International Association of Scientific Hydrology. Bulletin
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/thsj18

A NOTE ON THE ESTIMATION OF THE PARAMETERS IN LOGARITHMIC STAGE-DISCHARGE RELATIONSHIPS WITH ESTIMATES OF THEIR ERROR

C. VENETIS

Published online: 30 Dec 2009.

Bulletin of the International Association of Scientific Hydrology, XV, 2 6/1970

A NOTE ON THE ESTIMATION OF THE PARAMETERS IN LOGARITHMIC STAGE-DISCHARGE RELATIONSHIPS WITH ESTIMATES OF THEIR ERROR

C. VENETIS (*)

ABSTRACT

Under certain assumptions the stage-discharge relationship of a channel cross-section can be approximated by a logarithmic relationship. Observational pairs of stage and discharge plotted on log-log paper often cluster around a straight line and this suggests that the assumptions involved are often approximately satisfied.

In such cases the parameters of the logarithmic relationship are usually estimated graphically from the position and slope of the straight line on the log-log paper. In this paper principles and methods are outlined for the estimation of the parameters with estimates of their standard error, via regression analysis. Because the water level of zero flows is usually one of the unknown parameters, the regression is non-linear and least squares optimal estimates can be obtained by a step-by-step approximation. The variances of the parameter estimates can be obtained from the dispersion matrix of the joint distribution of the least squares estimators via the likelihood function. An estimate of the error in predictions of the discharge depending on the corresponding stage may be obtained.

1. INTRODUCTION

Most uniform flow formulae in practical use can be expressed in the form $V = CR^v S^w$ where V is the mean velocity, C the roughness coefficient, R the hydraulic radius, S the slope and v, w are exponential coefficients. Thus with A denoting the cross section, the discharge Q is given by

$$Q = ACR^v S^w \tag{1}$$

and with $CS^w = K_1$, where K_1 is a constant, we have

$$Q = K_1 AR^v. \tag{2}$$

A and R are functions of the depth y of the water in the channel. When the cross section is of a simple geometrical form it may be assumed that approximately $AR^v = K_2 y^z$ where K_2 and z are constants and equation (2) becomes

$$Q = Ky^z \tag{3}$$

where K is a constant.

Thus under the assumptions of approximately uniform flow, cross-section of a simple geometrical form and roughness coefficient not varying appreciably with y, the discharge Q may be approximately expressed as a function of the water level y by a power relation of the form of equation (3).

(*) Karaiskaki 23, Arta, Greece.

In practice the water stage H is recorded with reference to the mean sea level. Denoting by H_0 the theoretical water level above sea level as $Q \to O$, equation (3) becomes

$$Q = K(H - H_0)^z. \tag{4}$$

The exact value of H_0 is in general not known.

In order to estimate the parameters of equation (4) a series of observational pairs (Q_i, H_i) are obtained by direct measurements and usually these are plotted on log-log paper for various values of H_0 around a rough estimate of this parameter. The value of H_0 which brings the plotted points nearest to a straight line is chosen as the best estimate of H_0. Then a straight line is fitted by eye and the slope of this line constitutes an estimate of the parameter z whereas an estimate of the parameter K can be obtained from the value of Q corresponding to $H - H_0 = 1$.

The plotted points define a straight line on the log-log paper only in a statistical sense as they may be expected to show a certain spread mainly because of errors in measuring the flow but also because equation (4) is an approximation, the flow may not be strictly uniform and the roughness coefficient may not be completely independent of the water depth. Thus the straight line drawn by eye on the log-log paper in the middle of points with a certain spread may deviate significantly from the line of best fit obtained by setting objective criteria for optimal estimates via regression analysis. The fact that the straight line is fitted on the logarithms of the quantities involved, makes often errors which may appear insignificant to the eye take appreciable proportions when the logarithms are inverted.

Even if one fits the straight line by regression and obtains optimal estimates of the parameters, these estimates will depend on the available record of observational pairs (Q_i, H_i). If one could have a series of such records one would obtain a sequence of optimal estimates of the parameters K, z, H_0 in general different from one an other. Our purpose is not only to obtain the best parameters which fit the limited record of observations but from this record to make inferences upon the values of the parameters in equation (4) which will be valid for all possible samples. Therefore it may be rewarding not only to estimate the parameters by regression but to investigate also the distribution of the estimators both as a check for the goodness of fit of the straight line and as a basis towards estimating the standard errors and when necessary confidence limits, of the estimated flow.

2. Estimation of the parameters via regression Variance of the least squares estimators.

Taking logarithms both sides of equation (4) and assuming that the error e of $\log Q$, is additive and serially independent with constant variance σ^2 we have the statistical relationship

$$\log Q = \log K + z \log (H - H_0) + e \tag{5}$$

Equation (5) is non-linear in the parameters, as H_0 which is to be estimated appears as an additive part in the argument of a logarithm, therefore standard methods of linear regression cannot be directly applied. A step-by-step approximation is recommended: H_0 can be assumed to be known and least squares optimal estimates of the parameters K and z can be obtained corresponding to the assumed H_0 via linear regression. By varying the assumed values of H_0 around a first estimate suggested by general experience of the channel we can optimise the regression with respect to H_0, K and z. That is, having tried a sequence of values of H_0 and computed optimal estimates of K, z and σ^2 ($\sigma^2 = \mathrm{Var}\,(e)$) corresponding to each assumed H_0, we compare all estimates of σ^2 and the set of estimates of the parameters which minimizes σ^2 constitutes the set of optimal estimates. The principles and methods of linear regression are well known and we will not enter into this here. (For an introduction to linear regression see for instance Smillie 1966).

The approach is illustrated in Table 1. From the results of linear regression for each assumed

H_{0j} we have

$$\hat{z}_j = \sum_{i=1}^{N} (\log Q_i - \overline{\log}\, Q)(\log (H_i - H_{0j}) - \overline{\log}\, H_j) \Big/ \sum_{i=1}^{N} (\log (H_i - H_{0j}) - \overline{\log}\, H_j)^2 \quad (6)$$

$$\hat{K}_j = \log^{-1} [\overline{\log}\, Q - \hat{z}_j \overline{\log}\, H_j] \quad (7)$$

$$\hat{\sigma}_j^2 = \sum_{i=1}^{N} ((\log Q_i - \overline{\log}\, Q)^2 - \hat{z}_j^2 (\log (\hat{H}_i - H_{0j}) - \overline{\log}\, H_j)^2)/N - 2 \quad (8)$$

where $\overline{\log}\, Q$ is the mean of $\{\log Q_i\}$, $\overline{\log}\, H_j$ is the mean of $\{\log (H_i - H_{0j})\}$ and N is the number of available observational pairs (Q_i, H_i). For each chosen H_{0j} from an assumed grid which depending on the accuracy with which one works may become as dense as it appears meaningful, the least squares optimal estimates $\hat{z}_j$, $\hat{K}_j$ and the estimate $\hat{\sigma}^2$ of the variance of the error are computed via eqns. (6) to (8) and tabulated as in Table 1. Then for some $j = n$ the column of $\hat{\sigma}^2$ will reveal that $\hat{\sigma}_n^2 > \hat{\sigma}_n^2$ for all j $\neq n$, and so the set $(H_{0n}, \hat{K}_n, \hat{z}_n)$ corresponding to the minimum variance $\hat{\sigma}^2$ will be the set of optimal estimates in the least squares sense. Thus denoting by $(\hat{H}_0, \hat{K}, \hat{z}, \hat{\sigma}^2)$ the set of optimal estimates we can write

$$(H_{0n}, \hat{K}_n, \hat{z}_n, \hat{\sigma}_n^2) = (\hat{H}_0, \hat{K}, \hat{z}, \hat{\sigma}^2) \qquad (\hat{\sigma}_n^2 < \hat{\sigma}_j^2 \text{ for } j \neq n) \quad (9)$$

TABLE 1

General arrangement of the approximation procedure for the step-by-step least squares estimation of the parameters of regression (5)

Chosen	Least sq. estimates		Estimate of the variance of the error
H_{0j}	$\hat{K}_j$	$\hat{z}_j$	$\hat{\sigma}_j^2$
H_{01}	$\hat{K}_1$	$\hat{z}_1$	$\hat{\sigma}_1^2$
H_{02}	$\hat{K}_2$	$\hat{z}_2$	$\hat{\sigma}_2^2$
H_{03}	$\hat{K}_3$	$\hat{z}_3$	$\hat{\sigma}_3^2$
.	.	.	.
.	.	.	.
.	.	.	.
H_{0n}	$\hat{K}_n$	$\hat{z}_n$	$\hat{\sigma}_n^2$
$H_{0,n+1}$	$\hat{K}_{n+1}$	$\hat{z}_{n+1}$	$\hat{\sigma}_{n+1}^2$
$H_{0,n+2}$	$\hat{K}_{n+2}$	$\hat{z}_{n+2}$	$\hat{\sigma}_{n+2}^2$

Optimal estimates of K, z and H_0 can be obtained essentially by linear regression via the above step-by-step approximation procedure. When we come to the estimation of the variances of the least-squares estimators we cannot use the methods given in standard texts on linear regression as H_0 is not known, we only have an estimate of this parameter. To obtain estimates of the variances of the estimators one must compute the dispersion matrix of the joint distribution of the estimators involved referring to the general theory of estimation (see Kendall and Stewart, 1961, and Dutsch, 1965).

The least-squares estimators under the assumption of normality of the error are also maximum likelihood estimators. The joint distribution of maximum likelihood estimators tends, under regularity conditions, to the multivariate normal distribution with dispersion matrix V the inverse of which is given in terms of the derivatives of the logarithm of the likelihood function.

In general if $\vartheta = (\vartheta_1, \vartheta_2, \ldots \vartheta_n)$ is the vector of parameters to be estimated and $\hat{\vartheta} = (\hat{\vartheta}_1, \hat{\vartheta}_2, \ldots \hat{\vartheta}_n)$ is its maximum-likelihood estimate, then the dispersion matrix V of the joint distribution of the maximum likelihood estimators is given by

$$V^{-1} = -\left(\frac{\partial^2 l}{\partial\vartheta_r \partial\vartheta_q}\right)_{\vartheta=\hat{\vartheta}} \qquad (r, q = 1, 2, \ldots n) \tag{10}$$

where $l = \log L$ and $L = L(\vartheta)$ is the likelihood function.

In our case log Q_i are assumed to be independent and normally distributed with constant variance $\hat{\sigma}^2$. The parameters estimated by maximum-likelihood estimators (= least-squares estimators) are log K, z, H_0, σ^2. We have

$$L(\log Q_i\,;\, \log K, z, H_0, \sigma^2) = \prod_{i=1}^{N} \phi(\log Q_i\,;\, \log K, z, H_0, \sigma^2) \tag{11}$$

where N denotes the number of observational pairs (Q_i, H_i) and ϕ (.) denotes the normal density which substituted in equation (11) gives

$$L = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(\log Q_i - \log K - z\log(H_i - H_0))^2\right) \tag{12}$$

and with $l = \log L$ and N denoting the number of observations

$$l = -N\log(2\pi\sigma^2)/2 - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(\log Q_i - \log K - z\log(H_i - H_0))^2 \tag{13}$$

Differentiating expression (13) with respect to all parameters in pairs as equatiion (10) suggests and substituting the maximum likelihood estimates of the parameters we obtain from equation (10) the matrix V^{-1} and by inversion the dispersion matrix V. Since the estimates of the parameters maximize the likelihood function the matrix $(\partial^2 l/\partial\vartheta_r\partial\vartheta_q)_{\vartheta=\hat{\vartheta}}$ must be negative definite. This may be a useful check of the computations.

The diagonal elements of the dispersion matrix give the variances of each estimator whereas the off-diagonal elements give the covariances between each pair of estimators.

Substituting in the general eqn. (10) the parameters involved in the present case we have

$$V^{-1} = -\begin{bmatrix} \dfrac{\partial^2 l}{\partial \log K^2} & \dfrac{\partial^2 l}{\partial \log K\,\partial z} & \dfrac{\partial^2 l}{\partial \log K\,\partial H_0} & \dfrac{\partial^2 l}{\partial \log K\,\partial(\sigma^2)} \\[2ex] \dfrac{\partial^2 l}{\partial \log K\,\partial z} & \dfrac{\partial^2 l}{\partial z^2} & \dfrac{\partial^2 l}{\partial z\,\partial H_0} & \dfrac{\partial^2 l}{\partial z\,\partial(\sigma^2)} \\[2ex] \dfrac{\partial^2 l}{\partial \log K\,\partial H_0} & \dfrac{\partial^2 l}{\partial z\,\partial H_0} & \dfrac{\partial^2 l}{\partial H_0^2} & \dfrac{\partial^2 l}{\partial H_0\,\partial(\sigma^2)} \\[2ex] \dfrac{\partial^2 l}{\partial \log K\,\partial(\sigma^2)} & \dfrac{\partial^2 l}{\partial z\,\partial(\sigma^2)} & \dfrac{\partial^2 l}{\partial H_0\,\partial(\sigma^2)} & \dfrac{\partial^2 l}{\partial(\sigma^2)^2} \end{bmatrix} \tag{14}$$

It may be useful for applications to give here the double partial derivatives involved in eqn. (14). Eqn. (13) differenciated with respect to the parameters (log K, z, H_0, σ^2) gives:

$$\frac{\partial^2 l}{\log K \, \partial z} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \log (H_i - H_0)$$

$$\frac{\partial^2 l}{\partial \log K \, \partial H_0} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{z}{H_i - H_0}$$

$$\frac{\partial^2 l}{\partial \log K \, \partial(\sigma^2)} = -\frac{1}{(\sigma^2)^2} \sum_{i=1}^{N} (\log Q_i - \log K - z \log (H_i - H_0))$$

$$\frac{\partial^2 l}{\partial \log K^2} = -\frac{N}{\sigma^2}$$

$$\frac{\partial^2 l}{\partial z \, \partial H_0} = \frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{-1}{H_i - H_0} (\log Q_i - \log K - 2z \log (H_i - H_0))$$

$$\frac{\partial^2 l}{\partial z \, \partial(\sigma^2)} = -\frac{1}{(\sigma^2)^2} \sum_{i=1}^{N} (\log Q_i - \log K - z \log (H_i - H_0)) \log (H_i - H_0)$$

$$\frac{\partial^2 l}{\partial z^2} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} (\log (H_i - H_0))^2$$

$$\frac{\partial^2 l}{\partial H_0 \, \partial(\sigma^2)} = \frac{1}{(\sigma^2)^2} \sum_{i=1}^{N} \frac{z}{H_i - H_0} (\log Q_i - z \log (H_i - H_0) - \log K)$$

$$\frac{\partial^2 l}{\partial H_0^2} = -\frac{1}{\sigma^2} \sum_{i=1}^{N} \frac{z}{(H_i - H_0)^2} (z + \log Q_i - z \log (H_i - H_0) - \log K)$$

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{N}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3} \sum_{i=1}^{N} (\log Q_i - \log K - z \log (H_i - H_0))^2$$

The amount of computational work is such that for any real-scale applications the use of digital computers is absolutely essential.

3. Estimation of the error in the computed discharge due to the variances of the estimators

An estimate of the standard error of the computed discharge via equation (5) can be obtained from the estimates of the variances and covariances of the estimators of the parameters log K and z assuming $\hat{H}_0 = H_0$. To each value $y = H - H_0$ there corresponds a value of Q given by the relationship

$$\log Q = \log K + z \log y \tag{15}$$

We may consider the error e in equation (5) as originating mainly because of errors of measurement. The exact values of $\log K$ and z are not known. To predict $\log Q$ when y is given the estimates of these parameters will be used; thus we will have

$$\widehat{\log Q} = \widehat{\log K} + \hat{z} \log y \tag{16}$$

and the resulting error will be

$$(\widehat{\log Q} - \log Q) = (\widehat{\log K} - \log K) + (\hat{z} - z) \log y \tag{17}$$

Taking the expectation both sides of equation (17) squared we have

$$E(\widehat{\log Q} - \log Q)^2 = E((\widehat{\log K} - \log K) + (\hat{z} - z) \log y)^2 \tag{18}$$

or

$$\mathrm{Var}(\widehat{\log Q}) = \mathrm{Var}(\widehat{\log K}) + \log^2 y \,\mathrm{Var}(\hat{z}) + 2 \log y \,\mathrm{Cov}(\widehat{\log K}, \hat{z}) \tag{19}$$

Substituting in equation (19) the estimate of the variances of the parameters involved and any value of y, we obtain the variance of the error in the estimate of $\log Q$, and hence the standard error of the estimate Q corresponding to that value of y and confidence limits. It is worth while to note that the standard error in the estimate of Q depends on y.

Since $\widehat{\log Q}$ is a linear function of $\widehat{\log K}$ and $\hat{z}$ which are jointly normal-distributed, $\widehat{\log Q}$ will have the normal distribution with mean $\log Q$ and variance given by equation (19). From tables of the distribution of the standard normal variate we find that

$$\mathrm{Prob}\left\{1.96 \leqslant \frac{\widehat{\log Q} - \log Q}{\sqrt{\mathrm{Var}(\widehat{\log Q})}}\right\} = \mathrm{Prob}\left\{-1.96 \geqslant \frac{\widehat{\log Q} - \log Q}{\sqrt{\mathrm{Var}(\widehat{\log Q})}}\right\} = 5\% \tag{20}$$

and from equation (20) we can compute the 95 % confidence limits.

The error of the estimate $\widehat{\log Q}$ obtained by equation (16) is due to the error in the estimated parameters $\log K$ and z therefore in a series of predictions of $\log Q$ the error will not be random. For instance if $\widehat{\log K} > \log K$ then all predicted flows will be over-estimated in proportion to $\hat{K}/K$. We can be reasonably confident that over—or under—estimation of the flows will not exceed the 95 % confidence limits.

Figure 1 presents the general form of a logarithmic stage—discharge relationship with 95 % confidence limits.

Acknowledgements

This paper was essentially written while the writer was working at the City University in the Civil Engineering department on a sponsored project during 1968-1969 where he applied these ideas to practical problems. Miss J. Scott, mathematician at The City University, who did all the computer programming in connection with the application of the approach, discussed at several occasions with the writer topics on the theory of estimation and kindly checked some of the formulae that appear in the paper. The writer is also indebted to his collegue Mr. D. Me-

goulas of the Regional Direction of Public Works at Preveza for reading and discussing the draft and to the Director of the Regional Direction Mr. A. Artoumas for encouragement in connection with the writer's academic interests.

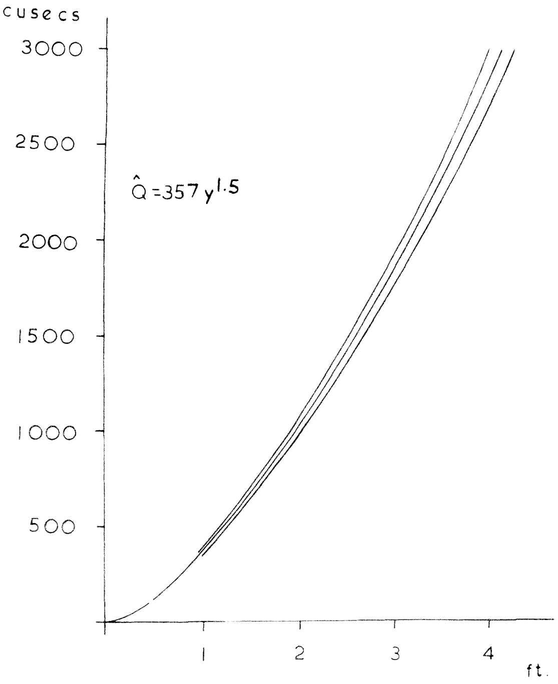

Fig. 1 — Stage-discharge curve with 95 % confidence limits.

REFERENCES

(1) CHOW, V. T. (1959), *Open-channel hydraulics*, McGraw Hill New York.
(2) SMILLIE, A. W. (1966), *An introduction to regression and correlation*, Academic Press, London.
(3) BLACKETT, R. L. (1960), *Principles of linear regression analysis*, Clarendon Press, Oxford.
(4) KENDALL, M. G. and STEWART, A. (1961), *The advanced theory of statistics*, vol. 2, Charles Griffin, London.
(5) DUTSCH, R. (1965), *Estimation theory*, Prentice Hall, New York.
(6) VENETIS C. (1969), A study for the determination of satisfactory estimates of the stage-discharge relationship of the River Leven (Old Recorder) consistent with all available information, Report No. 7/17.4.69. The City University.
(7) VENETIS C. (1969), Additional work for the determination of optimal stage-discharge relationships of the Rivers of the Morecambe Bay Catchment, Report No. 11/26.8.69, The City University.