

# GENERALIZED LINEAR MODELLING IN GEOMORPHOLOGY

PETER ATKINSON<sup>1</sup>, HESTER JISKOOT<sup>2\*</sup>, REMO MASSARI<sup>1</sup> AND TAVI MURRAY<sup>2</sup>

<sup>1</sup> Department of Geography, University of Southampton, Highfield, Southampton, SO17 1BJ, UK

<sup>2</sup> School of Geography, University of Leeds, Leeds, LS2 9JT, UK

Received 21 January 1997; Revised 20 May 1997; Accepted 18 July 1997

## ABSTRACT

Generalized linear modelling (GLM) is a statistical technique used to model the relation between a response variable and a set of explanatory variables. GLM is similar to the well known multiple regression. However, GLM is a powerful technique for exploratory data analysis with many advantages over more traditional techniques. For example, GLM allows the incorporation of categorical as well as continuous response and explanatory variables in the analysis. In this paper, GLM is explained and two examples of the application of the technique in geomorphology are given. The first example involves glacier surging and the second involves landslide susceptibility. The examples demonstrate the relevance of GLM to many common problems in geomorphology. © 1998 John Wiley & Sons, Ltd.

KEY WORDS: generalised linear modelling; logit regression; GLIM; glacier surging; landslide susceptibility

## INTRODUCTION

Understanding of physical processes in a geomorphological system can be aided by modelling the relations and interactions between the different components or variables in that system. Statistical black box approaches, such as multiple regression, provide predictive tools which can aid our understanding of physical relations. However, there are several limitations to the applicability of traditional linear models to geographical systems. One of the most important is that linear models which are based on ordinary least squares (OLS) regression are applicable to continuous response variables only. Certain geomorphological variables, such as type of lithology and type of morphology, cannot be quantified on a continuous scale. These variables are usually defined as categories, with either nominal or ordinal scales.

Generalized linear modelling (GLM) (Aitkin *et al.*, 1989; Collett, 1991) is similar to (multiple) linear regression. However, in linear regression the observed response variable and the values predicted by the linear model are linked linearly (they are the same). In GLM a link function is defined which allows the predicted values to be transformed to (or mapped onto) a variety of distributions including a binary response. Therefore, GLM allows the modelling of many different types of response variable with many different distributions. In simple terms, the variables involved (both response and explanatory) may be either continuous or categorical (ordinal or nominal), or any combination of both types.

As the name suggests, GLM is designed to model linear relations between a response variable and a set of explanatory variables: the underlying statistical model is linear and additive. Therefore, as for linear regression, transformation of the data prior to modelling may be necessary to fit non-linear relations. Other methods of distinguishing (dichotomous) groups are eigenvalue techniques such as principal component analysis, discriminant analysis, cluster analysis, factor analysis and correspondence analysis (Davis, 1986), but these techniques lack the predictive power of GLM. Useful introductions to GLM are given in Dobson (1990) and Collett (1991).

GLM has been applied in biostatistics since the 1940s, and more recently in medical statistics, econometry, medical geography and population geography. However, the technique has received relatively little use in

\* Correspondence to: H. Jiskoot, School of Geography, University of Leeds, Leeds, LS2 9JT, UK

E-mail: H.Jiskoot@geog.leeds.ac.uk

Contract/grant sponsor: University of Leeds, School of Geography

Contract/grant sponsor: San Marino Government

geomorphology (e.g. Uno *et al.*, 1994; Downs, 1995; Siegel *et al.*, 1995). In this paper, the applicability of GLM in geomorphology is explored and two examples from different areas of the discipline are presented; in each, the influence of several explanatory variables on the occurrence of a sudden morphological event is modelled.

### GENERALIZED LINEAR MODELLING

The description of GLM given in this section is restricted to the modelling of a binary response with the logistic regression.

#### *The logistic model*

In geomorphology, it is often necessary to explain the occurrence of some phenomenon (for example, landsliding) at several locations. Presence or absence can be modelled as a discrete binary process where 1='presence' and 0='absence'. Then, the objective is to link this binary response variable to, for example, environmental variables to ascertain what conditions affect presence or absence. We can do this in terms of the probability of occurrence of an event given several explanatory variables. For binary response variables it is convenient to model this probability in terms of the log odds of improvement, called the logit, and denoted as:

$$\text{logit}(P_i) = \frac{e^{f(X_i)}}{1 + e^{f(X_i)}} \quad (1)$$

where  $P_i$  is the probability of an event occurring associated with a given observation  $i$  (Dobson, 1990; Collett, 1991). The logit model can be used as a link function in GLM to produce the so-called logistic regression model. Equation 2 gives the form of the linear logistic model in which the left-hand side is the logistic link function and the right-hand side is the linear predictor:

$$f(X_i) = \alpha + \sum_{k=1}^n \beta_k X_{ik} \quad (2)$$

where  $\beta_k$  are the  $n$  coefficients to be estimated in the model, and  $X_{ik}$  are the  $n$  values for each of the explanatory variables. The logistic model can take two major forms, one expressing the model in logit (with the alternative of transforming this in odds), and one expressing the model in event probability (Liao, 1994). It is usually more convenient (in terms of plotting and interpreting the results) to express the estimated values as probabilities.

#### *Model fitting*

The GLM is best fitted in a step-wise manner. First, bivariate models are fitted to explore the relations between the response variable and the individual explanatory variables, and the most significant variable is added to the working model. Each further variable is then added in a step-wise fashion, including only the most significant variable at each step. In this way, each new variable is modelled while controlling for the effects of variables included previously. Eventually, no further variables will be significant at the chosen confidence interval.

The significance of a model may be determined from the model deviance  $D$ , which for binary data in a GLM may be expressed as:

$$D = -2 \sum_{i=1}^n P_i \text{logit}(P_i) + \log(1 - P_i) \quad (3)$$

The model deviance itself 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).

When terms are added to the model in a step-wise fashion, the significance of each term is checked at each step by finding the probability for a given reduction in model deviance and number of degrees of freedom. Only those terms which are significant at a given confidence level (for example, 95 per cent) are retained. Once a term has been found to be significant, the estimated coefficient (or coefficients in the case of categorical terms) must be tested for significance using Student's *t*-value (estimated coefficient/standard error). At the 95 per cent significance level a coefficient is significant when it is greater than approximately twice the standard error.

A potential problem exists for categorical variables which comprise several binary terms because some terms may be significant while others are not. It is common practice to aggregate such categorical variables into fewer, coarser classes to reduce the problem and also to simplify the resulting GLM. However, care should be taken as it is possible to lose vital information in this aggregating process.

It is common to test the significance of interaction terms as new main terms are added to the model (see Healy, 1988). For example, when a second main term is added to the model, the interaction between the two main terms (equal to the product of the two main terms) should be checked. When a third main term is added, the interaction between the three combinations of two main terms and the interaction between the three main terms should be checked. Clearly, the number of interaction terms increases exponentially with the number of main terms, and interaction terms between three or more variables can be difficult to interpret. For these reasons, interaction terms are usually limited to those between pairs of explanatory variables. For categorical variables which may be divided into several binary terms, the number of interaction terms may become large. Once again, the solution is to aggregate into fewer, coarser classes.

#### *Model interpretation*

The primary goal of GLM is prediction. However, in geomorphology, the cause and effect relation between response and explanatory variables is often unclear. Further, there are often many errors and much uncertainty in geomorphological primary (field) and secondary data, and outliers in the data are common. Therefore, while prediction remains the ultimate goal of GLM, one of the most important uses of GLM in geomorphology is as a means to quantify and explore the relations between response and explanatory variables with the eventual goal of understanding the underlying geomorphological processes.

In a linear model involving OLS regression, the coefficients of the regression model are interpreted on their sign and magnitude. A positive coefficient means that there is a direct relation between the response variable and the explanatory variable, and (for standardized variables) the magnitude of the model coefficient determines the strength of correlation. For linear regression this means that a unit increase in the explanatory variable results in a coefficient-times-unit increase in the response variable.

For GLM the interpretation is not as direct, but is also based on the sign and magnitude of the coefficient (Liao, 1994). For continuous variables, the coefficients are interpreted as in OLS regression except that the correlation is not linearly related to the magnitude of the coefficient. For GLM with the logit link the correlation is usually expressed as an odds ratio (Liao, 1994). For categorical variables, a positive coefficient implies an above-average correlation and a negative coefficient implies a below-average correlation. Interpretation of interaction terms is less straightforward, particularly if the interaction term is a mixture of data types (see Collett, 1991).

#### *Accuracy assessment*

While the physical interpretation of individual model coefficients is straightforward, the performance of the GLM can be difficult to assess. Residual analysis using a reliable set of test data is the best way to validate the GLM. Comparisons between the predictions and the validation data usually take the form of index plots and half-normal plots of the residuals (Collett, 1991, p. 131). These plots can lead to the identification of outliers and their possible removal from the model if a valid argument exists.

### *The GLIM software*

The GLIM (Generalized Linear Interactive Modelling) system software was used for GLM in both examples in this paper. GLIM was developed by the Royal Statistical Society and is a UNIX- and PC-based interactive statistical modelling software package. It has a concise command language with a specific syntax allowing the user to fit models interactively. Specific examples of applications of GLIM are given in Baker and Nelder (1978), Healy (1988) and Aitkin *et al.* (1989).

GLIM allows a range of link functions between the response and explanatory variables (for example, exponent, logit, probit, logarithm, reciprocal and square root) and provides several possibilities for testing significance. Further, GLIM allows the user to assume error distributions other than Gaussian, such as binominal, gamma and Poisson. A range of mathematical functions for transforming and manipulating variables is provided.

The modelling procedure in GLIM is standard for all models. The data are entered, the data types and the response variable are defined and the type of model is specified. Then the model can be fitted interactively. The model output is given as (i) estimated coefficient, (ii) standard error of the estimate, (iii) model deviance and (iv) degrees of freedom. The estimated probabilities and residuals can be calculated for each model and displayed graphically.

## EXAMPLES IN GEOMORPHOLOGY

In this section, two applications of GLM in geomorphological research are presented. First, GLM is used to model the relation between glacier surging and a set of glacier variables, and second, GLM is used to predict susceptibility to landsliding.

### *Case 1: Glacier surging*

Glacier surging is a cyclic flow instability in which slow flow of the glacier during a long period of quiescence is followed by extremely fast flow during a relatively short surge phase. The spatial distribution of glaciers that exhibit this phenomenon is non-uniform on both a global and a regional scale (Clarke *et al.*, 1986). In most areas there are no surge-type glaciers, but in other areas clusters of surge-type glaciers exist. Despite extensive observations on several surge-type glaciers, understanding of the mechanism(s) and trigger(s) that control this flow instability is incomplete.

The aim of this case study is to analyse a population of surge-type and non-surge-type glaciers to model the relation between surging and several glacier variables. Several theoretical treatments of glacier surging suggest that particular variables must cross critical thresholds for surging to occur. Examples of these variables are glacier length (e.g. Weertman, 1969; Kamb, 1987), glacier slope (e.g. Budd, 1975; Kamb, 1987), glacier thickness (e.g. Fowler and Johnson, 1995), and 'Fowler's index' – the product of glacier width squared and bed slope (Fowler, 1989, equation 4.8). Thus, it should be possible to use criteria such as the glacier geometry to distinguish surge-type glaciers from normal glaciers. Determining the variables that control surging should provide vital clues to the surge mechanism and trigger, and provide a test of present surge theories.

Previous studies have indicated that no single variable is able to explain the geographical distribution of surge-type glaciers adequately (Glazyrin, 1978; Clarke *et al.*, 1986; Wilbur, 1988; Hamilton and Dowdeswell, 1996). Moreover, bivariate analyses are difficult to interpret because of statistical interdependence between the explanatory variables. For example, glacier length and width have a strong positive correlation, and glacier length and slope a strong negative correlation (Clarke, 1991; Jiskoot *et al.*, 1998). Therefore, long glaciers tend to be large, and are thus wide and have low slopes. It is necessary to disentangle these multiple correlations via a fully multivariate analysis. Prior to the application of GLM to the present problem, traditional multivariate analysis was applied to model glacier surging in the St Elias Mountains, Yukon Territory, Canada (Clarke, 1991).

Table I. Significant multivariate logit models for glacier surging with the model deviance (D), degrees of freedom (DF), estimated coefficients (E) and standard errors (SE). The optimal multivariate logit model for glacier surging is shown in italics

Variable*	Subgroup	E	SE	D	DF
–	'Null model'	–3.179	0.1228	579.49	1725
L+W	base	–8.148	0.7302	306.84	1723
	log-L	7.027	0.7131		
	log-W	–2.338	0.8924		
L+w	base	–7.128	0.4852	293.83	1723
	log-L	7.045	0.6284		
	w	–0.00998	0.00232		
L+H	base	–9.487	0.9342	300.69	1723
	log-L	5.739	0.5271		
	H	0.00116	0.00032		
L+M	base	–11.16	1.072	284.68	1723
	log-L	7.820	0.7195		
	M	0.00215	0.00042		
<i>L+L.w+M</i>	<i>base</i>	<i>–12.08</i>	<i>1.161</i>	<i>262.75</i>	<i>1722</i>
	<i>log-L</i>	<i>9.649</i>	<i>0.9141</i>		
	<i>log-L.w</i>	<i>–0.0071</i>	<i>0.0014</i>		
	<i>M</i>	<i>0.0021</i>	<i>0.0004</i>		

\* L=glacier length, W=mean glacier width, w=Fowler's index, L.w=the interaction term between glacier length and Fowler's index, H=maximum altitude, M=minimum altitude

### Data

Data for the analysis were selected from the Yukon Glacier Inventory, which is a component of the Canadian Glacier Inventory (Ommanney *et al.*, 1973). The Yukon Glacier Inventory includes a surge index related to observed surges and morphological evidence of surging. Each glacier is assigned a code from 0 (no surge features) to 5 (definite surge evidence). Based on the surge probability schemes published in Clarke *et al.* (1986) the glaciers were divided into non-surge-type (digits 0–3) and surge-type (digits 4 and 5).

To compare the performance of GLM to Clarke's (1991) multivariate analysis we used a similar set of glaciers and variables, but added the categorical variable 'type of glacier front' to illustrate the power of GLM. Following Clarke (1991) we excluded all rock glaciers, remnants and non-real glacier features as well as all tributary glaciers from the 4675 entries in the Yukon Glacier Inventory, leaving 1726 glaciers for the GLM analysis. From these glaciers, 69 (4 per cent) glaciers were classified as surge-type. Since an object-based approach was adopted it was possible to represent the entire glacier population in the test data.

### Analysis

The GLM analysis was performed using a logit link and binomial error distribution. The response variable in the logit model is a dichotomous surge index,  $S$ , with  $S=1$  for surge-type glaciers and  $S=0$  for non-surge-type glaciers. The logit model was fitted for a set of continuous variables, that is, glacier length, mean glacier width, surface slope, minimum and maximum glacier altitude and 'Fowler's index' (Fowler, 1989), and the categorical variable, type of glacier front. The variables glacier length, width and surface slope were log-normally distributed and the logarithmically transformed variables sometimes resulted in a better linear fit. Therefore, the log-transformed data were used as separate variables in the model.

The null model (model without variables fitted) has a model deviance of 579.49 for 1725 degrees of freedom and the estimate is significantly negative (Table I). The log-transformed glacier length produced the greatest reduction in deviance of the bivariate models. Therefore, the other variables were added to this model in a step-wise fashion to determine which second explanatory variable led to the greatest further decrease in deviance. Each additional variable in the model was accompanied by a test of the interaction terms of the variables. All variables were tested, but only significant variables are shown in Table I.

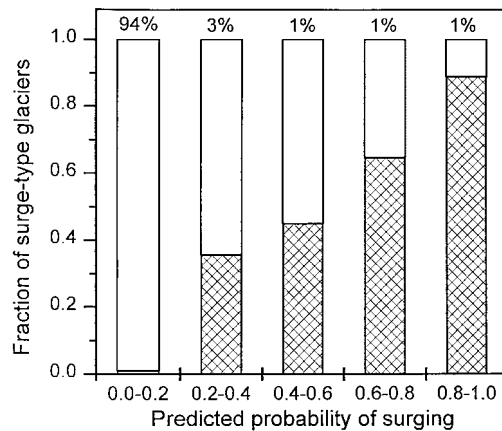


Figure 1. Comparison between the optimal multivariate logit model of glacier surging and observations of surge-type glaciers. Bar chart shows the fraction of observed surge-type (hatched) and non-surge-type (white) glaciers in each of five output classes. These output classes represent the predicted probabilities that the glacier is of surge-type from 0.0 (normal) to 1.0 (surge-type). The values above the bars give the percentage of total number of glaciers covered by each class

The variable that caused the greatest reduction in model deviance in combination with log-length was the minimum altitude. Thus, long glaciers with their margins at high altitudes are likely to be of surge-type. It is notable that when glacier width is fitted in the bivariate model its coefficient is positive, whereas in combination with log-length, the sign is negative (Table I). This indicates that long glaciers with relatively narrow width are more likely to surge.

Adding Fowler's index to the bivariate model involving log-length also caused a large reduction in model deviance. In this model, the coefficient for log-transformed length is positive, whereas the coefficient for Fowler's index is negative, but still significant as the standard error is less than half the value of the estimated coefficient. When fitting Fowler's index in a bivariate model the null model was not improved significantly, whereas when fitted in combination with glacier length, there is a significant correlation between glacier surging and Fowler's index. This may be explained by the interaction term between glacier length and Fowler's index: the interaction term appeared more significant than the relation between glacier surging and Fowler's index. As it is in most cases easier to interpret the model with only the interaction term (Healy, 1988), we retained this interaction term rather than Fowler's index in the final model.

The optimal logit model, with a reduction in model deviance of 316.6 for a loss of three degrees of freedom, indicated that the greatest correlation between glacier surging and the set of glacier variables tested consists of three variables: glacier length, the interaction term between Fowler's index and glacier length, and minimum altitude (Table I). The model coefficients for length and altitude are positive, and that for the interaction term is negative. A long, high altitude glacier is most likely to be of surge-type. The presence of the interaction term in the final model suggests that the influence of Fowler's index on glacier surging varies with length, and suggests that Fowler's index is more important for long glaciers.

### *Model validation*

To assess the model, probabilities were estimated for each individual glacier in the test data and the residuals examined. If the model was perfect the predicted probability of the response variable would be 1 for every surge-type glacier and 0 for every non-surge-type glacier. However, the probabilities are continuous, ranging from 0 to 1, so that exact correspondence is not possible. Outliers may indicate unusual cases which are not covered by the set of explanatory variables used in the model, or these outliers may be misclassified units, such as a surge-type glacier incorrectly classified as a normal glacier. Figure 1 shows graphically the relation

between the observed surge index (surge-type or non-surge-type) and the probabilities of surging predicted by the logistic model.

Clarke (1991) considered four geometric glacier variables in his multiple regression analysis: glacier length, glacier width, glacier slope and Fowler's index. Using traditional multiple regression, glacier surging was explained by glacier length only; the remaining variables had no additional influence (Clarke, 1991). Using a logit model we demonstrated that Fowler's index does influence glacier surging, but the influence varies with glacier length. Fowler's index was predicted by Fowler (1989) to be smaller than a certain critical value for surge-type glaciers. Our model agrees with this prediction, as surging and Fowler's index are negatively correlated.

### *Case 2: Landslide susceptibility*

Information on the spatial distribution of the probability of landsliding in the future (the landslide hazard) would provide a useful aid for a variety of planning and civil engineering applications. However, a landslide hazard map should include an indication of the time scale within which a particular landslide is likely to occur. In practice, data on the temporal dimension of the landslide hazard are difficult to obtain, since landslide triggering is usually due to external (extrinsic) causes (for example, rainfall or earthquakes). For this reason, landslide hazard maps are usually replaced by landslide susceptibility maps (Carrara *et al.*, 1996). Landslide susceptibility maps may be obtained directly in the field, but are more commonly obtained indirectly based on measurements of several (intrinsic) properties (such as type of geology and slope angle) which affect slope stability.

In Italy, landslides are often forested to consolidate the soil. This implies a direct correlation between susceptibility to landsliding and forestation. However, the relation between susceptibility to landsliding and forestation is actually likely to be inverse. Therefore, when producing a GLM it is important to identify conditions prior to landsliding. When the model constructed between landsliding and prior conditions is then applied to present conditions, the tendency is to map susceptibility to landsliding in the future.

Previously, researchers have taken many different approaches to mapping susceptibility to landsliding including heuristic and statistical approaches (Hansen, 1984; Carrara *et al.*, 1996). In the statistical approach, a statistical model of the relation between susceptibility to landsliding (the response variable) and a set of explanatory variables is constructed based on sample data. This model is then applied to map susceptibility to landsliding. Simple statistical techniques have been used widely (Gupta and Joshi, 1989; Siddle *et al.*, 1991; Maharaj, 1993), but relatively few researchers have used fully multivariate statistical approaches (Neuland, 1976; Carrara *et al.*, 1991, 1996; Bernknopf *et al.*, 1988; Wang and Unwin, 1992; Jäger and Wieczorek, 1994).

In this example, a geographic information system (GIS) was used in combination with GLM to model the relation between landsliding and several explanatory variables, chosen and defined to reflect conditions prior to landsliding. Two models were created: one for all landslides, and one for active landslides only. Further models are developed by Atkinson and Massari (1998).

### *Data*

A study area was chosen in the Umbro–Marchean Apennines in central Italy because the area was known to contain many landslides and a large amount of variation in the explanatory variables. The site lies between latitudes 43°30' and 43°35'N and longitudes 12°27' and 12°32'E and covers an area of 65 km<sup>2</sup>. The outcropping geological formations range in origin from the Jurassic–Cretaceous limestones and limestones with marls in the northeast to the Miocene flysch formations in the southwest.

Previous slope movements were mapped using air photo interpretation and field checking. A digitized map of landslide boundaries was input to the GIS and a vector-to-raster conversion undertaken to provide a raster image of slope movement with a spatial resolution of 20 m by 20 m. Slope movements were then classified by type following Varnes' (1978) and Wieczorek's (1984) classification systems. For the present analysis slope movements were separated into dormant and active landslides. Within each slope, movement feature, body rupture and deposit areas were identified since only the features of the rupture area were deemed relevant to mapping susceptibility to landsliding.

Table II. (a) Independent variables included in both models of landslide susceptibility. (b) Additional independent variables included in the model of active landslides only

Index	Variable	Code
<b>(a)</b>		
1	Geology	RGEOL
	Lithology (RGEOL(1,2)(3,4,5,6)(7))	Litho
	Limestones (RGEOL(3,4,5,6))	Lim
	Marls (RGEOL(1,2))	Marl
2	Dip	DIP
3	Strike	STR
4	Strata–slope	ATT
5	Aspect	RASP
6	Lineament density	L5
7	Lineament density squared	L5Q
8	Slope angle	LSP
9	Slope angle squared	LSQ
<b>(b)</b>		
1	Vegetation cover	RVEG
2	Soil thickness (categorical)	SOIL
3	Soil thickness (continuous)	SSOIL
4	Horizontal curvature	HC
5	Vertical curvature	VC
6	Concavity of slope	CXV
7	Local relief	REL
8	Roughness	R5
9	Roughness squared	R5Q

Several different explanatory variables were selected and mapped (see Table II a and b). Elevation data were digitized from a topographic map at scale 1:10 000 and data layers were obtained on slope aspect, slope angle, slope curvature (three measures of slope curvature were defined), local relief and surface roughness. Slope angle was determined separately for each landslide unit to provide slope conditions prior to failure. A 1:50 000 scale geological map of the area (published by the Servizio Geologico Italiano) was digitized to obtain data layers on the geological formations: the dip (inclination) and strike (orientation) of the strata, and a strata–slope relation. By means of air photo interpretation and further digitizing, data layers were also obtained on the density of lineaments (within areas of 1 km<sup>2</sup>) and on vegetation cover. Finally, categorical and continuous explanatory variables were produced representing soil thickness (Table IIb). The full definitions of the complete set of categorical variables used are given in Atkinson and Massari (1998).

To ensure that no preference was given to large or small landslides, each landslide unit received equal weighting. This allocation was achieved by extracting a single pixel from the centre of each of the 442 landslide rupture areas to give a set of 442 locations where landslides were present. Since the training data were obtained from single pixels, the limit of each landslide was not considered and thus one of the main errors in landslide survey, the exact location of the border line, was circumvented (Carrara *et al.*, 1992; van Westen, 1993). To balance the number of pixels for which landslides were present against the number for which landslides were absent it was necessary to obtain 1458 pixel locations from the remaining non-landslide area. These pixel locations were obtained using a random sampling scheme. For each of the total of 1900 pixel locations, values were obtained for the presence or absence of landsliding, and for each explanatory variable.

### Analysis

An initial GLM was constructed in a step-wise fashion for all landslides based on the intrinsic properties defined in Table IIa. The significant (at the 95 per cent confidence level) explanatory variables are given in Table III in the order of their inclusion in the model. Clearly, geology and slope angle are highly significant. Further, the interaction term between geology and slope angle is significant, indicating that the dependence of susceptibility to landsliding on slope varies with geology (Table III). The density of lineaments is not significant because this term is highly correlated with geology and is, therefore, largely redundant. The category



Table III. Independent variables (given in order of significance) included in the two models of landslide susceptibility. Variables are significant at the 95 per cent confidence level

Model	All landslides	Active landslides
1	RGEOL	RVEG
2	LSP	CXV
3	LSQ	RGEOL
4	STR	LSP
5	RASP	VC
6	RGEOL.LSQ	SSOIL
7	RGEOL.LSP	HC
8	RASP.LSP	STR
9	STR.RASP	REL
10	RASP.LIM	VC.REL
11		CXV.SSOIL
12		LSQ
13		RGEOL.HC
14		LSP.SSOIL
15		RASP
16		LSP.VC
17		LSQ.VC

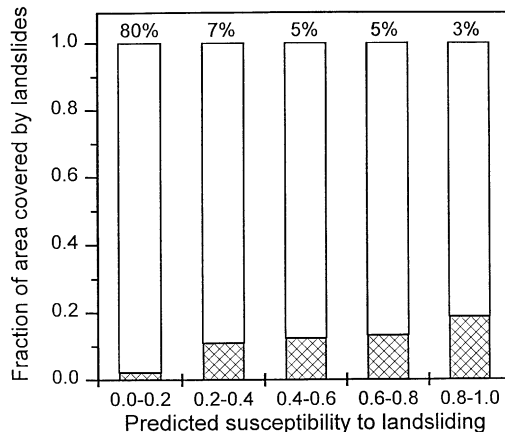


Figure 2. Comparison between the second model of active landslides (this model includes the variables in Table IIb) and observed landslides. Bar chart shows the fraction of area with observed landslides (hatched) and without landslides (white) in each of the five output classes. These output classes represent the predicted landslide susceptibilities for this model with a range from 0.0 (not susceptible to landsliding) to 1.0 (definitely susceptible to landsliding). The values above the bars give the percentage of total region covered by each class

within strike most related to landsliding is the one in the direction of the tectonic movement (northeast to east). In this orientation, the strata are most likely to be broken up by tectonic movement and are characterized by a greater dip and frequent overturning.

A second model for 'active' landslides only was produced, in which several additional explanatory variables applicable to active landslides only were included (Table IIb). The significant terms in this active landslide model are given in Table III. Now, vegetation cover and concavity (or convexity) of slope are highly significant, and more so than geology and slope angle. This significance is not surprising when one considers that active landslides are largely surface features. The relative significance of vegetation cover is interesting. In the study region, agricultural practices are related to the slope angle and soil type and, therefore, indirectly represent geology. However, we expect that vegetation cover also directly influences susceptibility to landsliding through various factors including (i) protection of the surface, (ii) consolidation of the soil and (iii) agriculture resulting

in bare soil for long periods. This expectation is reasonable because we have modelled the vegetation conditions *prior* to landsliding.

### *Model validation*

To map susceptibility to landsliding the GLM (constructed using conditions prior to landsliding) should be applied to present conditions. However, to validate this GLM, the model should be applied to conditions prior to landsliding and the estimates compared to the present landslide distribution.

To save space, the accuracy assessment of only the second (active) landslide model is presented in Figure 2. While there is not a direct one-to-one correlation between predicted susceptibility and the presence or absence of active landslides, the general pattern is the desired one (that is, more active landslides occur in areas of higher predicted susceptibility). Further, since not all susceptible areas will result in landsliding, the correlation should not be exact.

## CONCLUSIONS

GLM is a powerful quantitative technique which is finding increasing application in geomorphology. It provides the ability to model many different types of response variable including the binary response relevant to both of the examples presented above. Further, the results from a well designed GLM can elucidate the effects on a specific response variable of a set of complex interacting explanatory variables.

GLM is an exploratory technique in that the models are developed interactively, and this provides the experienced user with a real 'feel' for the data. While the many textbooks available on GLM cover the theory in substantial detail, and an understanding of this theory is necessary to obtain meaningful results, GLM is an important tool which could be useful in many areas of geomorphology.

## ACKNOWLEDGEMENTS

P. A. and R. M. thank the Geodata Unit at the University of Southampton for the provision of computing facilities and general support, and Dr Peter Smith for help with data analysis. H. J. and T. M. kindly acknowledge Dr Paul Boyle for his tremendous help with our first steps into GLIM and for many fruitful discussions. The GLIM software at MIDAS in Manchester was used to run the logit models on glacier surging. H. J. and R. M. thank the University of Leeds, School of Geography, and San Marino Government, respectively for studentship funding. The authors thank two anonymous referees whose comments resulted in substantial improvements to this paper.

## REFERENCES

- Aitkin, M. Anderson, D., Francis, B. and Hinde, J. 1989. *Statistical modelling in GLIM*, Oxford University Press, Oxford.
- Atkinson, P. M. and Massari, R. (1998) 'Predicting the relative likelihood of landsliding in the central Appennines, Italy', *Computers and Geosciences*, **24**(4), 373–385.
- Baker, R. J. and Nelder, J. A. 1978. *The GLIM System: Release 3*, Numerical Algorithms Group, Oxford.
- Bernknopf, R. L., Campbell, R. H., Brookshire, D. S. and Shapiro, C. D. 1988. 'A probabilistic approach to landslide hazard mapping in Cincinnati, Ohio, with applications for economic evaluation', *Bulletin of the Association of Engineering Geologists*, **25**, 39–56.
- Budd, W. F. 1975. 'A first model for periodically self-surging glaciers', *Journal of Glaciology*, **14**(70), 3–21.
- Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V. and Reichenbach, P. 1991. 'GIS techniques and statistical models in evaluating landslide hazard', *Earth Surface Processes and Landforms*, **16**, 427–445.
- Carrara, A., Cardinali, M. and Guzzetti, F. 1992. 'Uncertainty in assessing landslide hazard and risk', *ITC Journal*, **2**, 172–183.
- Carrara, A., Cardinali, M., Guzzetti, F. and Reichenbach, P. 1996. 'GIS-based techniques for mapping landslide hazard', in Carrara, A. and Guzzetti, F. (Eds), *Geographical Information Systems in Assessing Natural Hazards*, Kluwer Publications, Dordrecht.
- Clarke, G. K. C. 1991. 'Length, width and slope influences on glacier surging', *Journal of Glaciology*, **37**(126), 236–246.
- Clarke, G. K. C., Schmok, J. P., Ommanney, C. S. L. and Collins, S. G. 1986. 'Characteristics of surge-type glaciers', *Journal of Geophysical Research*, **91**(B7), 7165–7180.
- Collett, D. 1991. *Modelling Binary Data*, Chapman and Hall, London.
- Davis, J. C. 1986. *Statistics and Data Analysis in Geology*, 2nd edition, John Wiley, Chichester, 646 pp.
- Dobson, A. J. 1990. *An Introduction to Generalised Linear Models*, Chapman and Hall, London.

- Downs, P. W. 1995. 'Estimating the probability of river channel adjustment', *Earth Surface Processes and Landforms*, **20**, 687–705.
- Fowler, A. C. 1989. 'A mathematical analysis of glacier surges', *SIAM Journal of Applied Mathematics*, **49**(1), 246–263.
- Fowler, A. C. and Johnson, C. 1995. 'Hydraulic run-away: a mechanism for thermally regulated surges of ice sheets', *Journal of Glaciology*, **41**(139), 554–561.
- Glazyrin, G. E. 1978. 'Identification of surging glaciers by morphometric characteristics', *Mater. Glyatsiologicheskikh Issled. Khronika*, **33**, 136–137.
- Gupta, R. P. and Joshi, B. C. 1989. 'Landslide hazard zoning using the GIS approach – a case study from the Ramganga catchment, Himalayas', *Engineering Geology*, **28**, 119–131.
- Hamilton, G. S. and Dowdeswell, J. A. 1996. 'Controls on glacier surging in Svalbard', *Journal of Glaciology*, **42**(140), 157–168.
- Hansen, A. 1984. 'Landslide hazard analysis', in Brunsden, D. and Prior, D. B. (Eds), *Slope Instability*, John Wiley, Chichester, 523–602.
- Healy, M. J. R. 1988. *GLIM: an Introduction*. Clarendon Press, Oxford.
- Jäger, S. and Wiczorek, G. F. 1994. *Landslide susceptibility in the Tully valley area, Finger Lakes region, New York*, USGS Open File Report **94-615**.
- Jiskoot, H., Boyle, P. and Murray, T. 1998. 'Modelling the incidence of glacier surging in Svalbard', *Computers and Geosciences*, **24**(4), 387–399.
- Kamb, B. 1987. 'Glacier surge mechanisms based on linked cavity configuration of the basal water conduit system', *Journal of Geophysical Research*, **92**(B9), 9083–9100.
- Liao, T. F. 1994. *Interpreting probability models. Logit, Probit, and other Generalised Linear Models*, Sage Publication Series, Quantitative Applications in the Social Sciences, **101**, 96 pp.
- Maharaj, R. J. 1993. 'Landslide processes and landslide susceptibility analysis for an upland watershed: a case study from St. Andrew, Jamaica, West Indies', *Engineering Geology*, **34**, 53–79.
- Neuland, H. 1976. 'A prediction model of landslips', *Catena*, **3**, 215–230.
- Ommanney, C. S. L., Clarkson, J. and Strome, M. M. 1973. *Information Booklet for the Inventory of Canadian Glaciers*, Inland Waters Directorate, Environment Canada, Ottawa, Glacier Inventory Note, **4**, 1–7.
- Post, A. 1969. 'Distribution of surging glaciers in western North America', *Journal of Glaciology*, **8**(53), 229–240.
- Siddle, H. J., Jones, D. B. and Payne, H. R. 1991. 'Development of a methodology for landslip potential mapping in the Rhondda Valley', in Chandler, R. J. (Ed.), *Slope Stability Engineering*, Thomas Telford, London, 137–142.
- Siegel, E., Dowlatabadi, H. and Small, M. J. 1995. 'A probabilistic model of ecosystem prevalence', *Journal of Biogeography*, **22**, 875–879.
- Uno, T., Sugii, T. and Hayashi, M. 1994. 'Logit model for river levee stability evaluation considering the flood return period', *Structural Safety*, **14**, 81–102.
- Van Westen, C. J. 1993. *Application of Geographic Information System to landslide hazard zonation: a review of principles and practice*, ITC, Enschede, The Netherlands, ITC Publication **15**.
- Varnes, D. J. 1978. 'Slope movement types and processes', in Shuster, R. L. and Krizel, R. J. (Eds), *Landslides; Analysis and Control*, Transportation Research Board, Washington DC, 11–33.
- Wang, S.-Q. and Unwin, D. J. 1992. 'Modelling landslide distribution on loess soils in China: an investigation', *International Journal of Geographical Information Systems*, **6**, 391–405.
- Weertman, J. 1969. 'Water lubrication mechanism of glacier surges', *Canadian Journal of Earth Science*, **6**, 929–942.
- Wiczorek, G. F. 1984. 'Preparing a detailed landslide inventory map for hazard evaluation and reduction', *Bulletin of the Association of Engineering Geology*, **21**, 337–342.
- Wilbur, S. C. 1988. *Surging versus nonsurging glaciers: a comparison using morphometry and balance*, Unpublished MSc thesis, University of Alaska, Fairbanks.