Engineering Geology 173 (2014) 54–65

Contents lists available at ScienceDirect

Engineering Geology

journal homepage: www.elsevier.com/locate/enggeo

Development of a globally applicable model for near real-time prediction of seismically induced landslides

M. Anna Nowicki [a,b,*], David J. Wald [b], Michael W. Hamburger [a], Michael Hearne [b], Eric M. Thompson [c]

[a] Indiana University, Department of Geological Sciences, 1001 East 10th Street, Bloomington, IN 47405, United States
[b] U.S. Geological Survey Geologic Hazards Science Center, P.O. Box 25046, DFC, MS 966, Denver, CO 80225, United States
[c] San Diego State University, Department of Geological Sciences, 5500 Campanile Dr, San Diego, CA 92182, United States

ARTICLE INFO

Article history:
Received 9 August 2013
Received in revised form 28 January 2014
Accepted 3 February 2014
Available online 12 February 2014

Keywords:
Landslides
Earthquakes
Rapid response
Hazards

ABSTRACT

Substantial effort has been invested to understand where seismically induced landslides may occur in the future, as they are a costly and frequently fatal threat in mountainous regions. The goal of this work is to develop a statistical model for estimating the spatial distribution of landslides in near real-time around the globe for use in conjunction with the U.S. Geological Survey (USGS) *Prompt Assessment of Global Earthquakes for Response* (*PAGER*) system. This model uses standardized outputs of ground shaking from the USGS *ShakeMap* Atlas 2.0 to develop an empirical landslide probability model, combining shaking estimates with broadly available landslide susceptibility proxies, i.e., topographic slope, surface geology, and climate parameters. We focus on four earthquakes for which digitally mapped landslide inventories and well-constrained *ShakeMaps* are available. The resulting database is used to build a predictive model of the probability of landslide occurrence. The landslide database includes the Guatemala (1976), Northridge (1994), Chi-Chi (1999), and Wenchuan (2008) earthquakes. Performance of the regression model is assessed using statistical goodness-of-fit metrics and a qualitative review to determine which combination of the proxies provides both the optimum prediction of landslide-affected areas and minimizes the false alarms in non-landslide zones. Combined with near real-time *ShakeMaps*, these models can be used to make generalized predictions of whether or not landslides are likely to occur (and if so, where) for earthquakes around the globe, and eventually to inform loss estimates within the framework of the *PAGER* system.

Published by Elsevier B.V.

1. Introduction

Seismically induced landslides present costly and often deadly threats in many mountainous regions. Approximately 5% of all earthquake-related fatalities are caused by seismically induced landslides, in some cases causing a majority of non-shaking deaths (Marano et al., 2009). Substantial effort has been invested to understand where such landslides may occur in the future. Though some regional efforts have succeeded, no uniformly agreed-upon method is available for predicting the likelihood and spatial extent of seismically induced landslides. This study uses deterministic estimates of the ground motion from earthquake events (e.g., peak acceleration and velocity) produced by the U.S. Geological Survey (USGS) *ShakeMap* system (Allen et al., 2008; Garcia et al., 2012), combined with broadly available landslide susceptibility proxies (such as topographic slope and surface geology) to build a predictive model of the probability of landslide occurrence at a given location. The approach that we apply to landslides is similar to the strategy that Zhu et al. (in press) applied to liquefaction. We apply a logistic regression analysis (Peng et al., 2002) to a series of training events with well constrained ground shaking and landslide distribution data, which provide empirical observations to the model. The performance of the regression model is assessed with both statistical goodness-of-fit metrics and a qualitative review of the model's capability to capture the spatial extent of landslides for each training event, as well as for one test event (Niigata-Chuetsu, Japan, 2004) that was not used in the regression model. Combined with near real-time *ShakeMaps*, the model may be used to make generalized predictions of whether or not (and if so, where) landslides are likely to occur for earthquakes around the globe. The long-term goal is to incorporate this functionality into the USGS *Prompt Assessment of Global Earthquakes for Response* (*PAGER*) system (Earle et al., 2009).

2. Modeling methodology

The main objective of landslide hazard modeling is to predict areas prone to landslides either spatially or temporally (Brenning, 2005); here, we focus on short-term prediction of the spatial pattern of landslides triggered by an individual earthquake. Common approaches include both slope-stability methods, which incorporate physical models of landslide susceptibility, and statistical approaches, based on

* Corresponding author at: Indiana University, Department of Geological Sciences, 1001 East 10th Street, Bloomington, IN 47405, United States.
E-mail addresses: manowick@indiana.edu (M.A. Nowicki), wald@usgs.gov (D.J. Wald), hamburg@indiana.edu (M.W. Hamburger), hearne@usgs.gov (M. Hearne), ethompson@mail.sdsu.edu (E.M. Thompson).

http://dx.doi.org/10.1016/j.enggeo.2014.02.002
0013-7952/Published by Elsevier B.V.

empirical observations of landslide occurrence. Among the physical models the most prominent are methods relying on "Newmark displacement calculations," as developed by Newmark (1965), based on a simplified physical model representing the force balance of a landslide block sliding over an inclined slide plane. Studies applying this slope-stability method to landslide predictions include Jibson (1993), Jibson et al. (2000), and Kaynia et al. (2011). Statistical approaches to modeling seismically induced landslides have also been applied in many studies, including Jibson (2007), García-Rodríguez et al. (2008), Felicísimo et al. (2012), and Li et al. (2012). Both mechanical and statistical methods typically involve creation of a long-term landslide susceptibility map as their end product, and are usually focused on a small region where data are available at a relatively fine resolution. Both Li et al. (2012) and Kaynia et al. (2011) incorporate outputs from the *ShakeMap* system in their models, with their end product being a long-term landslide susceptibility map for their study areas.

Logistic regression is the dominant statistical technique currently used to predict landslides in the literature (Rennie, 2003). Logistic regression is appropriate for a process involving only a binary outcome (in this case, slide or no slide), and allows the observed outcomes to be fitted to the logistic function using data representing multiple predictor variables. The logistic function is given by

$$\mathrm{Logit(P)} = \ln(p/(1-p)) = a + bx_{1+}cx_2 + dx_3 + \ldots, \tag{1}$$

where x_1, x_2 and x_3 are the explanatory variables; and a, b, c, and d are the coefficients determined in the regression. The probability of a particular outcome is represented as

$$p(t) = 1/(1 + \exp(-z)), \tag{2}$$

where $z = a + bx_1 + cx_2 + dx_3 + \ldots$.

The Akaike Information Criterion (AIC) is a statistical measure used to represent the goodness of fit of a model to the data with which it was trained:

$$\mathrm{AIC} = -2 * \ln(m) + 2 * (n) \tag{3}$$

where m represents the maximum likelihood term and n the number of parameters in the model. The AIC value is dependent on the fit of the model estimates to the data (given by the maximum likelihood term), and the number of parameters the model contains. The AIC value decreases with better model fit, but increases with complexity of the model. In searching for the best fitting model for a given suite of parameters we then use the lowest AIC value (Akaike, 1987).

Comparisons of statistical methods previously used to model landslide hazards (Brenning, 2005) concluded that logistic regression results in the lowest rate of error. This study therefore uses logistic regression analysis as the main method for modeling seismically induced landslides.

Although the method of logistic regression has been widely used, a majority of the literature offers no broader geographic application of the technique than those developed within local projects, such as a model applicable to an area surrounding a city or within one country. The input of the shaking hazard varies from project to project, while two end products are commonly associated with these studies—either a landslide susceptibility map for a region based on shaking due to a particular earthquake, or a long-term landslide susceptibility map based on probable exposure to a specified amount of seismic shaking. Despite alternate end goals, many times no method is provided to test how well the model is performing. This study seeks to address these large gaps in the literature in order to present a globally applicable, short-term, deterministic estimate of the likelihood of landslides associated with a particular event.

2.1. Predictor variables

The spatial distribution of seismically induced landslides is dependent on certain physical characteristics of the area in which they occur. Empirical studies suggest that the bedrock lithology, slope, seismic intensity, topographic amplification of ground motion, fracture systems in the underlying bedrock, groundwater conditions, and also the distribution of pre existing landslides all have some impact on the landslide distribution, among other factors (Keefer, 2002). Predictor variables to test in the regression are chosen based on the relationships shown in previous studies, as well as the availability of global datasets that can be used as proxies for each of the studied variables.

Based on these guidelines, we include the following predictor variables in the regression: ground motion produced by the earthquake, topographic slope, material strength, and soil wetness. These variables will be compared with the spatial distribution of mapped landslides that occurred due to shaking produced in that particular event.

2.1.1. Landslide data

Mapped landslide data are available from various researchers in the field of earthquake-induced landslide research. Multiple methods are used to map landslides; these include field-based mapping of observed landslide deposits, and remote sensing techniques (such as mapping landslide deposits from satellite images). As our study uses an empirical model, a small number of case histories in the landslide literature are used to build the database of landslide observations that are used in the regression. These events have been selected based on the high quality and availability of data for these events (see Garcia et al. (2012) for further detail on case history characteristics). We focus first on using the Wenchuan, China (2008) earthquake, with data from the Chi-Chi, Taiwan (1999), Northridge, California (1994), Niigata-Chuetsu, Japan (2004), and Guatemala (1976) events incorporated during later stages of model testing. Table 1 provides a brief overview of each earthquake included in the database. The inventories are classified as complete or comprehensive, where complete indicates that all of the landslides were mapped only for a specified study area, and comprehensive indicates that all of the landslides were mapped that exceed a specified size. The estimated extent of shaking for each event is

Table 1
Summary of earthquake-induced landslides.

Earthquake	Date	Year	Time (UTC)	M	Depth (km)	EQ type	Inventory type	No. of landslide obs.	% (of cells) w/ landslide	Fatalities	Landslide fatalities	Ref.
Guatemala	4-Feb	1976	9:01	7.5	12.3	SS	C	6212	28.2	23,000	Unknown	1
Northridge (California)	17-Jan	1994	12:30	6.7	19	T	C*	11,111	20.1	61	Unknown	2
Chi-Chi (Taiwan)	20-Sep	1999	17:47	7.7	21	T	C*	9272	14.9	2465	78	3
Chuetsu (Japan)	23-Oct	2004	08:56	6.6	16	T	C	4615	76.7	68	6	4
Wenchuan (China)	12-May	2008	06:28	7.9	19	T	C*	197,481	12.0	87,633	20,000	5
Global								228,691	13.1	113,227	20,084	

M=Magnitude. Earthquake type: SS = Strike-slip, T = Thrust.
Inventory type: C = Complete; C* = Comprehensive.
References: 1: Harp et al. (1981); 2: Harp and Jibson (1996); 3: Liao and Lee (2000); 4: Sekiguchi and Sato (2006); 5: Dai et al. (2011).
Fatalities as reported in the Global Earthquake Model Earthquake Consequences Database (GEM ECD; http://gemecd.org).