External Award G11AP20150

REGIONAL-SCALE STATISTICAL MODELING OF THE NORTHRIDGE EARTHQUAKE LANDSLIDE DATABASE

Term covered by the award : May 15, 2011 - May 15, 2013

Principal Investigator:
Joseph Wartman

Primary Graduate Student Researcher:
Melissa Logan Gillespie

University of Washington
Department of Civil and Environmental Engineering
201 More Hall, Box 352700
Seattle WA 98195-2700

wartman@uw.edu

Research supported by the U.S. Geological Survey (USGS), Department of the Interior, under USGS award number G11AP20150. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

CONTENTS

TECHNICAL ABSTRACT…….....…….……ii

NON-TECHNICAL SUMMARY……..……..……iii

INTRODUCTION 1

METHODS 3

Ordinary Logistic Regression 4

Independent Variables Used in the Logistic Regression Model 5

Dependent Variables Used in the Logistic Regression Model 6

Sampling for Model Training 9

Evaluation of Residuals 10

RESULTS 11

Logistic Regression Model 11

Model Predictive Power 12

Landslide Susceptibility Map Within Northridge Study Area 13

Comparison with Existing Earthquake-Induced Landslide Inventory (San Fernando) 13

DISCUSSION AND CONCLUSIONS 16

REFERENCES 18

External Award G11AP20150

REGIONAL-SCALE STATISTICAL MODELING OF THE NORTHRIDGE EARTHQUAKE LANDSLIDE DATABASE

Principal Investigator:
Joseph Wartman

Primary Graduate Student Researcher:
Melissa Logan Gillespie

University of Washington
Department of Civil and Environmental Engineering
201 More Hall, Box 352700
Seattle WA 98195-2700

wartman@uw.edu

TECHNICAL ABSTRACT

Earthquake induced landslides often number in the thousands to tens-of-thousands and extend over regions of 10,000 km^2 or more. Due to their potentially large volumes, high velocities, long travel distances, and impact forces, the consequences of seismically induced landslides can be severe. Currently, the majority of landslide hazard maps are generated based on methods that rely primarily on slope angle and detailed strength parameters, which could be costly and difficult to obtain over large areas, for use in the evaluation. Because the results are highly influenced by slope angle, current hazard mapping procedures tend to predict high probability of failure over large areas characterized by steeply sloping terrain. This work is the first step in development of new low-cost, high-resolution tools to forecast the occurrence and consequences of geologic hazards over regional scales. The model described herein has shown to be capable of forecasting the likelihood and spatial distribution of earthquake-induced landsliding over an area of ~5,000 km^2 and requires relatively simple input consisting of widely available geophysical parameters, thus making this approach equally-suited for both industrialized and developing nations. By identifying and including controlling factors in addition to slope angle in the evaluation (i.e. slope aspect, topographic position, geologic unit, and PGA), the model is able to differentiate between landslide susceptibility within areas having similar slope inclination. In addition, validation of the model developed with a seismically-induced landslide database from the 1994 Northridge event with one from the 1971 San Fernando event suggests that the model is transferrable to areas of similar terrain and geologic formations.

External Award G11AP20150

REGIONAL-SCALE STATISTICAL MODELING OF THE NORTHRIDGE EARTHQUAKE LANDSLIDE DATABASE

Principal Investigator:
Joseph Wartman

Primary Graduate Student Researcher:
Melissa Logan Gillespie

University of Washington
Department of Civil and Environmental Engineering
201 More Hall, Box 352700
Seattle WA 98195-2700

wartman@uw.edu

NON-TECHNICAL SUMMARY

Earthquake induced landslides often number in the thousands to tens-of-thousands and extend over regions of 10,000 km^2 or more. Due to their potentially large volumes, high velocities, long travel distances, and impact forces, the consequences of seismically induced landslides can be severe. These consequences routinely include injuries, property losses, and damage to critical infrastructure systems such as transportation, water supply, and telecommunication networks. As such, resilience to natural hazards is a critical component of sustainability for metropolitan regions. Reducing community vulnerability to geologic hazards requires that the potential impacts of hazards be quantified so effective measures can be taken to reduce risk. Currently, the majority of landslide hazard maps are generated based on methods that rely primarily on slope angle and detailed strength parameters, which could be costly and difficult to obtain over large areas, for use in the evaluation. Because the results are highly influenced by slope angle, current hazard mapping procedures tend to predict high probability of failure over large areas characterized by steeply sloping terrain. This work is the first step in development of new low-cost, high-resolution tools to forecast the occurrence and consequences of geologic hazards over regional scales.

REGIONAL-SCALE STATISTICAL MODELING
OF THE NORTHRIDGE EARTHQUAKE LANDSLIDE DATABASE

INTRODUCTION

Landslides are prominent geomorphic features of earthquakes (e.g. Harp and Jibson 1996, Wartman et. al. 2006, Owen et al. 2008). Moderate-to-large magnitude earthquakes trigger thousands to tens-of-thousands of landslides over regions spanning tens-of-thousands of km^2 (Keefer 1984, Figure 1). The most abundant of these landslides are "disrupted" slope failures including rock and soil falls, slides, and avalanches (Keefer 1984). With their potentially large volumes, high travel velocities, and long runout distances, disrupted landslides can obstruct roadways, damage structures, produce large amounts of sediment, and deposit landslide dams. These represent a serious impact on the built and natural environments. Earthquake-induced landslides can have a significant economic impact in a region, in some cases resulting in financial losses that exceed combined losses from all other seismic effects (Kramer 1996). Recognizing the magnitude of this problem, the U.S. National Research Council recently recommended a 15-fold increase in funding for landslide research in the United States (NRC 2003). Similarly, the U. S. National Earthquake Hazard Reduction Program (NEHRP) has indentified earthquake-induced landslides a key research need in its recently released 2009-2013 strategic plan (NEHRP 2008).

Earthquake-induced landslides hazards have traditionally been analyzed at the site-specific scale. There is, however, growing interest in expanding seismic landslide hazard assessment to substantially larger spatial scales, a sentiment emphasized in the NEHRP 2009-2013 strategic plan. In addition to providing efficient and cost-effective seismic hazard assessments, regional-scale analysis tools have the potential to capture "system-level" performance and spatial propagation of risk within a defined study area. Such capabilities are important when considering the effects of landslides on geographically distributed critical infrastructure systems, which are highly susceptible to damage from slope failures.

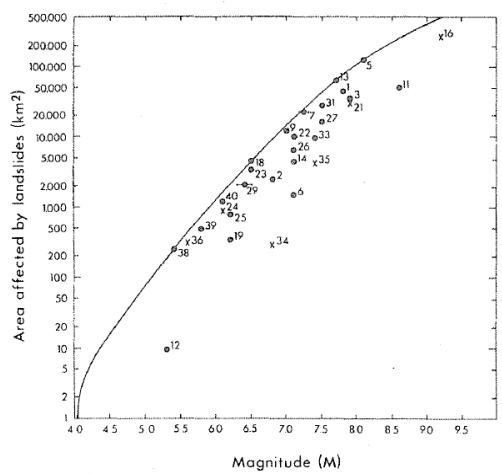

Figure 1: Area affected by landslides versus earthquake magnitude (Keefer 1984)

Seminal work by McCrink and Real (1996), Miles and Ho (1999), Jibson et al. (2000), and others represent important early steps toward regional-scale seismic landslide hazard assessment. Working independently, these researchers implemented the displacement-based Newmark (1965) sliding block model within a geographic information systems (GIS) framework, thus allowing the assessment of seismic slope hazards over regional scales. But while physically based landslide models such as Newmark (1965) have a good theoretical basis, obtaining required model parameters (e.g. friction angle) over large regions is a costly challenge that has hindered the widespread adoption of such approaches in practice (Guzzetti et al. 1999).

Over the past decade a fundamentally different but promising approach to regional-scale landslide assessment has emerged from the non-seismic landslide research community. This

approach formulates and "trains" robust statistical models using high quality landslide inventories. It has become a standard method for regional-scale precipitation-induced landslide hazard assessment throughout much of Europe (e.g. Brenning 2005, Van Den Eeckhaut et al. 2006, Greco et al. 2007) and Asia (e.g. Dai and Lee 2003, Lee and Pradhan 2007, Mathew et al. 2009). The statistical modeling is typically performed using one of several techniques including logistic regression (e.g. Greco et al. 2007), discriminate analysis (e.g. Baeza and Corominas 2001), and Bayesian-based weight-of-evidence approaches (e.g. Duhal et al. 2008). Although not purely statistical in nature, artificial neural networks have also been used for a similar purpose (Nefeslioglu et al. 2008). Advantages of regional-scale statistical models over their physically based counterparts include the following.

1) There is no need to directly quantify geotechnical mechanical properties (e.g. friction angle) over large spatial scales. These parameters are instead reflected in the underlying geophysical characteristics and landslide distribution of the training set, and are thus incorporated into the analysis in an indirect manner.

2) Potentially important slope-ground motion-landslide interaction factors such as three-dimensional slope stability and ground motion topographic amplification effects, which are difficult to model in physically based regional-scale physical models, are reflected in the training set and therefore captured by the statistical model.

3) Owing to the statistical nature of the technique, uncertainties in model predictions can be directly quantified and allowing probabilistic forecasts made.

Inspired by research demonstrating the feasibility, efficiency, and accuracy of statistical modeling for non-seismic landslides, we extended this technique to consider earthquake-induced landslide hazards. With the exception of recent work by a Taiwanese team (Lee et al. 2008), there is virtually no published research on this concept. Our proposed research builds on the substantial body of work in this area from the non-seismic landslide research community and represents an important step in the development of lower cost but higher resolution hazard mapping methods, which is recognized as a critical research need by the earthquake community (NRC 2004). Statistical modeling is a well-established and validated technique for predicting precipitation-induced landslides, thus suggesting a high probability that it can be successfully adopted for landslides triggered by other transient mechanisms such as earthquakes.

We formulated a logistic regression predictive model, which was trained and optimized using a high quality USGS database of approximately 11,000 landslides triggered by the 1994 Northridge earthquake. The combination of comprehensive landslide mapping, ample ground motion recordings, and high resolution topographic and geologic data make this uniquely suited for this work. We then independently assessed the model's validity and accuracy by comparing its predictions against the observed locations of landslide from the 1971 San Fernando earthquake, whose mesoseismal region overlaps that of the Northridge earthquake training set.

It is expected that continued work in this direction will help transform the way regional-scale geohazard assessment is conducted leading to additional research and an expansion of the work in future years to consider (i) integration with regional-scale hazard and risk codes such a

HAZUS or MAEviz, thus allowing fuller consideration of landslide consequences on networked infrastructure and urban systems, and (ii) forecasting of consequences for scenario earthquakes, (iii) integration with regional scale non-seismic (i.e. precipitation) landslide models for multihazard landslide risk assessment to consider, for example, impacts of climate change on earthquake-induced landslide hazards and (iv) use in land planning, hazard mapping, and policy making (e.g. development of California Geological Survey seismic hazard maps).

METHODS

Compared with physically based formulations, statistical modeling represents a profoundly different approach to regional-scale landslide hazard mapping. First proposed in the 1980s (Carrara 1983) and since aided by dramatic improvements in computational power and remote sensing, this approach has developed rapidly over the past several decades and has today become a popular tool for non-seismic landslide hazard mapping (Carrara and Pike 2008). Fundamental to this technique is the concept that past behavior is indicative of future performance. Statistical modeling of landslide hazards involves four basic steps (Guzzetti et al. 1999):

1) Landslides are mapped within a defined study region.

2) Geophysical parameters directly or indirectly correlated in landslides are mapped within the project study area. These parameters typically include factors related to topography, geology, vegetation, land use, and, in the case of non-seismic landslides, precipitation. The scale of the analysis is dictated by the resolution of the geophysical and/or landslide data, and is typically on the order of 1:10,000-20,000.

3) Statistical models are formulated, optimized and "trained." In this stage the relative contribution of the geophysical parameters is determined.

4) The resulting statistical model is applied over a large spatial region to develop localized measures of relative landslide susceptibility (e.g. high, moderate, low). A more common alternative today is to include a triggering parameter (e.g. rainfall intensity) to compute the localized probability of landslide occurrence.

Several techniques are used for statistical modeling: logistic regression, discriminate analysis, and Baysian (weight-of-evidence) approaches, among others. Several recent studies have applied each of these approaches over a study area and have generally found that the model predictions are remarkably consistent regardless of the technique used. This suggests that the quality of the training data (geophysical parameters and landslide inventory) is more important than the specific technique when developing a statistical model. As this proposed research will use logistic regression, only this technique will be described in detail in a separate section below.

The increasing availability of high-quality databases and advancements in spatial statistics has improved the reliability of statistical modeling. A number of recent studies have compared landslide predictions made using statistical techniques with observed slope performance in

regions outside of the training zone, or with observed performance during other triggering events (i.e. other precipitation events). These studies have shown a high degree of accuracy with success rate for correct cell classification typically in the range of 87% to 93% (Dai and Lee 2003, Mathew et al. 2009, Lee and Pradhan 2007).

Statistical modeling has received limited attention from the seismically-induced landslide research community. A notable and recent exception is the early work by Lee at al. (2008), who used a database of landslides triggered by the Chi-Chi Taiwan earthquake to train a discriminant analysis regional statistical model. The model captured landslide occurrence in a region neighboring the training set for the same earthquake, suggesting that the statistical approach can be successfully extended to seismically triggered events. Miles and Keefer (2009) proposed a different approach using fuzzy logic systems ("computing with words" Zadeh 1996) to predict the style and density of earthquake induced landslides in a region based on categorical descriptions (e.g. "extremely steep slopes," "very high intensity shaking," etc.). The model is not "trained" as a statistical formulation, but rather based on the developer's knowledge of landslide susceptibility factors. Miles and Keefer (2009) tested their model, known by its acronym CAMEL, against the Loma Prieta landslide inventory and found that it predicted landslide concentrations well.

Ordinary Logistic Regression

Logistic regression, a multiple regression technique well suited for situations where the outcome variable is categorical (in this case, landslide or no landslide) and the predictor variables (geophysical parameters) are either continuous (e.g. slope inclination) or categorical (e.g. type of vegetation) [Homer and Lemeshow 2000, Maynard 2002], was selected for use on this project. Logistic regression has several advantages over other multiple regressions techniques that make it well suited for statistical landslide hazard mapping: (i) predictor values do not have to be normally distributed, (ii) predictor values can be continuous, discrete, or dichotomous, and (iii) it features limited sensitivity to data outliers and homogeneity of statistical variances/covariances.

Logistic regression utilizes a logit function, defined as:

$$\text{logit}(y) - \ln[p/1\text{-}p]$$

where p is the probability of occurrence of the dependant variable y, and p/1-p is the odds of occurrence.

The functional form of the region is:

$$\text{logit}(y) = a + b_1x_1 + b_2x_2 + b_3x_3 + \ldots\ldots + e$$

Where y is as defined above, x1 is the i-th explanatory variable, a is a constant, b1 is the i-th regression coefficient and e is the error term.

The logit(y) function can be converted to probability by rewriting the first equation as;

$$p = \exp(a+b_1x_1+b_2x_2+b_3x_3+...)/1+\exp(a+n_1x_1+b_2x_2+b_3x_3+...)$$

Homer and Lemeshow (2000) provide a more complete description of logistic regression and its application.

Independent Variables Used in the Logistic Regression Model

Various combinations of potential explanatory variables such as slope angle, slope aspect, elevation, curvature, distance to ridge, rock type, land use classification, peak ground acceleration (PGA), distance to road, and road density were tested and their effect on the model's goodness of fit were studied. Independent variables for inclusion in the logistic regression model were selected based on their significance to the model, as evaluated based on the Wald test. The Wald test tests the null hypothesis that the regression coefficients of interest are equal to zero. If the test fails to reject the null hypothesis, this suggests that removing the variable from the model will not have a significant negative impact on model fit, since a predictor with a coefficient that is very small relative to its standard error is generally not doing much to help predict the dependent variable. Explanatory variables with a significance level greater than 0.05 were retained in the model. These include elevation, slope angle, slope aspect, position within the slope, PGA, and geologic unit. Slope aspect and geologic unit were treated as categorical. The reference category for each categorical variable is the baseline group against which all other groups are compared and was selected as the category for which the greatest incidence of mapped landslides was observed to occur for the study area. The class containing a combination of Modelo and Monterey Formations was selected as the reference category for geologic unit and southwest was selected as the reference category for slope aspect.

Explanatory variables were also evaluated for collinearity. Collinearity occurs when one variable is linearly related to another variable in the model. This was evaluated by checking for high standard errors, performing diagnostic testing for continuous variables through assessment of computed tolerance values, (Variance Inflation Factor) VIF values and through assessment of a collinearity diagnostics matrix. Collinearity between variables was not identified as a concern for variables retained in the regression model.

Independent variables used in the model, and sources of information are summarized in the following table.

Table 1 - Summary of Independent Variables Used in the Logistic Regression Model

Independent Variable	Data Source	Type of Feature, Resolution or Scale	Notes
Elevation	USGS National Elevation Dataset (http://seamless.usgs.gov)	1/3 arc-second (about 10 meters)	--
Slope Angle	Calculated from Elevation Data Using the Surface Toolset in ArcGIS	1/3 arc-second (about 10 meters)	--
Slope Aspect	Calculated from Elevation Data Using the Surface Toolset in ArcGIS	1/3 arc-second (about 10 meters)	1
Position Within Slope	Calculated from Elevation Data Using the Hydrology Toolset in ArcGIS	1/3 arc-second (about 10 meters)	--
PGA	USGS Earthquake Hazards Program http://earthquake.usgs.gov/eqcenter/shakemap	polygon	--
Geologic Unit	U.S. Geological Survey Open-File Report 2005-1019 http://pubs.usgs.gov/of/2005/1019/	polygon 1:100,000 map scale	2, 3, 4

Notes:

1. Reference category : Southwest. Other categories: North, Northeast, East, Southeast, South, West, Northwest.
2. Some formations had to be grouped into a category to reduce the total number of classes, thereby reducing the potential for too few observed landslide occurrences per class and large standard errors. Geographic location, frequency of landslide occurrence and predominant rock type were considered in determining groupings.
3. Reference category: Modelo and Monterey Fms. Other categories: a) Pico Fm, b) Towsley Fm, c) Saugus and Mint Canyon Fm, d) Sespe Fm, e) Granitic and Gneissic Rocks, f) Landslide Debris, Talus and Colluvial soil, g) Topanga, Tuna Canyon and Vaqueros Fms, h) Llajas Fm, i) Juncal and Matilija Sandstone Fms, j) Clalbasas, Castiac and Simi Conglomerate Fms, k) Sedimentary rocks having less than 1% of total landslide centroids per formation, l) Conejo and Zuma Volcanics, and m) Surficial Soil Deposits containing Landslide Centroids.
4. Digital geologic maps available for 22 of 32 of 30 x 60 degree quadrangles covered by the landslide study area. Remaining geologic quadrangles were georeferenced and digitized and for use in this research based on mapping by Thomas W. Diblee, Jr et al.

Research has demonstrated that selection of an appropriate grid resolution used in extraction of data for independent variables and sampling for model development can have a significant impact on the model's predictive power. Therefore, sources of thematic data were assessed to determine a suitable grid resolution (or cell size) for use in raster-based GIS, determined to be 10-m for this project. The grid resolution was selected with consideration factors such as the size of map delineations, inspection density (Hengl 2006). ArcInfo was subsequently used to convert vector data to raster data, perform calculations to generate variable data from thematic layers, and resample rasters as necessary to achieve uniform grid resolution and grid origins for all thematic layers, and, where applicable, reclassify data into categories for analysis.

Dependent Variables Used in the Logistic Regression Model

The binary dependent variable used in the logistic regression analysis is the presence (1) or absence (0) of a landslide initiated as a result of the 1994 Northridge, California earthquake. A digital database of over 11,000 such landslides was developed by USGS with contributions from

CDMG and Los Angeles County. The database was created based on results from detailed field investigations undertaken to (1) document earthquake-triggered landslides, (2) to locate areas of concentrated landsliding and (3) to find the farthest extent of landsliding. High-altitude aerial photographs (nominal scale 1:60,000) of the epicentral region by the U.S. Air Force, taken within hours of the earthquake, were also used in generating this detailed inventory of the earthquake induced ground failures.

The U.S. Air Force aerial photography was used to map fresh landslides on 1:24,000-scale USGS topographic base maps. Based on information from Harp and Jibson, landslides as small as 1-2 m across were visible where the slopes were sunlit. However, on the north side of steep slopes, where slopes were partly shaded, slides about 5-10 m across are the smallest that could be resolved. As such, the inventory is not complete. Field observations by members of the reconnaissance team indicate that south-facing slopes in most of the landslide area are generally steeper and produced far more landslides than north-facing slopes. Based on this, it is speculated that landslides on north-facing slopes that are not visible on the photos account for only a small proportion of the total landslides. Harp and Jibson estimate that no more than about 20 percent of the landslides that exceeded 5 m in maximum dimension and no more than 50 percent of those smaller than 5 m were overlooked. In terms of area, they estimate that more than 90 percent of the area covered by triggered landslides was mapped (because most of the landslides that are not visible on the photos are small).

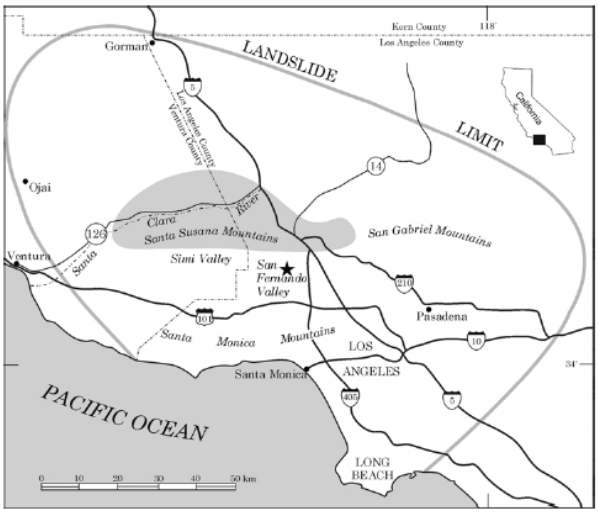

Figure 2: Map showing epicenter (star) of 1994 Northridge earthquake (M = 6.7) and areas affected by landslides

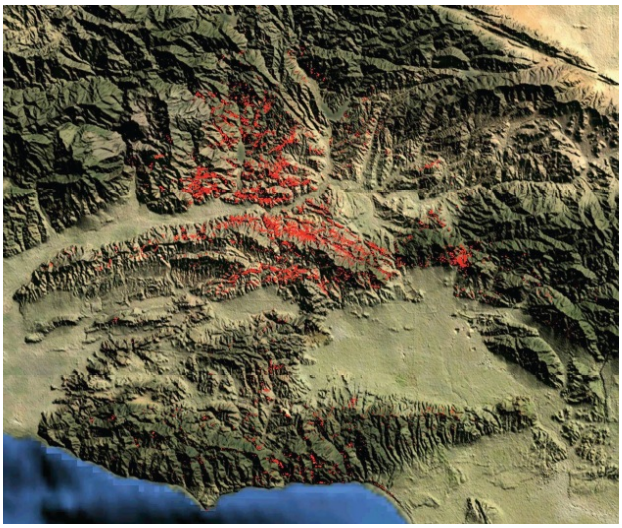

Figure 3 Digital database of over 11,000 landslides (shown in red above) initiated by the 1994 Northridge, California earthquake as mapped by Harp and Jibson. Topography of study area illustrated.

Upon completion of the mapping, identified landslide locations were manually digitized ARC Info geographic information system (GIS), resulting in creation of a ARC Info coverage file containing both line and polygon topology. Location accuracy of landslides mapped from

the air photos to the paper base maps was estimated to be within 10 m and no worse than 20 m. When the paper maps were registered on the digitizer, the computer calculated the root-mean-square (RMS) error in the base-map registration, which averaged 3.6 m and ranged from 0.2 to 10.4 m. Thus, landslide locations, as represented in the digital map file are generally accurate within about 15 m and are no more than 30 m mislocated (Harp and Jibson, 2005).

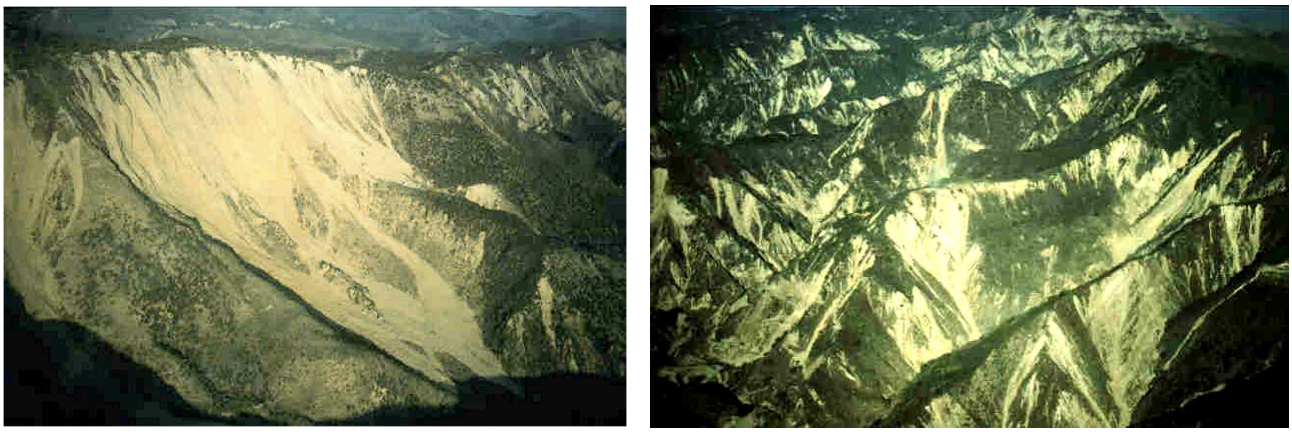

Figure 4 (a) Massive disrupted slide in very weakly cemented sediment (b) Light-colored areas are landslides that failed during the Northridge earthquake (Harp and Jibson 1995)

The digital database developed by Harp and Jibson was imported into Google Earth in order to allow for a visual observation of the terrain in which the earthquake-induced landslides occurred and observation of the overall accuracy with which the landslides were mapped. Observation of a number of the mapped Northridge landslides in Google Earth confirms the statement above by Harp and Jibson with regard to mapping accuracy. Imagery available with Google Earth from dates within months of the Northridge event were also used to develop a detailed mapping of 300 landslide locations, which included delineation of source, runout and deposition areas. The landslides to be included in the detailed mapping were randomly selected using the Sampling Design Tool for ArcGIS developed by the National Oceanic and Atmospheric Administration (NOAA) Biogeography Branch. The distribution of independent variables at the centroid of the 300 randomly selected landslide locations was compared with that for the approximately 11,000 landslide locations mapped by Harp and Jibson and the distribution of independent variables for each of these databases was noted to be approximately equal. Figures 5 and 6 illustrate examples of original landslide delineations and detailed landslide delineations overlain on three-dimensional Google Earth

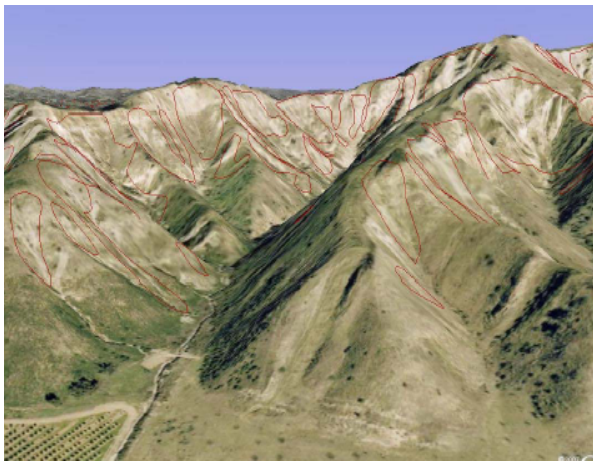

Figure 5: Example of landslide delineations (in red) as mapped by Harp and Jibson (1996) superimposed on three-dimensional Google Earth imagery

imagery.

Sampling for Model Training

The centroids for the source areas identified in the detailed landslide mapping described above were used to represent landslide cells in the model training set. The intent of this was to ensure that the values for independent variables used to represent landslide locations in the training set are representative of the conditions where landslides were initiated. Due to the fact that more than one source area was sometimes identified for a particular landslide location as mapped by Harp and Jibson, this resulted in a total of 411 landslide cells in the training set.

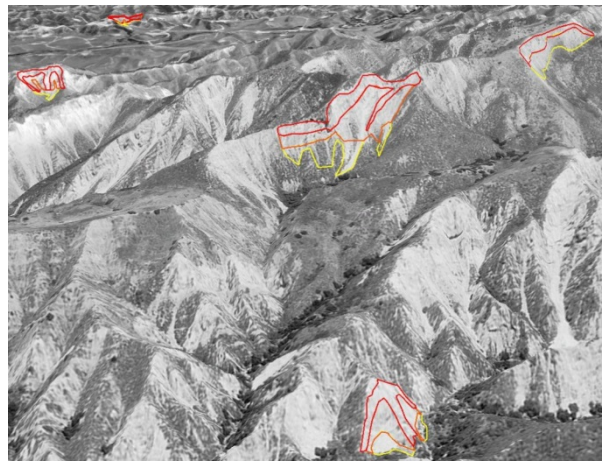

Figure 6: Example of detailed delineations depicting source (in red), runout (in orange) and deposition (in yellow) areas based on historic Google Earth imagery

Landslide free cells for use in the training set were randomly selected from 1) cells outside of a 50 m buffer from the landslide delineation and 2) cells where slope angle was greater than 15 degrees and PGA was greater than or equal to 0.2g. A 50 m buffer from the mapped landslide locations was used to account for potential inaccuracies in the landslide locations and to increase the likelihood of selecting cells that were truly landslide free. Minimum criteria for slope angle and PGA were used to prevent sampling of landslide free cells from areas where landsliding is not likely to occur. Sampling from flat areas and/or areas where less sever ground shaking occurred would over-estimate the importance of slope angle and PGA and underestimate the significance of other independent variables (Eeckhaut et al, 2011). The NOAA Sampling Design Tool was used for generating a random sample of landslide free cells meeting the above-mentioned criteria.

Based on review of the literature where binary statistical models are used in prediction or hazard mapping related to landslide occurrence, there does not appear to be a consensus on the proportion of landslide cells (1's) and landslide-free cells (0's) to be used in the raining set. Often the ratio of landslide to landslide-free cells of 1 to 1 or 1 to 2 appears to be used. For logistic regression analysis performed with rare events data (binary dependent variables where there are dozens to thousands of times fewer occurrences than non-occurrences), the model intercept, and in turn the predicted probabilities will be skewed if 1) the training set sample is not designed to reflect the proportion of 1's and 0's or 2) the intercept for the regression equation is not corrected to reflect these proportions (King and Zeng, 2001).

An attempt was made to estimate the proportion of landslide occurrence to non-occurrence cells based on a "study area" defined by cells having slope angle was greater than 15 degrees and

PGA was greater than or equal to 0.2g. Based on this study area, and assuming that all cells within the Harp and Jibson delineations are landslide cells, the ratio of landslide-free to landslide cells was estimated to be on the order of 10 to 1. Figure 7a and 7b below illustrates the difference in predicted probabilities of landslide occurrence for models generated based on ratios of landslide-free to landslide cells of 2 to 1 and 10 to 1, respectively.

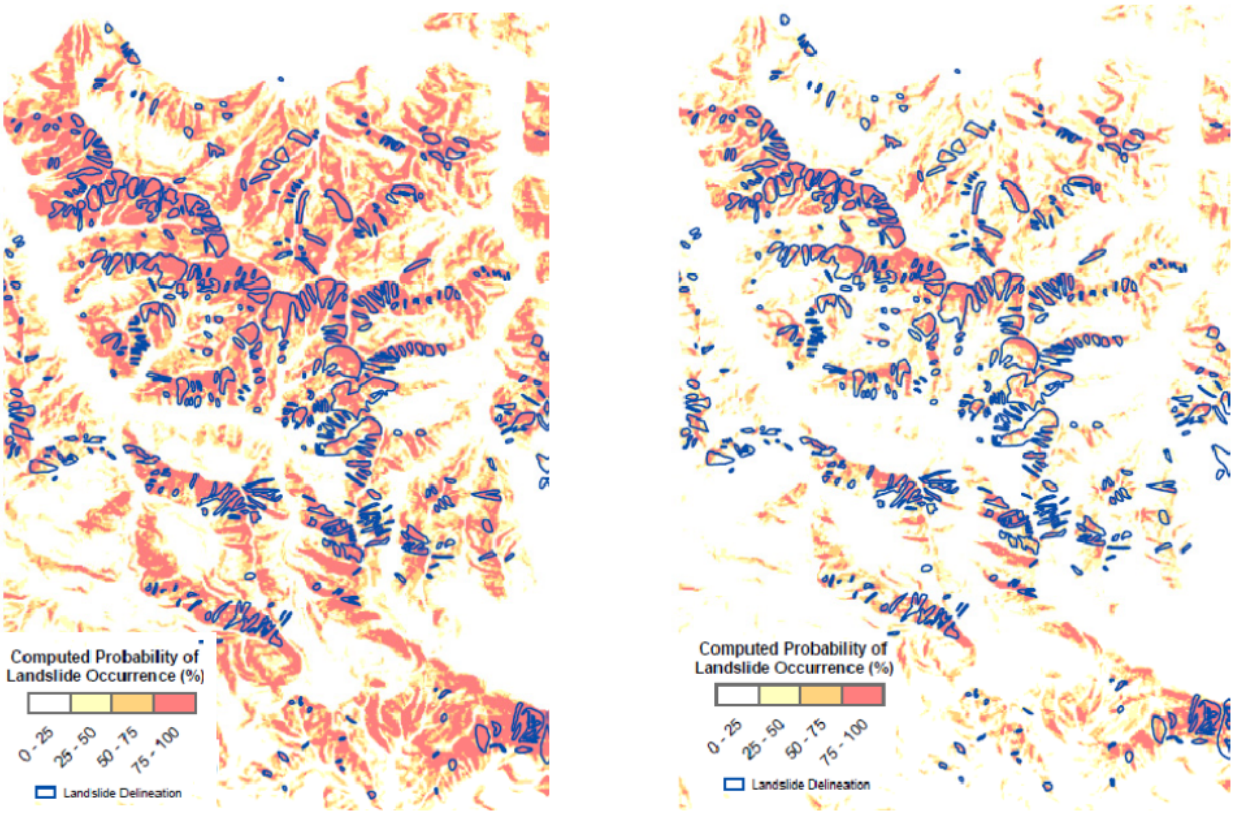

Figure 7: (a) Model prediction based on a 2 to 1 ratio of landslide-free to landslide cells (b) Model prediction based on a 2 to 1 ratio of landslide-free to landslide cells

Evaluation of Residuals

Residuals are the difference between the outcome predicted by the model and the outcome observed in the sample. Residuals calculated within SPSS such as DFBETA, the leverage statistic, and the studentized residual were evaluate in an effort to identify outliers or observations that could have an unusually large effect on the estimate of regression coefficients. No evidence of potential cause for conern was noted.

Residuals were also plotted in ArcInfo and evaluated visually for clustering of large residuals in a geographic region(s), which could be an indication of autocorrelation, or that an important variable(s) is missing from the model. Clustering of large residuals was not observed.

RESULTS

Logistic Regression Model

The model developed as part of this project indicates that increased slope angle, locations closer to the top of a slope, increased PGA, Pico and Llajas formations, and southern facing slope aspect increase the probability of landslide occurrence. Slope angle, geologic unit and slope aspect have the greatest influence on landslide occurrence or non-occurrence.

Table 2 - Summary of Logistic Regression Model for Producing Probability of Landslide Occurrence Map

Parameter	Description	B	Standard Error	Wald	Significance	Exp(B)
Intercept	--	-9.802	0.641	233.952	0.000	0.000
Elevation	Elevation (ft)	-0.003	0.000	33.941	0.000	0.997
Slope	Slope Angle (degrees)	0.210	0.011	367.629	0.000	1.233
Slope Position	See Note Below	2.277	0.425	28.752	0.000	9.752
PGA	Peak Ground Acceleration, %g	6.015	0.771	60.908	0.000	409.693
Geology	Modelo and Monterey Fm	--	--	182.772	0.000	--
Geology (1)	Pico Fm	1.149	0.271	17.927	0.000	3.154
Geology (2)	Towsley Fm	-0.331	0.314	1.109	0.292	0.719
Geology (3)	Saugus and Mint Canyon Fms	-0.885	0.325	7.438	0.006	0.413
Geology (4)	Sespe Fm	-0.936	0.378	6.141	0.013	0.392
Geology (5)	Granitic and Gneissic Rocks	-1.878	0.314	35.679	0.000	0.153
Geology (6)	Landslide, Talus and Colluvium	-0.601	0.373	2.595	0.107	0.548
Geology (7)	Topanga, Tuna Canyon and Vaqueros Fms	-2.091	0.398	27.620	0.000	0.124
Geology (8)	Llajas Fm	-0.799	0.581	1.893	0.169	0.224
Geology (9)	Juncal and Matilija Sandstone Fms	-0.509	0.609	0.697	0.404	0.601
Geology (10)	Clalbasas, Castiac and Simi Conglomerate Fms	-1.398	0.531	6.942	0.008	0.247
Geology (11)	Sedimentary rocks having less than 1% of total landslide centroids per formation.	-3.013	0.497	36.795	0.000	0.049
Geology (12)	Conejo and Zuma Volcanics,	-2.842	0.628	20.459	0.000	0.058
Geology (13)	Surficial Soil Deposits containing Landslide Centroids	-3.192	0.624	26.143	0.000	0.041
Aspect	Southwest	--	--	91.191	0.000	--
Aspect (1)	North	-2.549	0.432	34.900	0.000	0.078
Aspect (2)	Northeast	-2.598	0.432	36.193	0.000	0.074
Aspect (3)	East	-1.394	0.306	20.793	0.000	0.248
Aspect (4)	Southeast	-0.989	0.307	10.386	0.001	0.372
Aspect (5)	South	0.019	0.232	0.006	0.936	1.019
Aspect (6)	West	-0.586	0.254	5.304	0.021	0.557

Parameter	Description	B	Standard Error	Wald	Significance	Exp(B)
Aspect (7)	Northwest	-1.224	0.314	15.219	0.000	0.000
* Distance to Stream / Distance to Stream + Distance to Ridge. Higher numbers represent locations closer to the top of slope.						

Model Predictive Power

Predictive ability of the regression model was quantified through use of the receiver operating characteristic (ROC) Curve, which is a plot of 1-specificity (false positive) against sensitivity (true positive). The area under the ROC curve (AUROC) is an indication of predictive ability. If AUROC equals 0.5, the model predicts randomly. If AUROC is close to 1, that indicates good predictive ability (Hosmer and Lemeshow, 2000).

The ROC curve below was calculated in SPSS for a sample set consisting of 10,000 randomly selected landslide cells and 10,000 randomly selected landslide-free cells. All landslide-free cells were sampled from areas meeting the same criteria established for the training set (slope angle greater than 15 degrees and PGA greater than or equal to 0.2g). The computed AUROC is 0.951.

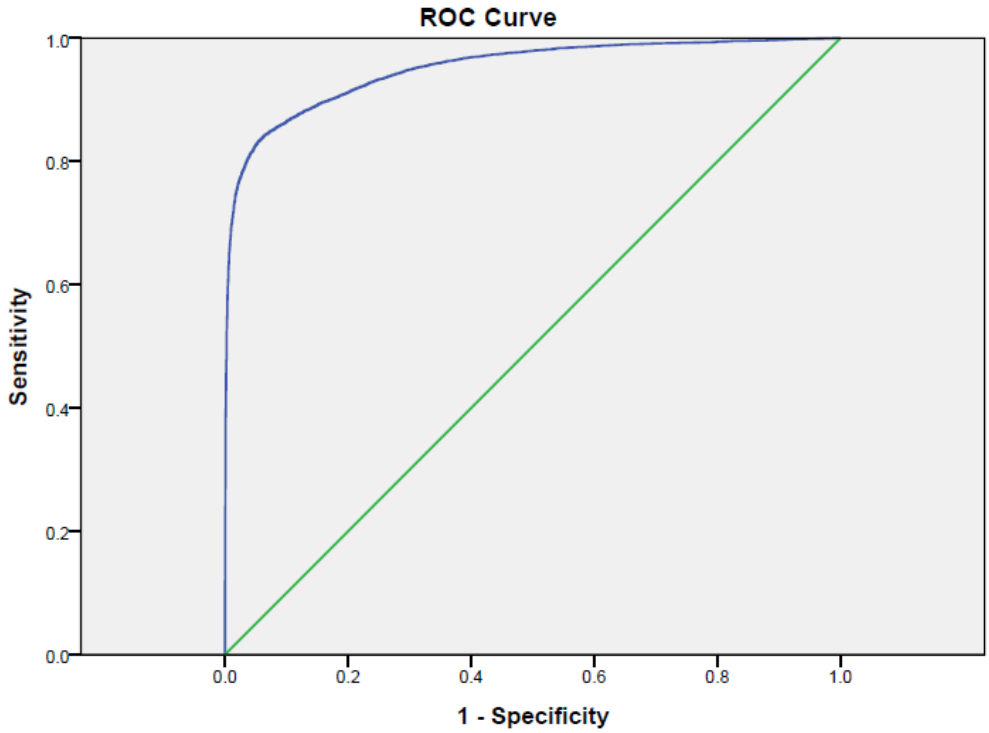

Figure 8: ROC curve for validation set consisting of 10,000 randomly selected landslide cells and 10,000 randomly selected landslide-free cells. All landslide-free cells were sampled from areas having slope angle greater than 15 degrees and PGA greater than or equal to 0.2g.

Landslide Susceptibility Map Within Northridge Study Area

The regression model was applied over the project study area to compute probability of landsliding for the Northridge event. The map excerpt illustrates accuracy of the prediction for a portion of the study area having high incidence of mapped landslides. This is generally representative of the overall results.

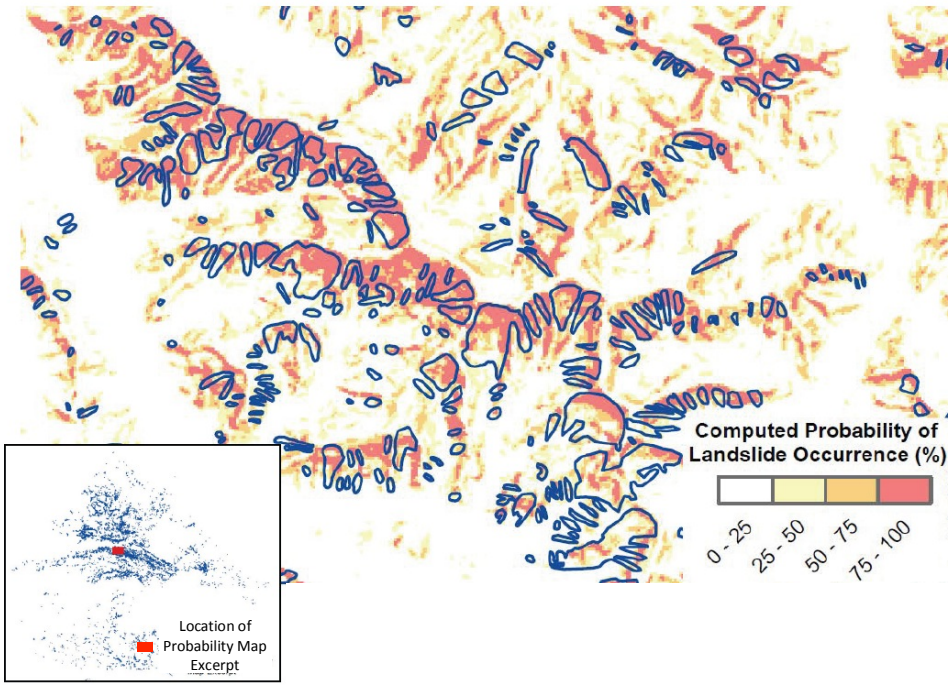

Figure 9: Predicted probabilities of landsliding for the Northridge event shown with the location of actual mapped landslides (blue outline). Inset diagram shows location of map excerpt.

Application Over the 1971 San Fernando Mesoseismal Area

The model described above was tested against a good quality landslide inventory developed by the USGS after the 1971 San Fernando earthquake (Grantz, 1971). This inventory consists of landslides mapped on USGS topographic maps and true-color airphotos and includes a legend that distinguishes between disrupted and coherent landslides. The San Fernando landslide inventory falls entirely within the mapped region from the Northridge earthquake and thus provides an extraordinary opportunity to independently test the model against a different earthquake within the study area.

PGA values available through USGS for the San Fernando event, along with the other independent variables described in previous paragraphs were incorporated into the model developed as part of this work and a map of probability of landslide occurrence was developed for the San Fernando mesoseismal area.

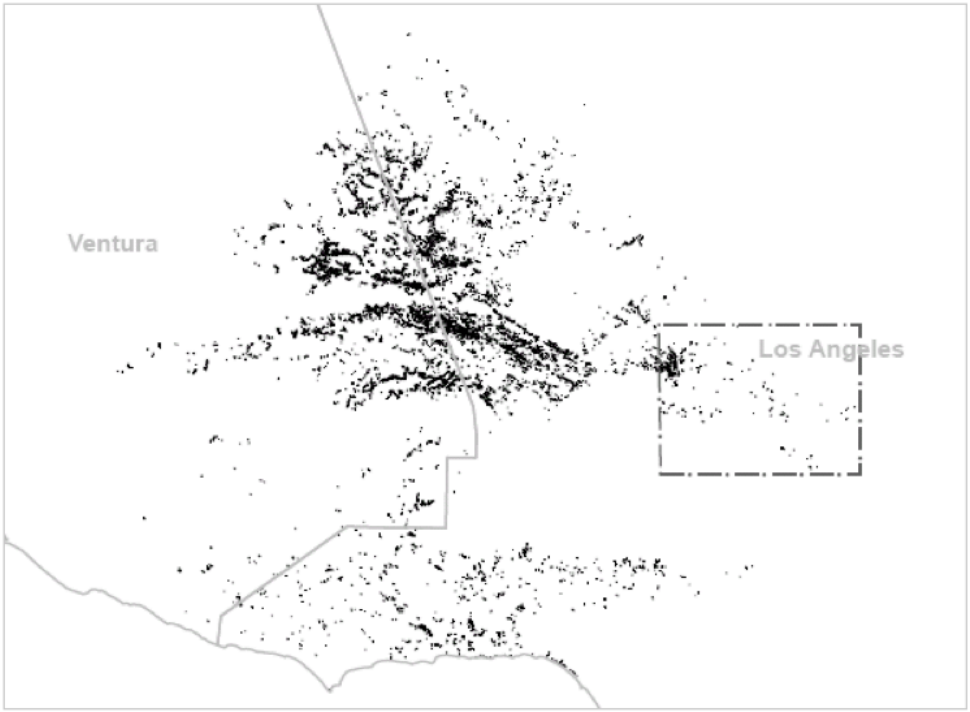

Figure 10a: Northridge inventory of ~11,000 landslides (shown in black). Area of 1971 San Fernando landslide inventory shown with dashed line.

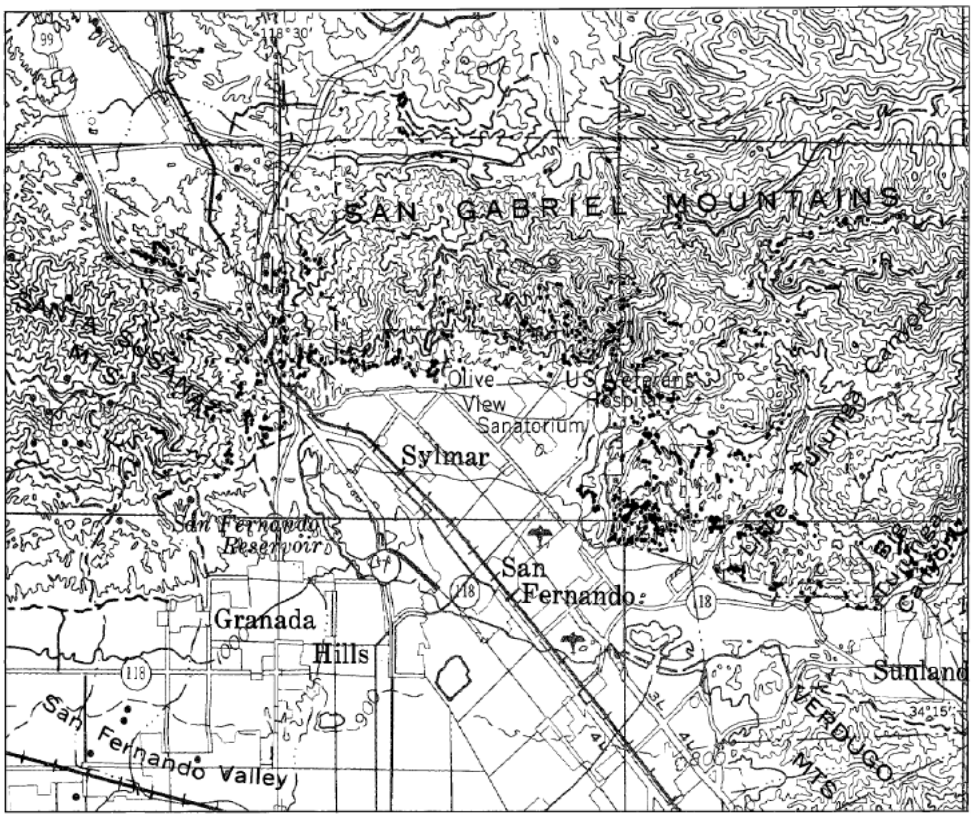

Figure 10b: Mapped Landslide locations, represented as black circles, for the 1971 San Fernando Event (Grantz 1971).

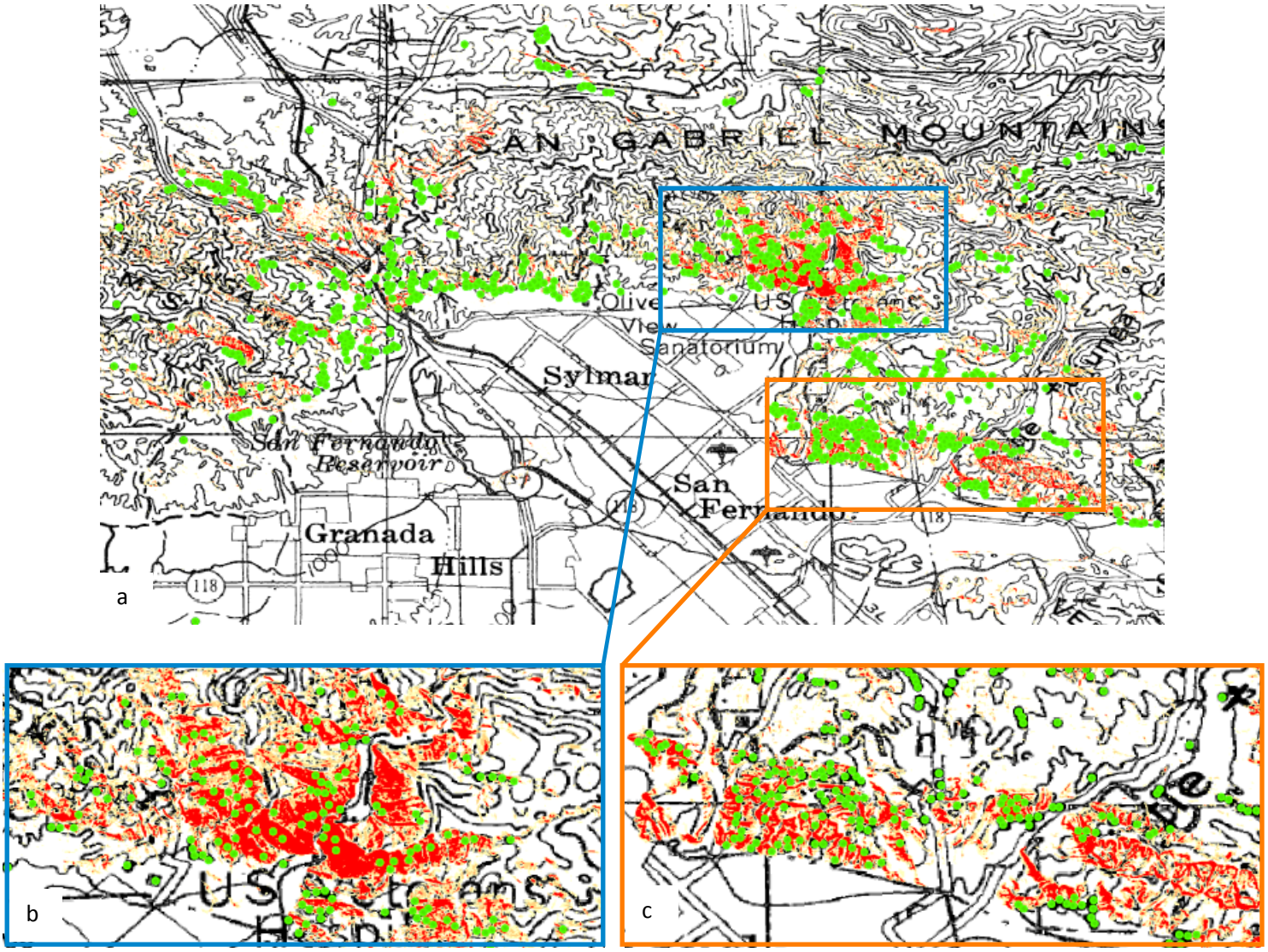

Figure 11: Predicted probabilities of landsliding of 75% to 100% (in red) for the Northridge event shown with the location of actual mapped landslides (green circles).

Assessment of the accuracy of the predications for the model as applied to the San Fernando earthquake mesoseismal area was based on observation. The predicted areas of high probability

of landsliding generally agree with areas of mapped landslide locations. In some portions of the San Fernando study area, the general shape of the predicted areas of high probability of failure match the shape clusters of observed landslide locations; however, they are offset slightly. This could potentially be attributed to the fact that the landslide locations for the San Fernando Event were not available in electronic format. They were manually digitized from a scanned image contained within Grantz, 1971. In addition, the source map used in digitizing was a small excerpt of a historic quadrangle map. Few coordinates were indicated on the map, and the coordinate system for the map was not known with certainty. An effort was made to identify coordinates for points on the map from USGS quadrangles available at the USGS Map Locator and Downloader. However, potential distortion in the scanning process and human error resulting from had digitizing of the landslide locations could have contributed in some of the discrepancy between mapped and predicted areas of landslide initiation.

Some over-prediction of high susceptibility areas are noted in Figure 11b when compared with locations of mapped landslides. However, it is not clear how complete the San Fernando earthquake-induced landslides database is because according to Grantz, 1971, very little field verification of landslide locations was performed and cloud cover obscured some areas in the northeast portion of the study area.

DISCUSSION AND CONCLUSIONS

Earthquake induced landslides often number in the thousands to tens-of-thousands and extend over regions of 10,000 km^2 or more. Due to their potentially large volumes, high velocities, long travel distances, and impact forces, the consequences of seismically induced landslides can be severe. These consequences routinely include injuries, property losses, and damage to critical infrastructure systems such as transportation, water supply, and telecommunication networks. As such, resilience to natural hazards is a critical component of sustainability for metropolitan regions. Reducing community vulnerability to geologic hazards requires that the potential impacts of hazards be quantified so effective measures can be taken to reduce risk.

Currently, the majority of landslide hazard maps are generated based on methods that rely primarily on slope angle and detailed strength parameters, which could be costly and difficult to obtain over large areas, for use in the evaluation. Because the results are highly influenced by slope angle, current hazard mapping procedures tend to predict high probability of failure over large areas characterized by steeply sloping terrain. This work is the first step in development of new low-cost, high-resolution tools to forecast the occurrence and consequences of geologic hazards over regional scales. The model described herein has shown to be capable of forecasting the likelihood and spatial distribution of earthquake-induced landsliding over an area of ~5,000 km^2 and requires relatively simple input consisting of widely available geophysical parameters, thus making this approach equally-suited for both industrialized and developing nations. By identifying and including controlling factors in addition to slope angle in the evaluation (i.e.

slope aspect, topographic position, geologic unit, and PGA), the model is able to differentiate between landslide susceptibility within areas having similar slope inclination. In addition, validation of the model developed with a seismically-induced landslide database from the 1994 Northridge event with one from the 1971 San Fernando event suggests that the model is transferrable to areas of similar terrain and geologic formations.

Research related to refinement and improvement of this model is ongoing. Modifications to the categorization of geologic units are currently underway such that they will be more general and will be based on engineering characteristics of the units, thus making the model more widely applicable to areas outside of the study area. A digital database of strike and dip data for the study area has been developed. This strike and dip data will be incorporated into future generations of the model if determined to be statistically significant to model prediction. Refinement to the "study area," an improved metric for representation of topographic position, and other minor adjustments are also planned.

Potential applications of this statistical model include:

- Estimation of regional losses resulting from landslides for scenario earthquake events.
- Development of seismic hazard zone maps, such as those currently being developed by the California Geologic Survey.
- Support of post-event emergency response and recovery though integration with real-time seismic monitoring networks such as PAGER: Prompt Assessment of Global Earthquakes for Response.
- Identification of geographically distributed infrastructure (e.g. roads, canals, communication and electric networks) susceptible to damage or destruction from movement of landslide-related debris.
- Delineation of areas with significant earthquake-induced sediment production, which represents an important secondary non-seismic slope hazard.
- Landslide forecasting for public policy analyses.

Our vision for future directions of related work include:

- Cross-validation for landslides in a given region based on other triggering events.
- Examination of the transferability of models to other regions and geologic settings.
- Modeling and simulation of other geologic hazards such as precipitation induced landslides, subsidence, and sinkholes.

- Integration with regional-scale hazard and risk codes such as HAZUS (by FEMA), thus allowing fuller consideration of landslide consequences on networked infrastructure and urban systems, and

- Modeling updating and evolution based on newly developed landslide databases.

REFERENCES

Ayalew, L., Yamagishi, H., Ugawa, N. (2004). "Landslide susceptibility mapping using GSI-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan." Landslides, 1, 73-81.

Ayalaw, L., Yamagishi, H. (2005). "The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan." Geomorphology, 65, 15-31.

Ayalew, L., Yamagishi, H., Marui, H., Kanno, T. (2005). "Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications." Engineering Geology, 81, 432-445.

Baeza, C., Corominas, J. (2001). "Assessment of Shallow Landslide Susceptibility by Means of Multivariate Statistical Techniques." Earth Surface Processes and Landforms, 26, 1251-1263.

Brenning, A. (2005). "Spatial prediction models for landslide hazards: review, comparison, and evaluation." Natural Hazards and Earth System Sciences, 5, 853-862.

Carrara, A., Crosta, G., Frattini, P. (2003). "Geomorphological and Historical Data in Assessing Landslide Hazar." Earth Surface Processes and Landforms, 28, 1125-1142.

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.

Castellanos Abella, E. A., Van Westen, C. J. (2007). "Generatioin of a landslide risk index map for Cuba using spatial multi-criteria evaluation." Landslides.

Chang, K., Chiang, S., Hsu, M. (2007). "Modeling typhoon- and earthquake- induced landslides in a mountainous watershed using logistic regression." Geomorphology, 89, 335-347.

Chau, K. T., Chan, J. E. (2005). "Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island." Landslides, 2, 280-290.

Clerici, A., Perego, S., Tellini, C., Vescovi, P., (2006). "A GIS-based automated procedure for landslide susceptibility mapping by the Conditional Analysis method: the Baganza valley case study (Italian Northern Apennines)." Environmental Geology, 50, 941-961.

Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Dhakal, S., Paudyal, P. (2008). "Predictive modelling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of-evidence." Geomorphology, 102, 496-510.

Dhakal, A. S., Amada, T., Aniya, M. (2000). "Landslide Hazard Mapping and its Evaluation Using GIS: An Investigation of Sampling Schemes for a Grid-Cell Based Quantitative Method." Photogrammetric Engineering & Remote Sensing, 66, 981-989.

Dai, F. C., Lee, C. F. (2003). "A Spatiotemporal Probabilistic Modelling of Storm-Induced Shallow Landsliding Using Aerial Photographs and Logistic Regression." Earth Surface Processes and Landforms, 28, 527-545.

Ercanoglu, M., Gokceoglu, C., Van Asch, T. W. J. (2004). "Landslide Susceptibility Zoning North of Yenice (NW Turkey) by Multivariate Statistical Techniques." Natural Hazards, 32, 1-23.

Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L. (2006). "Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium)." Geomorphology, 76, 392-410.

Eeckhaut, M. V. D., Reichenbach, P., Guzzetti, F., Rossi, M., Poesen, J. (2009). "Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium." Natural Hazards and Earth System Sciences, 9, 507-521.

Eeckhaut, M. V. D., Hervas, J., Jaedicke, C., Malet, J.P., Montanarella, L., and Nadim, F. (2011). "Statistical Modelling of Europe-wide landslide susceptibility using limited landslide inventory data." Landslides, 9(3), 357- 369.

Garcia-Rodriguez, M. J., Malpica, J. A., Benito, B., Diaz, M. (2008). "Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression." Geomorphology, 95, 172-191.

Harp, E. and Jibson, R. (1996) "Landslides triggered by the 1994 Northridge, California, earthquake," Bulletin of the Seismological Society of America, 86, S319-S332

He, Y., Beighley, E. (2007). "GIS-based regional landslide susceptibility mapping: a case study in southern California." Earth Surface Processes and Landforms.

Hosmer, D.W. and Lemedhow, S. (2000). Applied Logistic Regression. Wiley, New York, USA.

Işik, Yilmaz. (2008). "Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey)." Computers & Geosciences.

Jibson, R., Harp, W., Edwin, M. and Michael, J. (2000) "A method for producing digital probabilistic seismic landslide hazard maps," Engineering Geology, 58, 271-289.

Lee, S., Choi, J., Min, K. (2002). "Landslide susceptibility analysis and verification using the Bayesian probability model." Environmental Geology, 43, 120-131.

Lee, S., Choi, J. (2004). "Landslide susceptibility mapping using GSI and the weight-of-evidence model." International Journal of Geographical Information Science, 18, 789-814.

Lee, S., Ryu, J., Min, K., Won, J. (2003). "Landslide susceptibility analysis using GIS and artificial neutral network." Earth Surface Processes and Landforms, 28, 1361-1376.

Lee, S. (2004). "Application of Likelihood Ratio and Logistic Regression Models to Landslide Susceptibility Mapping Using GIS." Environmental Management, 34, 223-232.

Lee, S., Talib, J. A. (2005). "Probabilistic landslide susceptibility and factor effect analysis." Environmental Geology, 47, 982-990.

Lee, S., Ryu, J., Kim, I. (2006). "Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea." Landslides.

Lee, S. (2007). "Comparison of landslide susceptibility maps generated through multiple logistic regression for three test areas in Korea." Earth Surface Processes and Landforms, 32, 2133-2148.

Lee, S., Pradhan, B. (2007). "Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models." Landslides, 4, 33-41.

Lee, C. T., Huang, C. C., Lee, J. F., Pan, K. L., Lin, M. L., Dong, J. J. (2008). "Statistical approach to earthquake-induced landslide susceptibility." Engineering Geology, 100, 43-58.

Lee, C. T., Huang, C. C., Lee, J. F., Pan, K. L., Lin, M. L., Dong, J. J. (2008). "Statistical approach to storm event-induced landslides susceptibility." Natural Hazards and Earth System Sciences, 8, 941-960.

Luzi, L., Pergalani, F., Terlien, M. T. J. (2000). "Slope vulnerability to earthquakes at subregional scale, using probabilistic techniques and geographic information systems." Engineering Geology, 58, 313-336.

Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H. (2008). "An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps." Engineering Geology, 97, 171-191.

Mathew, J., Jha, V. K., Rawat, G. S. (2009). "Landslide susceptibility zonation mapping and its validation in part of Garhwal Lesser Himalaya, India, using binary logistic regression analysis and receiver operating characteristic curve method." Landslides, 6, 17-26.

Marques, R., Zezere, J. L., Queiroz, G., Coutinho, R. (2007). "GIS-based logistic regression method for susceptibility assessment of earthquake-triggered landslides: A case study from Fogo Volcano (S. Miguel, Azores)." Geophysical Research Abstracts, 9.

Miles, S. and Ho, C. (1999) "Rigorous landslide hazard zonation using Newmark's method and stochastic ground motion simulation," Soil Dynamics and Earthquake Engineering, 18,305-323

Miles, S. and Keefer, D. (2009) Evaluation of CAMEL - comprehensive areal model of earthquake-induced landslides, Engineering Geology, 104,1-15

Ramos, M. C., Cots-Folch, R., Martinez-Casanovas, J. A. (2007). "Effects of land terracing on soil properties in the Priorat region in Northeastern Spain: A multivariate analysis." Geoderma, 142, 251-261.

Süzen, M. L., Doyuran, V. (2004). "A comparison of the GIS based landslide susceptibilitiy assessment methods: multivariate versus bivariate." Environmental Geology, 45, 665-679.

Süzen, M. L., Doyuran, V. (2004). "Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey." Engineering Geology, 71, 303-321.

Santacana, N., Baeza, B., Corominas, J., Paz, A. D., Marturia, J., (2003). "A GIS-Based Multivariate Statistical Analysis for Shallow Landslide Susceptibility Mapping in La Pobla de Lillet Area (Eastern Pyrenees, Spain)." Natural Hazards, 30, 281-295.