Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Engineering Geology 69 (2003) 331–343
ENGINEERING GEOLOGY
www.elsevier.com/locate/enggeo

Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA

Gregory C. Ohlmacher*, John C. Davis

Kansas Geological Survey, 1930 Constant Avenue, Lawrence, KS 66047, USA

Received 8 February 2002; accepted 16 January 2003

Abstract

Landslides in the hilly terrain along the Kansas and Missouri rivers in northeastern Kansas have caused millions of dollars in property damage during the last decade. To address this problem, a statistical method called multiple logistic regression has been used to create a landslide-hazard map for Atchison, Kansas, and surrounding areas. Data included digitized geology, slopes, and landslides, manipulated using ArcView GIS. Logistic regression relates predictor variables to the occurrence or nonoccurrence of landslides within geographic cells and uses the relationship to produce a map showing the probability of future landslides, given local slopes and geologic units. Results indicated that slope is the most important variable for estimating landslide hazard in the study area. Geologic units consisting mostly of shale, siltstone, and sandstone were most susceptible to landslides. Soil type and aspect ratio were considered but excluded from the final analysis because these variables did not significantly add to the predictive power of the logistic regression. Soil types were highly correlated with the geologic units, and no significant relationships existed between landslides and slope aspect.

Keywords: Geologic hazards; Slope stability; Mass movement; Probability; Statistics; Hazard map

1. Introduction

Two general approaches exist for landslide prediction. The first is a deterministic or engineering approach based on mathematical models of the physical mechanisms that control slope failure (Kramer, 1996). A review of such models for landslides triggered by seismic events was given by Miles and Keefer (2000). A significant limitation of deterministic models is the need for material data (mechanical properties, water saturation, etc.) that are difficult to obtain over large areas (Terlien et al., 1995). The second general approach is statistical and thus does not posit mechanisms that control slope failure, but assumes rather that occurrences of past landslides can be related arbitrarily to measurable characteristics of the landscape. In turn, these characteristics can be used to predict future landslide occurrence. Because the requisite material data for the study area are not easily obtained, a statistical approach was chosen to develop a landslide-hazard map of Atchison, KS.

In 1995, a landslide in Overland Park, Kansas, a suburb of Kansas City, destroyed two houses valued at

* Corresponding author. Tel.: +1-785-8642194; fax: +1-785-8645317.
E-mail address: ohlmac@kgs.ku.edu (G.C. Ohlmacher).

0013-7952/03/$ - see front matter
doi:10.1016/S0013-7952(03)00069-3

more than $400,000 each (Fig. 1). This destruction highlights the need for development of landslide-hazard maps for the Kansas City Metropolitan Area. The Kansas Geological Survey is currently investigating landslides in the Kansas portion of the metropolitan area. The methodology is being developed and tested in Atchison, a city on the fringe of the metropolitan area (Fig. 2). The study area consists of 202.3 sq. km of mostly rural agricultural land and includes the city of Atchison, which has a population of about 10,000 people (Helyar, 2000). Currently, Atchison is not growing (Helyar, 2000); however, picturesque local vistas of the Missouri River valley may eventually attract more people.

The study area has the necessary conditions for landslides, including steep slopes, adequate precipitation, and weak soil and rock units. Although Kansas is in the flat to gently rolling central United States (Fig. 2), areas of hilly topography with steep slopes exist throughout the state. The topography in the Atchison area is dominated by the incision of the Missouri River and its tributaries into Pennsylvanian bedrock, consisting of sandstone, siltstone, shale, and limestone beds. Quaternary sediments overlie the bedrock and consist of glacial drift, loess, and alluvium. Weathering of the bedrock and overlying Quaternary sediments produces clay-rich soils that are prone to landslides.

The study area has both developed and undeveloped areas, providing a variety of natural and anthropogenic triggers for landslides. The natural triggers include severe storms, wet climatic cycles, and stream erosion of slope bases. Anthropogenic triggers include slopes that have been cut too steeply, improper construction or placement of fill, and poorly controlled surface drainage.

To express the potential for occurrence of landslides in a quantitative manner, maps must incorporate the concept of probability, which is an assessment of the relative frequency of occurrence. The term “risk” generally is used when mapping the probabilities of occurrence of landslides at specific locations or within small areas during some interval of time, perhaps 10 or 25 years, and includes an estimate of the vulnerability of the population and structures (Varnes, 1984). Estimation of probabilities necessary for mapping frequency of occurrence requires information about the dates of occurrence of past landslides, information that is not often available. Instead, probabilities may be estimated by enumerating past landslides without regard to when they occurred. The resulting map referred to as a “hazard” map shows probabilities of landslide occurrence without implying rate of occurrence. A map showing ordinal categories of probability (low, moderate, high) rather than explicit

Fig. 1. Photograph of a house damaged by a 1995 landslide in Overland Park, Kansas, a suburb of Kansas City. (Photo Credit: John Charlton, Kansas Geological Survey.)

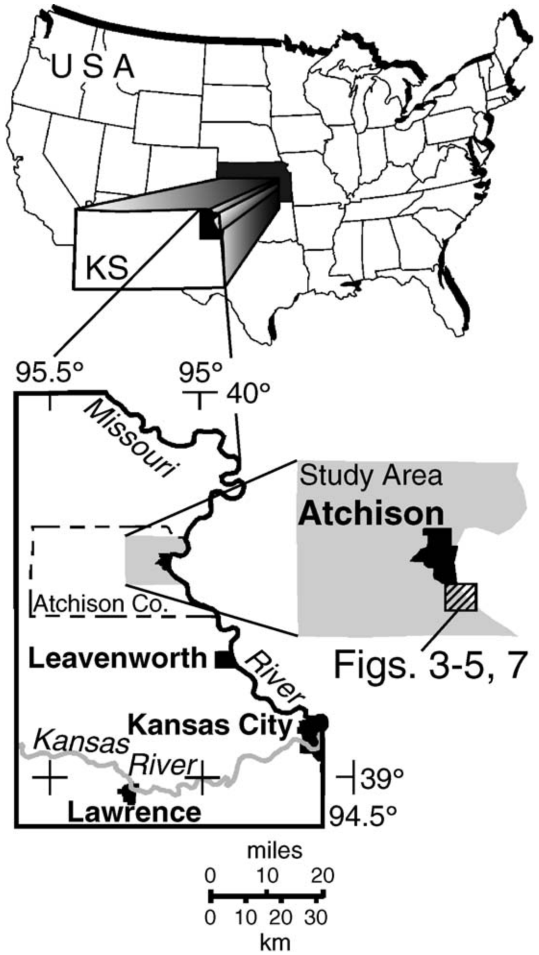

Fig. 2. Map of northeast Kansas showing the location of the study area relative to Kansas City, KS.

probabilities of occurrence is called a "susceptibility" map and can be based on subjective probability assessments (Hansen, 1984).

A landslide-hazard map in preparation for the Atchison study area will show the probability of future landslides, given the slope and geology of the site. This map will allow users, including planners, engineers, developers, and landowners, to assess levels of hazard in order to take appropriate action. This paper covers selection of input data, manipulation of data using GIS (geographic information system) technology, and the use of logistic regression to generate a landslide-hazard map.

2. Landslides and local geology of the Atchison study area

Atchison is located about 65 km northwest of Kansas City along the Missouri River (Fig. 2). The area west of the Missouri River is entirely within Kansas and consists of gently rolling hills to the west and more rugged hills to the east. Total relief in the study area is about 112 m; the greatest local relief, up to 73 m, occurs along the bluffs of the Missouri River and its tributaries where Pennsylvanian bedrock is exposed. Glacial deposits in the western portion of the study area have relatively gentle slopes. Slope angles in the study area range from nearly horizontal hilltops and floodplains to vertical cliffs.

2.1. Landslides

"Landslide" is a general term defined as a downslope movement of a mass of soil and rock material (Cruden, 1991). This paper uses the landslide classification developed for the Transportation Research Board (Varnes, 1978; Cruden and Varnes, 1996). Landslide types in the study area include earth flows, earth slides, rock falls, and rock topples.

An inventory map of landslides in the study area (Ohlmacher, 2000a) subdivides the landslide features into *recent* and *older landslides* (Fig. 3). Landslides described as "recent" have distinct features, clearly defined boundaries, and have moved in the past several years. They include active, suspended, and dormant earth flows and earth slides as defined by Cruden and Varnes (1996). Older landslides have hummocky topography, muted features, and indistinct boundaries. This category includes dormant, relic, and ancient earth flows and earth slides.

Data on recent and older landslides have been used to develop the landslide-hazard analysis discussed in this paper. These landslides are predominantly shallow failures with basal failure planes in the soil or weathered bedrock. Although deeper earth and rock slides also occur, such deep landslides overlap areas with shallow landslides.

2.2. Local geology

The geology of the study area consists of glacial drift, loess, and Quaternary alluvium overlying Penn-

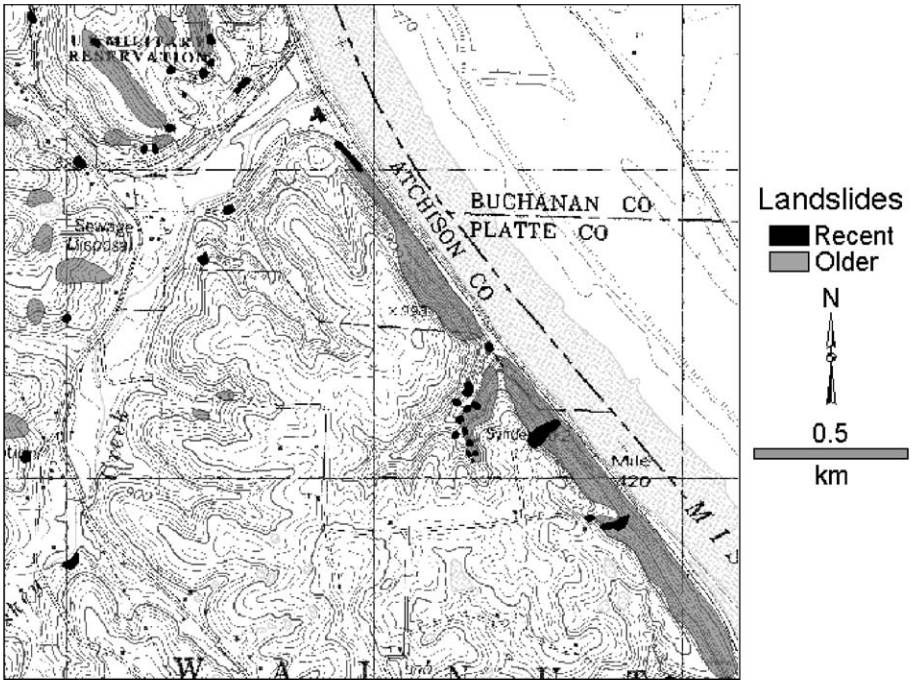

Fig. 3. Landslide inventory map. Recent landslides are earth flows and earth slides with distinct features and well-defined boundaries. Older landslides are earth flows and earth slides with muted features and indistinct boundaries. Location of area is shown on Fig. 2.

sylvanian bedrock (Ward, 1973; Ohlmacher, 2000b). Pennsylvanian rock units consist of marginal marine sandstone, siltstone, shale, and limestone beds deposited as repeated sequences or cyclothems that reflect changes in sea level during Pennsylvanian time (Moore, 1935; Merriam, 1963). The lithologic sequence in a normal cyclothem begins with terrestrial sandstones and siltstones and continues through a transgressive sequence of shallow marine shales and limestones into a regressive sequence.

The oldest unit in the study area is the Lawrence Formation, consisting of at least 8 m of green-gray shale, red silty shale, and coal. The Lawrence Formation occurs along the base of hills south of Atchison and is considered one of the most landslide-prone units in northeastern Kansas (F. Wilson, pers. comm., 1997). Numerous recent and older landslides have occurred in the Lawrence Formation.

The Oread Limestone overlies the Lawrence Formation and consists of about 16 m of limestone and gray to black shales. Four limestone members are recognized in the study area with the largest being 6 m thick. The Oread Limestone supports the steepest slopes in the study area. Above the 2-m-thick basal limestone member of the Oread Limestone is 4 m of gray marine shale. This marine shale weathers into clay-rich soil; small earth flows and earth slides have been observed in this member in highway cuts.

The Kanwaka Shale overlies the Oread Limestone and consists of 18 m of gray shale, siltstone, and sandstone with a 1-m-thick limestone bed near the middle of the formation. Landslides are observed throughout the Kanwaka Shale but especially above the 1-m limestone bed where the base of the upper shale member consists of a 20-cm-thick clay bed (Ohlmacher, 2000b, Fig. 14). The Lecompton Limestone overlies the Kanwaka Shale and consists of about 12 m of limestone and gray shale but is not well exposed in the study area. Blocks of Lecompton Limestone are incorporated into landslides in the underlying Kanwaka Shale.

The Tecumseh Shale overlies the Lecompton Limestone and consists of about 20 m of gray shale, siltstone, and sandstone. The unit occurs predominantly on gentle slopes. Landslides are observed in the Tecumseh Shale where it is exposed in highway cuts and on steep slopes. The Deer Creek Limestone overlies the Tecumseh Shale and consists of 11 m of limestone, and gray to black shales. Blocks of Deer Creek Limestone are incorporated into landslides in the underlying Tecumseh Shale.

The Calhoun Shale overlies the Deer Creek Limestone and consists of 5 m of gray siltstone and shale. The uppermost Pennsylvanian formation exposed in the area is the Topeka Limestone. Only the lowest limestone member of the Topeka Limestone is exposed. The Calhoun Shale and Topeka Limestone are exposed in the western portion of the study area where slopes are gentle and landslides are rare. No landslides have been mapped in either the Calhoun Shale or the Topeka Limestone, although these units may be involved in landslides in the hilly terrain along the Missouri River north of the study area.

The geologic units were subdivided into five categories. Bedrock units were separated into two categories based on their proportions of clastic material (clay, silt, and sand). The "shale category" includes the Lawrence Formation, Kanwaka Shale, Tecumseh Shale, and Calhoun Shale. The "limestone category" includes the Oread Limestone, Lecompton Limestone, Deer Creek Limestone, and the Topeka Limestone. The remaining three geologic units are glacial drift, loess, and alluvium. Glacial drift includes till, lacustrine deposits, and alluvium; material properties are highly variable. The alluvium within the glacial drift consists of isolated outcrops that were not mapped separately and thus cannot be added to the alluvium unit. Loess consists of gray to brown sandy silt that ranges in thickness up to 12 m. Alluvium is tan, brown, or gray silt, sand, and gravel and ranges in thickness up to 32 m. Thus, the geologic units used in this study are

1. Alluvium
2. Loess
3. Drift (glacial drift)
4. Limestone
5. Shale

Bedrock in the study area contains a high percentage of fine-grained clastic material and expansive clay minerals (Ohlmacher, 2000b). Soils developed on bedrock, glacial drift, and loess contain abundant clay. Agricultural soil maps of Atchison County (Sallee and Watts, 1984) show that the soil units closely match geologic units.

No folds or faults have been mapped in the study area (Ohlmacher, 2000b). In general, the Pennsylvanian bedrock dips at about 3 m/km ($\sim 0.2°$) to the west–northwest with some minor local variations.

3. Using GIS to produce landslide-hazard maps

The geographic information system ArcView 3.2 was used to compile and manipulate data for this study and to produce the final hazard maps. In addition to the presence or absence of landslides within 10-m cells, the data included *slope*, *slope aspect*, *geology*, and *soils*. A significant relationship exists between soil category and the occurrence of landslides. However, soil categories are highly correlated with geologic units, and soils do not contribute to the estimation of the probability of landslide occurrence when geology is also included. Because soils are redundant, they were not included in the analysis.

Although in some circumstances the direction that a slope faces may have an influence on the stability of the slope, in this study area no statistically significant relationship between slope aspect and the occurrence of landslides was found. The variable slope aspect was therefore excluded from the study.

Slope data were developed from digital elevation models (DEMs) produced by the U.S. Geological Survey. These DEMs have a 30-m spacing between values, and the quality of the data is good. Slopes were calculated from the DEMs using the slope routine in ArcView GIS, which calculates the maximum slope between each cell and the eight neighboring cells (Ormsby and Alvi, 1999). This results in a slope map with a 30-m spacing between values (Fig. 4). The probability analysis was performed using a 10-m grid spacing, so each 30-m cell was divided into nine equal cells on a 10-m grid, and the slope value from the 30-m cell was applied to each included 10-m cell.

A revised geologic map with additional field data (Ohlmacher, 1999), including data from an earlier geologic map (Ward, 1973), was used for this study. The revised geologic map was constructed as digital polygons for input into ArcView GIS (Fig. 5), which converted the polygons to a raster format with a 10-m grid spacing. Each grid cell was assigned a code

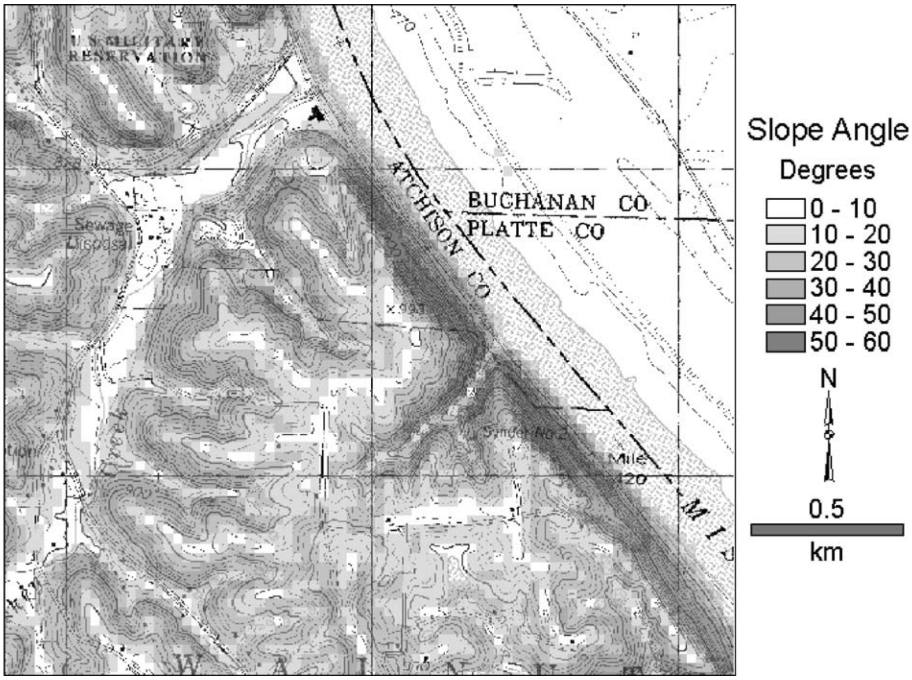

Fig. 4. Slope map showing 10° slope categories in a 30-m grid. Location of area is shown on Fig. 2.

representing the geologic unit at the center of the cell. The initial rasterized geologic map depicted all the formations of the study area, which were then assigned to the five geologic units discussed above.

The landslide-inventory map (Ohlmacher, 2000a), prepared originally as a digital vector file, was also converted to a 10-m grid file by ArcView GIS. Each cell was assigned "0" if no landslide was present or "1" if a landslide was present. A "no data" code was assigned if the cell was outside the study area. The geologic, slope, and landslide grid files were logically compared to ensure that the three grids covered a common area, then combined and converted to an ASCII-format file that included the UTM coordinates for the center of each cell. The ASCII file was then input into JMP v. 4 statistical software for probability analysis.

4. Probability methods and the multiple logistic regression approach

Most statistical studies of landslide prediction have used contingency-table analysis, multiple regression, discriminant analysis, or logistic regression. These procedures all relate one or more predictor variables (slope, geology, etc.) to some measure of landslide occurrence. The simplest statistical-prediction method uses contingency-table analysis, in which a cross-tabulation is made between two outcome states (landslide and nonlandslide) versus the discrete categories of a predictor variable. The proportions of tallies in the cells of the table can be interpreted as estimates of conditional probabilities of occurrence (or nonoccurrence) of a landslide, given a state of the predictor variable.

To use multiple regression, landslide occurrence must be expressed as a continuous variable. Carrara (1983) regressed the percentage of area affected by landsliding onto geomorphic variables. These measures relate to geographic areas (Chung et al., 1995), so the spatial resolution can be no finer than these areas. Because regression estimates are unconstrained, they may be negative or greater than 100% and hence cannot be directly interpreted as probabilities.

Using discriminant analysis, observations are classified into two mutually exclusive groups: they are a location or area where a landslide (1) has occurred, or

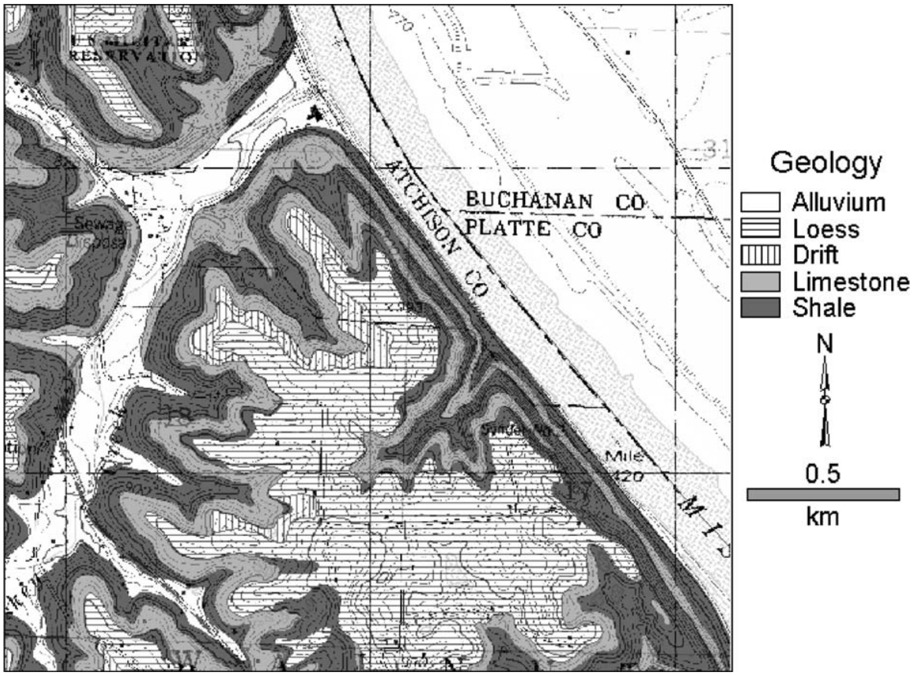

Fig. 5. Geologic map showing five geologic units. Location of area is shown on Fig. 2. *Shale* includes all Pennsylvanian bedrock formations that have a high percentage of clastic material. *Limestone* includes all Pennsylvanian bedrock units with a lower percentage of clastic material and a higher percentage of carbonates. *Drift* includes Quaternary glacial deposits consisting of till, alluvium, and lacustrine deposits. *Loess* contains Quaternary eolian deposits, and *Alluvium* consists of Holocene stream deposits.

(2) has not occurred. Each observation consists of a vector of independent variables. The differences between the landslide group and the nonlandslide group are expressed by coefficients that project every observation onto a line connecting the centroids of the two groups; an observation is classed as a landslide if its score is closer to the landslide centroid than to the nonlandslide centroid. Unfortunately, no way exists to express discriminant scores as probabilities of occurrence of landslides except by categorizing the scores and computing proportions of misclassifications (Carrara et al., 1991). In effect, discriminant analysis simply converts a multivariate problem into a univariate contingency-table analysis.

In logistic regression, the status of landslide within an area such as a 10-m cell is represented by a variable Y. An arbitrary code value such as "1" can be used to indicate the presence of a landslide and "0" can be assigned if no landslide is present, but these values have no intrinsic meaning. Rather than trying to predict an arbitrary code, it would seem more useful to estimate the probability that an observation will be classified into one category as opposed to the other. Because presence and absence categories are mutually exclusive and all-inclusive, if one probability is known, the other is also known; that is, $P(Y=1)=[1-P(Y=0)]$ where $P(Y=1)$ is the probability that the statement in parentheses is true.

Following the development by Menard (1995), the probability that $Y=1$ could be modeled by regression, using an equation such as $P(Y=1)=\alpha+\beta X$. However, a problem exists in that although the observed probabilities $P(Y=1)$ must lie between 0 and 1, nothing constrains the predicted probabilities, $\hat{Y}$, and they may be impossible values less than 0 or greater than 1. A partial way out of this problem is to replace the probabilities with *odds*. Odds are defined as the ratio of the probability that something occurs to the probability that it does not occur, or $Odds(Y=1)=P(Y=1)/[1-P(Y=1)]$. The odds are a ratio that has no fixed maximum, but does have a minimum value of 0. If the odds are modeled by regression, the possibility of estimating impossibly large values is removed, but the problem that the predicted odds may be an impossible value less than 0 still exists.

If, however, the odds are transformed by taking their natural logarithm, a variable is created that can vary from negative infinity to positive infinity. This value is the *logit*, defined as $(Y) = \ln\{P(Y=1)/[1-P(Y=1)]\}$, which goes toward negative infinity as the odds decrease from 0 to 1, and goes toward positive infinity as the odds increase beyond 1. Since logits may assume any value, the problem that values estimated by regression may exceed the maximum or minimum limits of probability is avoided.

It must be emphasized that the *probability*, the *odds*, and the *logit* are three ways of expressing exactly the same thing, and one can easily be converted into the other. However, by using logits, there are no constraints that otherwise would make it impossible to use regression in a predictive model. The coefficients that result from logit regression represent the change in log odds due to incremental-unit changes in the values of the predictors (DeMaris, 1992).

In logistic regression, the regression parameters must be estimated using maximum likelihood. To do this, it is necessary to know the probability distribution for the observed data, and in particular, the conditional distribution of the dependent variable given each combination of predictor variables, expressed in terms of the parameters. If the incidence of a landslide at a specific location i is 1, and the absence of a landslide at a location i is 0, then Y_i follows a binomial distribution with mean π_i and variance $\pi_i(1-\pi_i)$. To specify this distribution as a function of the α and β coefficients of regression, it also must be assumed that the relationship between π_i and the predictor variables follows a logistic distribution function. (This is exactly equivalent to the assumption in ordinary regression that the mean of Y is related to the predictors through a linear function.) In practice, the logistic distribution function usually is a good model for π_i, the conditional probabilities of occurrence.

Logistic regression proceeds by finding estimates for α and β that maximize the resulting conditional distribution, or *likelihood function,* for the set of observations $y_1, y_2, \ldots y_n$. This results in coefficients that make these observations the most likely values to have been observed. The coefficients cannot be estimated directly as in ordinary regression; instead, a tentative solution is initially chosen and revised slightly to see if the likelihood increases. The process is repeated until the increase in the likelihood function from one step to the next is negligible.

Maximizing the logarithm of the likelihood is equivalent to maximizing the likelihood itself and has certain advantages. If the log-likelihood is multiplied by -2, the resulting statistic, called the *negative log-likelihood*, has approximately a χ^2 (chi-square) distribution. The negative log-likelihood of a logistic regression model that contains only a constant term can be compared to the negative log-likelihood of a model that contains parameter estimates for all the predictive variables; the difference between the two models can be used in a χ^2 test of significance of the regression coefficients. If the difference is small, the conditional relationship between the predictive variables and the occurrence of landslides is not significant. That is, considering the magnitudes of the predictive variables does not improve the estimate of the probability of landslide occurrence. Variations of this procedure can be used to test individual predictive variables and to select those that are most effective. The tests are presented in a format that closely resembles analysis of variance tables used to test coefficients in ordinary regression, except that the test statistic follows a χ^2 distribution rather than an F distribution.

General logistic regression has been used to map landslide susceptibility using categorized predictor variables (Jäger and Wieczorek, 1994), continuous variables, and mixtures of the two (Bernknopf et al., 1988; Gorsevski et al., 2000). In the present study, a landslide-hazard map has been created using multiple logistic regression, a logistic regression based on two predictors, slope (a continuous variable) and geologic unit (a categorical variable).

For two possible response levels for a cell (the presence or absence of a landslide), the maximum likelihood regression model is:

$$P(Y_i = 1) = [1 + \exp(\boldsymbol{X}_i\boldsymbol{b})]^{-1}$$

which can be rewritten as

$$\boldsymbol{X}_i\boldsymbol{b} = \log\left[\frac{P(Y_i = 1)}{P(Y_i = 0)}\right]$$

where Y_i is the state of cell i, $\boldsymbol{X}_i$ is a vector of the predictor variables for cell i, and $\boldsymbol{b}$ is a vector of coefficients to be estimated. The term on the right side of the equation is the logit transformation, that is, the logarithm of the odds.

The predicted values $\hat{Y}$, which are the probabilities of landslide occurrence, will lie between 0 and 1 over the ranges of the $\boldsymbol{X}$'s (Hosmer and Lemeshow, 2000). Because the important variables for predicting landslides in this area appear to be topographic slope and geology, the appropriate multiple logistic regression equation can be written

$$P(\text{landslide}) = \frac{1}{1+\exp[-(\beta_0+\beta_1 slope+\Sigma\beta_k geology_k)]}$$

where $geology_k$ is an indicator variable equal to 1 if the geologic unit k is present in a cell and 0 otherwise. Note that the classes are mutually exclusive; that is, only one geologic unit can be present in a cell. This means that the general form of the regression is determined by the constant plus the $\beta_1 slope$ term, and for each geologic unit, the regression curve is offset by an amount equal to the β_k coefficient that corresponds to the geologic unit present in the cell. Because all the indicator variables are 0 in a cell where the geologic units are not present, the multiplication $\beta_k \times 0 = 0$. For the one geologic unit that is present in the cell, the indicator value is 1, so $\beta_k \times 1 = \beta_k$.

The Atchison study area contains 2,022,861 cells, each with a variable indicating the presence or absence of a landslide, a value for the slope in the cell, and a code for the geologic unit. The significance of the fitted logistic regression was tested by comparing the negative log-likelihood function for the complete regression to that for an alternative model consisting of a constant value only. The result is highly significant, as shown in Table 1.

Table 1
Maximum negative log likelihood values for testing the significance of multiple logistic regression of landslide probability on slope and geology

Whole model test				
Model	Log likelihood	*df*	Chi square	Significance
Difference	29,450.985	5	58,901.97	>0.9999
Full model	37,917.157			
Intercept only	67,368.142			

Table 2
Coefficients of the multiple logistic regression of landslide probability on slope and five geologic units

Variable	Coefficient	Standard error	Chi square value	Significance
Intercept	8.8117	0.0332	70,470	>0.9999
Slope	− 0.1472	0.0009	24,561	>0.9999
Geology			2257.04	>0.9999
Alluvium	0.0815	0.0632	1.66	0.8030
Loess	0.5043	0.0476	112.43	0.9998
Drift	1.0818	0.0428	639.16	0.9998
Limestone	− 0.6261	0.0278	507.01	0.9998
Shale	− 1.0415*	**	996.78*	0.9998

* Found by difference.
** Cannot be estimated.

The parameters of the fitted model are given in Table 2, with χ^2 values for testing the significance of the individual terms. The constant term and slope coefficient are both highly significant, with an infinitesimal probability that their values could be zero. The χ^2 test of geology as a predictor of landslides is also highly significant, although the coefficient for the category *Alluvium* is not significantly different than zero. The coefficients for the geologic units are constrained to sum to zero, so the final category (*Shale*) must be equal to the negative of the sum of the other coefficients and is found by difference. Because the final coefficient is predetermined, its standard error is not defined. Similarly, its χ^2 value is found as the difference between the χ^2 value for geology and the sum of the χ^2 values for the individual categories other than *Shale*.

It must be kept in mind that serious violations of the regression-model assumptions exist that render the exact significance levels suspect (Agresti, 1990). The most serious of these violations is that the observations are independent. Because the surface of the earth is physically continuous everywhere, the value in a geographic cell must be strongly related to values in nearby cells. This logistic analysis should more properly be done from a geostatistical perspective, but the tools for such an analysis have yet to be developed.

With this caveat, the fitted logit regression model can be evaluated for different slopes and geologic units to yield a family of curves that predict the probability of occurrence of landslides in the study

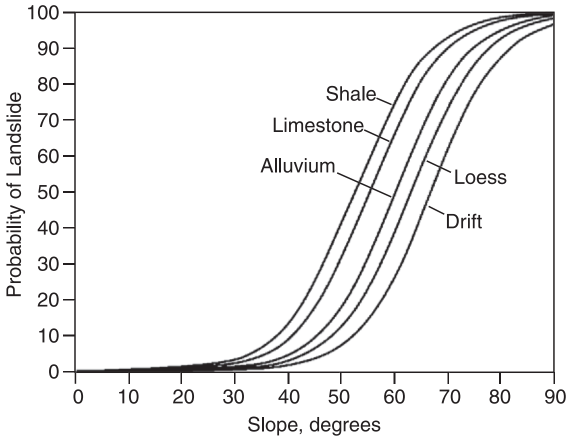

Fig. 6. Probability of landslide estimated by multiple logistic regression on slope and geologic unit. Curves in shaded region are extrapolations beyond the range of observed slopes in the Atchison study area.

area. The curves are identical in form, with a shape that depends upon the constant and slope terms, but offset from one another by amounts given by the geologic-unit terms (Fig. 6). The uncertainty in regression conventionally is shown by confidence bands around a fitted regression line; the interval between bands is interpreted as containing the true regression a specified percentage (usually 95%) of the time. Confidence bands are a function of the standard errors of the estimated regression coefficients and a specified probability level. Confidence bands are not shown around the logistic regression curves of Fig. 6 because they are too small to appear; the maximum uncertainty is for *Alluvium* at a 60° slope where the 95% confidence interval extends only from 45.4% to 51.7% probability.

5. Results of study

The geology and slope data for each grid cell were input into the logistic equation to generate a grid from which the landslide-hazard map was constructed using ArcView GIS (Fig. 7). The final landslide-hazard grid contains the probability that each cell contains a landslide, given the slope and geology in that cell.

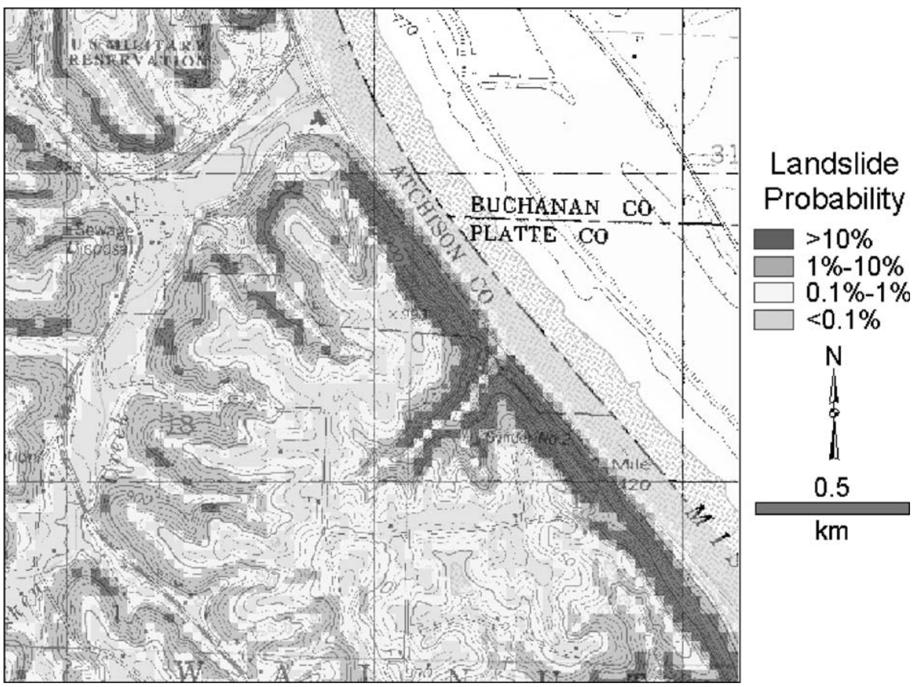

Fig. 7. Map showing landslide hazard. The map depicts the probability of a landslide given the slope and geology.

A visual comparison between the landslide-hazard map (Fig. 7) and the landslide-inventory map (Fig. 3) indicates good agreement between areas with mapped landslides and areas with high probability of landslides. Additionally, areas of high probability of a landslide exist outside areas of known landslides. Thus, the logistic regression is able to delineate areas of potential landslide hazard.

For display purposes the range of probabilities on the landslide-hazard map was subdivided into four intervals using an exponential scale. The highest interval includes areas with a greater than 10% probability of the occurrence of landslides. These areas include the steep bluffs and hilly topography along the Missouri River. The lowest interval indicates areas where the probability of landslide occurrence is less than 0.1%. In Fig. 7, these areas are relatively small zones on ridge tops and along valley bottoms. However, Fig. 7 shows a hilly subarea that is not typical of the entire study area; the extent of the lowest interval is greater in the western part of the study area. Examination of the data shows that all the intervals include landslides, but relatively few exist in the lowest interval. A possible concern is that the lowest interval includes small landslide-prone areas—for example land adjacent to cut banks along streams, and the 30-m DEM may be too coarse to define such small areas of steep slope. The four intervals displayed on this map may not be suitable for all purposes, and other users may define alternative levels of landslide hazard appropriate for their specific use.

The grid spacing of the DEM and the greatest local relief of the study area control the maximum slope angle that can be resolved. A measure of the maximum resolvable slope angle can be found by taking the arc tangent of the local relief divided by the grid spacing. The greatest local relief of this study area is 73 m, and the DEM has a 30-m spacing. Thus, the maximum resolvable slope is 68°. An examination of the slopes generated from the DEM using the slope routine in ArcView GIS indicates that the maximum measured slope is 66°. Thus, the ArcView GIS slope routine was able to determine slope angles near the maximum value. The slope values for the 10-m grid used in this study are the 30-m values applied to 10-m cells. A true 10-m DEM would have a maximum resolvable slope of 82°. However, since the hillsides would be divided into smaller cells, the maximum measured slope angle in ArcView GIS would be greater than 66° but not as high as 82°. This study does not address the issue of whether or not the effort to obtain a 10-m DEM would be warranted.

This application of logistic regression methodology tacitly assumes that the probability of occurrence of a landslide increases as the slope increases. This assumption is true for low to moderate slopes. However, as slope angles approach the vertical, it might be argued that because the observations consist of shallow landslides and soils are thinner on steep slopes, the probability of landslide occurrence might begin to decrease as slopes exceed some undetermined angle. The soils on very steep slopes may be nothing more than weathered bedrock, but such material can still fail. Shallow landslides were observed on the faces of steep highway cuts involving thin ($\sim$ 1 m) layers of weathered shale. The potential for rock falls and topples increases as the slope angle approaches vertical. Areas with shallow landslides may also have unrecognized deep landslides. These observations support the assumption of the methodology.

A simple sensitivity analysis of the probability equation using the coefficient values in Table 2 confirms that slope is the most important factor in determining the probability of landslide occurrence. If the geology is held constant and the degree of slope varied from 0° to 90°, the estimated probabilities range from 0% to almost 100% for all geologic units. If the slope is held constant and the geology is varied, the maximum difference in estimated probabilities is 49% at a slope value of 60°. Thus, the range in probabilities if slope is varied is twice the range in probabilities if geology is varied.

Table 2 indicates that bedrock units (*Shale* and *Limestone*) are more susceptible to landslides and that *Shale* has the highest susceptibility (largest negative coefficient). Negative coefficients shift the curves in Fig. 6 to the left indicating a higher probability of landslide occurrence at the same slope angle. These results support the observation that rock units such as the Lawrence Formation are more prone to landslides. It may seem surprising that limestone is more susceptible to landslide activity than unconsolidated units. However, the limestones occur in areas of steep slope and the limestone formations contain interbedded shales, which can constitute up to 50% of the total thickness of the formation.

Two possible circumstances may account for the fact that *Drift* is the geologic unit least susceptible to landslides. Glacial drift has highly variable physical properties because it is an inhomogeneous mix of materials and grain sizes. Additionally, glacial drift crops out in the western part of the study area where slopes are relatively low. *Loess* is more susceptible to landslide activity than glacial drift, but the 95% confidence intervals for *Loess* and *Drift* overlap, indicating that the susceptibility of these units to landslides may be statistically the same.

The coefficient for *Alluvium* in Table 2 is not significantly different from zero. This is interpreted to mean that alluvial units have an average susceptibility that is not significantly different than the slope value alone. However, the *Alluvium* coefficient has the largest standard error, reflecting the fact that in relation to the grid slope data, alluvium occurs in areas where the slopes are lowest. The stability of alluvium along streams is a concern.

Ground water and rainfall are two additional factors that were not considered in the analysis presented in this paper. Haneberg and Gokce (1994) stated that four factors control the rapid water-level fluctuations on clayey colluvial slopes: rainfall intensity, rainfall duration, hydraulic conductivity, and antecedent moisture conditions. Unlike debris flows where the probability of failure increases at some threshold rainfall amount (Keefer et al., 1987; Wilson and Wieczorek, 1995), the amount of antecedent moisture has a profound effect on the rate that pore pressures rise in earth flows (Haneberg and Gokce, 1994). Thus, the same storm that might cause a landslide when soils are wet may not initiate movement when soils are dry. Because slopes in the Atchison study area are clayey colluvial slopes, earth-flow mechanisms are more appropriate than debris-flow mechanisms. Unfortunately, the combinations of rainfall intensity, rainfall duration, and antecedent moisture conditions are impossible to map and would lead to an infinite number of multiple logistic regression models.

In conclusion, multiple logistic regression is a useful statistical method for the development of landslide-hazard maps. In the Atchison, KS, study area, slope and geology are the best predictor variables for estimating the probability of landslide occurrence. Soil type is closely related to geology and could be substituted for that predictor variable in areas where geologic maps lack sufficient resolution. Slope aspect was not a significant predictor of landslide activity in the study area, although it may be important in other locales.

Acknowledgements

The authors acknowledge the insightful reviews by Jean Michel Le Cleac'h, Jonathan W. Godt, and Martin Culshaw that improved the quality of this paper.

References

Agresti, A., 1990. Categorical Data Analysis. Wiley, New York. 558 pp.

Bernknopf, R.L., Campbell, R.H., Brookshire, D.S., 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.

Carrara, A., 1983. Multivariate models for landslide hazard evaluation. Mathematical Geology 15, 403–426.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., Reichenbach, P., 1991. GIS techniques and statistical models in evaluating landslide hazard. Earth Surface Processes and Landforms 16, 427–445.

Chung, C.F., Fabbri, A.G., Van Westen, C.J., 1995. Multivariate regression analysis for landslide hazard zonation. In: Carrara, A., Guzzetti, F. (Eds.), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 107–133.

Cruden, D.M., 1991. A simple definition of a landslide. Bulletin of the International Association of Engineering Geology 43, 27–29.

Cruden, D.M., Varnes, D.J., 1996. Landslide types and processes. In: Turner, A.K., Schuster, R.L. (Eds.), Landslides Investigation and Mitigation. National Research Council, Transportation Research Board, Washington, DC, pp. 36–75.

DeMaris, A., 1992. Logit Modeling: Practical Applications. Series on Quantitative Applications in the Social Sciences, vol. 86. Sage Publ., Thousand Oaks, CA. 87 pp.

Gorsevski, P.V., Gessler, P., Foltz, R.B., 2000. Spatial prediction of landslide hazard using logistic regression and GIS. 4th Int. Conf. on Integrating GIS and Environmental Modeling, Proceedings, Banff, Alberta, Canada. 9 pp.

Haneberg, W.C., Gokce, A.O., 1994. Rapid Water-Level Fluctuations in a Thin Colluvium Landslide West of Cincinnati. U.S. Geological Survey, Professional Paper 2059-C (16 pp.).

Hansen, A., 1984. Landslide Hazard Analysis. In: Burnsden, D., Prior, D.B. (Eds.), Slope Instability. Wiley, Chichester, UK, pp. 523–602.

Helyar, T., 2000. Kansas Statistical Abstract, 35th ed. Policy Research Institute, Univ. Kansas, Lawrence, KS.

Hosmer, D.W., Lemeshow, S., 2000. Applied Logistic Regression, 2nd ed. Wiley, New York. 373 pp.

Jäger, S., Wieczorek, G.F., 1994. Landslide Susceptibility in the Tully Valley Area, Finger Lakes Region, New York. U.S. Geological Survey, Open-File Report 94-615 (13 pp.).

Keefer, D.K., Wilson, R.C., Mark, R.K., Brabb, E.E., Brown III, W.M., Ellen, S.D., Harp, E.L., Wieczorek, G.F., Alger, C.S., Zatkin, R.S., 1987. Real time landslide warning during heavy rainfall. Science 238, 921–925.

Kramer, S.L., 1996. Geotechnical Earthquake Engineering. Prentice-Hall, Upper Saddle River, NJ. 653 pp.

Menard, S., 1995. Applied logistic regression analysis. Series on Quantitative Applications in the Social Sciences, vol. 106. Sage Publ., Thousand Oaks, CA. 98 pp.

Merriam, D.F., 1963. The Geologic History of Kansas. Kansas Geological Survey, Bulletin 162 (317 pp.).

Miles, S.B., Keefer, D.K., 2000. Evaluation of seismic slope-performance models using a regional case study. Environmental and Engineering Geoscience 6, 25–39.

Moore, R.C., 1935. Stratigraphic Classification of the Pennsylvanian Rocks of Kansas. Kansas Geological Survey, Bulletin 22 (256 pp.).

Ohlmacher, G.C., 1999. Geologic Map of the Atchison West and East 7.5 Minute Quadrangles, Atchison County, Kansas. Kansas Geological Survey, Open-file Map 99-38.

Ohlmacher, G.C., 2000a. Landslide Inventory Map of the Atchison West and East 7.5-Minute Quadrangles, Atchison County, Kansas. Kansas Geological Survey, Map M-89A.

Ohlmacher, G.C., 2000b. The relationship between geology and landslide hazards of Atchison, Kansas, and vicinity. Kansas Geological Survey, Current Research in Earth Sciences, Bulletin 244 (pt. 3) (9 pp., online only at http://www.kgs.ku.edu/Current/2000/ohlmacher/ohlmacher1.html/).

Ormsby, T., Alvi, J., 1999. Extending ArcView GIS. ESRI Press, Redlands, CA. 527 pp.

Sallee, K.H., Watts, C.E., 1984. Soil Survey of Atchison County, Kansas. U.S. Dept. Agriculture, Soil Conservation Service, Washington, DC. 105 pp.

Terlien, M.T.J., Van Westen, C.J., Van Asch, T.W.J., 1995. Deterministic modelling in GIS-based landslide hazard assessment. In: Carrara, A., Guzzetti, F. (Eds.), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 57–77.

Varnes, D.J., 1978. Slope movement types and processes. In: Schuster, R.L., Krizek, R.J. (Eds.), Landslides Analysis and Control. National Research Council, Transportation Research Board, Washington, DC, pp. 11–33.

Varnes, D.J., 1984. Landslide Hazard Zonation: A Review of Principles and Practice. UNESCO Press, Paris. 63 pp.

Ward, J.R., 1973. Geohydrology of Atchison County, Northeastern Kansas. U. S. Geological Survey, Hydrologic Investigations Atlas HA-467. 2 maps.

Wilson, R.C., Wieczorek, G.F., 1995. Rainfall thresholds for the initiation of debris flows at La Honda, California. Environmental and Engineering Geoscience 1, 11–27.