

A study of the stage-discharge relationship of the Okavango River at Mohembo, Botswana

F. T. K. SEFE

Department of Environmental Science, University of Botswana, Private Bag 0022, Gaborone, Botswana

Abstract The stage-discharge relationship or rating curve at a river cross-section is a fundamental technique in hydrology employed for determining discharge from catchments. While river cross-sections are inherently variable, it generally takes quite some time for a change in cross-section to occur. Thus, where good quality data on stage and discharge are available, a stable and representative rating curve may be established. However, all too often, the natural variability is aggravated by data errors of human origin. These measurement errors occur because the routine of periodic flow measurement may be undertaken by ill-trained personnel. Such errors resulting from this practice may lead to erroneous estimates of available water resources. This study discusses the derivation of a single rating curve for the Okavango River at Mohembo from measurements containing errors and outliers. It uses data transformation procedures, regression techniques and removal of outliers to derive a useful rating curve from suspect data. Finally, it investigates the possibility of fitting a stochastic model to the discharge series. The results suggest that it may be possible to estimate missing data with an ARIMA model.

Etude du rapport entre la hauteur et le débit du fleuve d'Okavango à Mohembo

Résumé Etablir pour une section de rivière le rapport entre la hauteur et le débit, c'est à dire la courbe de tarage, est en hydrologie une technique fondamentale destinée à évaluer les débits provenant des bassins hydrographiques. Même si les sections transversales des rivières sont essentiellement variables, il faut généralement un certain temps pour que ces modifications se manifestent. Là où des données de hauteur et de débit de bonne qualité seront disponibles, une courbe de tarage stable et représentative pourra être établie. Très souvent cependant, des erreurs d'origine humaine s'ajouteront à la variabilité naturelle. Ces erreurs sont dues aux mesures de débits réalisées en routine par un personnel insuffisamment formé et ce type d'erreur peut conduire à une estimation erronée des ressources en eau utilisables. La présente étude s'attache à tenter de comprendre les causes d'instabilité de la courbe de tarage du fleuve Okavango à Mohembo. Elle utilise des procédures de transformation de données, des techniques régressives et d'élimination des valeurs aberrantes afin d'établir une courbe de tarage utile à partir de données suspectes. Finalement, elle examine en détail les possibilités de réaliser une modélisation stochastique des séries de débits. Les résultats suggèrent la possibilité d'estimer les données manquantes à partir d'un modèle "ARIMA".

INTRODUCTION

The stage-discharge ($H-Q$) relationship is a fundamental technique employed in discharge calculation. Typically the relationship is established from periodic measurements of stream discharge and corresponding water surface elevation or stage. While recent developments (Gawne & Simonovic, 1994) may take the tedium out of rating curve determination, the fact still remains that a good and representative rating curve requires quality data. The $H-Q$ relationship at a particular river cross-section, even under conditions of meticulous observation, is not necessarily unique as rivers are often influenced by factors not always understood nor easy to quantify. This natural variability in the $H-Q$ relationship is often aggravated by considerable errors of human origin.

The Okavango River system in Botswana has, over the years, attracted considerable interest as a major source for water and other ecological resources. Recently a major water scheme focused on this system had to be abandoned by the Botswana Government as a result of widespread concern about the impact of the project on the unique ecosystem of the area (SMEC, 1987; IUCN, 1992). However, considering the projected increases in water demand in northeast Botswana, it is only a matter of time before the project is resurrected again. Undoubtedly, the management of any resource abstraction scheme in a unique environment such as this depends on a thorough understanding of the working of the entire system. In this regard, accurate discharge data on various time scales are required for resource assessment, environmental monitoring and flood forecasting. Given the primary importance of river gauging activities to discharge data gathering, this study focused on stage-discharge relationships for the Okavango at Mohembo which is the point where the flow from the Cubango system enters the Okavango Delta system.

The study reported here emanated from concern within the Department of Water Affairs, the government department responsible for the collection of hydrological data in Botswana, about the existence of "inexplicable" outliers in the stage-discharge plot. Outliers occur all too often in discharge data. However, in this particular case, huge discharge values were recorded which corresponded to stages for which much lower discharges were previously associated. Moreover, the records did not show any corresponding change in velocity or cross-section. Considering the fact that gauging at this remote site is largely carried out by inadequately trained staff, it seems likely that human error could be one of the causes of the occurrence of the outliers. This, however, cannot be assumed. The aim of this study, therefore, is to investigate the reason for the occurrence of these outliers and how a useful rating curve can be extracted from the available data. Given the absence of any information in the data files regarding change in cross-section or occurrence of unusually large floods, a single rating curve may be more suitable than multiple rating curves representing different flow regimes. The study describes and characterizes the nature of the $H-Q$ relationship, and then investigates the use of regression methods to verify the existence of errors in the data and to derive

INTRODUCTION

The stage-discharge ($H-Q$) relationship is a fundamental technique employed in discharge calculation. Typically the relationship is established from periodic measurements of stream discharge and corresponding water surface elevation or stage. While recent developments (Gawne & Simonovic, 1994) may take the tedium out of rating curve determination, the fact still remains that a good and representative rating curve requires quality data. The $H-Q$ relationship at a particular river cross-section, even under conditions of meticulous observation, is not necessarily unique as rivers are often influenced by factors not always understood nor easy to quantify. This natural variability in the $H-Q$ relationship is often aggravated by considerable errors of human origin.

The Okavango River system in Botswana has, over the years, attracted considerable interest as a major source for water and other ecological resources. Recently a major water scheme focused on this system had to be abandoned by the Botswana Government as a result of widespread concern about the impact of the project on the unique ecosystem of the area (SMEC, 1987; IUCN, 1992). However, considering the projected increases in water demand in northeast Botswana, it is only a matter of time before the project is resurrected again. Undoubtedly, the management of any resource abstraction scheme in a unique environment such as this depends on a thorough understanding of the working of the entire system. In this regard, accurate discharge data on various time scales are required for resource assessment, environmental monitoring and flood forecasting. Given the primary importance of river gauging activities to discharge data gathering, this study focused on stage-discharge relationships for the Okavango at Mohebo which is the point where the flow from the Cubango system enters the Okavango Delta system.

The study reported here emanated from concern within the Department of Water Affairs, the government department responsible for the collection of hydrological data in Botswana, about the existence of "inexplicable" outliers in the stage-discharge plot. Outliers occur all too often in discharge data. However, in this particular case, huge discharge values were recorded which corresponded to stages for which much lower discharges were previously associated. Moreover, the records did not show any corresponding change in velocity or cross-section. Considering the fact that gauging at this remote site is largely carried out by inadequately trained staff, it seems likely that human error could be one of the causes of the occurrence of the outliers. This, however, cannot be assumed. The aim of this study, therefore, is to investigate the reason for the occurrence of these outliers and how a useful rating curve can be extracted from the available data. Given the absence of any information in the data files regarding change in cross-section or occurrence of unusually large floods, a single rating curve may be more suitable than multiple rating curves representing different flow regimes. The study describes and characterizes the nature of the $H-Q$ relationship, and then investigates the use of regression methods to verify the existence of errors in the data and to derive

a useful rating curve. Finally, it investigates the possibility of using stochastic analysis for smoothing data into which estimates of missing values have been incorporated prior to fitting the rating curve by regression.

THE STUDY AREA

The Okavango River is the extension of the Cubango River into Botswana (Fig. 1). On entering Botswana the river becomes confined between two parallel faults at the end of which it spreads out to form the Okavango Delta. Mohembo is at the northern end of the panhandle (as the parallel faults are called). The panhandle is a broad well-defined channel with a clearly defined floodplain. The catchment area up to Mohembo is about 150 000 km² (IUCN, 1992) with most of this area contributing very little or no flow. The location of the gauging site at Mohembo provides natural control for river gauging, with negligible bed slope and a well defined channel. The gradient of the river between Mohembo and Shakawe (Fig. 1) is about 1:7000.

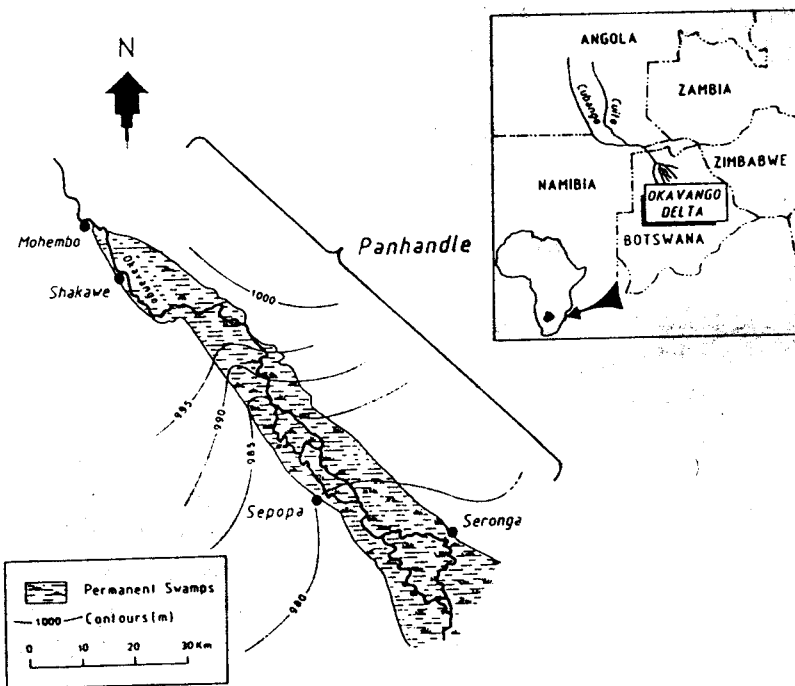


Fig. 1 Location of study area.

THE DATA

River gaugings at Mohembo started on a more or less regular basis in 1974. There are irregular data dating as far back as the 1930's (SMEC, 1987). Data of variable length are available on the following parameters: stage, discharge

a useful rating curve. Finally, it investigates the possibility of using stochastic analysis for smoothing data into which estimates of missing values have been incorporated prior to fitting the rating curve by regression.

THE STUDY AREA

The Okavango River is the extension of the Cubango River into Botswana (Fig. 1). On entering Botswana the river becomes confined between two parallel faults at the end of which it spreads out to form the Okavango Delta. Mohembo is at the northern end of the panhandle (as the parallel faults are called). The panhandle is a broad well-defined channel with a clearly defined floodplain. The catchment area up to Mohembo is about 150 000 km² (IUCN, 1992) with most of this area contributing very little or no flow. The location of the gauging site at Mohembo provides natural control for river gauging, with negligible bed slope and a well defined channel. The gradient of the river between Mohembo and Shakawe (Fig. 1) is about 1:7000.

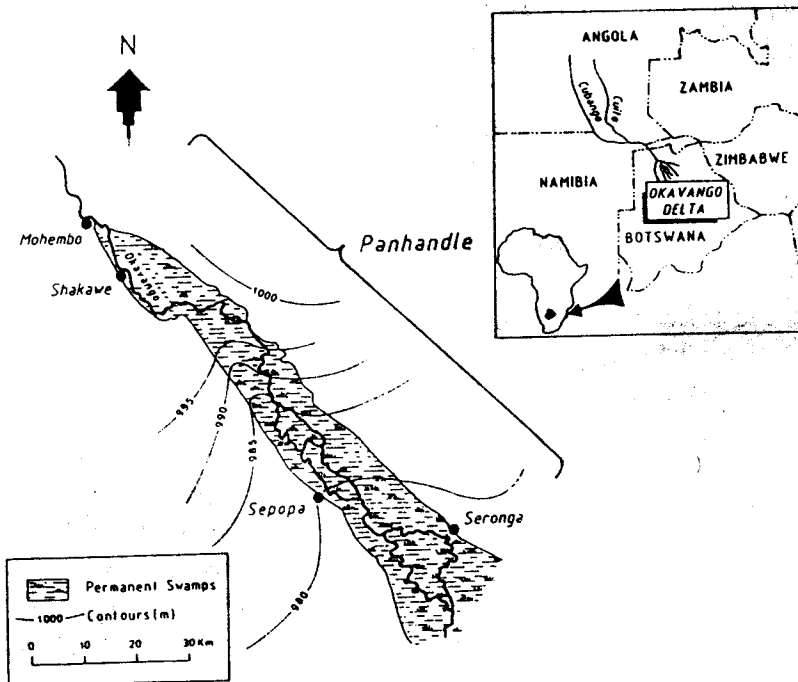


Fig. 1 Location of study area.

THE DATA

River gaugings at Mohembo started on a more or less regular basis in 1974. There are irregular data dating as far back as the 1930's (SMEC, 1987). Data of variable length are available on the following parameters: stage, discharge

(gauged or estimated from rating curve), cross-sectional area, surface width, mean velocity, mean depth and maximum depth. Discharge is gauged with an AOTT current meter from a boat, with velocity in the vertical measured by a mixture of the six-tenths and two-point methods. The sixth-tenths method was usually used when depth in the vertical was below 1 m. The number of verticals varied, although rarely, according to the width of the cross-section at the time of gauging. Usually the distance between verticals was typically 5 m, but this reduced to 2-3 m for the end verticals. The discharges were calculated using the mid-section method (Mosley & McKerchar, 1993). Although the number of verticals varied, the reliability of the gaugings was not assessed.

For this study, the primary gauging data sheets (the HY1 forms) were studied. In a few cases the discharges were recomputed. Overall, the following characteristics were observed about the data:

- (i) There are many years during which surface width was recorded as the same, irrespective of season. This is due to the fact that the river is incised between the two parallel faults, giving a stable gauging section. However surface width varied sometimes, and again irrespective of season. This may be attributed to the inability of gauging parties to follow the same line across the river.
- (ii) Most often the HY1 forms were not available and the records available (HY2 forms) showed only stage and discharge. Enquiries revealed that most of the discharge data on the HY2 forms may have been estimated from a previously defined $H-Q$ relationship, and the rest may have come from HY1 forms which may now have been lost. The mixing of data from the previous rating curve with actual gaugings in this manner may compound the errors and diminish the usefulness of any rating curve derived from the data as they are.
- (iii) The same magnitude of stage yielded different discharges at various times. There is no evidence to suggest that this could be attributed to variable channel storage, particularly as the plotted rating curve did not show a definite loop. Nor is there evidence of backwater effects. While channel aggradation and degradation cannot be ruled out, it seems likely that errors in discharge measurement might be a contributory factor. In one of the files kept on the gauging station, it was discovered that an attempt was made at one time to verify the different discharges at similar stages by undertaking a joint gauging exercise with the South African Department of Water Affairs using their equipment as well as that of the Botswana Department of Water Affairs. Over the two days of gauging, it was found that the Botswana equipment gave about 20% more discharge.
- (iv) Where the records indicated a change in stage during a gauging exercise, the discharge plotted or recorded is not the steady-state discharge.
- (v) The same surface width yielded different cross-sectional areas and discharges at various times. This could be due to a number of reasons, namely: (a) the depth soundings during gauging were not accurate; (b) the verticals were not located at the same distance apart for the same

cross-section width, which happened in a few of the data sheets examined; and (c) bed movement.

It was not possible to use all the available data for this study. The portion of data used for this study is listed in Table 1.

Table 1 Numbers of data used for various relationships

Relationship	No. of data points
<i>H vs Q</i>	1183 ¹
Mean depth vs <i>Q</i>	70
Mean vel. vs <i>Q</i>	354
Surface width vs <i>Q</i>	331

¹ Note that not all these are actual gaugings.

METHODOLOGY

Hydraulic geometry

One way in which to verify the quality of the available data is to examine the nature of the relationship between mean depth (*MND*), mean velocity (*MNV*) and surface width (*SW*) on one hand, and discharge (*Q*) on the other. Along with slope and friction factors, those parameters determine the nature of the cross-section and, therefore, the *H-Q* relationship. The relationships are expressed as follows (Leopold & Maddock, 1953):

$$MND = cQ^f \tag{1}$$

$$MNV = kQ^m \tag{2}$$

$$SW = aQ^b \tag{3}$$

with the condition that $f + b + m = 1$, and $a.c.k. = 1$ for consistent units. The degree to which these conditions are maintained is investigated by simple linear regression in which those conditions are maintained. The particular way in which the regression analyses are handled in this study is described later.

The stage-discharge relationship

The relationship between stage (*H*) and discharge (*Q*) is normally represented simply by a plot of *H* against *Q*. The plot can be either in natural or log space, and an equation is fitted to it as below:

$$Q = aH^b \tag{4}$$

or

$$\log Q = \log a + b \log H \tag{5}$$

These were the initial approaches adopted in the assessment of the stage-discharge data.

Linear regression

It is assumed that the data available contained errors of unknown magnitude that cannot be ascribed to the inherent variability in natural systems. It is also assumed that there is a 'reasonable' amount of 'correct' observations in the data set. Under these assumptions, the linear regression technique was employed in investigating the relationships among the various parameters as outlined in the above equations.

The basic method of achieving linearity was logarithmic transformation, although in the case of the $H-Q$ relationship, both power and Box and Cox transformations were used. Where it was necessary to choose between a number of regression equations, the decision was based on three statistics: standard error of estimate (SEE), the sample coefficient of determination (R^2) (Gawne & Simonovic, 1994), and the extent to which the distribution of residuals approached normality as indicated by the normal probability plot and the mean and standard deviation of the standard residuals. For equations (1) to (3) these criteria were replaced by the conditions that $f + b + m = 1$ and $a.c.k. = 1$.

Outliers were identified by the Mahalanobi's distance (D_i) routine (Norušis, 1990), which is a measure of the distance of an individual observation from average values of the independent variable in a regression analysis given as:

$$D_i = \left[\frac{X_i - MN_X}{S_X} \right]^2 \quad (6)$$

where X_i is the i th observation, and MN_X and S_X the mean and standard deviation of the observed values respectively. This method was supplemented by a physical examination of the plots of standardized residuals. The identified outliers were treated in a particular manner. It was assumed that they were observations which contained substantial error. They were therefore removed after each regression run until the results were found acceptable in terms of the three statistics outlined above.

It is recognized that the method of treatment of outliers adopted here may in fact lead to discarding of important information which may have a physical impact on the various relationships studied. However, it was judged that in the situation where the same magnitude of the independent variable is associated with two or more values of the dependent variable, the distortion of the relationship by retaining an outlier would far outweigh the loss of explanation when it is removed.

Stochastic modelling of gauging data

Although flow gaugings are not carried out daily throughout the year, the observed gaugings can be treated as a time series and it is possible to treat the days without gauging as days with missing values. These missing values can then be estimated by one of the routines incorporated in the SPSS statistical package (SPSS Inc., 1990). This was done in this case and the resultant time series subjected to an ARIMA (p, d, q) analysis.

Discussion of ARIMA models abound in the literature (Box & Jenkins, 1976; Kendall & Ord, 1990; Wei, 1993). The general ARIMA (p, d, q) model may be expressed as (Wei, 1993):

$$\phi_p(B)(1 - B)^d Z_t = \theta_o + \theta_q(B)a_t \tag{7}$$

where the stationarity AR operator $\phi_p(B) (1 - \phi_1 B - \dots - \phi_p B^p)$ and the invertible MA operator $\theta_q(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ share no common factors; Z_t is a realization of a stationary process and a_t is a zero mean white noise process; B is the backward shift operator. The modelling was done using the SPSS Trends package (SPSS Inc., 1990). This analysis was done with the purpose of investigating whether an ARIMA model could be fitted to the observed discharge data since, if this was possible, the technique could also be used to estimate missing data.

Effect of changing stage on the rating curve

The discharge plotted against stage to obtain the rating curve should be the steady-state discharge. Where stage changes during the period of gauging, the measured discharge is often adjusted to obtain a steady-state discharge. The methodology for carrying out such an adjustment is covered in standard texts in hydrology (Wilson, 1990).

The method is derived from the well-known Manning formula. For a rising or falling stage the steady-state discharge is computed as:

$$Q = \frac{Q_a}{\sqrt{\left[1 \pm \frac{Adh/dt}{1.3 Q_a S} \right]}} \tag{8}$$

where dh/dt is the change in stage over time (+ for rising stage and - for falling stage); Q_a is the observed discharge and Q the steady-state discharge; and S is the water surface slope.

The data files examined contained records of gaugings carried out under conditions of rising or falling stage, but the computed flows had not been adjusted. Accordingly, equation (8) was used to compute steady-state discharges in order to assess the impact of this omission on the rating curve. A

major limitation to the use of equation (8) was the absence of data to compute water surface slope. This limitation was overcome by using concurrent gauging data at Shakawe, about 10 km downstream (Fig. 1) to estimate an average water surface slope. This rather crude estimate was then applied to all the cases of falling or rising stage to estimate steady state discharge.

RESULTS AND DISCUSSION

Hydraulic geometry

The results of the regression of MNV , MND and SW on Q are summarized in Table 2 and Figs 2, 3, 4, 5, 6, and 7 (logarithms are to base 10). Figure 2 shows a plot of MNV vs Q using the original data. The first observation to make is about the single data point representing a discharge of more than 500 $\text{m}^3 \text{s}^{-1}$ and MNV of about 0.03 m s^{-1} . It is inconceivable that this data point was actually observed. It would require a cross-section area of about 16 000 m^2 to produce this flow. Clearly this point must be an error. A second area of interest on Fig. 2 is the small cluster of data points between the larger cluster above and the lone outlier below. The question that comes to mind is whether these points suggest different flow characteristics at the gauging station at some time. However, examination of the original data files did not reveal any evidence to suggest or support that view. The discharges were recorded at different times and so the possibility of a damaged current meter giving successive erroneous readings could not be suggested either.

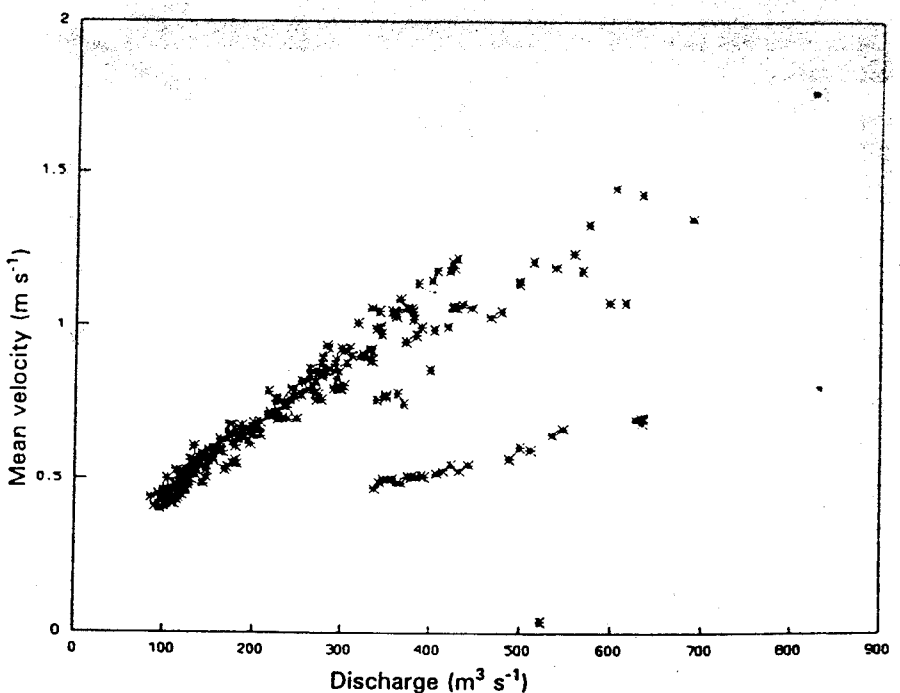


Fig. 2 Plot of mean velocity vs discharge.

Table 2 Results of regression of hydraulic parameters on discharge

Relationship	Regression equation	Correlation coefficient of logs
<i>MNV vs Q</i>	$MNV = 0.065Q^{0.43}$	$R = 0.67^1$
	$MNV = 0.047Q^{0.49}$	$R = 0.85$
	$MNV = 0.038Q^{0.54}$	$R = 0.91$
	$MNV = 0.025Q^{0.62}$	$R = 0.98$
<i>MND vs Q</i>	$MND = 0.56Q^{0.34}$	$R = 0.57$
	$MND = 0.46Q^{0.38}$	$R = 0.99$
<i>SW vs Q</i>	$SW = 73.0Q^{0.03}$	$R = 0.19$
	$SW = 77.0Q^{0.02}$	$R = 0.16$
	$SW = 76.0Q^{0.02}$	$R = 0.19$

¹ Note that *R* applies to the association of the logarithms, not the actual values.

Logarithmic transformation produced a better linear plot. Regression analyses identified those points referred to earlier as outliers. In line with the assumption that such outliers would be erroneous observations, they were removed on each occasion. The final relationship between mean velocity and discharge is shown in Fig. 3.

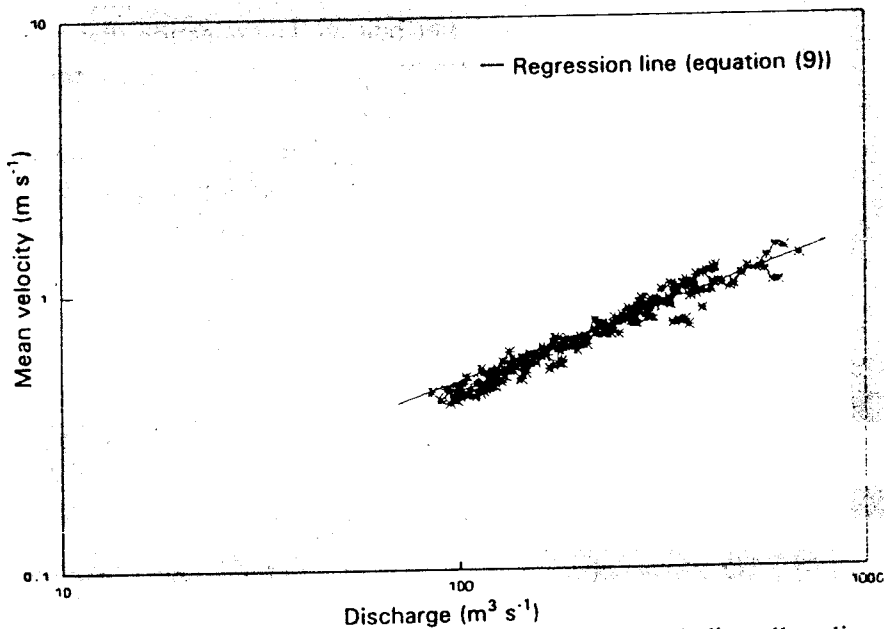


Fig. 3 Log-log plot of mean velocity vs discharge – excluding all outliers identified up to the third regression run.

Figure 4 shows the relationship between mean depth (*MND*) and discharge (*Q*). Here it can be seen that a discharge of about 300 m³ s⁻¹ is associated with two vastly different depths of flow. It is difficult to imagine the flow conditions that would have produced such a situation. Once again, there is a strong indication of errors in the data. The relationship between *MND* and *Q* was investigated in a similar manner as for mean velocity. The final relationship is shown in Fig. 5.

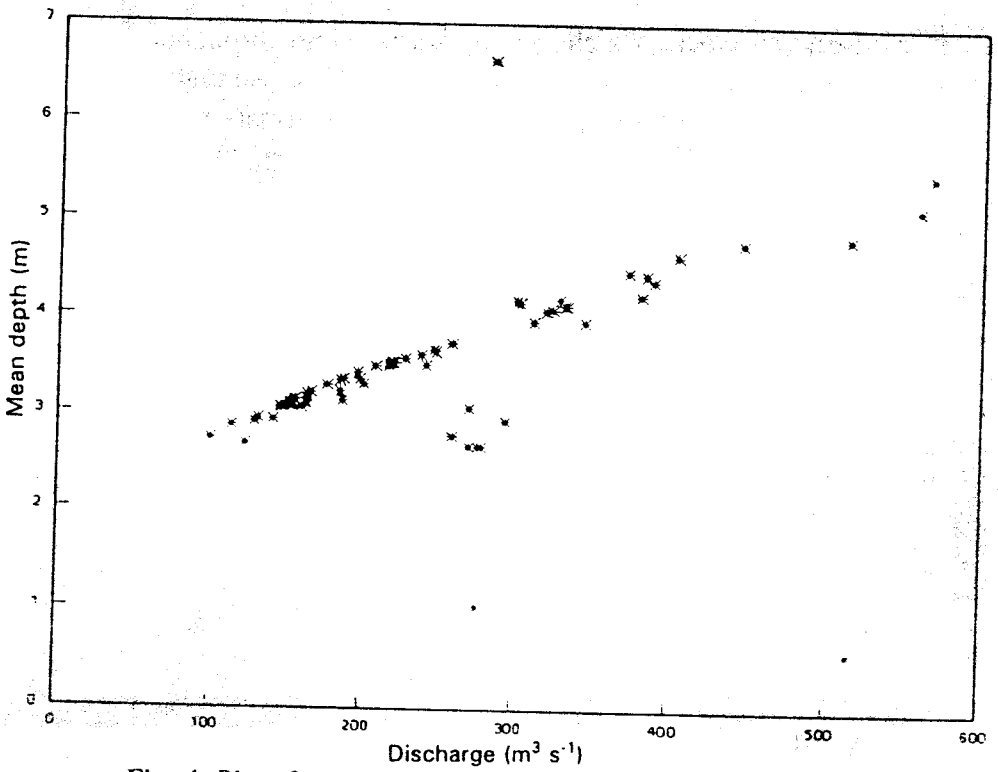


Fig. 4 Plot of mean depth vs discharge.

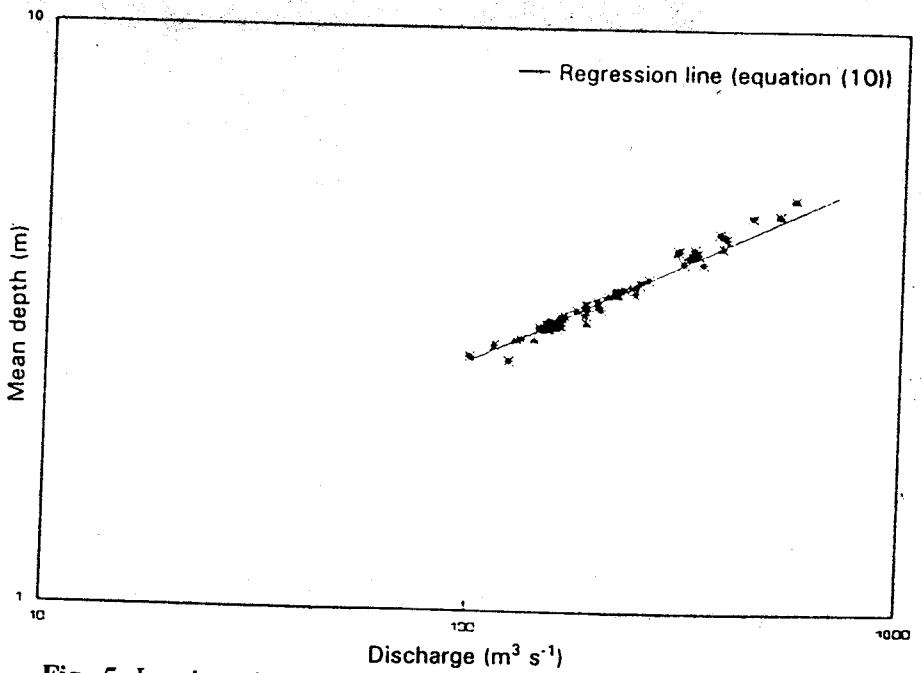


Fig. 5 Log-log plot of mean depth vs discharge – excluding all outliers identified at the first regression run.

Figure 6 shows a plot of surface width (SW) against discharge (Q). The cluster of points at the bottom of the figure lie across in two main bands, about 10 m apart. This raises the possibility of the existence of a systematic error of

measurement which occurred at random throughout the data. However, the relationship between SW and Q (Fig. 6), excluding the single outlier toward the top of the figure, provides further confirmation of a confined section. From Figs 2 and 4, it can be seen, for example, that a flow of $100 \text{ m}^3 \text{ s}^{-1}$ corresponds to a velocity and depth of approximately 0.4 m s^{-1} and 3 m respectively. From Fig. 6, with a width of 85 m , the discharge (computed as vwd) is $102 \text{ m}^3 \text{ s}^{-1}$. At a higher flow of $700 \text{ m}^3 \text{ s}^{-1}$, but with the same surface width of 85 m , the corresponding velocity of 1.4 m s^{-1} and depth of 6 m , the computed discharge is $714 \text{ m}^3 \text{ s}^{-1}$. That is, velocity has altered proportionately more than depth. Thus it seems that even with a low river gradient of about $1:7000$ in this area, the velocity of flow, rather than depth, is the dominant control on flow magnitude.

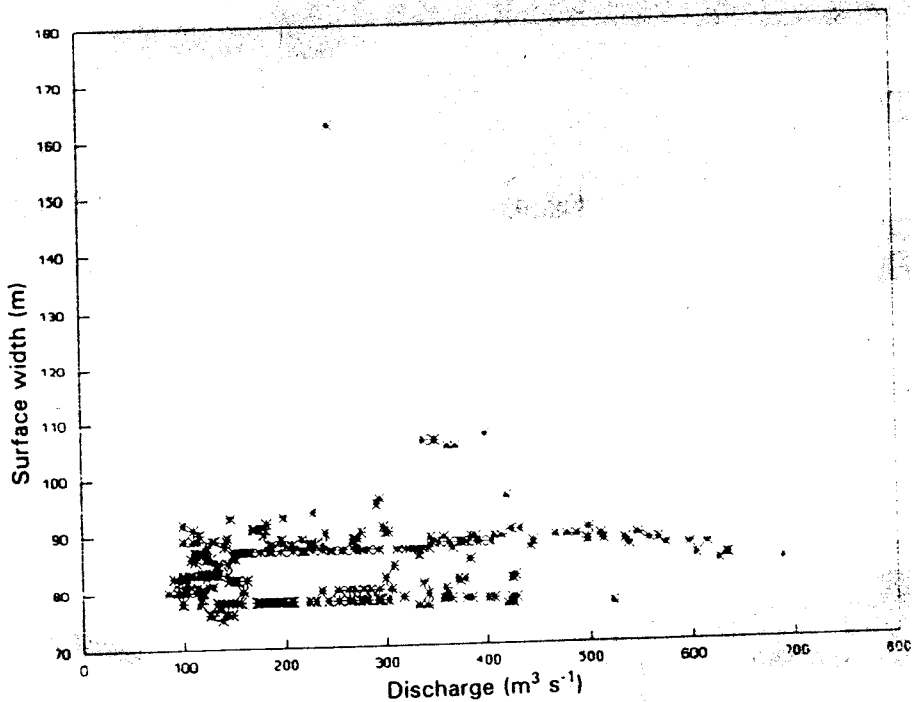


Fig. 6 Plot of surface width vs discharge.

An inspection of the gauging site indicated that at higher flows some discharge passes through the swamps on the left bank of the river. It is possible that this flow may be missed on a gauging mission, but that would not affect the relationships shown in Figs 2, 4 and 6 in any appreciable manner.

The relationship between SW and Q was analysed in a similar manner as for MNV and MND . Logarithmic transformation was carried out as before. Subsequent regression and removal of outliers did not improve the correlation and the exercise was abandoned after the third run.

The relationships between MNV , MND and SW on the one hand, and Q

on the other, which approximately satisfy the conditions of $f + b + m = 1$; $a.c.k = 1$ are:

$$MNV = 0.038Q^{0.54} \quad (9)$$

($R = 0.91$; standard errors: regression equation = 0.05, constant = 1.07, exponent = 0.01; degrees of freedom = 302).

$$MND = 0.56Q^{0.34} \quad (10)$$

($R = 0.57$; standard errors: regression equation = 0.08, constant = 1.38, exponent = 0.06; degrees of freedom = 68).

$$SW = 76.0Q^{0.02} \quad (11)$$

($R = 0.19$; standard errors: regression equation = 0.02, constant = 1.03, exponent = 0.01; degrees of freedom = 329).

Note that even though the relationships expressed in equations (9), (10) and (11) approximately satisfy the theoretical conditions, they are not the best regression equations in terms of R (Table 2). The relationship for surface width (equation (11)) is, in fact, not valid. The exponent could well be set to zero and the above conditions would still be approximately satisfied. These relationships indicate the existence of errors in the data.

The H - Q relationship

The H - Q plot characteristically shows curvature when plotted in natural units, but often the plot can be fitted with a linear regression equation after some form of data transformation thereby yielding a linear rating curve which greatly facilitates extrapolation. Figure 7 shows a plot of H against Q . A number of observations can be made about the stage-discharge relationship depicted. The figure suggests that there may be some curvature in the H - Q relationship. Secondly, in view of the foregoing discussion, several of the points plotted cannot be explained by normal or natural variation inherent in the discharge process. Particularly, the situation where a stage of less than 2 m yielded discharge in excess of $800 \text{ m}^3 \text{ s}^{-1}$ must be one of such errors that cannot be attributed to the inherent variability of the rating curve. Note also that a similar magnitude of discharge corresponds to a stage of about 3 m. Thirdly, it was found that logarithmic transformation did not remove the curvature in the H - Q relationship. This is quite interesting in view of the fact that some of the discharges plotted might have been obtained by linear extrapolation of an existing rating curve. However, it was not possible to determine from the available records what proportion of the data would have been obtained in this way.

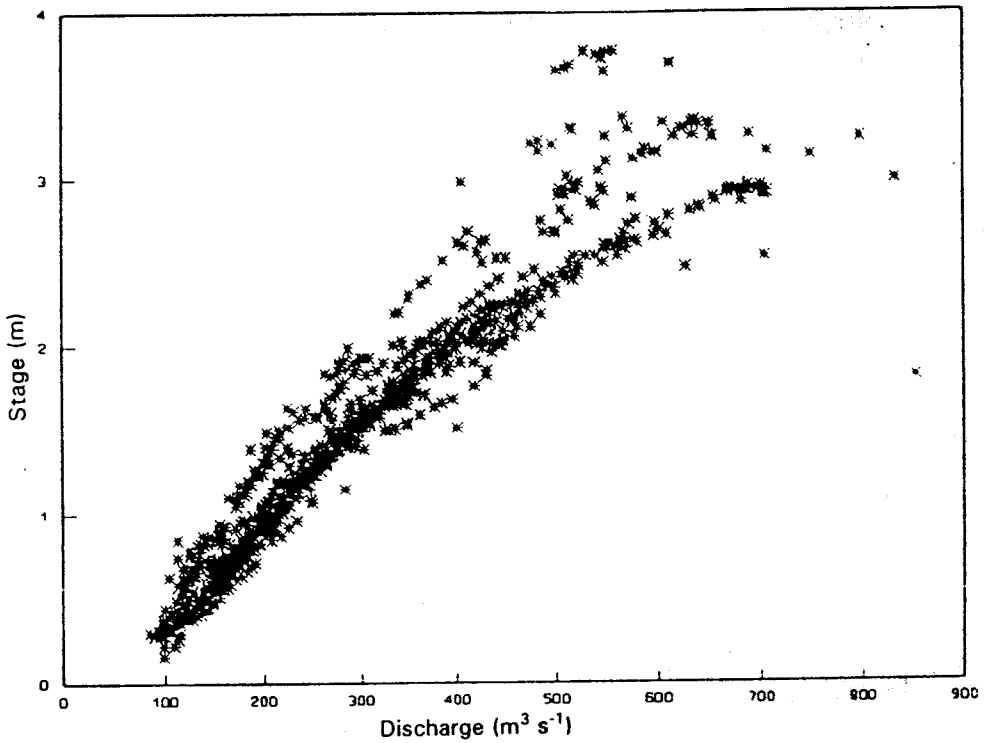


Fig. 7 Plot of stage vs discharge.

Transformations applied to the H - Q relationship

With the failure of the simple log transformation to yield a linear H - Q relationship, resort was made to other models. In particular, the power and Box-Cox transformations (Box & Cox, 1964; Gawne & Simonovic, 1994), and stochastic modelling were investigated.

Power transformation In the power transformation method, the transformation is applied to H . Thus, designating the transformed stage as H^λ :

$$QM = b_0 + b_1 H^\lambda \quad (12)$$

where QM is the estimate of the observed discharge computed from the regression model; b_0 and b_1 are parameters of the regression and λ is the exponent with which H was transformed.

As λ was evaluated by trial and error, there was the need to choose between the models produced by the various values of λ tried. This choice was aided by considering the closeness to normality of the residuals from each trial value of λ as indicated by the skewness and kurtosis coefficients (Gawne & Simonovic, 1994). The value of λ adopted was 1.5. Thus H was transformed to $H^{1.5}$, and as can be seen from Fig. 8, the relationship is approximately linear.

The transformed stage and discharge were then subjected to regression analyses and the outliers removed as previously described. The result is the

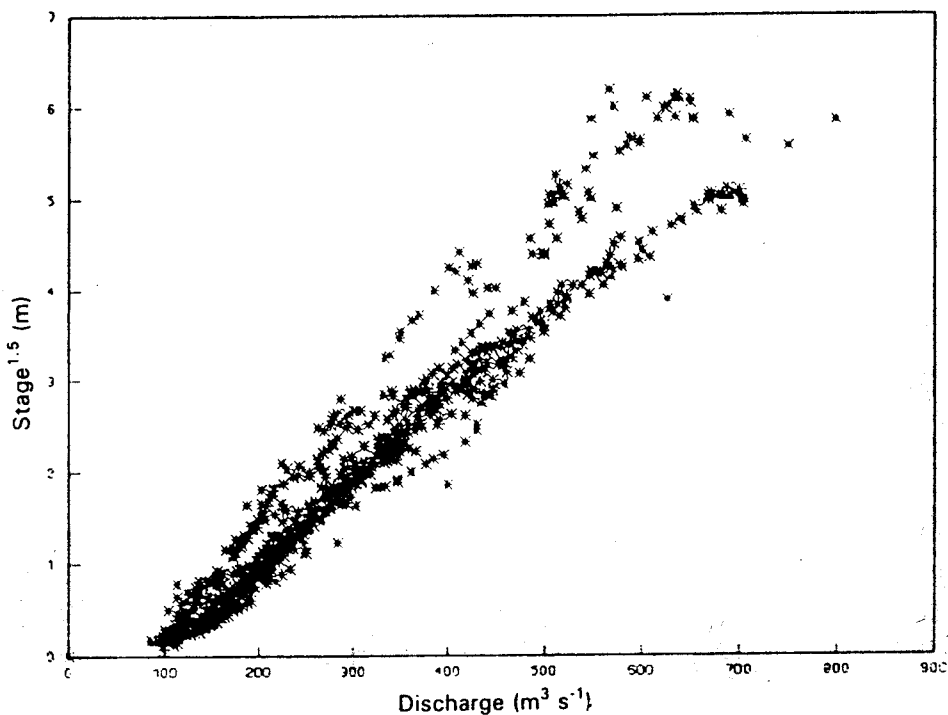


Fig. 9 Plot of power transformed stage and discharge — excluding all outliers identified up to the second regression run.

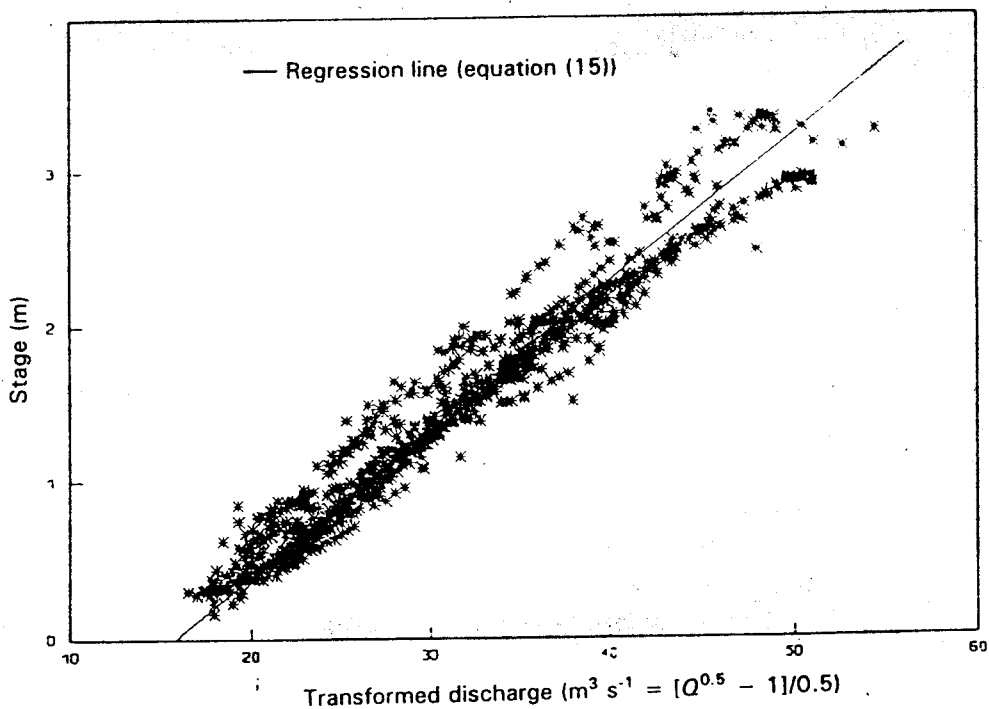


Fig. 10 Plot of Box-Cox transformed stage and discharge — excluding all outliers identified up to the second regression run.

Stochastic modelling of discharge Owing to technical limitations it was not possible to use the entire discharge data in the stochastic analysis. A sample consisting of 635 (Box-Cox transformed) observations was selected from the available data in such a way that gaps in the data sequence were kept to a minimum. The missing values, totalling 51 cases, were estimated by the linear interpolation routine incorporated in the SPSS statistical package (SPSS Inc., 1990). It is likely that linear interpolation would introduce some bias to the resulting ARIMA model. In order to link the observed stage to the corresponding discharge measurements, the sampled series of (Box-Cox) transformed discharges were divided by the corresponding stage, yielding a transformed discharge-stage ratio series.

Preliminary non-seasonal differencing led to an ARIMA (3,1,1) model being applied to the transformed discharge-stage ratio series. The final parameters are listed in Table 3. It can be seen that the model fits the data adequately. The Durbin-Watson statistic of 1.98 for the residual series fulfils the requirement of absence of autocorrelation, and the Box-Ljung statistic (SPSS Inc., 1990) is nowhere significant.

Table 3 Final parameters of the ARIMA (3,1,1) model

Variables in the model				
	<i>B</i>	SEB	<i>T</i> -ratio	Approx. prob.
AR(1)	-0.4020	0.0473	-8.5038	0.0000
AR(2)	-0.2156	0.0507	-4.2525	0.0000
AR(3)	-0.0997	0.0452	-2.2070	0.0277
MA(1)	0.8560	0.0275	31.1405	0.0000
Constant	-0.0009	0.0029	-0.3053	0.7603
Covariance matrix:				
	AR(1)	AR(2)	AR(3)	MA(1)
AR(1)	0.0022	0.0013	0.0008	0.0007
AR(2)	0.0013	0.0026	0.0012	0.0008
AR(3)	0.0008	0.0012	0.0020	0.0006
MA(1)	0.0007	0.0008	0.0006	0.0008
Correlation matrix:				
	AR(1)	AR(2)	AR(3)	MA(1)
AR(1)	1.0000	0.5548	0.3822	0.5307
AR(2)	0.5548	1.0000	0.5286	0.5473
AR(3)	0.3822	0.5286	1.0000	0.4674
MA(1)	0.5307	0.5473	0.4674	1.0000

Fit error statistics: number of cases: 630; degrees of freedom: 625; mean error: -0.0150; mean absolute error: 0.3912; RMS: 0.8613; Durbin Watson statistic: 1.9819.

The ARIMA-modelled discharges (as transformed by the Box-Cox procedure) are plotted against stage in Fig. 11. It can be seen that the ARIMA model preserved the range of variability in the original data as shown by the preservation of the 'outliers'. The problem of a lower stage producing an

inexplicably large discharge is also noticeable in Fig. 11. The rating equation associated with this plot is:

$$QTM = 15.64 + 11.29H \quad (16)$$

where QTM is the Box-Cox transformed ARIMA-derived discharge.

($R = 0.98$; standard errors: regression equation = 1.35, constant = 0.13, slope of regression line = 0.09; degrees of freedom = 633).

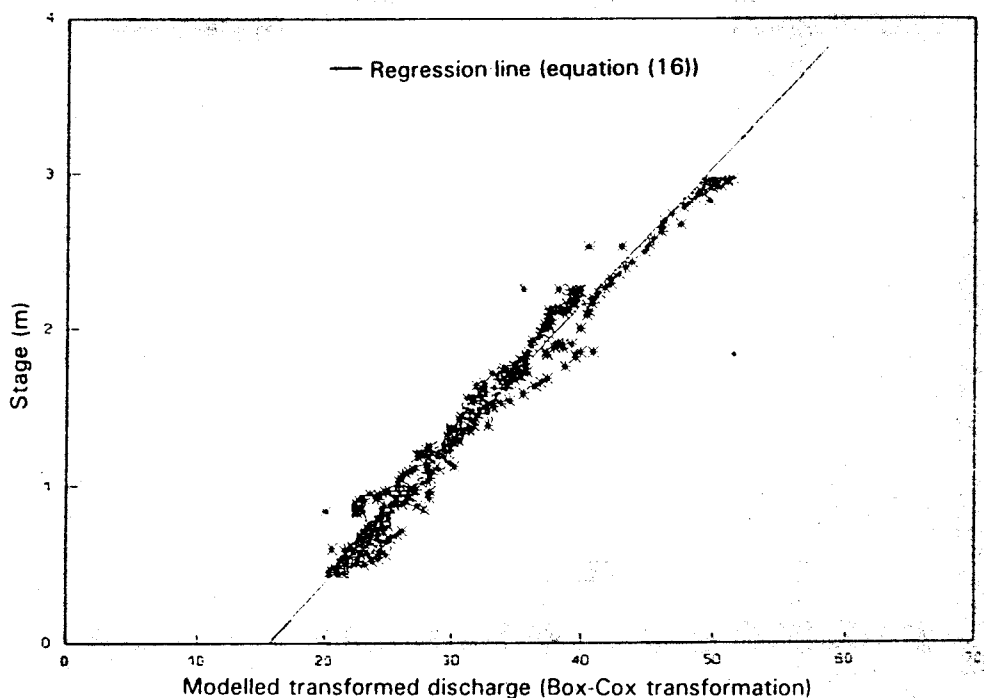


Fig. 11 Plot of transformed modelled discharge vs stage.

Although the regression results indicated the existence of outliers, it was not considered worthwhile to remove them and rerun the analysis as before because of the high value of R . Two conclusions can be drawn from the above results. Firstly, they suggest that discharge in the Okavango at Mohembo on intervening days when no measurements were undertaken can be estimated from an ARIMA (3,1,1) model. Secondly, the retention of the outliers in the ARIMA modelled discharge series implies that if the outliers were genuinely observed values, flash floods, for example, the ARIMA model would reproduce them reasonably well. This is a finding that may have significant implications in a region in which flash floods are prevalent, yet difficult to gauge. The ability to estimate missing discharge values with an ARIMA model may in the long run improve derivation of the rating curve. However, this conclusion, and subsequent ones, pertaining to the ARIMA model should be considered tentative as an independent test of the fit of the model has not yet been undertaken. For such a test a new set of data which had not been used in the derivation of the model would be required.

Influence of changing stage on discharge measurements

An averaged value of water surface slope was estimated by using occasionally available water level observations at Shakawe (see Fig. 1), about 10 km downstream of Mohembo. Average water surface slope was estimated as 0.01%. Equation (8) was then used to compute the steady-state discharges for the few occasions on which changes in stage were recorded. All the instances dealt with represented rising stage and steady-state discharge would have been 0.12-4.31% below the measured discharge. Although only a crude estimate of water surface slope was used, it is likely that, overall, the non-adjustment of the measured discharges will have only minimal direct impact on the rating curve. Nevertheless, because the existing rating curve is used at times to convert observed stage to discharge, this impact may become compounded with time leading to overestimation of discharges.

Choice of model

Firstly, it is important to observe that the analyses undertaken here did not remove all the erroneous values of stage and discharge. The treatment of outliers undertaken here removed only those glaring "unreasonable" values that exerted considerable influence on the regression. The remaining data points may be taken as containing "allowable error" (Gawne & Simonovic, 1994). Secondly, it should also be noted that the transformations investigated provide only alternatives to conventional means of fitting an algebraic relationship to rating curves. Their value lies in the greater flexibility they provide and for choosing a more appropriate relationship, especially in situations as this one when there are doubts about the quality of the data.

It is essential to take another look at equations (15) and (16). Equation (15) is the rating relationship with the basic Box-Cox transformation of discharge while equation (16) gives the relationship between the ARIMA model-derived discharge and stage. Both equations are, in fact, the same (at two significant figures). However, note that equation (15) was arrived at using the observed data set and after the outliers had been removed. On the other hand, equation (16) was obtained from the regression of stage on the ARIMA-derived data set with no outliers removed. Each of these equations would suit a different purpose. For example, if the assumption that outliers represent observations with considerable error, and therefore it is desirable to remove them, then equation (15) would yield an $H-Q$ relationship with considerably less scatter. On the other hand, if the outliers were unusual discharge events, say, flash floods, then it would be desirable to retain them in the $H-Q$ relationship. In that case, the ARIMA-derived relationship would be more suitable.

Reasons for the existence of outliers on the rating curve

The apparent instability of the rating curve depicted by the occurrence of

outliers, as described in this study, can result from human and physical causes. From the evidence so far presented, especially the review of the data and the results of the analysis shown in Figs 2-6, a number of observations regarding the reasons for this apparent instability of the rating curve can be made. The gauging site is a stable confined section in which velocity and depth of flow are the dominant hydraulic variables controlling discharge magnitude. Although bed movement cannot be completely ruled out, it is difficult to attribute to it the large magnitude outliers observed in the data series. It seems, therefore, that physical causes can largely be excluded. The implication is that the outliers on the rating curve can be attributed to human errors. Some of the evidence in support of this position has already been presented in the review of the data. Additional evidence that can be cited includes the assumption of linearity of the logarithmic rating curve, the failure to adjust the measured discharge for falling or rising stages, and the use of faulty equipment at times. One can also cite the loss of field sheets (HY1 forms) from which doubtful values can be verified.

CONCLUSION

This study has described the nature of available river gauging data for the Okavango River at Mohembo. Many data points were found to be out of the range of a possible stage-discharge relationship. By repeated regression and removal of outliers, a relationship was established between hydraulic parameters and discharge that approximately fulfilled the conditions that $f + m + b = 1$ and $a.c.k = 1$ in equations (1) to (3). Of three hydraulic parameters investigated, the data on surface width did not conform to the theoretical exponential relationship with discharge. The relationship between the respective hydraulic parameters and discharge, however, conformed to a confined gauging section.

The relationship between the original stage and discharge measurements was found to be non-linear and to require transformation of the data to become linear. The Box and Cox transformation in which $\lambda = 0.5$ was found to be suitable. The use of linear regression and removal of outliers led to the establishment of a rating curve represented by equation (15). It was also found that the series of gauged and stage-estimated daily discharges can be fitted with an ARIMA (3,1,1) model. The quality of the fit obtained means that missing data for this site can be estimated with an ARIMA (3,1,1) model.

Overall, the conclusion can be reached that the cumulative effects of the errors outlined are likely to lead to overestimation of discharges for the Okavango at Mohembo. Reasons for this conclusion can be found in the following: (i) the fact that linear extrapolations (in log-space, hopefully) were used to estimate discharges for some observed stages, when, in fact, as shown by the results of this study, there is curvature in the $H-Q$ plot even in log-space; (ii) the values of discharge plotted on the rating curve were not necessarily the steady-state discharges as no adjustments for rising or falling

stages were carried out; and (iii) the discrepancy in discharge measurements reported during the gauging exercise undertaken jointly with the South African Department of Water Affairs.

Acknowledgements The author is grateful to the reviewers for their very helpful comments and to the Swedish Agency for Research Cooperation (SAREC) for its support for hydrological research in Botswana.

REFERENCES

- Box, G. E. P. & Cox, D. R. (1964) An analysis of transformations. Royal Statistical Society, Series B, 26, 211-252.
- Box, G. E. P. & Jenkins, G. M. (1976) *Time Series Analysis, Forecasting and Control*. Holden Day, McGraw-Hill, New York, USA.
- Gawne, K. D. & Simonovic, S. P. (1994) A computer-based system for modelling the stage-discharge relationships in steady state conditions. *Hydrol. Sci. J.* 39(5), 487-506.
- IUCN (1992) The IUCN Review of the Southern Okavango Integrated Water Development Project.
- Kendall, Sir M. & Ord, J. K. (1990) *Time Series*. Edward Arnold, UK.
- Leopold, L. B. & Maddock, T. (1953) The hydraulic geometry of stream channels and some physiographic implication. US Geol. Survey Prof. Paper no. 252.
- Mosley, P. M. & McKerchar, A. I. (1993) Streamflow. In: *Handbook of Hydrology*, ed. D. R. Maidment. McGraw-Hill, New York, USA.
- Norušis, M. J. (1990) SPSS/PC+ Statistics 4.0 for the IBM PC/XT/AT and PS/2. SPSS Inc., Chicago, USA.
- SMEC (1987) Southern Okavango Integrated Water Development Phase 1. Department of Water Affairs, Government of Botswana.
- SPSS Inc. (1990) SPSS/PC+ Trends for the IBM PC/XT/AT and PS/2. SPSS Inc., Chicago, USA.
- Wei, W. S. (1993) *Time Series Analysis*. Addison-Wesley, Redwood City, California, USA.
- Wilson, E. M. (1990) *Engineering Hydrology*. MacMillan, UK.

Received 3 January 1995; accepted 19 September 1995