

**Peter M. Atkinson**  
**Sally E. German**  
**David A. Sear**  
**Michael J. Clark**

## ***Exploring the Relations Between Riverbank Erosion and Geomorphological Controls Using Geographically Weighted Logistic Regression***

*The relations between riverbank erosion and geomorphological variables that are thought to control or influence erosion are commonly modelled using regression. For a given river, a single regression model might be fitted to data on erosion and its geomorphological controls obtained along the river's length. However, it is likely that the influence of some variables may vary with geographical location (i.e., distance upstream). For this reason, the spatially stationary regression model should be replaced with a non-stationary equivalent. Geographically weighted regression (GWR) is a suitable choice. In this paper, GWR is extended to predict the binary presence or absence of erosion via the logistic model. This extended model was applied to data obtained from historical archives and a spatially intensive field survey of a length of 42 km of the Afon Dyfi in West Wales. The model parameters and the residual deviance of the model varied greatly with distance upstream. The practical implication of the result is that different management practices should be implemented at different locations along the river. Thus, the approach presented allowed inference of spatially varying management practice as a consequence of spatially varying geomorphological process.*

The authors thank the Environment Agency, Welsh Region and the Department of Geography, University of Southampton for research funding to Sally German. The GeoData Institute, in particular, Chris Hill and James Feaver are thanked for general support and for the provision of computing facilities. Finally, we thank the anonymous referees for their constructive comments.

*Peter M. Atkinson is a professor in the Department of Geography at the University of Southampton. E-mail: P.M.Atkinson@soton.ac.uk. Sally E. German is a research scientist at the Geodata Institute at the University of Southampton. Email: seg@geodata.soton.ac.uk. David A. Sear is a senior lecturer in the Department of Geography at the University of Southampton. Email: D.Sear@soton.ac.uk. Michael J. Clark is a professor in the Department of Geography at the University of Southampton. Email: M.J.Clark@soton.ac.uk.*

*Geographical Analysis*, Vol. 35, No. 1 (January 2003) The Ohio State University  
Submitted: July 1, 2001. Revised version accepted: July 10, 2002.

## INTRODUCTION

The evolution of river planform is an important area of research in fluvial geomorphology, and is of significance to river managers and owners of riparian land. Bank erosion, a key process of planform change, continues to be a focus for process and applied research. The erosion of river banks not only causes change in channel planform and cross-section but is a key process whereby floodplain sediments are remobilised to become a part of the basin sediment yield (Bull 1997).

The literature associated with channel planform change is now extensive, with research largely focusing at two scales: (i) mesoscale analysis of channel planform and bank retreat rates over periods of up to 200 years (Gurnell 1997) and (ii) more detailed site-specific studies of the processes of bank retreat (Hooke 1995; Lawler, Thorne, and Hooke 1997; Bull 1997). Relatively few studies attempt to link such different scales of analysis, and even fewer attempt this at the basin scale. Notable exceptions include the work of Lawler (1992) who identified theoretical process domains for bank erosion, and Abernathy and Rutherford (1998) who attempted a similar analysis including the additional influence of vegetation. However, even these studies involve fragmentary information and make no explicit attempt to link historical channel activity with the full suite of contemporary controls on the physical processes of bank erosion.

Recent geomorphological investigations into the processes of bank erosion have been undertaken through the use of erosion pins and PEEPS (Lawler 1993; Lawler, Thorne, and Hooke 1997), investigations into bank stability (Thorne 1982; Simon and Hupp 1992; Simon and Darby 1995), and the use of bank reconnaissance methods (Thorne, Allen, and Simon 1996). However, these investigations tend to focus on particular banks and specific erosion issues, and the detailed data sets required preclude investigation of the entire catchment. Further, the multivariate nature of the controls and the variability of these in space can prevent identification of the actual processes of bank erosion. In particular, geomorphologists are frequently challenged to explore relationships that may vary in their nature from place to place (they require spatially non-stationary models) and that involve categorical data.

Bank erosion is a significant problem for many people who derive income from, or live on, land next to rivers. Bank erosion management has typically been site-specific (tending to treat the problem where it arises) and simplistic (using techniques centred on the assumption that the erosion is caused by fluvial scour) (Sear, Newson, Brookes 1995; Thorne 1997). Recent application of geomorphological theory to bank erosion management has focused on, first, identifying the cause of erosion and, second, developing a more sustainable bank management plan in partnership with engineers and stakeholders. At the heart of this philosophy is the ability to identify the dominant processes and the controls on these processes operating at the bank scale, whilst recognising that the causes may lie at the reach or catchment scale. To date, this process has been largely field-based and qualitative, with interpretation of the evidence relying on the professional capability of those involved. What is needed is a more objective framework for integrating and assessing different types of data across a broad range of spatial scales.

In this paper, new survey and statistical techniques are presented that can be applied at the basin scale, and that integrate spatially distributed historical data on the timing of river channel planform change with spatially distributed data on the properties controlling bank failure. A method for the analysis of spatial variation in the relations between bank erosion and its controlling factors is presented that incorporates non-stationary parameters and categorical data. This method is termed geographically weighted logistic regression (GWLR). The objective of this paper is to apply

GWLR to model the spatial variation in the relation between erosion (presence or absence) and several controlling variables for the Afon Dyfi in West Wales.

## THEORY

Statistical modelling has been applied extensively to fluvial systems (see Rhoads 1992). In this section, the basic linear regression model and its extension to (i) spatially varying parameters (GWR) (Brunsdon, Fotheringham, and Charlton 1996) and (ii) prediction of a binary response variable (logistic regression) are described.

### *Linear Regression*

Multiple linear regression is commonly applied in geomorphology and the steps in modelling and the assumptions involved are well understood. Therefore, only the basic linear regression model is given here,

$$y_i = \beta_0 + \sum_{k=1}^{m-1} \beta_k x_{ik} + \varepsilon_i \quad (3)$$

predicting the target variable  $y_i$  at location  $i$  as a function of  $m$  parameters  $\beta_0$  and  $\beta_k$  ( $k = 1, 2, \dots, m-1$ ) and  $m-1$  explanatory variables  $x_{ik}$ . The  $\varepsilon_i$ s are independent normally distributed error terms with zero means. Usually, the least squares method is used to estimate the  $\beta_k$ s. This is most readily expressed using matrix notation as

$$\hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y} \quad (4)$$

where  $\hat{\beta}$  is a single column vector of coefficients, the independent variables fill the columns of  $\mathbf{x}$ , the dependent variable fills the single column of  $\mathbf{y}$  and  $t$  indicates the transpose of a matrix (Brunsdon, Fotheringham, and Charlton 1996). To estimate  $\beta_0$   $\mathbf{x}$  must contain a first column of 1s.

Equation 3 is implicitly a *stationary* model of the relations between  $\mathbf{y}$  and  $\mathbf{x}$ . A stationary model is one in which the parameters of the model do not change with geographical location. Thus, it is only appropriate where the relations may reasonably be modelled as invariant over space. Where the relations may be expected to change with location it is necessary to apply a non-stationary model: one that allows some freedom in the parameters of the model as one moves from place to place. Brunsdon, Fotheringham, and Charlton (1996) developed such a non-stationary regression model and termed it geographically weighted regression (GWR). A similar model was applied in relation to fluvial systems by Rhoads (1991). GWR offers a basis upon which to tackle some of the issues mentioned in the introduction and represents a significant potential enhancement over conventional regression approaches.

### *Geographically Weighted Regression*

GWR extends the traditional regression model of Equation 1 to the non-stationary case by allowing the  $\beta$  coefficients to vary with location  $i$ ,

$$y_i = \beta_{i0} + \sum_{k=1}^{m-1} \beta_{ik} x_{ik} + \varepsilon_i \quad (5)$$

where  $\beta_{ik}$  is the value of the  $k^{\text{th}}$  parameter at location  $i$ . It is worth noting that Equation 3 is a special case of Equation 5 where the parameters do not vary across space.

A problem with Equation 5 is that there is only one observation at  $i$  for each of  $y_i$  and the  $m-1$   $x_{ik}$ . To provide more observations with which to build a regression model for each  $i$ , Brunsdon, Fotheringham, and Charlton (1996) chose to consider the relation between  $\mathbf{y}$  and  $\mathbf{x}$  within a region defined as geographically proximate to each  $i$ , while acknowledging that such an approach involves approximation.

The weighted least squares approach to calibrating regression models provides a means by which to vary the influence of observations on the resulting model. Specifically, in weighted least squares a weight  $w_i$  is applied to each squared difference before minimization so that the inaccuracy of some predictions carries more weight than that of others. If  $\mathbf{w}$  is the diagonal matrix of  $w_i$ s then the estimated coefficients satisfy:

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{w}\mathbf{x})^{-1}\mathbf{x}'\mathbf{w}\mathbf{y} \quad (6)$$

Since the weight  $\mathbf{w}$  varies with location  $j$ , such that a different calibration exists for each  $j$ , it is more appropriate to write Equation 6 more explicitly as:

$$\hat{\boldsymbol{\beta}} = (\mathbf{x}'\mathbf{w}(\mathbf{j})\mathbf{x})^{-1}\mathbf{x}'\mathbf{w}(\mathbf{j})\mathbf{y} \quad (7)$$

Note that location has now been referenced by  $j$  since the location at which a model is calibrated need not equal the location of available data  $i$ .

Brunsdon, Fotheringham, and Charlton (1996) consider a range of possible weighting functions for use in Equation 7. These include a simple step function,

$$\begin{aligned} w_{ij} &= 1 & \text{if } d_{ij} < d_j \\ w_{ij} &= 0 & \text{otherwise} \end{aligned} \quad (8)$$

where  $d_{ij}$  is the distance between the location  $i$  of any observation and the location  $j$  at which the model is being calibrated. This weighting function has the advantage that the number of data used in the regression model is limited resulting in increased speed of computation. However, it has the disadvantage that its discontinuous nature can result in artefacts in the spatial variation in the regression coefficients. An alternative is to specify  $w_{ij}$  as a *continuous* function of  $d_{ij}$ . The exponential function is an obvious choice,

$$w_{ij} = \exp\left(\frac{-d_{ij}}{a}\right) \quad (9)$$

where  $a$  is a non-linear parameter. As a compromise between the computational saving of Equation 8 and the continuity of Equation 9, Brunsdon, Fotheringham, and Charlton (1996) suggest using the bisquare function:

$$\begin{aligned} w_{ij} &= [1 - d_{ij}^2/d_j^2]^2 & \text{if } d_{ij} < d_j \\ w_{ij} &= 0 & \text{otherwise} \end{aligned} \quad (10)$$

The essential idea of GWR, whatever weighting function is chosen, is to give more weight to observations close to the location  $j$  at which the regression model is desired than further away. In this sense, the method draws strong parallels with geostatistical techniques for local prediction (Matheron 1965, 1971; Goovaerts 1997).

*Logistic Regression*

Generalised linear modelling (GLM) (Aitkin et al. 1989; Collett 1991) allows the formation of a (multivariate) regression relation between a response variable and explanatory variable(s). All variables may be either continuous or categorical, or any combination of both types. GLM has at least two advantages over traditional regression. First, GLM allows one to include variables that are categorical and many, if not most, variables in geomorphology are categorical. Second, parameters do not need to be normally distributed about their mean and no assumptions are made about their error distributions. Thus, most of the limitations of traditional regression are removed. Examples of this type of modelling in geomorphology include Wang and Unwin (1992), Uno, Sugii, and Hayashi. (1994), and Atkinson and Massari (1998).

In this paper, our interest is the presence or absence of erosion. The most natural regression model for this type of binary response (with 1 = “erosion” and 0 = “no erosion”) is the logistic where the usual linear predictor is augmented by a logistic link function. The full logistic regression model is thus,

$$\log\left(\frac{p_i}{1-p_i}\right) = \gamma_i \quad (11)$$

or equivalently,

$$p_i = \frac{\exp(\gamma_i)}{1 + \exp(\gamma_i)} \quad (12)$$

where  $p_i$  is the probability of an event occurring associated with a given location  $i$  and  $\gamma_i$  is the linear model (equation 3). The linear model formed is then a logistic regression of success or failure of a given binary variable (for example, presence or absence of erosion) on the explanatory variables. Useful introductions to GLM are given in Baker and Nelder (1978), Healy (1988), Aitkin et al. (1989), Dobson (1990), and Collett (1991).

The model deviance  $D$  for binary data in a GLM may be expressed as:

$$D = -2 \sum_{i=1}^n p_i \log(p_i) + \log(1-p_i) \quad (13)$$

This model deviance cannot be used directly as an indicator of goodness-of-fit since one cannot compare the fitted probabilities and the binary observations (Collett 1991). However, since the difference in deviance for a given difference in degrees of freedom between separate models can be approximated to have a Chi-squared distribution, the reduction in model deviance can be used as an indicator of the significance of each term added to the model (Baker and Nelder 1978). Therefore, terms are normally added to the model in a stepwise fashion, checking the significance of each term at each step by finding the probability for a given reduction in model deviance and number of degrees of freedom. Only those terms that are significant at a given confidence level (for example, 95%) should be retained. Once a term has been checked as significant, each parameter estimate should be tested for significance against its standard error using the  $t$ -value (estimate/s.e.). At the 95% significance level a parameter estimate is significant when it is greater than approximately twice the standard error (Atkinson and Massari 1998).

A potential problem exists for categorical variables that are divided into binary terms because only certain of these terms may be significant while others may be insignificant. It is common practice to aggregate such categorical variables into fewer, coarser classes to reduce the problem and also to simplify the resulting GLM. Interpreting the resulting GLM is quite straightforward. For categorical variables, a positive value implies an above average correlation and a negative value implies a below average correlation. For continuous variables, a positive value implies a direct correlation and a negative value implies an inverse correlation, while the magnitude of the value gives the strength of the correlation.

In this paper, GWR and GLM (specifically the logistic model) are combined to form a geographically weighted logistic regression (GWLRL). This combination of a spatial model with logistic regression has similarities with, but is conceptually different to, autologistic regression (Augustin, Muggleston, and Buckland 1996). The GWLRL will be used to explore the dependence on geographic location of the relation between erosion and geomorphological controls for the Afon Dyfi, Wales.

#### FIELD SITE AND SURVEY DATA

##### *The Afon Dyfi*

The Afon Dyfi is a gravel-bed river situated in North Wales, running partly through Snowdonia National Park. The river drains from the Aran Mountains and enters the sea through an estuary of international importance (designated as a Ramsar site and a Biosphere Reserve). The river and its catchment have a glacial legacy and are composed of two parts. The upper reaches flow through bedrock, and there is very little floodplain due to glacial terraces and valley walls. The middle and lower reaches run through alluvial and glacial deposits (Jones 1995). Here, wide floodplains and grazing land take over from bedrock gorges and wooded riverbanks, although several terraces, both glacial and fluvial, are still present in the middle reaches that confine channel movement.

Riparian landowners have identified several bank erosion issues within the Afon Dyfi that relate primarily to loss of valley-bottom farmland. In response to a series of requests for bank protection, the local river management agency initiated a study to identify a more rational, catchment-scale management strategy that accounted for the processes controlling bank erosion on the Afon Dyfi. To determine the controls and influences on the processes of erosion, large-area (both temporal, or historical, and spatial) investigations were undertaken (German 2000). The “historical” investigation provided data on the long-term stability of the planform and the “spatial” investigation identified and measured properties that defined the fluvial environment at the time of the investigation. All the data sets were contiguous (spatially) for left and right bank, from the mouth of the river to its source.

##### *Historical Data*

A historical stability index was constructed by digitising historical maps and aerial photographs from 1890 to 1992. These planforms were imported into the SPANS geographical information system (GIS) and annotated with the age of the adjacent floodplain and the number of times the channel had remained in its present position.

The age of the floodplain adjacent to the present bank line (1998) provided information on the age of the present bank material. The banks were graded according to the age of the most recent channel as shown in Table 1. The number of times the channel had been present in the same position was determined by GIS overlay and was between one and five times. From this information a historical stability index was created and each individual bank was assigned a number (between 1 and 50), as follows:

TABLE 1  
The Allocation of Scores to Bank Lines of Different Dates

Date	No.
Before 1890	10
1890	8
1902	6
1948	4
1980	2

$$\text{Historical stability} = \begin{array}{l} \text{age of the adjacent floodplain} \\ \times \text{number of times channel has been} \\ \text{in the same position} \end{array} \quad (14)$$

The larger the number, the more historically stable was the bank. In addition, the width of the planform movement (swath width) was measured every 100 m to determine the magnitude of the channel change.

### *Field Survey*

Field data were acquired during the 1998 field season using a continuous walk-through survey. Data were acquired on several variables including the location of current erosion, vegetation, and bank material. Both banks were surveyed from mouth to source. In the field, banks were classified according to the active erosion process acting on the bank (for example, overhang, cantilever, or slab failure). This information was annotated on a base map and subsequently attributed to digitised vectors within a GIS package. The same process of data acquisition was used for vegetation and bank material. To determine the relationship between the factors controlling bank erosion and the physical locations of this erosion, it was considered appropriate to reclassify the detailed erosion types into the presence or absence of erosion. Vegetation data included type, location on the bank, density, and age. All of these properties provided information related to the stability of the bank on which the vegetation grows. These data were used to create a vegetation stability index, and individual banks were allocated a number (between 1 and 97), as follows:

$$\begin{aligned} \text{Vegetation stability} = & \{( \text{woody age} \times 3) \\ & + ( \text{woody density} \times 3) \\ & + ( \text{woody location} \times 3) \} \\ & + \{( \text{herbaceous age} \times 2) \\ & + ( \text{herbaceous density} \times 5) \\ & + ( \text{herbaceous location} \times 3) \} \\ & + \{( \text{grass density} \times 4) \\ & + ( \text{grass location} \times 3) \} \end{aligned} \quad (15)$$

Equation 2 defines the weights given to different types of vegetation. The larger the number, the greater the contribution to bank stability. Woody vegetation has the same weight for each variable (age, density, and location), whereas for herbaceous vegetation the weights differ reflecting the view that age is a less important influence on bank stability than density (Thorne 1990). Age has two categories (young and mature): the older the vegetation the more stable the bank. Density has three categories (sparse, average, and dense): the denser the vegetation the more stable the bank. Location has three categories (top, middle, and bottom of the bank in any combination).



Each of the three locations has the same weighting; the more locations that the vegetation occupies the larger the score. Table 2 describes the scores assigned to each variable in Equation 2.

Bank material was classified into five categories by field observation. These were (i) cohesive, (ii) composite (30% gravel), (iii) composite (60% gravel), (iv) non-cohesive (100% gravel), and (v) bedrock. These data were allocated to every bank within the river.

Other variables used in this research were stream power (Bagnold 1966) and the location of meander impact points on the bank lines (Thorne 1997). Direct measurement of stream power was not possible for the whole river due to the absence of survey data. Instead, an approach was adopted whereby the cross-sectional area every 100 m down river was predicted from the bank height and map-derived channel widths. Channel gradient was predicted from a 1:25000 map and bankfull discharge was calculated using the regime equations of Wharton (1995). Controls on these predictions were available at five points along the channel where direct survey had been made, or where a gauging station existed. Typically, the values obtained under-predicted the locally derived values of stream power by 20%.

Locations of meander impact points were assigned to the bankline within the GIS as 0 = no meander impact point and 1 = meander impact point. Each meander impact point was spatially distributed, that is, represented a sequence of points along the river.

Data Organisation

Data associated with each bank were sorted by location. The sequence of locations was used to determine the mid-point of each individual length for each bank. The histograms for each of the five variables are shown in Figure 1 (left bank) and Figure 2 (right bank). Each variable is also plotted against position upstream in Figure 3 (left bank) and Figure 4 (right bank).

The data were entered into a GIS to allow visualisation of the individual spatial variables. However, the number of variables available and the length of the river (over 42 km) made it difficult to view the different relations between control and effect. In addition, for management purposes, it is important to determine any differences in these controls at a local level, as this information can provide suitable guidance on the type of bank protection that may need to be adopted locally. Therefore, a methodology was required that could deal with a large data set and quantify the relations between contemporary erosion and several variables at the local scale.

METHODOLOGICAL CHOICES

This section describes several of the choices made in applying GWLR to model the spatial variation in the relation between erosion (presence or absence) and geomorphological variables for the Afon Dyfi in West Wales.

TABLE 2  
Scoring System Used to Construct the Vegetation Index

	(a) woody vegetation		(b) herbaceous vegetation		(c) grass	
	Score	Weight	Score	Weight	Score	Weight
Age	1 2	3	1 2	2		
Density	1 2 5	3	1 2 5	5	1 2 5	4
Location	1 2 3	3	1 2 3	3	1 2 3	3

NOTE: A score is assigned per variable (e.g., young age = score of 1) and all weighted scores are then summed to construct the vegetation index.



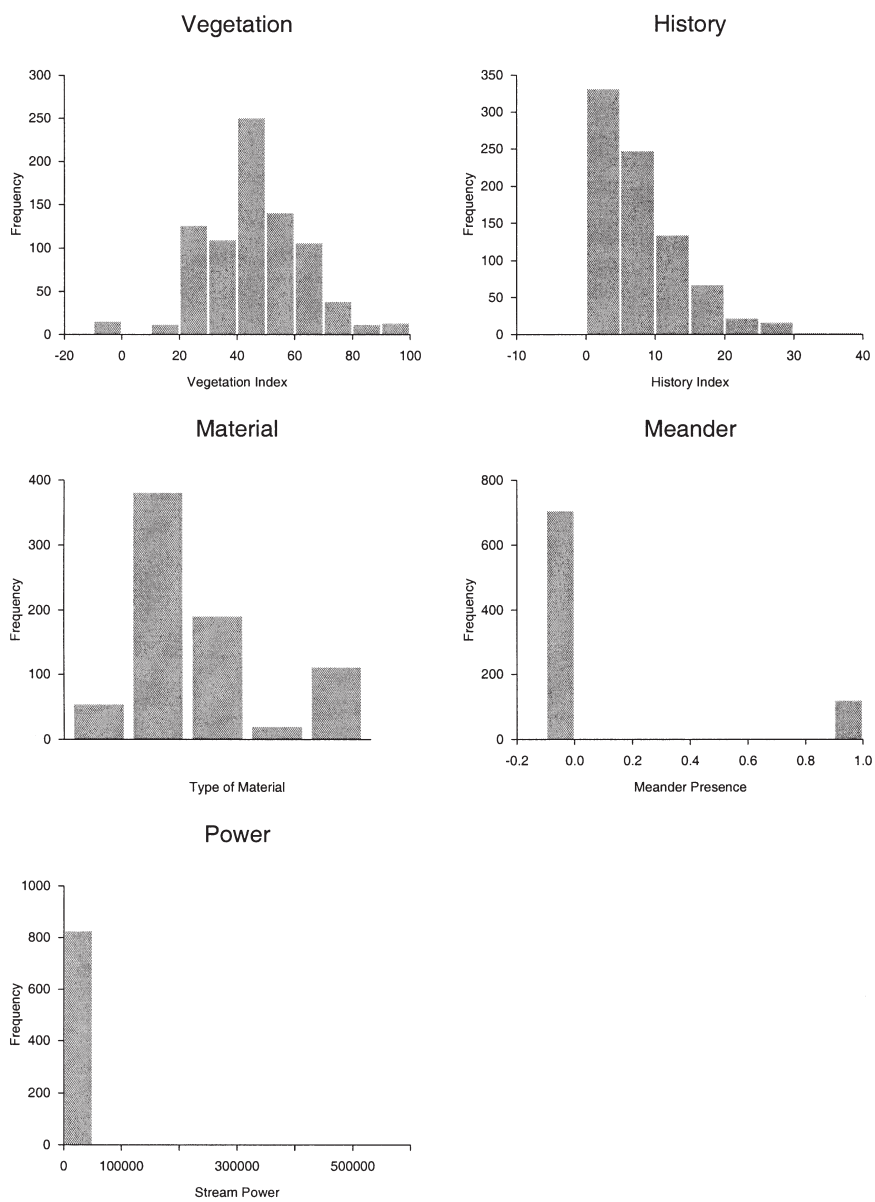


FIG. 1. Histograms of Vegetation, History, Material, Meander and Power for the Left Bank. The variables are defined in the text.

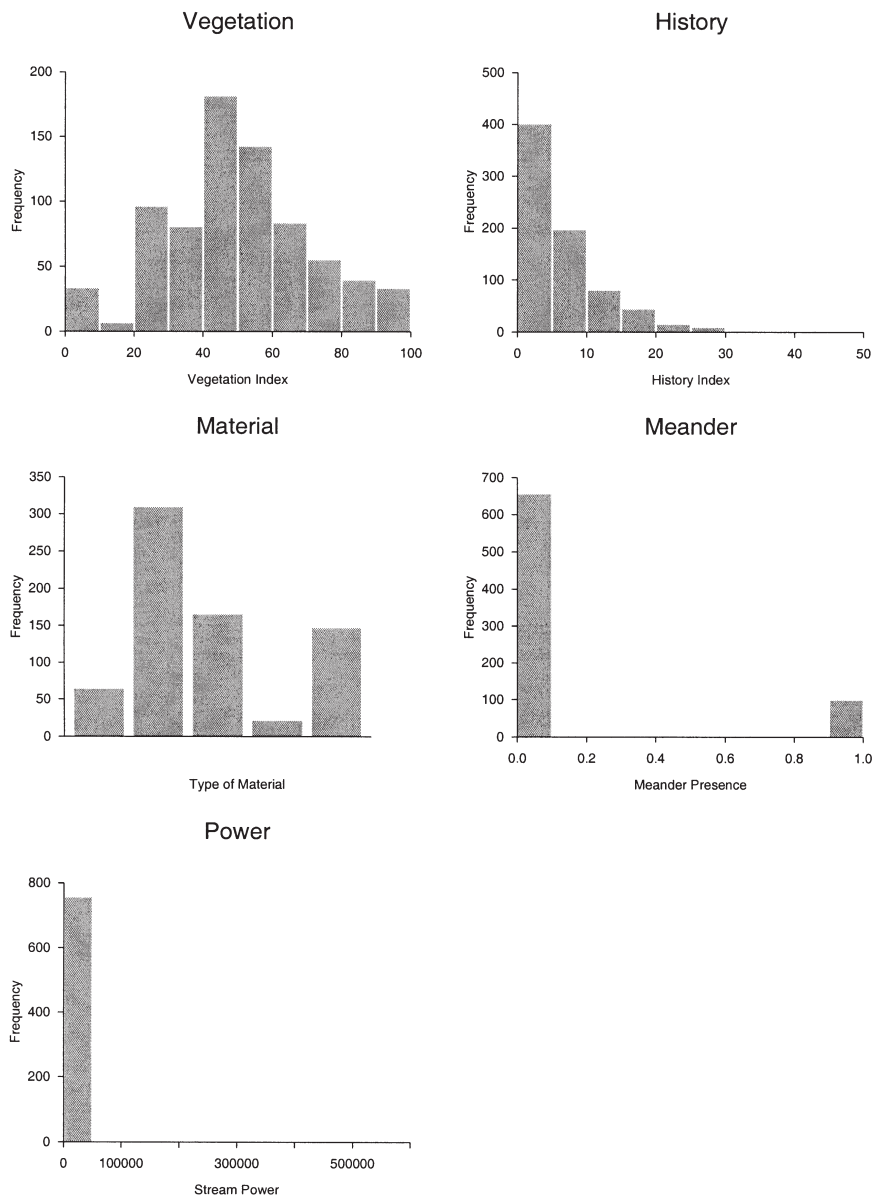


FIG. 2. Histograms of Vegetation, History, Material, Meander and Power for the Right Bank. The variables are defined in the text.

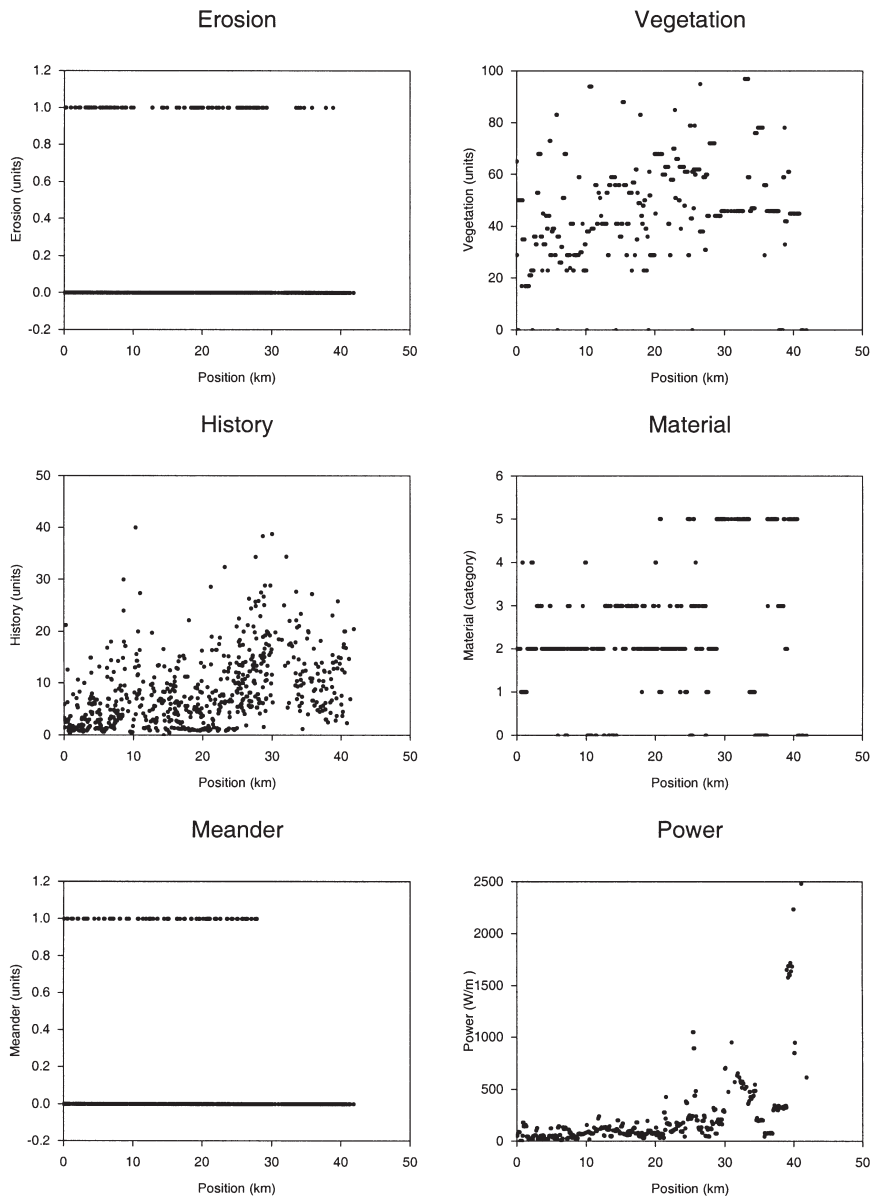


FIG. 3. Erosion, Vegetation, History, Material, Meander and Power Plotted Against Distance Upstream for the Left Bank.

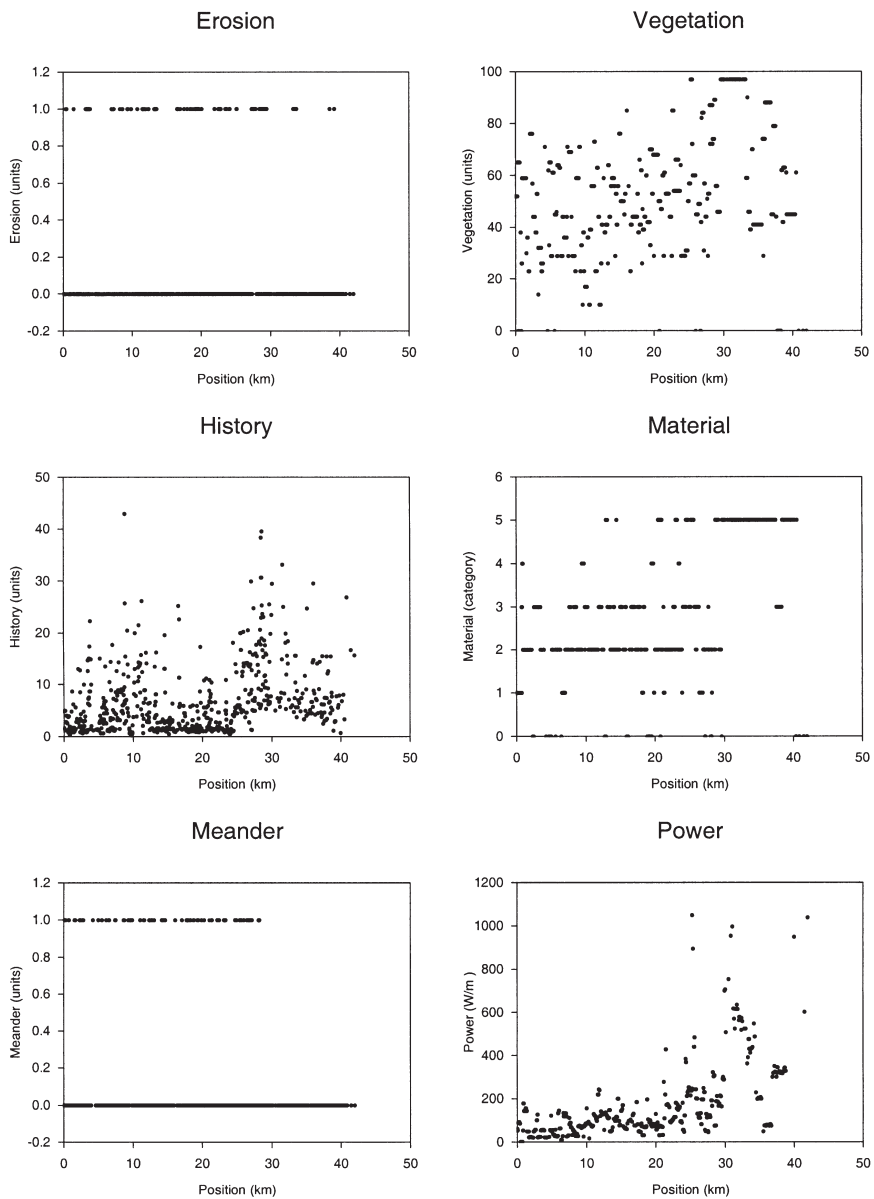


FIG. 4. Erosion, Vegetation, History, Material, Meander and Power Plotted Against Distance Upstream for the Right Bank.

*Univariate Modelling*

In this paper, the multivariate GWLR was not fitted. Rather, univariate GWLR models were fitted to each explanatory variable in turn. The basic rationale for this choice was that the main objective of the paper was not prediction (using the geomorphological data to determine with greatest precision the level of erosion likely), but rather to explore the relations between erosion and several individual geomorphological variables. Multivariate regression is such that meaningful relations with certain explanatory variables can become downweighted and potentially omitted from the multivariate model as other correlated variables enter the model first. To avoid this situation, variables must, at least initially, be treated individually. Additional benefits of univariate modelling are that the procedure is kept simple, the models are easier to interpret, and there are no interaction terms. However, one should remember that univariate models do not involve independent effects (i.e., controlling for the effects of other explanatory variables). Thus, the fitted parameters in a set of univariate models may be very different to those in the equivalent multivariate model.

*Choice of Weighting Function*

The exponential weighting function (Equation 10) with  $a = 1$  km was chosen for this research because the experience of the authors in geostatistics suggests that for physical properties in geomorphology and related subjects the exponential model often describes well the autocorrelation structure. In geostatistics, and specifically the linear prediction technique known as (co)kriging, interpolation weights are chosen by reference to a model fitted to the experimental autocorrelation function (or similar) and the exponential model often provides a good fit to the experimental plot. Thus, the exponential model represents a reasonable *a priori* choice. The Gaussian model (one of several models used by Fotheringham, Brunsdon, and Charlton 2000) is, by comparison, rarely used in geostatistics and describes only the autocorrelation structure of variables that are locally very smooth (for example, digital elevation data).

*Choice of Number of Data*

Initially the exponential model was used to weight *all* 830 (or 758) data along the left (or right) banks of the river. However, the computational burden when running the model-fitting procedure in Splus was prohibitive, each variable for each bank taking around one hour to run. Therefore, the number of data used in the regression modelling was limited to the nearest 101 points, that is, 50 points either side of the central point  $i$ . This number was chosen as a compromise between having many data for statistical regression and few data distant from  $i$ . The computational time required for model fitting was reduced to around one-fifth of that required when using all data.

A disadvantage of using a limited number of data was that when combined with the exponential model the weights do not decrease to zero at the most distant point (as for the bisquare function), but rather decrease to a minimum. This may lead to artefacts in the maps of regression coefficients as described earlier. However, we chose this strategy because (i) the exponential model is commonly fitted to autocorrelation functions in physical geography and (ii) we wanted to fix the number of data used in modelling at each location  $i$ , (except for the ends of the river course where the number decreased). The bisquare function would involve different numbers of data being used at each location  $i$ . A comparison between the results using all data and the results using only 101 data was made and the differences were only slight.

## ANALYSIS

The GWLR model was applied to each of the five explanatory variables, and for each of the left and right banks. This generated two sets of results. The first is the usual slope coefficient plotted against position. The second is a plot of  $\{(\text{null deviance-residual deviance})/\text{null deviance}\}$ , here called a deviance ratio. This was provided to give some measure of the goodness-of-fit of the model at each location. The results for each of the explanatory variables are described below.

*Vegetation*

The model coefficients for vegetation are plotted against position in Figures 5a (left bank) and 6a (right bank), while the deviance ratio is plotted against position in Figures 7 (left bank) and 8 (right bank). There are several interesting features to note. Overall, there is a negative relation between vegetation and erosion. That is, the greater the index of vegetation the less erosion; the implication is that vegetation makes banks more stable (or, in principle, that more vegetation grows on banks that are stable) (Thorne 1990; Millar 2000). However, the nature of the relation between vegetation and erosion varies along the river channel. In particular, vegetation appears to have a greater influence on erosion for the downstream reaches (less than 20 km upstream on the plots). The influence seems to peak at around 10 km and 15–18 km upstream for both the left and right banks. This again is what might be expected since the downstream sections of the river run through alluvial deposits where the presence of additional protection from fluvial scour or increased stability of the bank materials would be more pronounced. It is worth noting that around 13 km upstream there is no relation between vegetation and erosion (for both banks) suggesting that either there is no erosion and no vegetation as bank protection is present in this area or the controlling variables do not explain erosion in this reach.

There is a large peak in the influence of vegetation on erosion at around 30 km upstream, especially for the right bank. For the upper reaches of the river, the channel is bedrock, but large glacial terraces overlain on the bedrock adjacent to the channel provide sufficient material for vegetation to influence erosion.

Not only does the strength of the relation vary with position, but also there is a direct relation between the two variables around 20 km upstream. At this point, a high vegetation index appears to correspond to large amounts of erosion. This unexpected result appears at a position where erosion is dominated by scour of the upper banks arising from the turbulence developed around widely spaced trees.

*History*

The most notable historical feature is the small correlation with erosion (Figures 7 and 8). Generally, one might expect banks that are eroding now to have been eroding in the past. However, contemporary erosion does not necessarily contribute greatly to historical lateral change in river planform. Thus, banks that are eroding at present are not necessarily related to planform change over the last 108 years. There are two main sets of positions at which the relation between history and erosion is inverse as expected: one at 10 km and 38 km upstream (left bank) (Figure 5) and one at about 23 km upstream (right bank) (Figure 6). At 10 km and 38 km (left bank) we expect positive erosion and a low historical index value (implying less stability). At 23 km (right bank) we expect zero erosion and a high historical index value (implying greater stability). At around 38 km upstream (right bank) there appears to be an exception to the generally inverse relation between history and erosion. We currently have no explanation for this exception, but it is worth noting that the magnitude of this effect is very small (Figure 8).

*Material*

The deviance ratio for material (Figure 7, left bank and Figure 8, right bank) shows that, overall, variation in material has the largest effect on erosion along the Afon Dyfi. However, as for vegetation, the influence is not constant along the river. The largest peak occurs at around 30 km, with a second peak at around 20 km (both left and right banks). These positions correspond to changes from composite banks with 30% gravel and 60% gravel to bedrock (30 km) and from composite banks with 30% gravel to composite banks with 60% gravel (20 km). There is also a notable peak for the left bank at around 10 km (Figure 7c). This unexpected relation is most likely to be the result of bank protection centred on this location.

The coefficients of the fitted model are shown in Figures 5b (left bank) and 6b (right bank). The plots for composite banks with 30% gravel and composite banks with 60% gravel are similar. These two binary variables are competing at positions less than 30 km upstream and, therefore, convey similar information. Bedrock is important upstream of 30 km as might be expected. Cohesive banks have some influence between 20 km and 30 km upstream while banks with 100% gravel, which rarely occurs, has limited influence.

*Meander*

Figures 5a, 7 (left bank), 6a, and 8 (right bank) demonstrate that, as for history, whether the channel is meandering or not appears to have limited influence on bank erosion. Examining the deviance ratios closely reveals that this small influence is greatest between 10 km and 30 km upstream. The coefficients for meandering relate closely to the deviance ratios. In particular, two areas with large negative coefficients occur at around 10 km (and 16 km, right bank) and 24 km (and 28 km, right bank) upstream. As for material (above), the relation observed at around 10 km is likely to be the result of bank protection. The relation observed at 24–28 km may be explained by the controlling influence of bank material, in this case bedrock that outcrops where the meander cuts into the valley side. Overall, the Afon Dyfi has relatively few meandering sections, and in many cases these have been the focus for bank protection measures. In the headwater reaches, meanders are often confined by cohesive glacial drift or bedrock. Together these reduce the expected magnitude of correlation between meandering and the presence of bank erosion.

*Stream Power*

The most notable feature of the plots of the coefficients and deviance ratios of the GWLR for stream power against position upstream (Figure 5a, 6a, 7, and 8) is that the influence is greatest for the upper reaches (around 23–27 km to 34 km upstream). However, interestingly the relation is inverse for these positions, which was not anticipated. This can be explained in terms of the large values of unit stream power for the upper reaches. Where stream power is very high, there is little erodible material present. Where stream power is lower, the material present includes glacial terraces that are more likely to be eroded than the bedrock. In the lower reaches, an inverse relation is present for several positions. These relations arise in sections of the river where gravel bars result in both relatively low stream power (since channel width is maximised) and bank erosion. It is necessary to remember that the values of bankfull stream power are all above the critical threshold for the median bed material. Also, bank protection, which has not been included in the model data, results in reaches of high stream power (hence the need for bank protection) and low erosion.

There are several positions where a direct relation between stream power and erosion results, the most important of which appears at 18 km upstream. For these locations, weaker alluvial material supports such a direct relation.



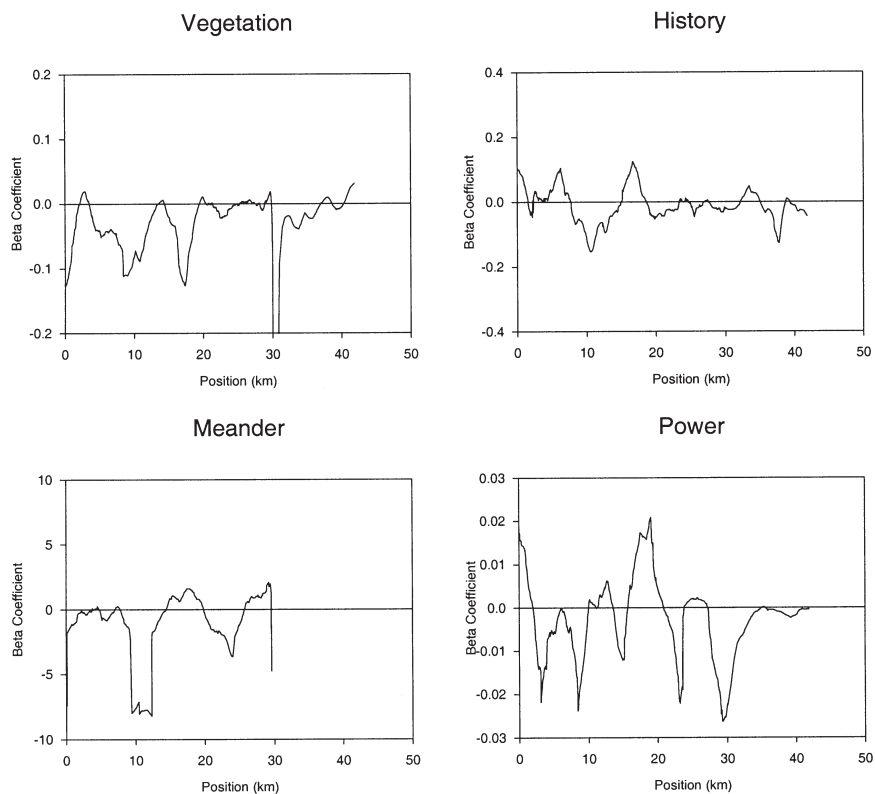


FIG. 5a. The Regression Coefficients for Vegetation, History, Meander, and Power Plotted Against Distance Upstream for the Left Bank.

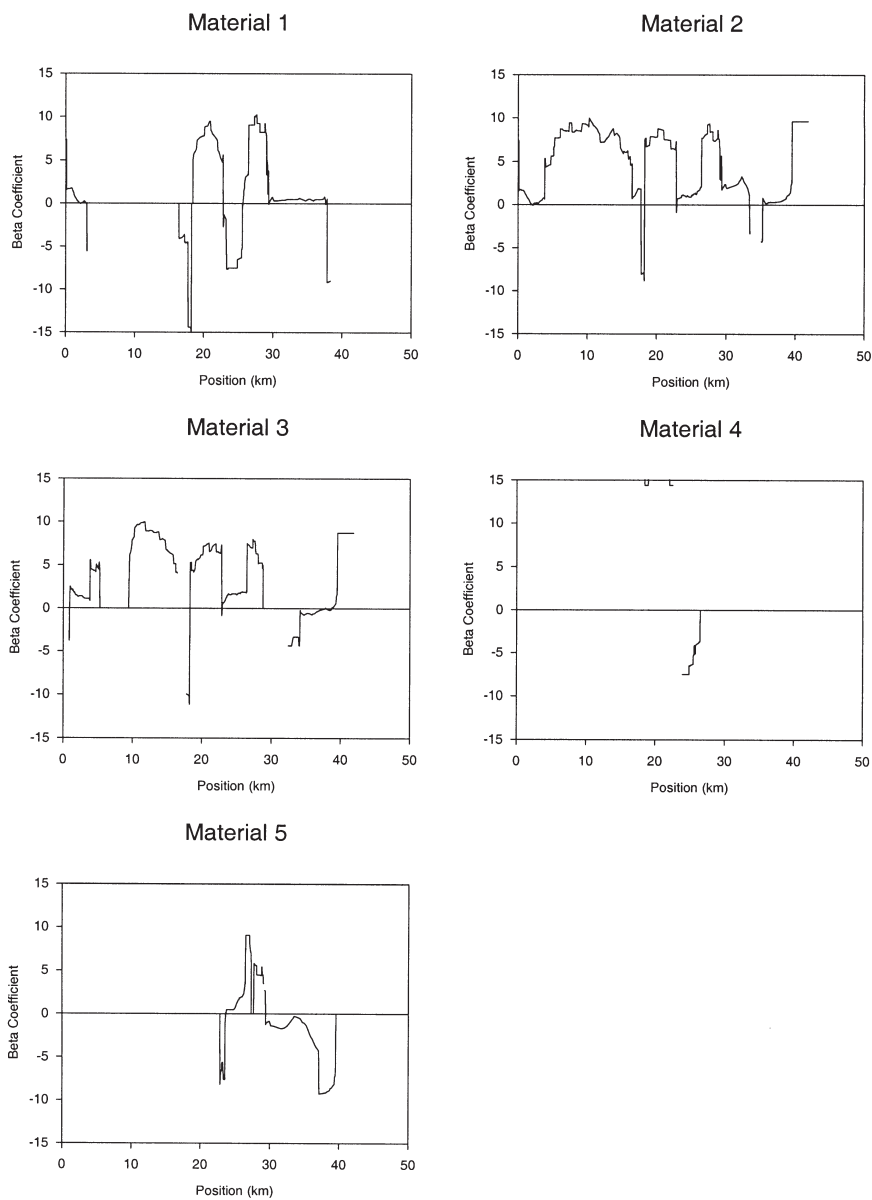


FIG. 5b. The Regression Coefficients for Material Plotted Against Distance Upstream for the Left Bank. Material is divided into five binary components as described in the text.

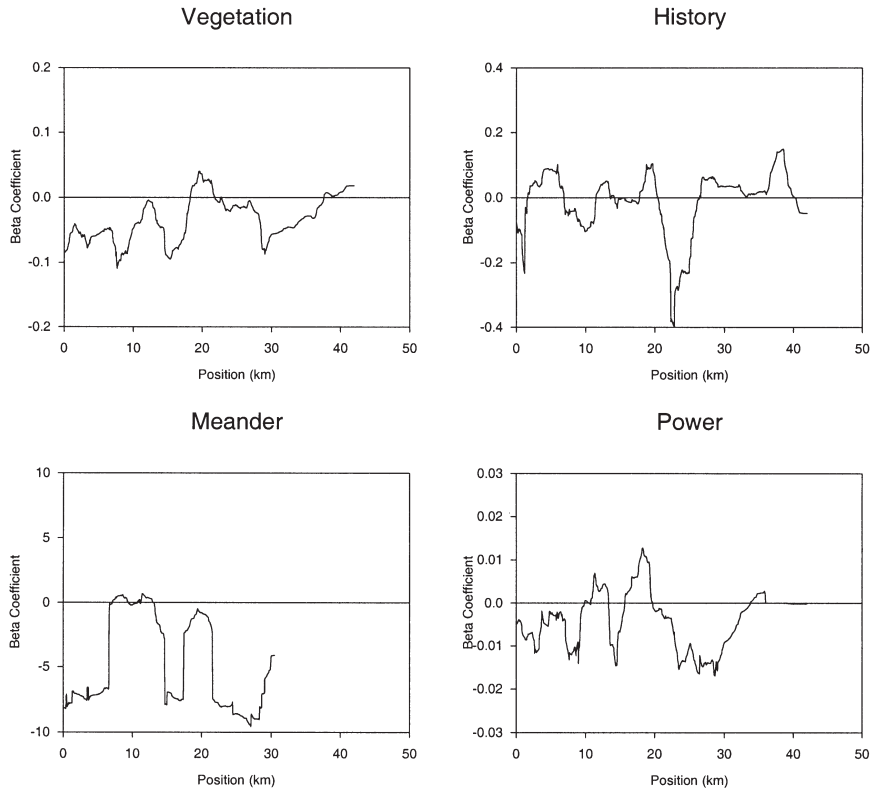


FIG. 6a. The Regression Coefficients for Vegetation, History, Meander, and Power Plotted Against Distance Upstream for the Right Bank.

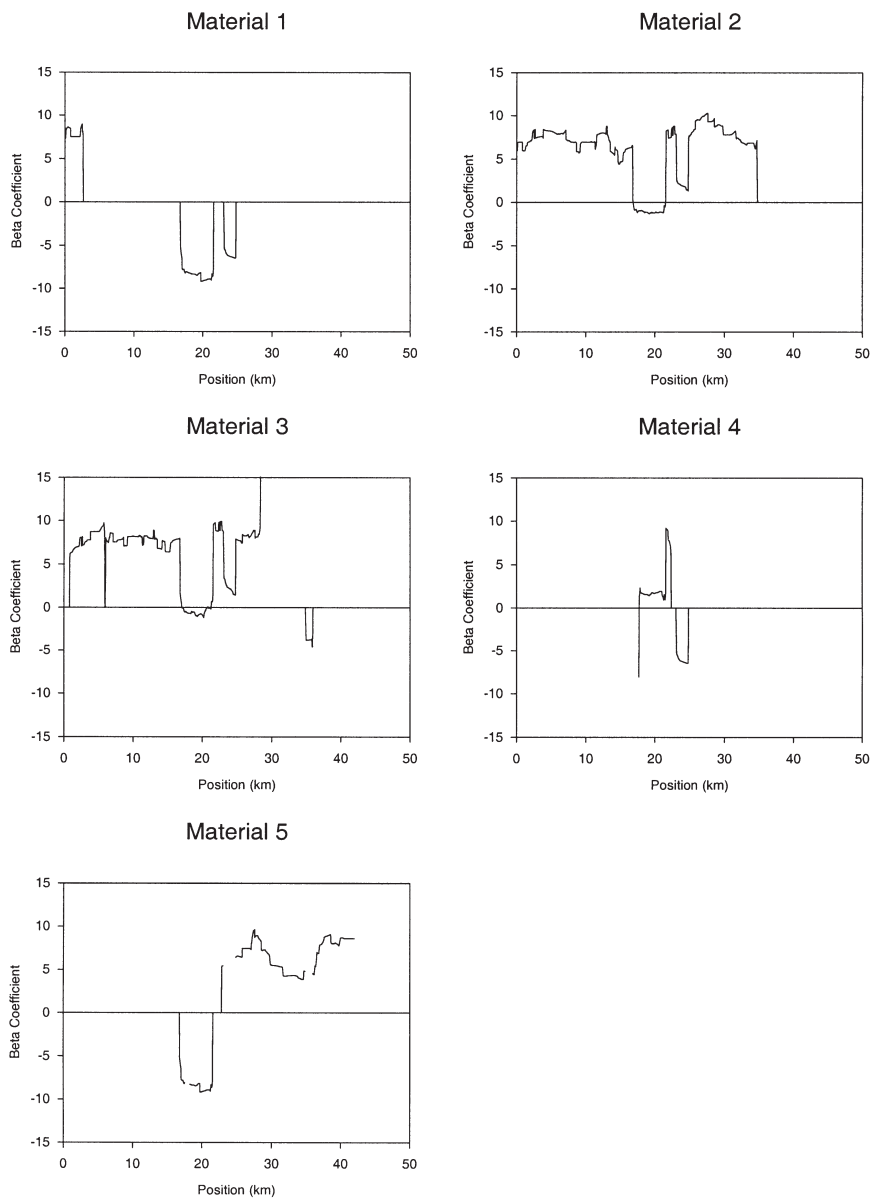


FIG. 6b. The Regression Coefficients for Material Plotted Against Distance Upstream for the Right Bank. Material is divided into five binary components as described in the text.

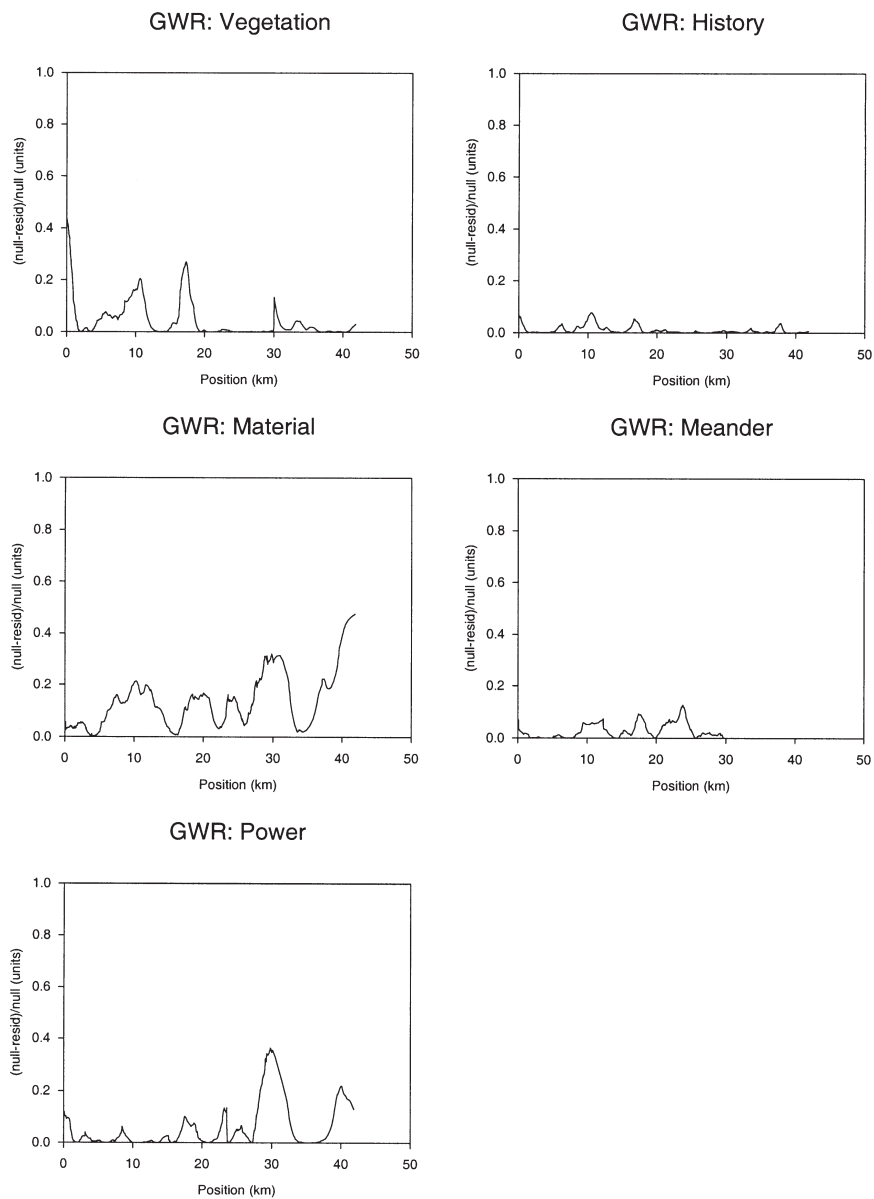


FIG. 7. Deviance Ratio ( $[\text{Null deviance}-\text{Residual deviance}]/\text{Null deviance}$ ) Plotted Against Distance Upstream for the Left Bank.

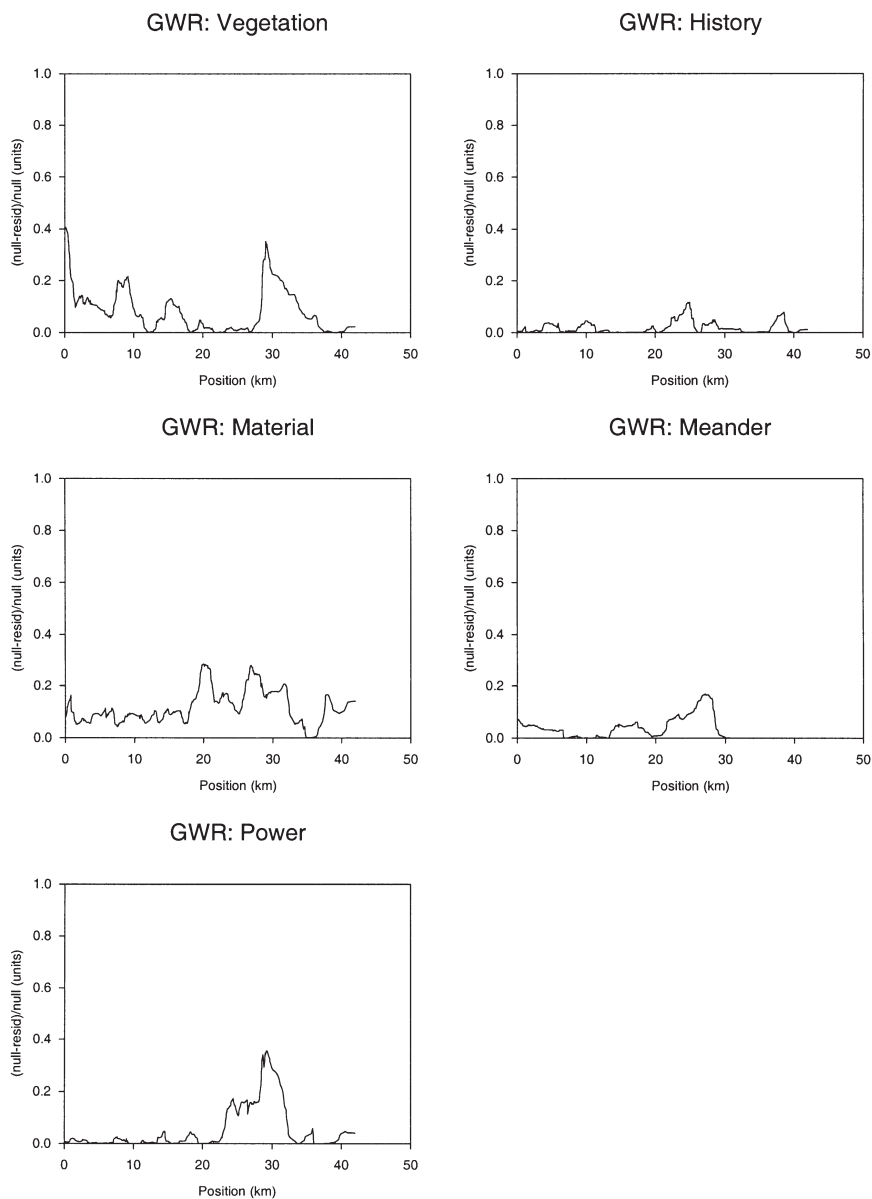


FIG. 8. Deviance Ratio ( $[\text{Null deviance}-\text{Residual deviance}]/\text{Null deviance}$ ) Plotted Against Distance Upstream for the Right Bank.

## DISCUSSION

*GWLR in Physical Geography*

GWR was developed primarily for application in human geography, for example, to socioeconomic data such as provided by the UK census. In human geography, it is reasonable to expect geographical variation in relations. For example, house prices in the urban area of Southampton might depend on an entirely different set of variables to house prices in its rural surroundings. One example would be the relation between house price and proximity to playing fields. In the urban area this might increase house price, whereas in the rural area it might decrease it. The most obvious measurable quantity that would describe such a difference is geographical location.

In physical geography, it is generally assumed that variation in some property over space cannot be due to space itself, but rather is due to some control that varies over space. The positivist view commonly held by physical geographers suggests that observations can be generalised into laws (Fotheringham, Brunsdon, and Charlton 2000). Thus, the goal of much research in physical geography is to find the underlying causal factors responsible for variation in some property (or in the present case, its relation with other properties) over space. In terms of regression modelling, it should be possible to build a single model that describes the entire set of possible circumstances that may arise. However, a general model describing such a large set of possible circumstances would include a large number of variables, possibly non-linear relations and a large number of interaction terms. Such a model would be cumbersome and difficult to interpret.

An advantage of GWR (or GWLR) is that by fitting the model locally, certain variables may be redundant in the local model because they are effectively held constant over space. Such terms have low-frequency or long-range variation. For example, in the case of the Afon Dyfi it is possible that elevation or some second-order derivative of elevation has the greatest influence on erosion. Since this property has not been measured, the *global* regression model would fail to describe well the relation between erosion and other geomorphological variables. However, within a region for which elevation varies little, the *local* model may explain most of the *local* variation in erosion. Thus, in an applied context, there is merit to suspending the search for causal factors, instead producing predictive models that have value locally and reduce the need to acquire large numbers of data in the field.

*GWLR in Geomorphology*

The attraction of GWLR analysis to geomorphology lies in the ability to explore spatial variation in the relations between driver variables and process—in this case, bank erosion. Such analysis permits the identification of zones of process control, which might be expected to vary between reaches of a river, but also between catchments. Using the Afon Dyfi as a case study, it becomes possible to comment on the widely held theoretical view that a specific sequence of process domains exist along a watercourse (Lawler 1992; Abernathy and Rutherford 1998). The GWLR analysis of the Dyfi data set demonstrates that the scale and sequencing of process domains may be much finer and more varied than existing models suggest. Rather, controls on bank erosion processes vary at a fine spatial frequency, primarily according to the available energy to drive erosion (stream power) and the resistance to that erosion provided by bank materials and vegetation. For the Dyfi at least, channel planform and the historical activity of the channel do not appear to influence the process or location of contemporary bank erosion directly, although both may control the location and nature of bank materials and vegetation.



*GWLR for Management*

The results from GWLR provide an insight into the local relations between erosion and the controls on that erosion. This information can be used, as part of a management strategy, to focus decisions regarding bank protection options. In a particular reach, the fact that, say, vegetation has an influence over erosion locally can be used to define the most suitable management recommendations *for that reach*. This allows managers to work with the natural fluvial system on a local basis. By providing understanding of the physical processes that create areas of erosion locally and of how these relations can change within a small area of the river, GWLR may allow river management to be more targeted and effective.

The remainder of this discussion is concerned with several issues which must be addressed when implementing GWLR.

*Other Issues*

Although some important advances of understanding and management application have been achieved, it is important to exercise caution when interpreting the results of GWLR, and this is particularly highlighted by the example of bank material. Since the categories of material exhibit relatively low frequency spatial variation, the successive peaks and troughs in the influence of material on erosion along the Dyfi may reflect local variation (large influence) and lack of local variation (small influence) between the classes of material. Thus, the deviance ratio will tend to be small where there is little or no variation in the category of material locally, and has the potential to be large only where variation in material exists locally.

An issue that is closely related to that above is significance. GLM is normally an exploratory technique: models are developed interactively, which provides an experienced user with a real “feel” for the data. This interactive procedure for fitting a logistic model sits awkwardly alongside GWR for which it is a requirement that fitting is automatic. In GWR, models must be calibrated for each desired location  $j$  of which there are likely to be between one hundred and one million, precluding interactive fitting. A consequence of automatic fitting is that some model coefficients are likely to be insignificant for certain locations, and for some locations it is possible that the entire model may not be significant. The issue of lack of significance is discussed by Brunsdon, Fotheringham, and Charlton (1996), but is effectively avoided when they map the entire set of  $m$  coefficients from GWR.

A further problem with GWLR that should be apparent from Equations 8 to 11 is how to choose an appropriate non-linear parameter  $a$  for the weighting function. In this paper, a choice was made following experimentation. However, an alternative is the cross-validation approach suggested by Cleveland (1979), in which the values at each  $i$  are omitted when calibrating the regression model at  $i$ :

$$\sum_{i=1}^n [y_i - y_{\neq i}^*(a)]^2 \quad (16)$$

where  $y_{\neq i}^*(a)$  is the fitted value of  $y_i$  with the observation for location  $i$  omitted from the regression calibration. The optimal value of  $a$  can be chosen as the one that minimises the cross-validation score. An important point, however, is that where the objective is to explore relations (rather than to predict), variation in the regression model with  $a$  can be informative.

## CONCLUSION

The major attraction of the GWLR approach lies in its ability to permit objective quantification of spatial variation in the magnitude and direction of the relations between some process-related response variable (bank erosion in this example) and its controlling variables. In the present example, for the Afon Dyfi, Wales, the model parameters and the residual deviance of the model varied greatly with distance upstream. The practical implication of this result is that different management practices should be implemented at different locations along the river. Thus, the approach presented allowed inference of spatially varying management practice as a consequence of spatially varying geomorphological process.

Future research will focus on extending the model to (i) the prediction of categories and magnitudes of erosion (c.f., a binary response), (ii) multivariate modelling, and (iii) new application areas in geomorphology.

## LITERATURE CITED

- Abernethy, B., and I. D. Rutherford (1998). Where along a River's Length Will Vegetation Most Effectively Stabilise Stream Banks? *Geomorphology* 23, 55–75.
- Aitkin, M., D. Anderson, B. Francis, and J. Hinde (1989). *Statistical Modelling in GLIM*. Oxford: Oxford University Press.
- Atkinson, P. M., and R. Massari (1998). Predicting the Relative Likelihood of Landsliding in the Central Apennines, Italy. *Computers and Geosciences* 24, 373–85.
- Augustin, N. H., M. A. Mugglestone, and S. T. Buckland (1996). An Autologistic Model for the Spatial Distribution of Wildlife. *Applied Ecology* 33, 339–47.
- Bagnold, R. A. (1966). An Approach to the Sediment Transport Problem from General Physics; *U.S. Geol. Survey Prof. Paper* 422-J.
- Baker, R. J., and J. A. Nelder (1978). *The GLIM System: Release 3*. Oxford: Numerical Algorithms Group.
- Brunsdon, C., A. S. Fotheringham, and M. E. Charlton (1996). Geographically Weighted Regression: a Method for Exploring Spatial Nonstationarity. *Geographical Analysis* 28, 281–98.
- Bull, L. J. (1997). Magnitude and Variation in the Contribution of Bank Erosion to the Suspended Sediment Load of the River Severn. *Earth Surface Processes and Landforms* 22, 1109–1123.
- Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* 74, 829–36.
- Collett, D. (1991). *Modelling Binary Data*. London: Chapman and Hall.
- Dobson, A. J. (1990). *An Introduction to Generalised Linear Models*. London: Chapman and Hall.
- Fotheringham, S., C. Brunsdon, and M. Charlton (2000). *Quantitative Geography. Perspectives on Spatial Analysis*. London: Sage.
- German, S. E. (2000). A Geomorphological Approach to the Strategic Management of River Bank Erosion: A Case Study of the Afon Dyfi. Ph.D. diss., University of Southampton.
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. New York: Oxford University Press.
- Gurnell, A. M. (1997). Channel Change on the River Dee Meanders, 1946–1992, from the Analysis of Air Photographs. *Regulated Rivers: Research and Management* 13, 13–26.
- Healy, M. J. R. (1988). *GLIM: An Introduction*. Oxford: Clarendon Press.
- Hooke, J. M. (1995). Processes of Planform Change on Meandering Channels in the UK. In *Changing River Channels*, edited by A. M. Gurnell and G. E. Petts, 87–115. Chichester: John Wiley and Sons Ltd.
- Jones, A. P. (1995). Provenance and Palaeoenvironments of Late Quaternary Sediments and Their Associated Landforms. Ph.D. diss., University of Wales, Aberystwyth.
- Lawler, D. M. (1992). Process Dominance in Bank Erosion Systems. In *Lowland Floodplain Rivers: Geomorphological Perspective*, edited by P. A. Carling and G. E. Petts. Chichester: John Wiley and Sons Ltd.
- Lawler, D. M. (1993). The Measurement of River Bank Erosion and Lateral Change: A Review. *Earth Surface Processes and Landforms* 18, 777–821.
- Lawler, D. M., C. R. Thorne, and J. M. Hooke (1997). Bank Erosion and Instability. In *Applied Fluvial Geomorphology for River Engineering and Management* edited by C. R. Thorne, R. D. Hey, and M. D. Newson, 137–72. Chichester: John Wiley and Sons Ltd.
- Matheron, G. (1965). *Les Variables Régionalisées et Leur Estimation*. Paris: Masson.
- Matheron, G. (1971). *The Theory of Regionalized Variables and its Applications*. Fontainebleau: Centre de Morphologie Mathématique de Fontainebleau.
- Millar, R. G. (2000). Influence of Bank Vegetation on Alluvial Channel Patterns. *Water Resources Research* 36, 1109–18.

- Rhoads, B. L. (1991). A Continuously-Varying Parameter Model of Downstream Hydraulic Geometry. *Water Resources Research* 27, 1865–72.
- Rhoads, B. L. (1992). Statistical Models of Fluvial Systems. *Geomorphology* 5, 433–55.
- Sear, D. A., M. D. Newson, and A. Brookes (1995). Sediment Related River Maintenance: The Role of Fluvial Geomorphology. *Earth Surface Processes and Landforms* 20, 629–47.
- Simon, A., and C. R. Hupp (1992). Geomorphic and Vegetative Recovery Processes along Modified Stream Channels of West Tennessee. *U.S. Geological Survey Open-File Report*. Nashville, Tenn.: USGS, 91–502.
- Simon, A., and S. Darby (1995). Preliminary Analysis of Bank-Stability Problems on the Upper Missouri River below Ft. Peck Dam. *Report to the USDA-Agricultural Research Service*. Oxford, Miss.: USD National Sedimentation Laboratory.
- Thorne, C. R. (1982). Processes and Mechanisms of River Bank Erosion, in *Gravel-bed Rivers*, edited by R. D. Hey, J. C. Bathurst, and C. R. Thorne, 227–71. Chichester: John Wiley and Sons.
- Thorne, C. R. (1990). Effects of Vegetation on Riverbank Erosion and Stability. In *Vegetation and Erosion*, edited by J. B. Thornes, 125–43. Chichester: John Wiley and Sons Ltd.
- Thorne, C. R. (1997). Channel Types and Morphological Classification. In *Applied Fluvial Geomorphology for River Engineering and Management* edited by C. R. Thorne, R. D. Hey, and M. D. Newson, 175–222. Chichester: John Wiley and Sons.
- Thorne, C. R., R. G. Allen, and A. Simon (1996). Geomorphological River Channel Reconnaissance for River Analysis, Engineering and Management. *Transactions of the Institute of British Geographers* 21, 469–83.
- Uno, T., T. Sugii, M. Hayashi (1994). Logit Model for River Levee Stability Evaluation Considering the Flood Return Period. *Structural Safety* 14, 81–102.
- Wang S.-Q., and D. J. Unwin (1992). Modelling Landslide Distribution on Loess Soils in China: An Investigation. *International Journal of Geographical Information Systems* 6, 391–405.
- Wharton, G. (1995). Information Form Channel Geometry-Discharge Relations. In *Changing River Channels*, edited by A. Gurnell and G. Petts, 325–46. Chichester: John Wiley and Sons Ltd.