STATISTICAL MODELING OF LANDSLIDE HAZARD USING GIS

By Peter V. Gorsevski, Department of Forest Resources College of Natural Resources – University of Idaho; Randy B. Foltz, U.S. Forest Service, Rocky Mountain Research Station, Moscow, Idaho; Paul E. Gessler, Department of Forest Resources College of Natural Resources – University of Idaho; and Terrance W. Cundy, Potlatch Corporation, Lewiston, Idaho

Abstract: A model for spatial prediction of landslide hazard was applied to a watershed affected by landslide events that occurred during the winter of 1995-96, following heavy rains, and snowmelt. Digital elevation data with 22.86 m x 22.86 m resolution was used for deriving topographic attributes used for modeling. The model is based on the combination of logistic regression and principal component analysis (PCA) of road related landslides and assumes that the meteorologic factors which caused failures in the past will cause failures in the future. This model was intended to provide forest land managers with additional information on landslide hazards associated with roads. The advantage of the model is that it allows rapid assessment of spatial correlation of the topographic attributes on large areas. A Geographic Information Systems derived map using the principal component analysis model delineates areas with relatively high probabilities of failure in the basin.

INTRODUCTION

Forest roads play a key role in the vitality of any managed forest area. These roads provide access to the forest for multiple purposes including timber harvesting, recreation and fire suppression. Without forest roads, millions of acres of forest lands are unreachable. Beside providing many positive benefits, roads can have negative impacts as well. Concern about erosion and landslide failure from roads is high especially in highly dissected mountainous topography such as the Clearwater Basin in Idaho. Damage from landslides is a critical issue concerning fish habitat and water quality. Concerns about landslide failure have created a need for developing quantitative hazard models and prediction of landslide hazard.

In the literature there are various approaches for developing quantitative hazard models for prediction of landslide hazard (Carrara et al., 1991; Montgomery and Dietrich, 1994; Mark and Ellen, 1995, Gorsevski et al. 2000a, Gorsevski et al. 2000b). They are mostly based on the combination of Geographic Information Systems (GIS) and either an infinite slope stability model (Montgomery and Dietrich, 1994; Wu and Sidle, 1995; Okimura and Ichikawa, 1995) or statistical models, which link environmental attributes based on spatial correlation (Carrara et al., 1991, 1995; Chung et. al., 1995, 1999; Mark and Ellen, 1995; Chung and Fabbri, 1999).

Multivariate methods combined with GIS have been applied since the early 1980s (Carrara et al., 1983, 1991, 1992, 1995). The strategy of sampling in these applications was based either on a large-grid basis or in morphometric units. In either case the presence or absence of landslides were represented in a binary manner. All attributes believed to be relevant were sampled. These data sets were used to construct a model by applying multiple regression or deterministic analysis.

More recently Gorsevski et al. (2000a) applied logistic regression for spatial prediction of landslide hazard. A Receiver Operator Characteristic curve (Williams et al., 1999) was used to assist in the interpretation of the logistic regression results.

The approach taken in this paper uses a high resolution digital elevation model (DEM), a combination of logistic regression and Principal Component Analysis (PCA), and GIS. The DEM provides a representation of the topography of the study area and allows topographic attributes to be determined. The logistic regression analysis models the probability that an individual grid pixel will contain a landslide. The PCA identifies the most significant topographic variables that influence its occurrence. Although interpretation of the most significant topographic variables is difficult when PCA is applied, it can help to better screen the data and transform a set of correlated variables into a new set of uncorrelated variables called principal components. The principle components are also used for locating and identifying abnormalities in the data, and to check assumptions that may be required for certain statistical analysis to be valid. The principle components are used directly into the model for a backward elimination logistic regression. A table was derived from the logistic regression showing the classification of the landslides hazard. The classification table can be used as a tool for deciding a proper thresholds for hazard classification.

Although there are several solutions for categorizing ranges of probabilities, for this study area we used a scatterplot derived from the results of the logistic regression. The scatterplot shows a range of different outcomes based on the correctly identified landslides and non-landslide pixels from the classification table. From the scatterplot or the classification table we chose the influential point, which is the optimal combination for maximizing the correctly identified pixels from both groups. Departing from this influential point in one direction would classify better one of the binary groups, and misclassify in the other direction. Errors in one direction might be more serious than errors in the other direction. Managers can choose the level of risk they are prepared to accept based on the scatterplot or the classification table.

METHODS

In this paper we focused on the 85.9 km^2 Silver Creek watershed (Figure1) located northwest of Headquarters, Idaho in the North Fork of the Clearwater River Basin. The topography is highly dissected with elevations ranging from 485 m to 1635 m and slopes vary between 0 and 48 degrees. Precipitation averages about 980 mm annually. The density of roads is 5.5 km of road per square kilometer. A total of 81 road related landslides were inventoried by both photo interpretation and field inventory. The landslide triggering mechanism in the winter of 1995-96 was a combination of heavy rains and snowmelt.

DEM data with 22.86 m x 22.86 m resolution was used for deriving the topographic attributes. The original data for developing the DEM was collected on 38.1 x 76.2 m grid, then resampled to 22.86 m x 22.86 m. Topographic attributes included slope, elevation, aspect, profile curvature (slope profile curvature of a surface at each pixel center), tangent curvature (curvature of line formed by intersection of surface with plane normal to flow line), plan curvature (contour curvature), flow path (distance from watershed divide to the point of interest), and contributing area. The topographic attributes were generated from a DEM using TAPES-G software (Gallant et al., 1996). The parent material coverage was supplied by Potlatch Corporation.

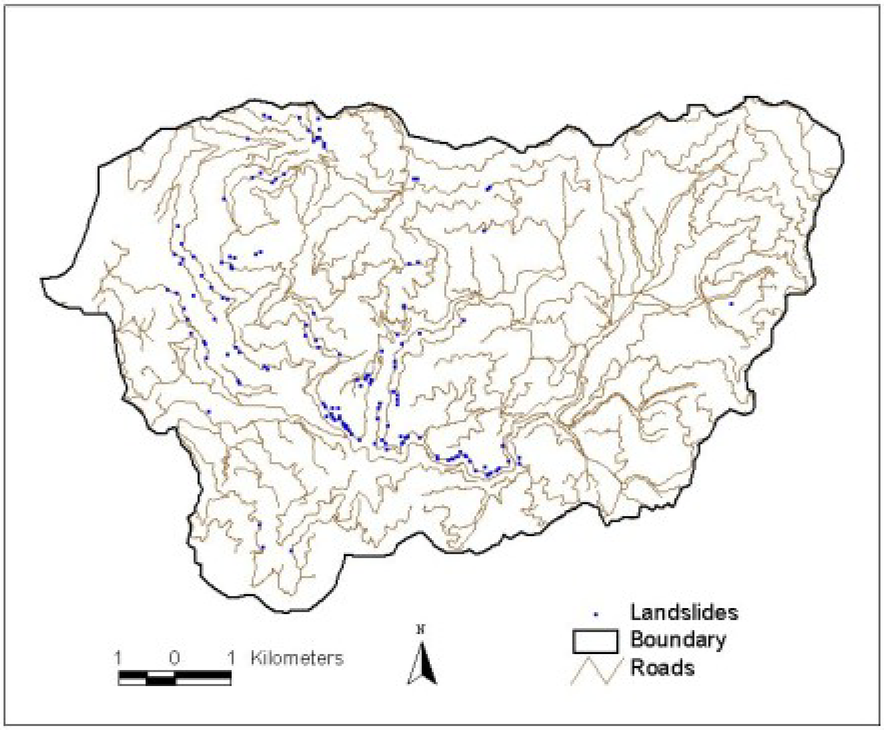

Figure 1. Distribution of Landslides and Road Network Over the Silver Creek Watershed

The topographic attributes of 15 percent of the roaded, non-landslide pixels in the watershed (24,372) were randomly sampled. All of the road-containing landslide pixels (81) were selected. These data sets were combined for further analysis.

PCA were used to screen the combined data, to produce a smaller set of uncorrelated variables, to determine the dimensionality of the space in which the data fall, and to determine the number of principal components to be used. The dimensionality of the space in which the data lies was decided by scree plot and accounting of at least 80% of the total variability in the original topographic attributes. Following the PCA, logistic regression with a backward elimination procedure was applied to the principal components. Backward elimination procedure helped to remove those principal components that were not statistically significant. After the logistic regression equation was derived, intermediate maps were generated for each of the principle components to assist in deriving the final landslide hazard map using the logistic regression equation.

RESULTS

A total of 25 original topographic attributes were reduced to 15 new, uncorrelated variables (principal components) by applying PCA. The backward elimination procedure used with the logistic regression produced six variables that were statistically significant at 95% significance level. The backward elimination method resulted in the equation shown in Table 1.

Log (p/(1-p)) = -7.0732 – PRIN1 * 1.0975 - PRIN2 * 0.6742 + PRIN3 * 0.2430 - PRIN4 * 0.3946 - PRIN5 * 0.5033 – PRIN6 * 0.5839

p – probability of landslide hazard
PRIN1 – first principle component (slope and schist)
PRIN2 – second principle component (elevation and basalt)
PRIN3 – third principle component (east and northeast)
PRIN4 – fourth principle component (south, southeast and quartzite)
PRIN5 – fifth principle component (northeast, north, granite and quartzite)
PRIN6 – sixth principle component (north, gneiss, quartzite, and alluvium)

Table 1. Logistic Regression Equation for Predicting Landslide Hazard

Interpreting the principle components with high confidence is difficult. Upon examination of the normalized eigenvectors, the elements with the largest absolute value suggested that the first principal component (PRIN1) in the logistic regression has a strong relationship with slope and the parent material schist. The variables that tend to have strong relationship with the second principal component (PRIN2) are elevation and the parent material basalt. The variables that tend to have strong relationship with the third principal component (PRIN3) are the east and northeast aspect. The fourth principal component (PRIN4) has a strong relationship with south and southeast aspect and quartzite parent material. The fifth principal component (PRIN5) has a strong relationship with northeast and north aspect and granite and quartzite parent material. Finally, the sixth principal component (PRIN6) has a strong relationship with north aspect and the following parent materials gneiss, quartzite, and alluvium. Except for the third principal component (PRIN3) variable, all other variables in the logistic regression have negative coefficients. When interpreting the logit coefficients, caution should be taken, because the logit model assumes a nonlinear relationship between the probability and the explanatory variables (Allison, 1999). For example, the estimated coefficient for the second principle component (PRIN2) is –0.6742. If our interpretation that the second principal component (PRIN2) have strong relationship with elevation and basalt is correct, then the risk of landslide hazard decreases with higher elevation for basalt.

The scatterplot which show the trade off between correctly identified landslides and non-landslides is shown in Figure 2. We chose an influence point of 67 for the landslides pixels, and 20,132 for non-landslide pixels. This point corresponded to a probability value of 0.005 from equation 1. Our selection of the influence point was determined by the desire to set the probability of making an error equal for both landslides and non-landslides.

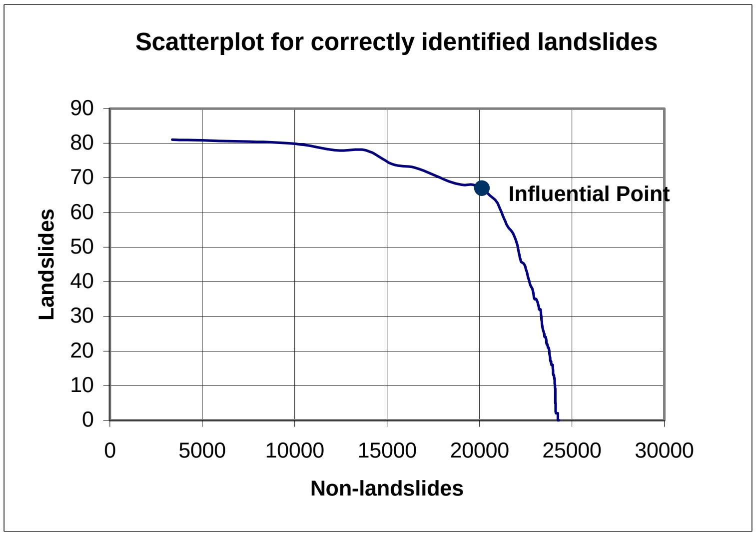

Figure 2. Scatterplot From the Correctly Identified Pixels From the Classification Table

Probability lower than the threshold (0.005) were mapped as Low Hazard, and probability higher than (0.005) were mapped as High Hazard. The resulting map is shown in Figure 3. It appears that 82.7% of the landslide pixels and 82.6% of the non-landslide pixels were correctly identified by the model.

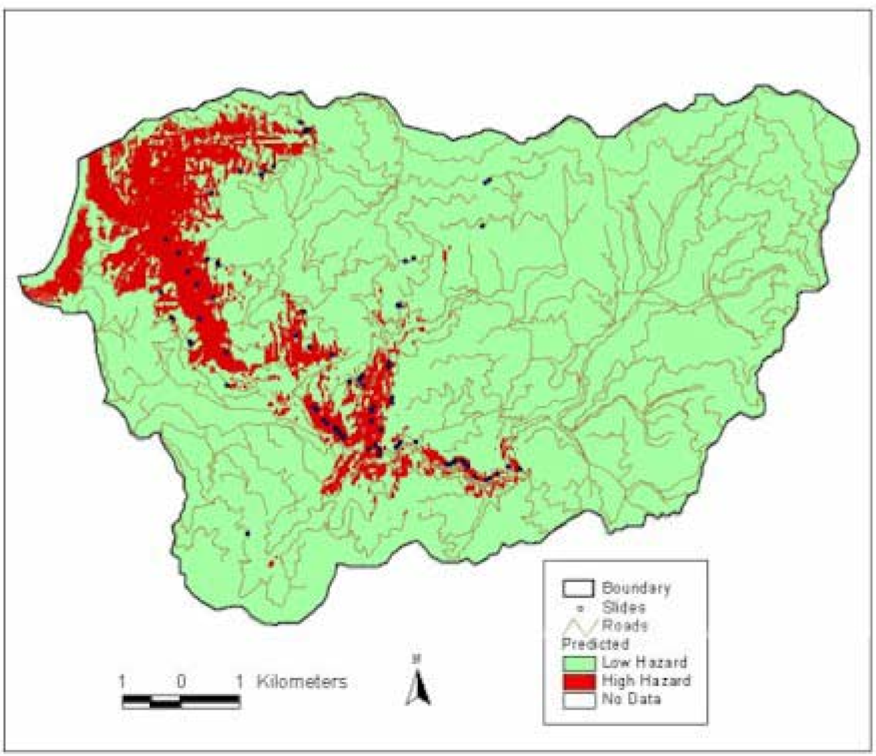

Figure 3. Predicted Landslide Hazard Assuming a Road in Each Pixel

DISCUSSION

The landslide hazard map presented in Figure 3 represents the road-related landslide hazard assuming that each pixel contained a road. While this is not the case, it presents, at a glance, the hazard associated with having a road in each pixel. The same analysis method presented in this paper, but using non-road containing pixels could be used to generate a similar map. That map would present the inherent hazard under the assumption of no roads. Comparison of the two maps would reveal risk reductions associated with road removal.

Our analysis did not include any attributes related to road construction technique which has been identified as important (Sessions et al., 1987). A further refinement of the proposed model would include this important attribute.

CONCLUSION

We developed a landslide hazard map of a 85.9 km^2 basin using a combination of logistic regression and principal component methods. Landscape characteristics were derived from a DEM. The map showed high and low risk areas for road placement. This tool can be used by land managers to focus limited resources toward the prevention of future landslides. The next step in our analysis is to include an attribute reflecting the road construction technique.

ACKNOWLEDGEMENTS

The research and work presented in this paper was supported through data provided by Potlatch Corporation and the Rocky Mountain Research Station in Moscow, Idaho.

REFERENCES

Allison P.D., 1999. Logistic Regression Using the SAS System: Theory and Application, Cary, NC: SAS Institute Inc., 1999. 304 pp.

Carrara A., 1983. Multivariate Models for Landslide Hazard Evaluation, Mathematical Geology, Vol. 15, No. 3, 1983

Carrara A, M. Cardinali, and F. Guzzetti 1992. Uncertainty in Assessing Landslide Hazard and Risk. ITC Journal, No. 2, pp. 172-183.

Carrara A., M. Cardinali, R. Detti, F. Guzzetti, V. Pasqui, and P. Reichenbach 1991. GIS Techniques and Statistical Models in Evaluating Landslide Hazard. Earth Surface Processes and Landforms, Vol. 16, 427-445 (1991)

Carrara A., M. Cardinali, F. Guzzetti, and P. Reichenbach, 1995. GIS Technology in Mapping Landslide Hazard, Geographical Information Systems in Assessing Natural Hazards (A. Carrara and F. Guzzetti, editors, Kluwer Academin Publishers, Dordrecht, The Netherlands, pp. 135-175.

Chung C.F., A.G. Fabbri and C.J. van Westen, 1995. Multivariate Regression Analysis for Landslide Hazard Zonation, Geographical Information Systems in Assessing Natural Hazards (A. Carrara and F. Guzzetti, editors, Kluwer Academin Publishers, Dordrecht, The Netherlands, pp. 107-133.

Chung C.F., and A.G. Fabbri, 1999. Probabilistic Prediction Model for Landslide Hazard Mapping. Photogrammetric Engineering and Remote Sensing. Vol. 65, No. 12 1999

ESRI Inc, 1998. Environmental Systems Research Institute, Inc. Computer Software Version 7.2.1, Redlands, CA: ESRI Inc.

Gallant J.C. and J.P. Wilson 1996. TAPES-G: A Grid-based Terrain Analysis Program for Environmental Sciences, Computers & Geosciences, Vol. 22, No. 7, pp. 713-722, 1996

Gorsevski P.V., P.E. Gessler, and R.B. Foltz 2000a. Spatial Prediction of Landslide Hazard Using Logistic Regression and GIS, 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4):Problems, Prospects and Research Needs. Banff, Alberta, Canada, September 2 - 8, 2000.

Gorsevski P.V., P.E. Gessler, and R.B. Foltz 2000b. Spatial Prediction of Landslide Hazard Using Discriminant Analysis and GIS, GIS in the Rockies 2000 Conference and Workshop: Applications for 21st Century. Denver, Colorado. September 25 – 27, 2000.

Mark R.K., and S.D. Ellen 1995. Statistical and Simulation Models for Mapping Debris-Flow Hazard, Geographical Information Systems in Assessing Natural Hazards (A. Carrara and F. Guzzetti, editors, Kluwer Academin Publishers, Dordrecht, The Netherlands, pp. 93-106.

Montgomery D.R., and W.E. Dietrich 1994. A physically based model for the topographic control on shallow landsliding. Water Resources Research. Vol. 30. No. 4 Pages 1153-1171. April 1994.

Okimura T. and R. Ichikawa 1985. A prediction method for surface failures by movements of infiltrated water in a surface soil layer. Natural Disaster Science 7:41-51

SAS Institute Inc., 1996 SAS/STAT Software: Changes and Enhancements Through Release 6.12. Cary, NC: SAS Institute Inc.

Sessions J., J.C. Balcom, and K.Boston 1987. Road Location and construction practices: Frequency and size in the Oregon Coast Range. Western Journal of Applied Forestry, 2(4):119-124.

Williams C.J., S.S. Lee, R.A. Fisher, and L.H. Dickerman 1999. A Comparison of Statistical Methods for Prenatel Screening for Down Syndrome. Applied Stochastic Models and Data Analysis 15. 000-000 (1999)

Wu W., R.C. Sidle 1995. A distributed Slope Stability Model for Steep Forested Basins. Water Resources Research, 31 2097-2110.