of variable describing »the state of nature«.

C) Derive stochastic properties of state variables.

D) Select best alternative by
   - i calculating the expected value of the goal function for each alternative, and
   - ii choosing the alternative which minimizes the expected value of the goal function.

E) Evaluate uncertainties and find the worth of additional data by
   - i determining the expected opportunity loss, *EOL* (due to uncertainty), and
   - ii finding the reduction in *EOL* either by
     - a) collecting more primary data by postponing the project, or
     - b) transmitting information from a secondary set of data by a regression model.

F) Find the net worth of additional information in monetary terms as the decrease in *EOL* minus the cost of obtaining this information.

The present study deals with the design of a wastewater treatment plant discharging its effluent into a stream. Thus the set of decision alternatives to be considered are the possible low flows of the stream M^{alt} to be used for the design. The hydrologic parameters treated as uncertain, the state variables, are the mean μ and the variance σ^2 of the annual minimum daily discharge Y. The goal function $G(M^{alt}|\mu,\sigma^2)$ is a penalty function. It indicates the excess cost that has to be paid because of either a realized overdesign or underdesign during the lifetime of the plant. We return in more detail to the construction of the goal function later on. The uncertainty about »the state of nature« is reflected by assigning a probability density function to the state variables μ and σ^2. Assuming that the annual minimum of daily discharge, Y, is a random stochastic variable following a normal distribution with parameters μ and σ^2, the joint a posteriori distribution of μ and σ^2 may be obtained by assuming a diffuse a priori knowledge. Given a set y_i of independent primary observations of Y the distribution is found to be

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

$$\left\{\sqrt{\frac{1}{2\pi\,\sigma^2/n}}\ \exp\left(-\frac{1}{2}\ \frac{(\mu-\bar{y})^2}{\sigma^2/n}\right)\right\}\left\{\frac{\left(\frac{n-1}{2}\right)^{\frac{n-3}{2}}}{\Gamma\left(\frac{n-3}{2}\right)}\ \frac{1}{s^2}\left(\frac{s^2}{\sigma^2}\right)^{\frac{n-1}{2}}\ \exp\left(-\frac{n-1}{2}\ \frac{s^2}{\sigma^2}\right)\right\} \tag{1}$$

where $\bar{y}$ and s^2 are the sample statistics

$$\bar{y} \equiv \frac{1}{n}\sum_1^n y_i \quad ; \quad s^2 \equiv \frac{1}{n-1}\sum_1^n (y_i-\bar{y})^2 \tag{2}$$

A decision is made by choosing the alternative M^*, that minimizes the expected value of the goal function

$$R(M^*) = \underset{M^{alt}}{\text{Min}} \iint G(M^{alt}|\mu,\sigma^2)\,f(\mu,\sigma^2|n,\bar{y},s^2)\,d\mu\, d\sigma^2 \tag{3}$$

$R(M^*)$ is called the Bayes Risk and M^* the Bayes Solution. If the »true« state variables μ_t and σ_t^2 were known the alternative chosen M_t would be the one that minimizes the goal function

$$G(M_t|\mu_t,\sigma_t^2) = \underset{M^{alt}}{\text{Min}} \{G(M^{alt}|\mu_t,\sigma_t^2)\} \tag{4}$$

Having used M^* instead of M_t an opportunity loss *(OL)* is the result because of the nonoptimal choice. The expected opportunity loss *(EOL)* is calculated as

$$EOL = \iint [G(M^*|\mu,\sigma^2) - G(M'|\mu,\sigma^2)]\, f(\mu,\sigma^2|n,\bar{y},s^2)\,d\mu d\sigma^2 \tag{5}$$

where M' is the design alternative that minimizes the goal function for given values of μ and σ^2.

It is seen that *EOL* is a function of the available data sample.

Additional information from secondary data

So far, having performed step *A* through *E.i*, primary data have been our only concern. As previously mentioned direct observations of low flows for a particular stream under investigation are very often sparse. It is not unusual that data for only a couple of years are available when the decision has to be made on the design of the treatment plant. One way to overcome the problem is by augmenting the primary data by means of a secondary set of hydrological data. The secondary set of data may be low flow data from a nearby stream for which direct observations of streamflow exists for a much longer period, or it may be calculated streamflows based on a rainfall-runoff model in combination with longer series of meteorological observations. A linear regression model in combination with the secondary set of data may be utilized for lengthening the primary set of data and thereby transmitting information. Immediately the following question turns up: What is the worth of the secondary set of data?

Matalas and Jacobs (1964) treated the question by comparing the estimates of the mean and the variance based on the short and the lengthened series, respectively. They showed that if the correlation coefficient, which measures the strength of the linear regression, exceeds about 0.5, then the estimates of the mean and the variance based on the lengthened series are better than the estimates based on the short series. For a specific problem related to the design of sediment control reservoirs Jacobi (1974) applied the Bayesian approach in an approximate calculation of the economic worth of a secondary data set of water discharges. The approximations made were, however, rather crude.

From the foregoing section it is recalled that an important feature in making economic evaluations is the probability distribution $f(\mu,\sigma^2 \,|\, n_1, \bar{y}_1, s^2{}_1)$ of the state parameters μ and σ^2 given a set of sample statistics based on a series of independent primary observations. From a strictly Bayesian point of view we ought to derive the revised distribution of μ and σ^2 based on the primary plus the secondary sets of data in order to take into consideration the experienced gain in information concerning the state paramters μ and σ^2. A revision without some unacceptable simplifying assumptions turns out to give raise to insurmountable calculations why this approach has to be abandoned.

In the present study a plausible approximate method for the calculation of the economic value of secondary data is suggested. The method is based on a heuristical combination of the Bayesian approach and the previously mentioned results obtained by Matalas and Jacobs (1964).

In order to account for the suggested approximate method for the revision of the distribution the observed set of primary and secondary data is considered

$$y_1, \ldots\ldots, y_{n_1}$$
$$x_1, \ldots\ldots, x_{n_1}, x_{n_1+1}, \ldots\ldots, x_{n_1+n_2}$$

Thus the primary data exist for n_1 years while the secondary data besides the same n_1 years are at hand for some other n_2 years. Utilizing a linear regression model based on the n_1 concurrent values the primary data are supplemented by n_2 calculated values according to the linear regression equation.

$$\hat{y}_i = \bar{y}_1 + b(x_i - \bar{x}_1) \; ; \quad i = n_1 + 1, \ldots\ldots, n_1 + n_2 \tag{6}$$

where the sample statistics are defined as follows

$$\bar{x}_1 = \frac{1}{n_1} \sum_1^{n_1} x_i \qquad s_{x_1}^2 \equiv \frac{1}{n_1 - 1} \sum_1^{n_1} (x_i - \bar{x}_1)^2$$

$$\bar{y}_1 = \frac{1}{n_1} \sum_1^{n_1} y_i \qquad s_{y_1}^2 \equiv \frac{1}{n_1 - 1} \sum_1^{n_1} (y_i - \bar{y}_1)^2$$

$$b = \frac{\sum_1^{n_1} (x_i - \bar{x}_1)(y_i - \bar{y}_1)}{(x_i - \bar{x}_1)^2} \qquad r = b \frac{s_{x_1}}{s_{y_1}}$$

b and r are estimates of the regression coefficient and the correlation coefficient ρ, respectively. For the primary data series lengthened by the n_2 calculated values obtained by Eq. (6) Matalas and Jacobs (1964) derived the following revised estimates for the state parameters μ and σ^2

$$\bar{y}_{1+2} \equiv \bar{y}_1 + \frac{n_2}{n_1 + n_2} b (\bar{x}_2 - \bar{x}_1) \tag{7}$$

$$s_{y_{1+2}}^2 = \frac{1}{n_1 + n_2 - 1} \Big[(n_1 - 1) s_{y_1}^2 + (n_2 - 1) b^2 s_{x_2}^2 + \frac{n_1 n_2}{n_1 + n_2} b^2 (\bar{x}_2 - \bar{x}_1)^2 + (n_2 - 1) \alpha^2 (1 - r^2) s_{y_1}^2 \Big] \tag{8}$$

where

$$\alpha^2 = \frac{n_2 (n_1 - 4)(n_1 - 1)}{(n_2 - 1)(n_1 - 3)(n_1 - 2)} \qquad \bar{x}_2 = \frac{1}{n_2} \sum_1^{n_2} x_{n_1 + i}$$

$$s_{x_2}^2 \equiv \frac{1}{n_2 - 1} \sum_1^{n_2} (x_{n_1 + i} - \bar{x}_2)^2$$

Furthermore they derived expressions for the variance of the estimates given by Eqs. (7) and (8)

$$\text{var}\{\bar{y}_{1+2}\} = \frac{\sigma^2}{n_1}\left[1 - \frac{n_2}{n_1+n_2}\left(\rho^2 - \frac{1-\rho^2}{n_1-3}\right)\right] \tag{9}$$

$$\text{var}\{s^2_{y_{1+2}}\} = \frac{2\sigma^4}{n_1-1} + \frac{n_2\sigma^4}{(n_1+n_2-1)^2}(A\rho^4+B\rho^2+C) \tag{10}$$

A, B, and C are constants given by rather lengthy expressions of n_1 and n_2 (Matalas and Jacobs. 1964), ρ is the linear correlation coefficient.

By comparing the variances given by Eqs. (9) and (10) with the corresponding expressions for the variances of estimates obtained on the basis of n independent observations

$$\text{var}\{\bar{y}\} = \frac{\sigma^2}{n} \qquad \text{var}\{s^2\} \equiv \frac{2\sigma^4}{n-1} \tag{11}$$

we may define an equivalent number of independent observations with respect to estimation of the mean

$$n_{em} = n_1 + n_{2em} = n_1\left[1 - \frac{n_2}{n_1+n_2}\left(\rho^2 - \frac{1-\rho^2}{n_1-3}\right)\right]^{-1} \tag{12}$$

and with respect to estimation of the variance

$$n_{ev} = n_1 + n_{2ev} = 1 + 2\left[\frac{2}{n_1-1} + \frac{n_2}{(n_1+n_2-1)^2}(A\rho^4+B\rho^2+C)\right]^{-1} \tag{13}$$

Firstly we realize that the equivalent number of independent observations with respect to the mean and the variance are in general different functions of n_1, n_2 and ρ. Secondly we realize that only when the right hand side of Eq. (12) is greater than n_1 the lengthened series contains more information about the mean than does the short series. Similarly we realize from Eq. (13) that only when the right hand side of this equation is greater than n_1 the lengthened series contains more information about the variance than does the short series.

The first immediate reaction in relation to the desired revision of the probability density function $f(\mu,\sigma^2|n_1, \bar{y}_1, s^2)$ due to the incorporation of n_2 secondary observations could be to substitute for $\bar{y}_1$ and $s^2{}_1$ the corresponding revised expressions given by Eqs. (7) and (8). At the same time n_1 should be substituted by an equivalent number of primary observations that contains as much information as n_1 primary observations plus n_2 secondary observations. As described above this last question has, however, no simple and unique answer. Thus n_2 secondary observations contain as much information about the mean as n_{2em} (Eq. (12)) primary observations while they contain as much information about the variance as n_{2ev} (Eq. (13)) primary observations where in general $n_{2em} \neq n_{2ev}$. In the following is suggested a heruristically based approach to circumvent the problem of how to make relevant substitutions for n_1 in the probability distribution $f(\mu,\sigma^2|n_1,\bar{y}_1,s^2{}_1)$.

We observe that the probability distribution given by Eq. (1) can be interpreted as a product of two independent distributions. The terms in the first bracket is the well-known Normal distribution for the sample mean with parameters $(\mu, \sigma^2/n)$. Except for a factor (n-3)/(n-1) close to unity the terms in the second bracket is the well-known Gamma distribution for the sample variance with parameters ((n-1)/2,2$\sigma^2/(n$-1)). These observations in combination with the possibility of calculating the equivalent number of independent primary observations corresponding to the sample mean n_{em} as well as the sample variance n_{ev}, (Eqs. (12) and (13)) contain the clue to the problem. By substituting n_{em} for n in the first bracket of Eq. (1) and n_{ev} in the second bracket we take approximately account of the different amount of information contained in a secondary data set of length n_2 with respect to the sample mean and the sample variance of the primary observations.

Thus in summary it is suggested in this study to arrive at a probability density function of the state variable μ and σ^2 given a set of n_1 primary data and $n_1 + n_2$ secondary data by substituting for $\bar{y}_1$ and $s^2{}_1$ in $f(\mu,\sigma^2 n_1, y_1, s^2{}_1)$ the improved expressions given by Eqs. (7) and (8) and at the same time substituting for n_1 different expressions (Eqs. (12) and (13)) dependent on the place of occurrence of n_1. Hereby the following revised density function is obtained

$$g(\mu,\sigma^2 | x_1, \ldots\ldots, x_{n_1+n_2}, y_1, \ldots\ldots, y_{n_1}) =$$

$$\sqrt{\frac{1}{2\pi\sigma^2/n_{em}}} \exp\left(-\frac{1}{2}\frac{(\mu-\bar{y}_{1+2})^2}{\sigma^2/n_{em}}\right) \frac{\left(\frac{n_{ev}-1}{2}\right)^{\frac{n_{ev}-3}{2}}}{\Gamma\left(\frac{n_{ev}-3}{2}\right)} \frac{1}{s^2_{y_{1+2}}}\left(\frac{s^2_{y_{1+2}}}{\sigma^2}\right)^{\frac{n_{ev}-1}{2}}$$

$$\cdot \exp\left(-\frac{n_{ev}-1}{2}\frac{s^2_{y_{1+2}}}{\sigma^2}\right) \qquad (14)$$

Having obtained this distribution it is possible to perform all the Bayesian risk calculations described in the foregoing section. By substituting $g(\ )$ for $f(\ )$ we may for instance by means of Eq. (5) calculate $EOL(n_1, n_2, \rho)$ corresponding to the situation with n_1 primary data and $n_1 + n_2$ secondary data, where the primary and secondary data are linearly correlated with a correlation coefficient ρ.

Having obtained $EOL(n_1,n_2, \rho)$ an equivalent number of primary observations n_e may be defined as shown in Fig. 1. Thus we may conclude that for the particular problem considered n_2 secondary observations are as valuable as $n_{2e} = n_e - n_1$ primary observations. It has to be emphasized that because the weighing is performed in economic terms the calculations depend on the applied goal function why the obtained results are not universal but intimately related to the specific case.

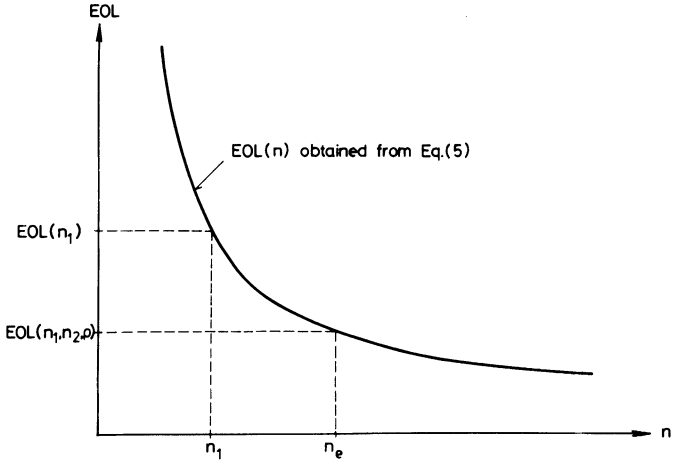

Fig. 1. Definition of the equivalent length of a secondary set of data, $n_2e=ne-n_1$, when $EOL(n_1,n_2,\rho)$ is obtained from Eqs. (5) and (14).

Case Study. Assumption and Results

Probability distribution of low flows

As previously mentioned the economic value of low flow data is considered in the present study with special reference to the design of treatment plants for typical Danish conditions.

Table 1

River	Time period	Median minimum (l/s)	Mean $\bar{y}$ (l/s)	Standard deviation s_y (l/s)	$s_y/\bar{y}$
Ringsted Å Middle Zealand	1950-59	100	113	50	0.44
Halleby Å Western Zealand	1932-59	143	162	64	0.40
Ryom Å Eastern Jutland	1933-59	180	178	75	0.42
Uggerby Å Northern Jutland	1931-59	347	358	118	0.35
Årup Å Jutland	1936-59	525	533	123	0.23
Guden Å at Egeballe Middle Jutland	1931-55	705	725	108	0.15

Low flow data for six Danish streams have been selected and their distributions examined. By plotting on normal probability paper the distributions have been found to be approximately normal. An example is shown in Fig. 2. Some main characteristics are given in Table 1.

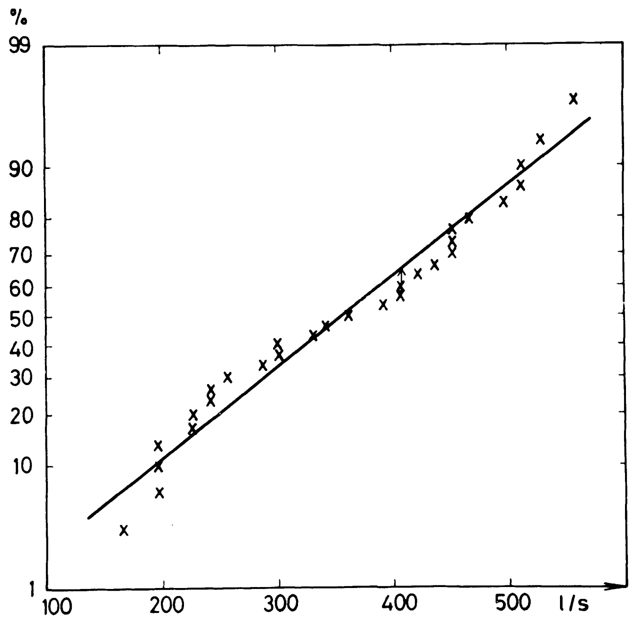

Fig. 2. Plotting of annual minimum streamflows for Uggerby Å 1931-59.

The goal function

The goal function is constructed on the basis of the diagram in Fig. 3, showing the cost of a wastewater treatment plant as a function of the design streamflow, the amount of wastewater, and the degree of purification. The diagram is prepared by civil engineer Jan Hassing, Cowiconsult and it is constructed under the following assumptions:

i mechanical-biological treatment,

ii one person equivalent *(p.e.)* is the same as 250 l/day with a BOD_5 concentration of 200 mg/l. The maximum hourly flow into the plant is assumed to be 1/12 of one day's flow,

iii the critical output concentration of BOD_5 from the plant is determined by mass balance calculations on the basis of the design discharge of the stream in combination with a critical BOD_5 concentration of 2 mg/l in the stream just downstream the outlet, as recommended by the Agency of Environmental Protection (1974).

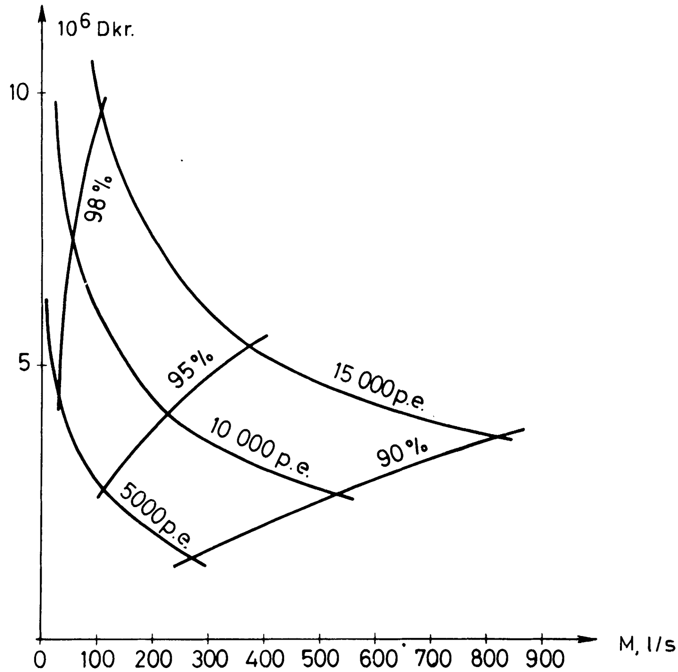

Fig. 3. Cost of sewage treatment plant (construction + capitalized working expenses). Interest rate i = 15% and design lifetime L = 30 years. prices in 1974 D.kr. (Prepared by civil engineer Jan Hassing, Cowiconsult).

The case considered in the present study is a community corresponding to 10.000 *p.e.* From the diagram in Fig.3 a cost function $COST(M, M^{alt})$ is constructed, showing the extra cost of the plant designed for a median minimum streamflow value M^{alt}, while the realized median minimum within the design lifetime, L, turns out to be M.

As stated earlier the annual minimum Y follows a normal distribution with parameters μ and σ^2. The distribution of the median M of L values of Y, $\varphi(M|\mu, \sigma^2)$, is shown by Gumbel (1958) to be approximately normal, with mean μ and variance $1.56\,\sigma^2/L$.

The goal function is the expected value of the cost function, given the state parameters μ and σ^2. The functional form is

$$G(M^{alt}|\mu,\sigma^2) = \int_0^\infty COST(M,M^{alt}) \cdot \phi(M|\mu,\sigma^2)\,dM \qquad (15)$$

By the construction of the cost function two cases have to be considered, namely underdesign and overdesign.

Case 1: Underdesign, $M^{alt} > M$

In this case more pollution is made than society can accept. Therefore the cost ought to be connected with the consequences of this extra pollution. This is, however, a very difficult problem (a study in itself) because tangibles as well as intangibles are involved. Instead $COST(M,M^{alt})$ is chosen somewhat arbitrarily as the cost by extending the plant to a capacity large enough to meet the requirements of society. These extra costs are said to be the difference between the costs of a plant designed for M^{alt} and a plant designed for M, plus an overhead of 20%.

Case 2: Overdesign, $M^{alt} < M$

In this case a bigger capital investment than necessary is made. Therefore $COST(M,M^{alt})$ is the difference between the costs of a plant designed for M and a plant designed for M^{alt}.

In this way $COST(M, M^{alt})$ is found as the curves shown in Fig.4. For computational purposes the cost function is wanted in analytical form so the cost function has been approximated by linear expressions as indicated in Fig.4.

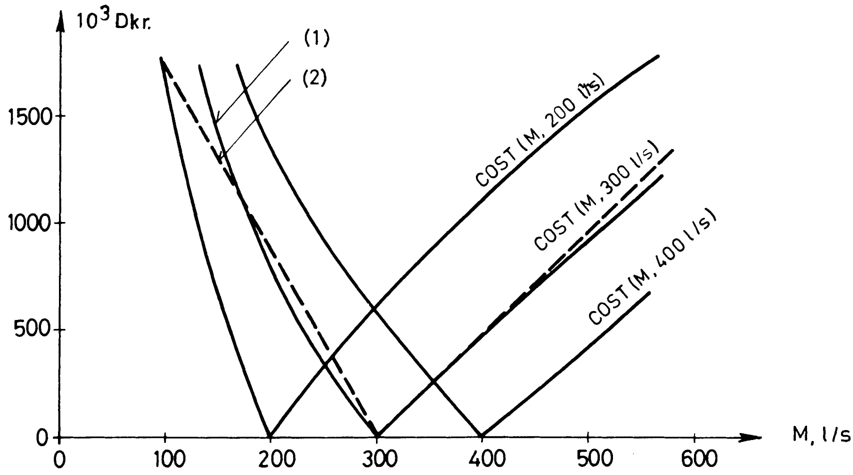

Fig. 4. The cost function cost (M,M^{alt}) showing the cost of either a realized overdesign or underdesign: (1) cost curve as computed from Fig. 3. (2) approximated cost curve used in the computations.

Results

The computations of Bayes Risk, Bayes Solution and EOL (Eqs. (1) - (5)) have been made on a digital computer. The necessary computer programs have been placed at our disposal by civil engineer Sven Jacobi, Ph.D. The following values of the sample statistics have been used

$$\begin{aligned} \bar{y} &= 300 \text{ l/s} \\ s^2 &= (60 \text{ l/s})^2,\ (105 \text{ l/s})^2,\ (150 \text{ l/s})^2 \\ n_1 &\in [6,50] \text{ years} \end{aligned} \qquad (16)$$

In Eq.(16) the range of ratio $s/\bar{y}$ is chosen to be 0.2-0.5 in accordance with typical Danish conditions (see table 1).

The results of the *EOL* computations are shown in Fig.5, where the importance of the sample length and the sample variance are seen.

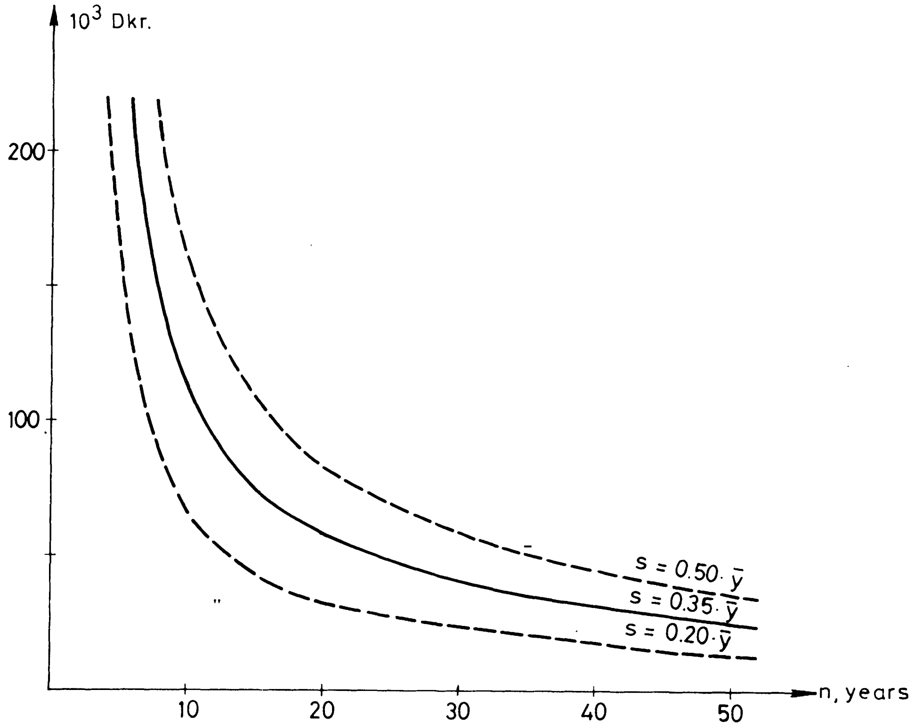

Fig. 5. *EOL* as a function of the length, n_1, of the primary set of data for different ratios between the sample mean ($\bar{y}$ = 300 l/s) and the sample variance, s^2.

By means of the approach outlined in the foregoing section similar computations have been made including secondary sets of data with the sample statistics

$$\begin{aligned} n_1 &= 10 \text{ years}; \qquad n_2 = 50 \text{ years} \\ \bar{y}_1 &= \bar{x}_1 = \bar{x}_2 = 300 \text{ l/s} \\ s^2_{y_1} &= s^2_{x_1} = s^2_{x_2} = (105 \text{ l/s})^2 \\ \rho &\in [0.30, 1.00] \end{aligned} \qquad (17)$$

The sample mean and the sample variance have without loss in generality been assumed to be identical in the primary and the secondary series. By use of the formulas derived in the theoretical section it is now possible to calculate the equivalent number of primary observations, n_{2e}, corresponding to n_2 secondary observations. n_{2e} is seen to be a function of n_1, n_2, and ρ. As an example Fig.6 shows n_{2em}, n_{2ev}, and n_{2e} as functions of ρ for given values of n_1, and n_2. It is seen that the equivalent length of a secondary set of data with respect to the mean, n_{2em}, is much greater than the equivalent length with respect to the variance, n_{2ev}. Besides it is realized that by the economic weighing n_{2em} is dominating relative to n_{2ev}. This last result is in good agreement with the study of Jacobi (1974) who found that information about the mean was more important than information about the variance.

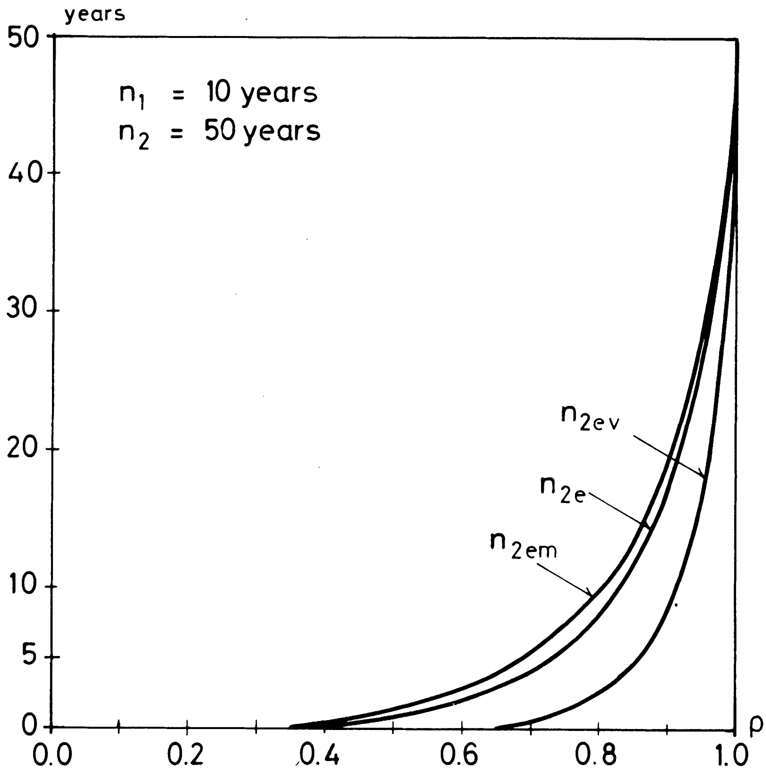

Fig. 6. Equivalent length of a secondary set of data as a function of the correlation ρ between primary and secondary data.

Using Figs. 5 and 6 it is now possible to find the worth of a secondary set of data $B(n_1,n_2,\rho)$ as a decrease in *EOL*

$$B(n_1,n_2,\rho) = EOL(n_1) - EOL(n_1,n_2,\rho) = EOL(n_1) - EOL(n_1+n_{2e}) \quad (18)$$

Thus for the specific case considered in this study corresponding to n_1 = 10 years of available primary observations supplemented by n_2 = 50 years of secondary observations it is observed that the worth of the secondary data depends heavily on the value of the linear correlation coefficient ρ, between primary and secondary data. The cost of obtaining the secondary information depends on which procedure one applies. Utilizing for instance a numerical rainfall-runoff model like the one developed by Nielsen and Hansen (1973) computations of daily values of streamflow may be performed after calibration of the model on the basis of input of daily values of precipitation and temperature together with monthly values of potential evapotranspiration.

The cost of collecting the necessary meteorological observations and performing the computer calculations has to be compared with the worth of the obtained information as given in Table 2. We realize that for the case considered it is economically sound to spend up to D.kr. 10,000 - 70,000 in obtaining secondary information depending on the correlation coefficient ρ.

Table 2

ρ	0.50	0.60	0.70	0.80	0.90
$B(10,50,\rho)$ (D.kr.)	10.200	19.400	34.100	51.200	70.800

In order to get an idea of the potential economic benefits of transmitting information on low flows by means of for instance a rainfall-runoff model let us consider a not too unusual situation where primary observations of minimum discharges are only available for n_1 = 3 years. If a treatment plant for 10,000 *p.e.* is built on the basis solely of this information of the low-flow conditions in the stream an *EOL* of approximately D.kr. 500,000 may result. Now unfortunately it is not possible to apply the technique of transmitting information from a secondary set of data by means of a linear regression model. This is due to the fact that a regression model based on as few as n_1 = 3 values contains by far too much uncertainty to transform any real information. Of course this does not mean that secondary data calculated by means of a rainfall-runoff model are without any value. If the applied model from experience with other streams is known to simulate low-streamflow values in an appropriate way it might be possible to transmit from previous meteorological observations an information equivalent to for instance 7 years of primary low-streamflow values. Utilizing again Fig. 5 we realize that the resulting *EOL* has decreased to approximately D.kr. 100,000. Thus a total decrease of approximately D.kr. 400,000 is experienced by utilizing a rainfall-runoff model in this case. This result indicates that the economic value of secondary data under certain circumstances may have a value of an order of magnitude larger than that given in Table 2.

Summary and Conclusions

Nonoptimal decisions due to lack of basic data are frequently occurring within the field of water resources management. In the present study low streamflows and their importance for the design of treatment plants have been considered and evaluated in economic terms by means of the Bayesian decision approach.

For a specific case the worth of primary data in form of direct observations of low streamflows have been investigated in terms of an expected opportunity loss (*EOL*) as a function of the number of primary observations.

Indirect information concerning the low-flow properties of interest may be obtained by means of a secondary set of data in combination with a regression model. In this study is suggested a heuristic method for the economic evalution of a secondary set of data taking into account the uncertainty embedded in the regression model. It is shown that corresponding to a situation with n_1 primary observations and $n_1 + n_2$ secondary observations an equivalent number of primary observations n_{2e} may be defined so that the n_{2e} primary data give raise to the same decrease in *EOL* as the n_2 secondary observations.

For typical Danish conditions it is shown that the economic value of secondary information concerning low-streamflow values as obtainable by means of a rainfall-runoff model lies in a very wide range of D.kr. 10,000 - 400,000 depending upon the available number of primary observations and the functioning of the rainfall-runoff model. If for instance 10 years of primary observations are available a rather moderately operating rainfall-runoff model may produce calculated secondary low-flow values worth approximately D.kr. 20,000. If on the other hand only 3 years of primary data are available a good rainfall-runoff model may produce calculated low-flow values worth approximately D.kr. 200,000.

References

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, Arizona.

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

Gumbel, E. J. (1958) *Statistics of extremes.* Colombia University Press, New York and London.

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

Jacobi, S. (1975) Economic optimum record length. *Nordic Hydrology, 6,* pp. 28-42.

Matalas, N. C. and Jacobs, B. (1964) A correlation procedure for augmenting hydrologic data. U.S. Geological Survey, Professionel paper 434-E, Washington.

Nielsen, S. A. and Hansen, E. (1973) Numerical simulation of the rainfall-runoff process on a daily basis. *Nordic Hydrology, 4,* pp. 171-190.

Vejledende bestemmelser for udledning af spildevand. Vejledning nr. 6/1974. København (Recommendations for discharge of wastewater. Agency of Environmental Protection).

Vicéns, G. J., Rodriguez-Iturbe, J., and Schaake jr., J. C. (1974) A Bayesian approach to hydrologic time series modelling. M.I.T. Ralph M. Parsons Laboratory for water resources and hydrodynamics. Report No. 181, Cambridge, Mass.

Wood, E. F., Rodriguez-Iturbe, J., and Schaake jr., J. C. (1974) The methodology of Bayesian inference and decision making applied to extreme hydrologic events. M.I.T. Ralph M. Parsons Laboratory for water resources and hydrodynamics. Report No. 178, Cambridge, Mass.

Received: 12 December, 1975

Address:
Institute of hydrodynamics and hydraulic engineering
ISVA
Technical University of Denmark
Bldg. 115,
DK-2800-Lyngby, Denmark