THE QUANTITATIVE ESTIMATION OF EARTHQUAKE-INDUCED GROUND FAILURE HAZARDS IN THE SAN FERNANDO VALLEY, CALIFORNIA

A THESIS
SUBMITTED TO
THE DEPARTMENT OF GEOLOGICAL AND ENVIRONMENTAL SCIENCES
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

AMY ELISABETH CRAVEN
AUGUST 1997
Approved for the Department:

W. Gary Ernst

Paul Switzer

Richard Bernknopf
Abstract

Given that large populations reside in areas subject to strong ground motion, ways to live with earthquake hazards must be found. Scientists understand factors which control earthquake effects, but this knowledge is rarely in a form that can be used to promote seismic safety. This thesis provides a model for translating geological influences on ground failure into a quantitative estimate of hazard. A multivariate statistical approach, which relies heavily on GIS capabilities, is used to produce probabilistic maps of ground failure potential in the San Fernando Valley, CA. An unparalleled dataset, which pinpoints the locations of ground failures induced by the 1994 Northridge earthquake, is present in the form of damaged sewer pipes, water pipes, and streets. These precisely located ground failure sites are related to geologic, hydrologic, topographic, and seismic factors which are known to influence ground failure. Predictor variables include: age of the geologic unit, average percent clay of the geologic unit, average shear wave velocity, slope, depth to shallow ground water, and peak ground velocity from the Northridge earthquake. All the variables are compiled in a GIS and converted into grids with 1 hectare cells. Each cell constitutes an observation for the regression analysis. Using a multivariate regression model, the probability of ground failure in the San Fernando Valley is specified alternatively as:

\[
\text{probability} = \Phi \{1.9522 - 0.00673 \text{ water depth} + 0.0123 \text{ clay} - 0.549 \ln(\text{shear wave velocity})\};
\]

or, \[
\text{probability} = \Phi \{-0.00858 \text{ water depth} + 0.0169 \text{ clay} - 0.354 \ln(\text{shear wave velocity}) + 0.0133 \text{ ground velocity}\}.
\]

The first equation provides a probabilistic estimate of the susceptibility of ground failure since only site parameters are used; the second equation describes the ground failure potential during a specific earthquake because susceptibility factors are combined with a seismic parameter. Age and slope variables are dropped from the final equations because they are not statistically significant in conjunction with the included variables which were all significant at the 99% confidence level. The coefficients make sense in terms of geologic processes. When probabilities are mapped, the spatial distribution matches expectations. In addition, the probability distributions correspond to actual ratios of failed cells to total cells. The development of these probabilistic hazard maps is a substantial improvement over traditional hazard maps because they provide quantitative estimates of
ground failure hazard. Quantitative results are desirable because they facilitate communication between scientists and policy makers, are useful for risk assessment, and lend themselves to economic evaluation. Probabilistic maps of seismic hazards have broad implications for hazard mitigation, land-use zonation, and emergency response planning.
Acknowledgments

This thesis would not have been possible without the support of several individuals at Stanford University and at the United States Geological Survey. Rich Bernknopf provided the initial project ideas and offered encouragement during the entire study. I am indebted to Gary Ernst for his guidance and support not only with this project but throughout my two years at Stanford University. Paul Switzer contributed much needed statistical insight, always offering great suggestions when I was perplexed. I thank Lori Dinitz for helping me understand difficult statistical concepts and explaining the workings of ARC/INFO grids and LIMDEP regressions. More importantly, I thank Lori for sharing hours of laughter and fun in the office. Both John Tinsley and Dan Ponti kindly spent countless hours with me, offering geologic expertise and thoughtful discussions about the science behind my models. John provided digital maps delineating the depth to ground water and the age and grain size of the geologic units in the San Fernando Valley. Dan made the sewer pipe, water pipe, and street repair data available to me. I also thank Caren Chaika and Raphael Thomadsen for numerous conversations about statistical modeling. Other individuals who provided support include Dave Wald, Art Tarr, and Mike Bennett. This material is based upon work supported under a National Science Foundation Graduate Fellowship. Any opinions, findings, conclusions, or recommendations expressed in this thesis are those of the author and do not necessarily reflect the views of the National Science Foundation.
Table of Contents

Abstract .................................................................................................................. iii
Acknowledgments.................................................................................................... v
Table of Contents.................................................................................................... vi
List of Tables .......................................................................................................... ix
List of Illustrations ................................................................................................. x
List of Plates ........................................................................................................... xi

Chapter One: Introduction ..................................................................................... 1

Chapter Two: Earthquake Hazards ....................................................................... 4
GROUND SHAKING .......................................................................................... 5
Predicted Shaking Intensity Maps .......................................................................... 6
SURFACE FAULTING ......................................................................................... 7
GROUND FAILURE ........................................................................................... 7
Liquefaction .......................................................................................................... 7
Liquefaction Hazard Evaluation ............................................................................ 10
  Geologic and Hydrologic Criteria .................................................................... 10
  Geotechnical Methods ...................................................................................... 12
  Stochastic Methods ......................................................................................... 14
Landslides ............................................................................................................. 14
Landslide Hazard Evaluation ............................................................................... 15
  Susceptibility under Static Conditions .......................................................... 15
  Susceptibility under Dynamic Conditions .................................................... 15
SUMMARY .............................................................................................................. 17

Chapter Three: The Study Area .......................................................................... 18
TECTONIC SETTING ......................................................................................... 18
THE SAN FERNANDO VALLEY .......................................................................... 21
THE NORTH RIDGE EARTHQUAKE ............................................................... 23
FUTURE EARTHQUAKES ................................................................................... 24
Chapter Six: Discussion ................................. 63
Comparision with Other Hazard Studies .............. 63
Applicability of Models ................................... 64
Uses of the Hazard Maps ................................... 65

Chapter Seven: Conclusions .............................. 66

References .................................................. 69
List of Tables

1. Type of sewer pipe breaks and their frequency................................. 33
2. Detailed descriptions of the damaged sewer and water pipes................. 34
3. Geologic units and their corresponding age ranges............................ 41
4. Percent clay assigned to the geologic units .................................... 42
5. Average shear wave velocity ($V_s$) for alluvial deposits ..................... 46
6. Average shear wave velocity ($V_s$) of Tertiary and older units ............. 47
7. Average depth to ground water ..................................................... 49
8. Descriptive statistics of dependent and independent variables .............. 54
9. Results of the susceptibility model .............................................. 55
10. Results of the earthquake-specific model ...................................... 58
11. Descriptive statistics of probability models .................................... 58
12. Mean squared error for both models ........................................... 61
List of Illustrations

1. Vertical section showing a lateral spread before and after failure .................. 9
2. Vertical section showing the ground oscillation mode of liquefaction ............ 9
3. Curves showing particle sizes which are most liquefiable and potentially liquefiable ................................................................. 11
4. Method of evaluating liquefaction potential ............................................. 13
5. Relationship between penetration resistance and cyclic stress ratio ............... 13
6. Plots of critical acceleration vs. slope angle for different lithologic categories . 16
7. Earthquake displacement curves showing landslide movement vs. critical acceleration ................................................................. 16
8. Generalized geology of the Los Angeles region .......................................... 19
9. Plate tectonic evolution of coastal California ............................................ 20
10. Map of the San Fernando Valley ............................................................ 22
11. The probability distribution of $\varepsilon$ .................................................... 28
12. Normal and logistic cumulative distribution functions ................................ 28
13. Map of the state of the sewer pipes after the Northridge earthquake ............ 35
14. Histogram of sewer pipe breaks ............................................................. 36
15. Map of the locations of water pipe breaks and street repairs ...................... 37
16. Histogram of water pipe breaks ............................................................. 38
17. Histogram of street repairs ................................................................. 38
18. Diagram showing the particle size classification system used by CALTRANS 43
19. Depth to ground water map ................................................................. 50
20. Relative histogram of probabilities for model one ..................................... 56
21. Comparison of actual failures with predicted probabilities for model one .... 56
22. Relative histogram of probabilities for model two .................................... 59
23. Comparison of actual failures with predicted probabilities for model two ..... 59
List of Plates

1. Age of Geologic Units in the San Fernando Valley, California
2. Percent Clay in the San Fernando Valley, California
3. Average Shear Wave Velocity in the San Fernando Valley, California
4. Slope of the San Fernando Valley, California
5. Depth to Ground Water in the San Fernando Valley, California
6. Modeled Peak Ground Velocity from the 1994 Northridge Earthquake
7. Probability of Ground Failure in the San Fernando Valley, California: Susceptibility Version
8. Probability of Ground Failure in the San Fernando Valley, California: Earthquake-Specific Version
CHAPTER ONE: INTRODUCTION

Earthquakes, touted as unpredictable causes of death and destruction throughout the ages, can be explained and modeled employing recent scientific and technological advances. Scientists understand many aspects of earthquake triggers and the factors which influence earthquake effects such as shaking, faulting, and ground failure. Questions remain, however, regarding the best way to specify and communicate seismic risk.

Traditional studies either use site-specific geological and geotechnical investigations to make deterministic (yes/no) decisions about hazards, or they take a regional approach and assign qualitative ratings based on certain factors. Site-specific investigations are extremely limited spatially, only considering the area of the borehole, which is a few centimeters in diameter. If enough of these boreholes are drilled, then an accurate determination of hazard for a certain building site can be made, but such investigations are labor-intensive and expensive. Since detailed studies cannot be done everywhere, regional, qualitative approaches are often used. These studies are informative, but qualitative estimates of hazard are not easily incorporated into decisions about seismic safety. Consider an analogy; if you were about to undergo surgery and the doctor said that you have a medium rather than a high chance of survival, wouldn't you want more information? You would ask for actual numbers and data describing success in past procedures. You would want to know the difference between medium and high ratings. In short, you would want quantitative results.

The value of detailed, quantitative information about earthquake hazards cannot be understated. Scientific information is available but is often too vaguely defined to be incorporated into seismic safety planning (Olshansky et al., 1991). Quantitative estimates of hazard, on the other hand, lend themselves to risk communication and risk assessment. When different levels of hazard for different locations are specified in numeric terms, there is little confusion about the meaning of hazard categories. Decision rules for mitigation requirements can be considered. Quantitative estimates can be combined with property values to determine dollars at risk, so that cost-effective safety strategies can be devised. These are only a few examples of the many uses of quantitative information about earthquake hazards.

This thesis produces quantitative estimates of ground failure hazard using a method developed by previous workers (Bernknopf et al., 1988, 1993; Mark, 1992; Bernknopf
and Soller, 1994; Pike et al., 1994). This cross-disciplinary method combines information about predictor factors to produce probabilistic estimates of hazard. Geologic knowledge is required to understand the physical processes affecting the hazard and to know what predictor factors to include. Computer expertise is necessary because all the relevant variables are compiled and manipulated within a Geographic Information System (GIS). Statistical knowledge is needed to run a multivariate regression model—the means by which information about predictor factors is translated into probabilistic estimates of hazard.

The specific earthquake hazard this study delineates is ground failure, defined as permanent ground deformation resulting from seismic activity but not including disruption owing to surface faulting mechanisms. Ground failure thus defined can result from a variety of processes, such as liquefaction, compaction, landsliding, or shear failures. History has shown that ground failure can be a major cause of damage and destruction. It was ground failure which damaged all the water pipes during the 1906 San Francisco earthquake and left the city powerless to fight the devastating fires which ensued (Youd, 1978). Flow failures induced by the 1920 Kansu, China earthquake killed an estimated 200,000 people (Youd, 1978), and ground failure was responsible for 60% of the estimated $300 million (1964 value) damage resulting from the 1964 Alaska earthquake (Youd, 1978). Obviously, ground failure is an earthquake hazard which demands the attention of seismic safety policies.

Seismically active areas which are densely urbanized have the greatest potential for earthquake damage and casualties, and thus are in desperate need of hazard assessments. The Los Angeles region, home to 9.2 million people and almost 100 active faults, fits into this category. Accordingly, the San Fernando Valley, the northern part of the Los Angeles metropolitan region and the site of two moderate (M = 6.7) earthquakes (1971 San Fernando and 1994 Northridge) within the last 26 years, is the study area for this thesis.

The 1994 Northridge earthquake, the most recent and perhaps the best studied earthquake in the United States, induced widespread ground failure. More than 11,000 landslides within a 10,000 km² area were triggered, and ground failures on level ground were observed up to 58 km from the epicenter. Ground failures in the San Fernando Valley alone caused over 80,000 sewer pipe breaks, caused over 600 water pipe breaks, and necessitated the repair of streets at over 500 locations. The locations of these damages provide an extremely detailed dataset with which to study ground failure occurrence. These observed ground failures are compared to geologic, hydrologic, topographic, and seismic
variables using regression analysis. The final products of the analysis are maps which display quantitative estimates of ground failure potential in the San Fernando Valley.

This thesis describes the methodology of translating earth science information into quantitative estimates of ground failure hazard. Chapter 2 reviews the physical processes affecting earthquake hazards and discusses the techniques commonly used to evaluate them. Shaking and surface faulting, as well as ground failure, are considered for completeness and because evaluation techniques useful for one hazard type can be useful for others. Chapter 3 summarizes the tectonic setting of southern California, describes the geology of the San Fernando Valley, and discusses details of the Northridge earthquake. Chapter 4 describes the methodology of producing the ground failure maps. The use of regression analysis and GIS is explained, and detailed descriptions of the variables incorporated into the model are given. Chapter 5 presents the results of the hazard models, in terms of equations and probability maps. Chapter 6 compares the models to other hazard studies and discusses future applications of the hazard maps. Finally, Chapter 7 provides conclusions and suggestions for future work.
CHAPTER TWO: EARTHQUAKE HAZARDS

It was not long ago that earthquakes were considered spatially and temporally intractable. Presently, a knowledge of tectonics, mapped faults, and historical seismicity allow the identification of broad hazard regions. These hazard regions can be further delineated by the use of geologic, topographic, and hydrologic parameters. In this way, it is possible to predict the locations where seismic hazards are likely.

In recent years, temporal predictions have been attempted. For example, the Working Group on California Earthquake Probabilities (1988, 1990, 1995) has forecast earthquake probabilities in California for 30 year time spans. These studies identify fault segments expected to produce large earthquakes and estimate the time until the next earthquake on a particular segment. Median recurrence times (T) are calculated using $T = \frac{D}{V}$, where $D$ is the estimated displacement along a fault segment and $V$ is the estimate of a long-term slip rate along a fault segment. Historical earthquake catalogs are used to determine the date of most recent movement on a particular segment. A lognormal probability density function is then used to determine the probability of earthquake recurrence. While these studies are intriguing, they confirm only that certain areas are within hazard zones. Their actual probability estimates are questionable since they include only faults on which data are available, the uncertainties involved in calculations are underrated (Bolt, 1991), and the use of fault slip rates is potentially flawed (Schell, 1991). For this reason, this thesis will focus on predicting the spatial distribution rather than the temporal aspects of earthquake hazards.

Earthquake effects can be divided into three categories: shaking, surface faulting, and ground failures. This section provides a general overview of these hazards and also discusses the techniques used to determine areas at risk. This is by no means an exhaustive review of existing literature but is intended to put the current hazard model in its proper context. It should be noted that these hazard studies have the potential to affect public safety, since seismic considerations have become an ever increasing component of California law. The Alquist-Priolo Special Studies Zones Act of 1972 mandates that human occupancy structures should not be built within 50 feet of an active fault. More recently, the Seismic Hazards Mapping Act of 1990 calls for the delineation of regions (Special Study Zones) at risk for strong ground shaking, liquefaction and other ground failure, landslides, and other seismic hazards caused by earthquakes (Tobin, 1991). Since
hazard studies have far-reaching implications, it is imperative that they are accurate and easy to apply as policies.

GROUND SHAKING

In terms of dollar value, ground shaking is the most damaging of all earthquake effects because energy radiated from the source can affect a widespread area. Energy released during an earthquake travels in the form of high-frequency body waves (compressional and shear waves) and low-frequency surface waves (Rayleigh and Love waves). Shear waves are the most damaging because buildings are most vulnerable to their horizontal motion (Hays, 1981). In addition, surface waves can be particularly damaging because they are efficient at causing tall buildings to vibrate and they decay less rapidly than high-frequency waves (Hays, 1981).

A single earthquake has a range of associated shaking intensities, usually measured by the modified Mercalli intensity scale (Richter, 1958). Factors determining the severity of ground shaking at a particular location can be divided into three categories: source effects, path effects, and site effects. Source effects depend on the earthquake magnitude, the length of fault rupture, the amount of displacement, and the type of faulting (Holden and Real, 1990). Obviously, the greater the area of fault rupture, the larger the amount of energy released in an earthquake. This also causes a greater area to be exposed to shaking effects. The larger the length of a fault, typically the larger the expected earthquake on that fault will be. This logic has been used to determine characteristic earthquakes for certain faults, and hence is the basis for hazard maps using scenario earthquakes (discussed in the following section). The severity of shaking is also related to the type of faulting; as reaffirmed by the Northridge event, earthquakes on thrust faults typically produce higher ground accelerations than those on strike-slip faults (USGS Staff, 1996).

Dominant path effects involve the attenuation of energy with increasing distance from the fault and directivity of energy propagation. Empirical relationships which quantify the decrease in intensity with increasing distance from the causative fault have been developed (Borcherdt et al., 1975). In addition, models which account for the focusing of energy along the fault in the direction of rupture have been proposed (Perkins and Boatwright, 1995). Such models maintain that the intensity of shaking decreases with distance from the fault and that it decreases more rapidly in locations perpendicular to the fault, in the simplified case of a strike-slip fault.
Site effects, related to topography, basin structure, or the nature of surficial soil, can be very important in the amplification or attenuation of seismic waves. Minor hills can amplify seismic waves. The interactions between seismic waves and basin structure can focus energy, thereby increasing shaking intensities at some sites, while diminishing intensities at other sites (USGS Staff, 1996). Depending on the nature of the surficial deposits, the intensity of shaking can be amplified or reduced. The combination of site amplification and inadequate construction was responsible for the destruction of about 400 tall buildings in Mexico City, located 380 km from the epicenter of the 1985 Mexico earthquake (Hays, 1986).

**Predicted Shaking Intensity Maps**

The following is a brief sampling of studies which predict shaking intensities employing a GIS database. Often, the studies must rely on a scenario earthquake since actual earthquake occurrences cannot be forecast. Thomson and Evernden (1986) predicted shaking intensities for a repeat of the 1906 San Francisco earthquake in San Mateo County, CA. A mathematical model for earthquake source effects and calculations for attenuation and site geologic effects were used. Perkins (1987) produced a series of intensity maps, one for each potential fault in the San Francisco Bay area, following Borcherdt et al. (1975). These intensity maps were then weighted, using recurrence intervals and damage expectations to create maps of the cumulative damage potential that could be expected. Borcherdt et al. (1991) used strong and weak motion data and a geologic map to define amplification capability groups in the San Francisco Bay region. These groups were combined with earthquake opportunity maps to produce maps which give the potential for shaking intensities to exceed a specified level. This map can be used to delineate Special Study Zones which require more detailed investigation. Finally, Perkins and Boatwright (1995) developed a model relating earthquake parameters, such as magnitude, distance from source, directivity, and geologic materials, to shaking intensities. They tested their intensity model by comparing the number of actual red tagged housing units (those buildings deemed unsafe to enter) from past San Francisco Bay area earthquakes to the number of red tagged units that the model predicted.

Such predicted shaking intensity maps are useful in that they give an idea of the range of intensities which possibly could affect a particular site. Given this information, design requirements for buildings may be determined based on the maximum shaking intensity they should withstand. Having these maps in a GIS format makes them easy to
manipulate and update as new information is obtained. One drawback to such maps is that they are often limited to the specifications of the chosen earthquake scenario.

SURFACE  FAULTING

Although tales of people getting swallowed up by the earth during an earthquake abound in films and novels, the actual effects of surface faulting are much less dramatic. Surface faulting, the differential movement of two sides of a fracture, often causes damage to structures and to lifelines, and thus indirectly leads to injuries and casualties. Fault displacements in the United States range from a fraction of an inch to more than 20 feet per single event, the length of rupture ranges from less than a mile to more than 200 miles, and fault zones typically range from 6 ft to 1000 ft in width (Bonilla, 1981).

The best way to reduce losses related to surface faulting is to 1) identify active faults and 2) avoid building near active faults or at least appropriately design structures which must straddle faults. The Alquist-Priolo Special Studies Zones Act, which mandates that human occupancy structures should not be built within 50 feet of an active fault in California, is an example of such a loss-reduction measure. This Act has been successful in that it has established seismic safety as an element in the land-use decision-making process (Holden and Real, 1990).

GROUND  FAILURE

Ground failure, defined here as permanent ground deformation resulting from seismic activity, can be caused by numerous processes, such as soil compaction, shear deformation, liquefaction, and landslides. Liquefaction and landslides will be discussed here because these processes are the best studied and the most frequently evaluated. Other types of ground failure, such as shear failure in soft clays, also occurred during the Northridge earthquake and will be discussed in Chapter 3.

Liquefaction

Liquefaction is defined as the "transformation of cohesionless material from a solid state to a liquefied state as a consequence of increased pore pressure and reduced effective stress" (Youd, 1978). Liquefaction is often induced by earthquakes. Seismic shear waves disrupt grain to grain contacts, leaving water in void spaces little time to escape. This
increase in pore-water pressure results in a loss of strength in the soil, eventually transforming the soil from a solid into a viscous fluid. The process of liquefaction often results in ground failure which may be categorized into four types: lateral spread, ground oscillation, loss of bearing strength, and flow failure (Youd, 1978, 1984).

Lateral spreads involve the horizontal movement of soil layers at the surface as a result of a liquefied layer at depth (Figure 1). These movements usually occur on gentle slopes (typically 0.3° to 3°) and may be enhanced by proximity to a free face such as a river channel because the liquefied mass is not laterally confined (Tinsley et al., 1985). Horizontal displacements commonly amount to several meters but may be as much as tens of meters if conditions are favorable (Youd and Keefer, 1981). Although lateral spreads typically are not perceived as particularly dangerous, they may cause extensive damage. Youd (1978) notes that lateral spreads from the 1906 earthquake were responsible for every major pipeline break in San Francisco, and the disruption of these pipelines allowed the ensuing fires to go unchecked and destroy a significant portion of San Francisco.

Ground oscillation occurs in very level ground which is not near a free face. Cracks develop between blocks of sediment when seismic waves cause the sediment layers to vibrate at different frequencies (Figure 2). This mode of ground failure results in smaller displacements (15-20 cm) than lateral spreading (Obermeier, 1996).

Loss of bearing strength occurs whenever soil liquefies, but it is perceived as a problem only when a structure is present to demonstrate the effect. The loss of strength which accompanies liquefaction enables differential settlement and may cause buildings to tilt as in the case of the 1964 Niigata, Japan earthquake, by as much as 60° (Youd, 1978). In addition, if buried structures are buoyant, they may rise upward; the 1964 Niigata, Japan earthquake caused buried septic tanks to rise as much as one meter (Tinsley et al., 1985).

Flow failures are basically landslides of severely liquefied sediment. These failures occur on slopes greater than 3°. This mode of failure is the most destructive of all the categories. Displacements of several tens of meters or even several tens of kilometers are common and can occur at velocities up to tens of kilometers/hour (Tinsley et al., 1985). Submarine flow failures were responsible for the loss of the port facilities of Seward, Whittier, and Valdez Alaska during the 1964 earthquake (Youd, 1978), and flow failures
Figure 1. Vertical section showing a lateral spread before and after failure. Liquefaction occurs in the zone marked by diagonal lines. The surface layer then moves horizontally toward the free face and divides into blocks separated by fissures. Sand is vented at some locations. Triangle points to water table. Originally from Youd (1984) and modified by Obermeier (1996).

Figure 2. Vertical section showing the ground oscillation mode of liquefaction. The zone marked by diagonal lines liquefies and vibrates at a different frequency than the surrounding sediment, causing fissures to form. Ground oscillation occurs on level ground away from breaks of slope. Triangle points to water table. Originally from Youd (1984) and modified by Obermeier (1996).
triggered by the 1920 Kansu, China earthquake killed an estimated 200,000 people (Youd and Keefer, 1981).

Various authors have defined factors known to influence liquefaction potential (Seed and Idriss, 1971; Tinsley et al., 1985; Youd and D.M. Perkins, 1987; Obermeier, 1996). The dominant geologic and hydrologic factors include: grain size, relative density, age of the sediment, and depth to water. In general, clean sands to silty sands are most susceptible to liquefaction, although some gravels, silts, and sensitive clays are known to liquefy (Figure 3). The looser the sediment, the easier it is to build up pore-water pressure and initiate liquefaction. Age is an important factor since older deposits often are more dense and thus more difficult to liquefy. Sediment must be water saturated to liquefy, so the depth to water is very important. Liquefaction susceptibility decreases as depth to water increases because 1) overburden pressure increases with depth and 2) age, cementation, and compactness of sediment tends to increase with depth (Tinsley et al., 1985). Even if liquefaction does occur at depth, the displacement at the ground surface is small and hence does not pose a safety hazard.

Liquefaction Hazard Evaluation

Given that the factors influencing liquefaction potential are known, it should be possible to delineate areas at risk. Methods of evaluation include 1) the identification of susceptible zones based on mappable geologic and hydrologic conditions, 2) geotechnical measures to distinguish liquefiable sediment, and 3) stochastic models which predict the probability of susceptibility to liquefaction.

Geologic and Hydrologic Criteria

Current hazard maps often distinguish between liquefaction susceptibility, liquefaction opportunity, and liquefaction potential maps (Youd and Perkins, 1978). Liquefaction susceptibility maps delineate areas which contain units which could liquefy on the basis of geologic and hydrologic factors. Liquefaction opportunity maps are a function of a region's seismicity and the rate of occurrence of shaking sufficient to produce liquefaction (Youd and Perkins, 1978). When superimposed, these two maps produce a liquefaction potential map, which delineates areas where liquefiable sediments are present and where an earthquake large enough to liquefy the sediment could occur.
Figure 3. Curves showing particle sizes which are most liquefiable and potentially liquefiable. From Tsuchida and Hayashi (1971) in Obermeier (1996).
The technique of estimating liquefaction susceptibility based on the geologic character of the area was first proposed by Youd and Perkins (1978). They suggested using the type of deposit (the geologic environment of deposition), the general distribution of cohesionless sediment, and the age of sediments to estimate the susceptibility of a particular region to liquefaction. A liquefaction opportunity map is produced by using locations of known seismic sources and the maximum distance from these sources within which liquefaction can occur. Finally, liquefaction susceptibility and opportunity maps are combined to produce liquefaction potential maps.

Youd and J.B. Perkins (1987) applied these techniques to San Mateo County, CA. Geologic units were deemed susceptible based on lithology and age; Standard Penetration Test (SPT) data\(^1\) were used as a measure of the cohesiveness of the unit, and depth to water data were used to determine if the sediment was saturated. For each map unit, the probability that each of the three factors (geology, lack of cohesion, and saturation) is favorable to liquefaction was estimated. Tinsley et al. (1985) employed a similar method, using age and lithology of deposit, SPT data, and depth to water to assign qualitative susceptibility rankings (very high, high, moderate, low, and very low) to mapped units. More recently, Real et al. (1996) delineated hazard zones as those where ground water is above 40 feet and Holocene alluvium, Holocene river deposits, or artificial fill is present.

**Geotechnical Methods**

Perhaps the most widely used procedure for the site-specific evaluation of liquefaction potential is that developed by Seed and Idriss (1971). This method involves the comparison of the shear stresses induced by an earthquake to the stresses required to induce liquefaction, thereby defining a zone where liquefaction might be expected to occur. (Figure 4).

The cyclic stress ratio, the "ratio of the average cyclic shear stress developed ... as a result of cyclic or earthquake loading to the initial vertical effective stress acting on the sand layer before the cyclic stresses were applied" (Seed et al., 1983), is used to describe the stress induced by an earthquake. This ratio incorporates information about the depth of the soil layer, the depth to water, the intensity of ground shaking, and other cyclic loading phenomena. This earthquake-induced cyclic stress ratio is then compared to the stress

---

\(^1\) The Standard Penetration Test (SPT) procedure measures the number of blows (of a 140 lb hammer) falling freely through a height of 30 inches, required to drive a standard sampling tube 1 foot into the ground (ASTM, 1983). Thus SPT data, reported as N values (resistance, measured in blows/foot) describe the looseness of a particular sediment.
Figure 4. Method of evaluating liquefaction potential. The stress induced by earthquake motions is compared to the stress required to cause liquefaction to define a zone at depth where liquefaction could occur. From Seed and Idriss (1971).

Figure 5. Relationship between penetration resistance and cyclic stress ratio. From Seed et al. (1983).
ratio required to initiate liquefaction, which is determined using 1) field observations of the performance of similar sediments in previous earthquakes, or 2) laboratory data of representative samples (Seed et al., 1983). Since undisturbed representative samples are often difficult to obtain, field observations are often used. Soil index parameters, such as Standard Penetration Resistance, Cone Penetration Resistance, electrical properties, and shear wave velocity can be related to past liquefaction incidents. Typically, the correlation between Standard Penetration Resistance and field liquefaction behavior is used (Figure 5). After the expected cyclic stress ratio is calculated, the Standard Penetration Resistance can be measured at the site of interest. Depending on the blow count (N) of the sediment in question, liquefaction may or not be expected in the future.

Stochastic Methods

While traditional ways of evaluating liquefaction are useful in delineating hazard zones, it is not obvious how to translate qualitative susceptibility rankings into hazard policies, and it is not feasible to perform detailed geotechnical analyses at every site. It is for these reasons that workers recently have developed probabilistic liquefaction maps. Ostadan et al. (1991) describe a method to estimate liquefaction potential using empirically derived equations (Liao et al., 1988) relating SPT data, site acceleration, and shaking duration to the probability of liquefaction. Juang and Elton (1991) used fuzzy set theory to estimate liquefaction susceptibility quantitatively. Pike et al. (1994) created probabilistic maps of lateral spread susceptibility, estimated by threshold criteria and a multivariate regression; percent sand, distance from a stream, age of the geologic unit, and slope were used as predictors of liquefaction susceptibility in their analysis.

Landslides

Landslides, the downslope movement of material under gravitational influence, are often triggered by earthquakes. The 1994 Northridge earthquake caused more than 11,000 landslides within an area of about 10,000 km² (Harp and Jibson, 1996). Not only are such landslides damaging and widespread, but they can be extremely dangerous; more than 18,000 people were killed as a result of a single rock avalanche that was triggered by the 1970 Peruvian earthquake (Youd and Keefer, 1981).

For a slope to be stable, the forces available to resist sliding must be greater than the forces which induce sliding. Resistance is largely controlled by the material of a hillside; pristine bedrock units are less likely to fail than clays or weakly cemented, weathered
rocks. Vegetation also provides stability. The slope and geometry of the hillside is another important factor. Common events which increase the stress on a hillside include earthquake shaking and heavy precipitation. Whether a particular slope fails depends on the balance between factors resisting sliding and those increasing the stress on a slope.

Landslide Hazard Evaluation

Susceptibility under Static Conditions

Before considering the complexity inherent in proposing triggering mechanisms such as earthquakes or heavy precipitation, it is instructive to examine slope stability under static conditions. Areas susceptible to landslides may be identified on the basis of relevant information such as the topography, hydrology, and geology of the region. Within a GIS framework, Real et al. (1996) grouped spatial geologic and topographic data into material-strength and slope-gradient categories. Where weak materials intersected high slope gradients, hazard zones were identified. Mejia-Navarro et al. (1994) used a GIS and a series of algorithms relating topography, hydrology, precipitation, geology, soils, vegetation, and land-use to calculate the debris flow susceptibility of various locations.

Other workers have used statistical methods to combine criteria relevant to slope stability. Bernknopf et al. (1988) used a multivariate regression model to delineate areas where landslide hazards are a concern. Slope, soil shear strength, and the presence of recent construction were used as explanatory variables; the locations of historical landslides were used as the dependent variable. Mark (1992) used a logistic regression to predict the probability of debris flow initiation in the future. The distribution of debris flows from a 1982 storm in San Mateo County, CA was related to physical factors (slope, precipitation, hillside material, and vegetation).

Susceptibility under Dynamic Conditions

Once the stability of a slope under static conditions is estimated, the effect of an earthquake trigger may be examined. Newmark (1965) developed theoretical equations which can be used for dynamic slope stability assessments. His methods were found to accurately describe actual seismically induced slope failures (Wilson and Keefer, 1983) and hence, have been applied to further analyses (Wilson and Keefer, 1985; Wieczorek et al., 1985). These studies involve two steps. First, static slope stability is examined using topographic, geologic, and hydrologic conditions to determine the critical acceleration required to initiate movement (Figure 6). Second, earthquake displacement curves,
Figure 6. Plots of critical acceleration \( (a_c) \) versus slope angle for different lithologic categories of San Mateo County, CA. Dashed lines indicate wet conditions, solid lines indicate dry conditions. From Wieczorek et al. (1985).

Figure 7. Earthquake displacement curves showing landslide movement versus critical acceleration \( (a_c) \) for several earthquakes in the San Mateo County, CA region. Patterned area shows upper and lower boundary limits for design earthquakes. From Wieczorek et al. (1985).
relating landslide movement to critical acceleration for various scenario earthquakes are constructed (Figure 7). Given the minimum tolerable displacement (5 or 10 cm), the acceleration expected for a specific earthquake can be determined from Figure 7. This expected acceleration is then compared to the critical acceleration required to generate movement in order to assign qualitative slope stability categories (Wieczorek et al., 1985).

Static techniques combine relevant factors (usually present in existing maps) to produce estimates of the stability of slopes. This can be taken one step further by comparing critical acceleration to the acceleration expected in an earthquake. Landslide hazard maps provide knowledge of where avoidance or mitigation efforts should be focused. Stochastic methods, in particular, give quantitative results which can be used in cost/benefit analyses for land-use and mitigation decisions (Bernknopf et al., 1988; Bernknopf and Soller, 1994).

SUMMARY

Much is known about earthquake-induced hazards and the processes which control them. Accordingly, factors which influence hazards can be identified and combined to create hazard maps. A variety of techniques have been developed which combine these factors in different ways. Methods range from qualitative studies to mathematical algorithms to stochastic models. Statistical analyses are favored since they provide quantitative estimates of hazards which account for uncertainties in the analysis. Such techniques translate what scientists know into a form that can provide regional guides for seismic zonation and safety policies. Besides providing a general background of earthquake hazards and methods of evaluation, this chapter highlights ground failure as a potentially damaging phenomenon which is worthy of further study. The next chapter will discuss earthquake hazards in the context of the study area (the San Fernando Valley) and the study earthquake (the 1994 Northridge earthquake).
CHAPTER THREE: THE STUDY AREA:
SAN FERNANDO VALLEY, LOS ANGELES COUNTY, CA

TECTONIC SETTING

The Los Angeles metropolitan region (Figure 8) sits astride a zone of deformation related to the boundary between the Pacific and North American lithospheric plates. As such, this area is the host of almost 100 active faults (Ziony and Yerkes, 1985) and the site of 38 moderate to large earthquakes in the historical period2 (Toppozada, 1995). Additional effects of deformation include mountain building, basinal development, deformation of upper Quaternary terrace deposits, and regional uplift and strain accumulation (Yerkes, 1985).

Southern California has been the site of tectonic activity for millions of years (Figure 9). Before 30 million years ago (Ma), the Farallon oceanic plate was subducting beneath the North American continental crust-capped plate (Atwater, 1970). At about 29 Ma, the East Pacific Rise collided with the subduction zone, and transform movement along the evolving plate boundaries was initiated. Transform slip occurred on faults west of the modern San Andreas until 5 Ma (Yerkes, 1985), when the right-lateral transform boundary moved inland. Since this time, coastal California has been sliding northwestward along the San Andreas fault system at a rate of approximately 50 mm/year (California Working Group on Earthquake Probabilities, 1995).

A geologic map of the Los Angeles region exhibits not only the northwest trending strike-slip faults of the San Andreas system, but also the east-west trending thrust faults and folds of the Transverse Ranges (Figure 8). These faults and folds formed in response to the convergence caused by the big bend in the San Andreas fault, where the fault changes its northwest trend, parallel to plate motion, to a more westerly trend, at an angle of 30° to plate motion (California Working Group on Earthquake Probabilities, 1995). The Transverse Ranges province is one of the most active in the United States, with regard to seismicity and crustal mobility (USGS Staff, 1996). Earthquakes originating within this province involve faults which can be mapped at the surface (1971 San Fernando M = 6.7 and 1952 Kern County M = 7.5), or blind thrust faults (1987 Whittier Narrows M = 5.9 and 1994 Northridge M = 6.7). The presence of blind thrust faults is particularly

---

2 This refers to earthquakes identified since 1769 of M≥6 within 160 km of Los Angeles and M≥5 within 50 km of Los Angeles.
Figure 8. Generalized geology of the Los Angeles region. From Greenwood (1995).
Figure 9. Plate tectonic evolution of coastal California. From Dickinson (1981).
disturbing because they pose significant dangers but are unknown with regard to location, geometry, past behavior, and seismic potential.

THE SAN FERNANDO VALLEY

The Los Angeles region is a sprawling metropolis, including not only Los Angeles proper, but the surrounding areas of the San Fernando Valley, the San Gabriel Valley, and parts of Orange County. The densely urban San Fernando Valley, site of two moderate earthquakes in the last 26 years, is the study area for this thesis.

The San Fernando Valley and its surrounding mountains (Figure 10) are a part of the Transverse Ranges physiographic province. The 1200 km² valley is an alluvial basin, surrounded by the tectonically uplifted rocks of the Santa Monica Mountains to the south, the Simi Hills to the west, the Santa Susana and San Gabriel Mountains to the north, and the Verdugo Mountains to the east. These mountains provide the source for the Pleistocene and Holocene alluvium which fill the structural basin. The Holocene alluvium is generally 5 m to 8 m thick (Tinsley, 1997). The thickness of the Pleistocene alluvium is known to be highly variable, ranging from thousands of feet to nonexistence. The nature of the Holocene alluvium is dependent on its provenance. Sediment of the eastern half of the valley (east of Interstate 405) is derived from the granitic and metamorphic rocks of the San Gabriel Mountains and the Verdugo Mountains; therefore, this alluvium tends to be coarse textured, consisting primarily of sand, gravel, and cobbles (Gibbs et al., 1996). Water can easily pass through this sediment, so the water table is typically greater than 50 feet deep in the eastern portion of the valley (Tinsley et al., 1985). Sediment of the western half of the valley is derived from the sandstones, siltstones, and mudstones of the Santa Monica and Santa Susana Mountains and the Simi Hills (Gibbs et al., 1996), so the western valley is predominantly fine-grained. Clayey alluvium is present in floodplain deposits, and sand, silty sand, and minor gravel appear in buried stream channels (Tinsley et al., 1985). As a consequence of the reduced permeability of these fine-grained deposits, the water table in the western valley tends to be close to the surface (Tinsley et al., 1985). These characteristics of the valley are important to keep in mind since they relate to ground failure susceptibility and partly controlled the sites of damage in the Northridge earthquake.
The Northridge earthquake struck the Los Angeles region at 4:31 am PST on January 17, 1994. The hypocenter was 17.5 km beneath the city of Northridge in the San Fernando Valley (Figure 10). The earthquake was caused by the rupture of a previously unknown blind thrust fault, striking N58°W and dipping 42°S (Wald and Heaton, 1994). The rupture propagated upward and northwestern along the fault plane from a 17.5 km depth to a 5 to 6 km depth; the rupture area measured approximately 15 x 20 km (USGS Staff, 1996).

In terms of the earthquake hazards discussed in Chapter 2, the Northridge earthquake caused strong shaking and widespread ground failure but no surface faulting. The Los Angeles region was subjected to strong shaking for 10 to 20 seconds. Peak ground accelerations of nearly 1.8g were recorded at a Tarzana site, and several instruments recorded peak ground accelerations over 0.9g. Shaking effects were felt over an area of more than 200,000 km² in the United States and Mexico; even Las Vegas, Nevada, 375 km away from the epicenter felt the earthquake (USGS Staff, 1996). Occurrences of ground failure were prevalent. More than 11,000 landslides over a 10,000 km² area were triggered (Harp and Jibson, 1996), and ground failures on level ground were observed up to 58 km from the epicenter (Stewart et al., 1996).

USGS scientists have conducted detailed studies at four sites of ground failure in the San Fernando Valley and have linked the failures to two primary mechanisms: liquefaction and shear failure in weak clays (Holzer et al., 1996). The inference of liquefaction is based on the presence of saturated, liquefiable sediment in regions of ground failure, rather than the observation of vented liquefied material³ (Stewart et al., 1996; Holzer et al., 1996). At a site where shear failure is suspected, the only 'weak link' in the sediment column is a layer of saturated clay (Holzer et al., 1996). If unconsolidated, saturated weak clays are subjected to high enough accelerations, they can fail (Tinsley, 1997). By extension, one could argue that other ground failures which occurred in the valley were caused by the same mechanisms of liquefaction and shear failure in weak clays, believed to be responsible for failure at the four sites which were studied.

³ Stewart et al., 1996 note the presence of some sand boils near the Los Angeles River, but such occurrences are rare and not found in connection with most regions of ground failure.
The Northridge earthquake was one of the most, if not the most expensive disaster in United States history. Estimated damages range from $20 billion (USGS Staff, 1996) to $45 billion (McCarthy, 1997). Fifty seven people died, more than 9,000 people were injured, and more than 20,000 people were displaced from their homes (USGS Staff, 1996). Thousands of sewer, water, and oil pipelines were damaged due to the widespread ground failure.

Even though damages were severe, the situation could have been much worse. The early morning timing of the earthquake was fortuitous; few people were on the freeways, some of which collapsed, or in the buildings which were damaged. In addition, the seismic energy was directed and focused along the fault plane beneath the more sparsely populated regions north of the San Fernando Valley. Thus, the majority of the people and buildings of the San Fernando Valley were spared the full force of the shaking (USGS Staff, 1996).

FUTURE EARTHQUAKES

There is a common belief that seismic risk decreases in the aftermath of an earthquake. While this may be true for great earthquakes such as the 1906 San Francisco event, it is not the case for present-day southern California. Since Northridge was only a moderate sized earthquake, it was insufficient to significantly decrease the accumulated strain in the region (Hauksson, 1995). In addition, Dolan et al. (1995) note that during the past 210 years of historical events, there have been no large earthquakes\(^4\) (\(M_w > 6.7\)) and conspicuously few moderate earthquakes (only 1971 San Fernando and 1994 Northridge earthquakes) occurring on faults within the metropolitan region. They conclude that faults within the region capable of generating large earthquakes have a collective recurrence interval of 140 years, considerably less than the 210 years of historical quiescence (Dolan et al., 1995). Models developed by the California Working Group on Earthquake Probabilities (1995) predict an 80% to 90% probability of a \(M \geq 7\) earthquake within southern California before 2024. All of this information indicates that the Los Angeles metropolitan region needs to prepare for a large earthquake in the future—perhaps the very near future.

\(^4\) Dolan et al. are referring only to earthquakes occurring within the Los Angeles metropolitan region. Large earthquakes (\(M_w > 6.7\)) have occurred in the area surrounding Los Angeles region and have been felt by Los Angeles (i.e. 1857 Fort Tejon \(M = 7.8\) and 1952 Kern County \(M = 7.5\)).
CHAPTER FOUR: METHOD

This thesis uses a multivariate statistical approach to predict ground failure potential within the San Fernando Valley. The null hypothesis is that ground failure occurrences are random; the alternative hypothesis is that specific geologic, hydrologic, and topographic attributes of the valley or specific seismic factors of an earthquake scenario can be identified which are significant in predicting ground failure locations. These predictor variables are selected based on an understanding of the physical processes which may cause ground failure. Locations of sewer pipe and water pipe breaks and street damages are used as indicators of ground failure occurrence. Data describing the predictor factors and earthquake damage locations are compiled in the form of digital maps. Within a GIS framework, the information contained in these maps can be georeferenced, manipulated, and analyzed. The map corresponding to each variable is converted from a vector to a raster form, with a grid cell size of 1 hectare. Each one hectare cell, containing information about the predictor factors and damage occurrences, serves as a single observation for the analysis. Finally, a multivariate regression is estimated. This analysis tests the significance of each predictor variable, estimates the weight of each variable on predicting ground failure, and when converted to a standard normal variate can be used to estimate the probability of ground failure occurring within each cell.

ORIGIN OF THE METHOD

Geologists have been combining information about physical factors to produce estimates of hazard for decades (e.g. Brabb et al. 1972; Tinsley et al., 1985; Youd and J.B. Perkins, 1987). These studies typically superimpose relevant factors to spatially delineate qualitative hazard zones. Another way to combine information is through the use of regression analysis. The method applied here, which relates a binary hazard outcome (the hazard did or did not occur) to relevant earth science information has been developed by previous workers (Bernknopf et al., 1988; Mark, 1992; Bernknopf and Soller, 1994; Pike et al., 1994). Bernknopf et al. (1988) related past landslide locations in the Cincinnati, Ohio area to variables describing slope, soil shear strength, and the presence of recent construction using a binary choice regression. Mark (1992) related mapped debris flows to predictor variables of slope, precipitation, hillside materials, and vegetation to produce estimates of the probability of debris flow initiation in San Mateo County, CA.

---

5 The term georeferencing refers to the registration of various earth science data to a common coordinate system so that spatial analyses can be performed.
Bernknopf and Soller (1994) extended this method to look at earthquake-triggered landslides in the Santa Cruz Mountains of California. They produced probabilistic estimates of landslides in a two-step procedure. First, susceptibility to landsliding was estimated using topographic variables; susceptibility estimates were then combined with earthquake probability estimates of the Working Group on California Earthquake Probabilities (1988) to produce probabilistic maps for the earthquake-triggered landslides. Most recently, Pike et al. (1994) estimated the susceptibility of a portion of the central coast of California to earthquake-induced lateral spreading. This thesis applies and refines the method of these studies to estimate the probability of ground failure in the San Fernando Valley.

TOOLS: REGRESSION ANALYSIS AND GIS

Regression Analysis

There are several reasons for using regression analysis in this study. Most importantly, it provides a way to translate earth science information into a quantitative measure of hazard. Quantitative results are desirable since they facilitate risk communication, are useful for risk assessment, and lend themselves to economic evaluation (Bernknopf et al., 1988, 1993). Regression analysis provides statistical inferences about predictor variables, such as their significance and the weight of their predictive power. In addition, uncertainty is incorporated and quantified in this approach. Stochastic methods can be rapidly implemented since they use existing information and maps, in contrast to labor-intensive geotechnical methods (e.g. Seed and Idriss, 1971). For all of these reasons, regression analysis is seen as a superior method of hazard assessment.

Choosing an Appropriate Model

Multiple linear regression analysis attempts to explain the behavior of a dependent variable in terms of a suite of independent variables. A typical equation can be written as:

\[ y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon, \]

where \( y \) is the dependent variable, the \( x \)'s are the independent variables, \( p \) expresses the number of variables in the model, \( \alpha \) is the intercept term, and the \( \beta \)'s are the coefficients which specify the weight of a particular \( x \) in predicting \( y \). \( \epsilon \) is an error term which
represents all unexplained variations in y, caused by important but omitted variables or by unexplainable random phenomena. For every value of \( x_i \), there exists a probability distribution of \( \epsilon \) and therefore a probability distribution of \( y \) (Figure 11). A model is fit to the data such that the sample variation in \( y \) can be attributed to using the \( x \)'s to predict \( y \). The method of ordinary least squares is commonly used to fit the data to a model.

In this study, the geologic, hydrologic, topographic, and seismic factors are the independent variables, and the ground failure occurrences specify the dependent variable. The predictor variables are continuous quantities, but the form the dependent variable takes is not as clear. The earthquake damages could be represented as frequencies (the number of pipe breaks or street repairs within a given cell) or as a dichotomous variable indicating whether ground failure did or did not occur within a given cell. The latter option is chosen because 1) the occurrence, not the frequency of ground failure is of interest and 2) it is unclear how to quantize ground failure frequencies.

Conventional regression methods are inappropriate when the dependent variable has a discrete outcome, so an alternative model must be sought. Three options are the linear probability model, the probit model, and the logit model. The linear probability model takes the form \( y_i = \beta_i x_i + \epsilon_i \) for \( i = 1 \) to \( n \) observations; \( y_i \) is estimated using the method of ordinary least squares. Problems with this model occur because \( \epsilon_i \) can only take two values and hence cannot be normally distributed, \( \epsilon_i \) is heteroscedastic, and probabilities outside the 0-1 interval can be obtained (Greene, 1990). The last of these problems is a serious flaw, which necessitates the use of another model.

Both the probit and logit models are alternatives. They lead to probabilities that are confined to the unit interval by transforming the original model using a cumulative probability function. The probit model uses a standard normal distribution function, while the logit model uses the logistic cumulative distribution function. The difference between the two models is slight (Figure 12), so the choice is often based on personal preference or computational convenience (Greene, 1990). A probit model is used here since it has thinner tails than the logit model. This makes the null hypothesis harder to reject for each coefficient, thereby making the model as conservative as possible.

**Mechanics of the Probit Model**

Numerous statistical texts describe the probit model (Greene, 1990; Gujarati, 1988;
Figure 11. The probability distribution of $\epsilon$. From Mendenhall and Sincich (1993).

Figure 12. Normal and logistic cumulative distribution functions. From Greene (1990).
Maddala, 1983). The summary given here is drawn largely from Pindyck and Rubinfeld (1981, p. 280-283). The probit model assumes that there exists some unobservable index, $Z$, which is a linear function of the explanatory variables, such that:

$$ Z_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_p x_{pi} + \varepsilon_i, $$

where $i = 1$ to $n$, and $n$ is the number of observations, and $p$ expresses the number of variables in the model. $\varepsilon_i$ is assumed to be a standard normal random variable, so it follows that $Z_i$ is a continuous random variable which is normally distributed. In this study, $Z_i$ is estimated using the maximum likelihood technique (see Mood et al., 1974, p. 276-286 for details). The theoretical $Z$ index is related to the ground failure dependent variable ($y$) through the probit model. The probability of $Z$ leading to ground failure ($y = 1$) or no ground failure ($y = 0$) can be computed using the cumulative normal probability function, which is written:

$$ P_i = F(z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} e^{-\frac{s^2}{2}} ds, $$

where $s$ is a random variable which is normally distributed with a zero mean and unit variance. $P_i$ represents the probability of ground failure occurring and is measured by the area under the standard normal curve from $-\infty$ to $Z_i$. $P_i$ is confined to the unit interval and thus avoids the problem of the linear probability model.

**GIS Implementation**

Geographic Information Systems can be defined as "a technology designed to capture, store, manipulate, analyze, and visualize the diverse sets of georeferenced data that are required to support accurate modeling of the earth's environmental processes" (Goodchild et al., 1993). Much of the power of a GIS comes from its ability to manage spatial data and to link those data to nonspatial attributes. This allows for the compilation, manipulation, and analysis of diverse datasets with ease. Many workers have realized this power and have employed GIS in seismic hazard studies (Brabb, 1995; Real et al., 1996; King and Kiremidjian, 1994; Wentworth et al., 1991; Pike et al., 1994).

**Elements of a GIS**

GIS can be divided into four basic components which allow efficient task completion: data input, data management, data manipulation and analysis, and data output (see Star and Estes, 1990). The following describes these components and how they were used in this study.
Data Input

This is often the most difficult and time-consuming step in any GIS project. Diverse datasets are needed, and they must be converted from their original form into a digital form that the GIS can employ. Original forms of the data describing the dependent and independent variables included paper maps, digital maps, ASCII files, raster files, and tables. Paper maps were digitized, existing digital, ASCII, and raster files were converted to appropriate formats, and tabular data were linked to geographic coordinates or spatial elements.

Data Management

Data management is central to the GIS functionality. It provides for efficient storage and retrieval of the spatial and nonspatial data needed for this project. Spatial data may be stored using vector or raster data structures. Vector data structures represent features using points, lines (arcs), and polygons; these features are linked topologically. Raster data structures represent features as continuous and regular grids, with each cell containing information about a given location. The structure employed depends of the type of data. For example, it is logical to store a network of roads as vectors but an elevation surface as a raster structure. Either of these two structures may be linked to information, or attributes, which are specific to a geographic location. Both vector and raster data structures were used in this study.

Data Manipulation and Analysis

A GIS allows spatial and attribute data from different sources to be merged, manipulated, and analyzed. Maps existing in different projections can be transformed and reprojected. Distances and areas can be measured. The database can be queried to find data matching certain criteria. Statistical analyses, including histograms, correlations, and regressions, can be performed. Finally, new maps can be created using spatial operations. Maps may be overlaid or intersected with one another. Distance operators can specify regions a certain distance away from an object of interest, and neighborhood operators can be used in ways such as deriving a slope map using a roving window on an elevation grid. These functions are only a sampling of GIS capabilities for data manipulation and analysis that were used during the course of this project.

Data Output

Results can be quickly output in numerous forms. Multiple versions of maps on a computer screen can be produced, data can be converted into other digital forms which may
be used in other applications, and maps ranging from page to poster size can be printed. The ease of data output means that products may be viewed not only after project completion but at multiple stages in the development and modeling process. This facilitates new ideas for analysis and explanations for observed patterns.

**Essential Capabilities**

Three GIS capabilities which were paramount to the success of this analysis warrant further emphasis. They include 1) georeferencing, 2) the conversion from vector to raster data structures, and 3) the transformation of spatial data into ASCII files. Georeferencing ensures that all data exist in the same coordinate system; since the results of this analysis depend on the coexistence of physical factors likely to produce ground failure, spatial accuracy is essential. One misregistered map could alter all of the results. Georeferenced vector data then must be converted into a raster format. Grid cells provide a convenient unit of analysis for the regression. Finally, a GIS can output this gridded, spatial data into ASCII files which can be understood by statistical computer programs. Together, these functions ensure the ability to conduct statistical analyses of spatial data.

Two GIS software packages, MapInfo and ARC/INFO, were used in this study. MapInfo was used for the georeferencing and processing of the dependent variable. MapInfo was useful in the early stages since it allows for the easy manipulation and display of data. The more sophisticated gridding capabilities of ARC/INFO were needed later. Therefore, the data describing the dependent variable were converted to ARC/INFO, and the independent variables were all compiled and manipulated in ARC/INFO.

**VARIABLE SELECTION AND PROCESSING**

**Dependent Variable**

The Northridge earthquake caused a great deal of damage to underground pipelines and streets. Damage to these types of structures are good indicators of permanent ground deformation (Stewart et al., 1996). The exact mechanism of ground failure is not known but was probably due to liquefaction, soft clay shear failure, or landsliding. Data describing pipeline and street damage are preferable to inventories of building damage for studying ground failures because they are uncomplicated by the variability of building construction. The locations of sewer pipe, water pipe, and street damages are used as
surrogates for the occurrence of earthquake-induced ground failures, and the occurrence of ground failure is the dependent variable in this study.

Sewer Pipe Damage

Tens of thousands of sewer pipes in the San Fernando Valley sustained damage from the Northridge earthquake. This is not surprising, since most of the sewer pipes are ceramic, and hence brittle. The City of Los Angeles has done extensive work to locate and fix this sewer pipe damage. They have surveyed all the areas of expected damage with television cameras. This television footage was viewed by city workers who recorded the location and the nature of the sewer damage. The type of pipe damage is given by the sewer defect classification system (Table 1). Unfortunately, there is no obvious link between type of pipe damage and the mechanism responsible for failure. Since sewer pipes are so brittle, damage could have resulted from either strong shaking or ground failure (liquefaction, shear failure, or landslides). Of all the pipe damages, the ones coded as JSL (Joint Offset Light) and JCS (Joint Cracked Spiral) are most likely related to shaking phenomena rather than permanent ground deformation (Ponti, 1997); therefore, these damages were excluded from the study. This resulted in a 58% decrease in the number of sewer pipe breaks in the study area. In addition to structural codes, detailed information about the pipe size, material, age, and depth of burial is available (Table 2).

Tables indicating damage locations with respect to manhole numbers were acquired, and a digital sewer map containing manhole locations and numbers was obtained so that the damage records from the table could be georeferenced. The task of transferring the tables of damage locations to a map was tedious but provided the fundamental data needed for analysis. Basic calculations were used:

$$x_d = umh_x + \left(\frac{\text{footage}}{\text{length}}\right)(dmh_x - umh_x)$$

$$y_d = umh_y + \left(\frac{\text{footage}}{\text{length}}\right)(dmh_y - umh_y)$$

where \(x_d\) and \(y_d\) are the \(x\) and \(y\) coordinates of the damage location, \(umh\) and \(dmh\) are the uphill and downhill manholes, footage is the distance from the uphill manhole to the damage, and length is the length of the sewer pipe between the manholes.

Once the damage locations were mapped, three pockets of disruption were
<table>
<thead>
<tr>
<th>Type of Break</th>
<th>number of breaks</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipe (P)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cracked (C)</td>
<td>Longitudinal (L)</td>
<td>227</td>
</tr>
<tr>
<td></td>
<td>Radial (R)</td>
<td>1144</td>
</tr>
<tr>
<td></td>
<td>Spiral (S)</td>
<td>379</td>
</tr>
<tr>
<td></td>
<td>Multiple (M)</td>
<td>318</td>
</tr>
<tr>
<td>Fractured (F)</td>
<td>Longitudinal (L)</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Radial (R)</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Spiral (S)</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Multiple (M)</td>
<td>217</td>
</tr>
<tr>
<td>Holes (H)</td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>Broken (B)</td>
<td></td>
<td>327</td>
</tr>
<tr>
<td>Deformed (D)</td>
<td></td>
<td>30</td>
</tr>
<tr>
<td>Collapsed (X)</td>
<td></td>
<td>44</td>
</tr>
<tr>
<td>Joint (J)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cracked (C)</td>
<td>Longitudinal (L)</td>
<td>1936</td>
</tr>
<tr>
<td></td>
<td>Radial (R)</td>
<td>246</td>
</tr>
<tr>
<td></td>
<td>Spiral (S)</td>
<td>30171</td>
</tr>
<tr>
<td></td>
<td>Multiple (M)</td>
<td>24495</td>
</tr>
<tr>
<td>Fractured (F)</td>
<td>Longitudinal (L)</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>Radial (R)</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Spiral (S)</td>
<td>794</td>
</tr>
<tr>
<td></td>
<td>Multiple (M)</td>
<td>2003</td>
</tr>
<tr>
<td>Offset (S)</td>
<td>Light (L)</td>
<td>18536</td>
</tr>
<tr>
<td></td>
<td>Medium (M)</td>
<td>293</td>
</tr>
<tr>
<td></td>
<td>Severe (S)</td>
<td>57</td>
</tr>
<tr>
<td>Broken (B)</td>
<td></td>
<td>758</td>
</tr>
<tr>
<td>Open (O)</td>
<td></td>
<td>89</td>
</tr>
<tr>
<td>Alignment (A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal (H)</td>
<td></td>
<td>434</td>
</tr>
<tr>
<td>Vertical (V)</td>
<td>Light (L)</td>
<td>1452</td>
</tr>
<tr>
<td></td>
<td>Medium (M)</td>
<td>314</td>
</tr>
<tr>
<td></td>
<td>Severe (S)</td>
<td>219</td>
</tr>
</tbody>
</table>

Table 1. Type of Sewer Pipe Breaks and their Frequency.

Pipe descriptions are from the Sewer Defect Classification Table of the City of Los Angeles' Sewer Assessment Program. Total number of breaks is 84,837.
apparent (Figure 13). These sites represent the loci of damage found in the 15 months after the earthquake. Damages occurred on a range of pipelines, indicating that age or depth of burial did not determine the likelihood of pipe failure. In the areas where no damage was noted, pipelines were either inspected and found to be undamaged or were not inspected at all (Figure 13). It is assumed for this study that the areas of no data are areas of no damage. This assumption is reasonable since the City of Los Angeles only conducts detailed surveys in areas of known damage and since broken pipelines elsewhere would have been reported in 15 months' time.

The maps of sewer damage were gridded using 1 hectare cells. The number of damages within grid cells ranges from 0 to 138 (Figure 14), but in order to run the probit model, any cell with one or more damage location was assigned a 1 (ground failure occurred), and any cell without damage or without data was assigned a 0 (ground failure did not occur).6

6 Because the frequency of breaks within a cell is not preserved in the regression analysis, the exclusion of the sewer pipe codes of JSL and JCS did not make much difference in the final models. Even though 58% of the sewer pipe breaks were excluded from the study, the number of cells with ground failure only decreased from 5,071 cells to 4,499 cells. However, the exclusion of certain sewer codes is still justified for scientific reasons.

Table 2. Detailed Descriptions of the Damaged Sewer and Water Pipes

<table>
<thead>
<tr>
<th></th>
<th>SEWER PIPES</th>
<th>WATER PIPES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>distribution</td>
<td>trunk lines</td>
</tr>
<tr>
<td>Number of Damage</td>
<td>36,130</td>
<td>531</td>
</tr>
<tr>
<td>Occurrences</td>
<td></td>
<td>71</td>
</tr>
<tr>
<td>Pipe Size</td>
<td>8 inches (90%)</td>
<td>4 to 12 inches</td>
</tr>
<tr>
<td></td>
<td>10 inches (2%)</td>
<td>(96%)</td>
</tr>
<tr>
<td></td>
<td>12 to 54 inches (8%)</td>
<td>2 to 10 feet</td>
</tr>
<tr>
<td>Pipe Material</td>
<td>clay (99%)</td>
<td>iron (75%)</td>
</tr>
<tr>
<td></td>
<td>minor concrete and plastic</td>
<td>steel (23%)</td>
</tr>
<tr>
<td></td>
<td>minor cement and concrete</td>
<td>minor iron</td>
</tr>
<tr>
<td></td>
<td></td>
<td>and concrete</td>
</tr>
<tr>
<td>Pipe Age</td>
<td>ranges from 1914 to 1990</td>
<td>ranges from 1917 to 1991</td>
</tr>
<tr>
<td></td>
<td>1950 to 1965 dominant</td>
<td>1920s and 1950s dominant</td>
</tr>
<tr>
<td>Depth of Burial</td>
<td>6 to 10 feet dominant</td>
<td>3 and 4 feet dominant</td>
</tr>
</tbody>
</table>
Figure 13. Map of the state of the sewer pipes of the San Fernando Valley after the Northridge earthquake. Note that sewer pipe damage is concentrated in three main pockets. Sewer pipes are primarily limited to the flat ground of the valley, although some pipes extend into the mountainous regions. The study area is delineated by a black line.
Water Pipe Damage

Hundreds of water distribution and trunk lines were damaged by the earthquake. Digital maps of water pipe locations were obtained from O'Rourke and Toprak (1997). Because these point data were appropriate for a smaller scale than was used in this study, some of the water pipe locations did not match their supposed street addresses. To make the dataset as accurate as possible, some point data were moved based on the street addresses of water pipe repair locations. Detailed information about the pipe size, material, age, and depth were available (Table 2).

The water pipe breaks are concentrated in the same three pockets as sewer pipe breaks (Figure 15). Water pipe damages are less abundant than sewer pipe damages because water pipes are made from stronger materials. Water pipe damage does occur in areas outside of the damaged sewer pipe pockets, however. This is taken as evidence that the water pipe dataset compliments the spatially limited sewer pipe dataset. Water pipe breaks were gridded according to the sewer pipe method. The number of breaks within a cell range from 0 to 13 (Figure 16).
Figure 15. Map of the locations of water pipe breaks and street repairs. Note that water pipe and street damage is concentrated in roughly the same areas as the sewer pipe breaks.
Street Damages

Digital maps locating street repairs that were made as a consequence of the Northridge earthquake were obtained from the Bureau of Street Maintenance of the City of Los Angeles (Figure 15). These data were available only for the area west of Interstate 405, so only two-thirds of the study area is represented. Areas of no data were assumed to be areas of no damage. These data were gridded like the other damage maps. The number of street repairs required in a cell range from 0 to 5 (Figure 17).
Assumptions about the Dependent Variable

Where available, the sewer pipe breaks, water pipe breaks, and street repair locations offer very detailed information about where ground failure occurred. Unfortunately, sewer pipe data and street repair data were not available for all of the valley, and this has necessitated the assumption that areas of no sewer or street data are areas of no sewer or street damage. This assumption is justified for the following reasons. The survey of sewer pipe damage was not conducted in a random fashion; rather, surveys were concentrated in areas of known damage. Areas of no sewer data were purposely not surveyed because there were no other reasons to expect ground failure in those regions. Sewer pipe investigations radiated outward from areas of known damage, stopping when no more damage was detected. The use of water pipe data, which are available throughout the valley, compliments the sewer and street data. Cells assumed to have no sewer pipe breaks or street repairs may still be classified as cells in which ground failure occurred if at least one water pipe break is present. For the most part, areas of known ground failure are represented in at least one of the three datasets. Minor exceptions include areas where no pipes or streets are present, such as the Van Norman Complex, the Sepulveda flood control basin, and some locations within or immediately adjacent to active drainages (Ponti, 1997); these areas are very limited in extent. In sum, the three datasets of sewer pipe breaks, water pipe breaks, and street repair locations together provide an accurate picture of ground failure regions.

Independent Variables

The independent variables include geologic, topographic, hydrologic, and seismic factors which could act as predictors of ground failure occurrence. The chosen variables are 1) age of the geologic unit, 2) average percent clay in the surficial geologic units, 3) average shear wave velocity, 4) slope, 5) depth to shallow ground water, and 6) peak ground velocity. These factors were selected based on their demonstrated ability to influence the processes leading to ground failure. Data describing all variables were obtained and compiled in the form of digital maps for the San Fernando Valley. In the end, seven grids, one grid for each variable with two versions of ground water conditions, were produced (Plates 1 to 6 and Figure 19).
The age of a geologic unit is often considered an important determinant of liquefaction susceptibility; historical occurrences of liquefaction indicate that the younger the sediment, the more likely it is to liquefy (Youd and Perkins, 1978; Youd and D.M. Perkins, 1987; Tinsley et al., 1985). This relationship reflects the fact that older deposits are typically better cemented and more deeply buried than their younger counterparts. In the western United States, Holocene sediment is the most likely to liquefy; Pleistocene sediment rarely liquefies, and pre-Pleistocene sediment has never been known to liquefy (Obermeier, 1996).

Traditional geologic mapping does not adequately characterize the surficial geologic units. Fortunately, detailed mapping (1:24,000 scale) of the Quaternary deposits of the Los Angeles region has been done (Tinsley et al., 1985). This mapping, based on historical flooding records, the topographic position of the units, and the relative degree of soil profile development, distinguished upper Holocene deposits (Qya2), lower Holocene deposits (Qya1), and upper Pleistocene deposits (Qoa) (Tinsley et al., 1985). Age ranges for these units are given in Table 3.

While quite detailed for the valley sediments, the Tinsley et al. (1985) maps generalize the older deposits which rim the valley. Geologic maps of the eight quadrangles of the San Fernando Valley (Yerkes, 1996a, 1996b, 1996c, 1996d; Yerkes and Campbell, 1993a, 1993b, 1995; Yerkes and Showalter, 1993) were used to differentiate these older units. Specifically, the Saugus (Qs) and the Pico (Tp) formations were preserved as distinct units, because they straddle the Pleistocene/Pliocene time boundary (Table 3). Since pre-Pleistocene sediment is unknown to liquefy in the western United States and since soft clay failures would not occur in older, consolidated rocks, it was not deemed necessary to further differentiate the Tertiary units (Table 3).

There were numerous steps involved in compiling the maps of depositional age. The San Fernando Valley quadrangles were in digital form, but they had to be edgematched and mosaicked. This 'geologic mosaic map' was then intersected with the digital form of the Quaternary map (Tinsley et al., 1985). The geologic units and related age ranges were maintained as attributes of the derived map (Plate 1).
In order to use depositional age as a variable in the regression analysis, a single age, rather than a range, had to be associated with each geologic unit. Following Pike et al. (1994), the midpoint of the logarithms of the minimum age and the maximum age for a given unit was assigned:

\[
\text{Average Age} = \frac{\log_{10}(\text{minimum age}) + \log_{10}(\text{maximum age})}{2}
\]

This transformation places emphasis on the younger, more liquefiable sediments of a geologic unit (Pike et al., 1994).

Grain Size of the Surficial Geologic Units

Sediment grain size has long been recognized as a factor influencing liquefaction susceptibility (Figure 3). According to Tinsley et al. (1985):

"Sand and silty sand are the textural classes that have the greatest likelihood of liquefaction. Gravelly sand or deposits containing less than 15 percent clay are less likely to liquefy. Bouldery and cobbly gravels or deposits containing more than 15 percent clay are not known to liquefy."

Given this information, it seems logical to include a variable describing the percent sand or percent clay of the surficial deposits of the valley. Percent sand is viewed as an imperfect indicator, since silt and silty sands as well as clean sands can liquefy. The percent clay in a sediment is considered to be a better indicator of liquefaction susceptibility because sediment with more than 15 percent clay is too cohesive to liquefy. The use of percent clay

<table>
<thead>
<tr>
<th>Geologic Unit</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Af, Qya2</td>
<td>0 to 1,000 years</td>
</tr>
<tr>
<td>Qya1</td>
<td>1,000 to 10,000 years</td>
</tr>
<tr>
<td>Qoa</td>
<td>10,000 to 500,000 years</td>
</tr>
<tr>
<td>Qs, Tp</td>
<td>500,000 to 2,000,000 years</td>
</tr>
<tr>
<td>Ts, TQ (undifferentiated)</td>
<td>2,000,000 to 65,000,000 years</td>
</tr>
<tr>
<td>Tv, Mx</td>
<td></td>
</tr>
</tbody>
</table>

Table 3. Geologic units and their corresponding age ranges.
as a variable has the added advantage of describing the susceptibility of deposits to soft clay failures. For these reasons, average percent clay$^7$ is chosen as a predictor variable.

The development of a map describing the average percent clay of the sediments of the San Fernando Valley was not an easy task. First, a digital map which divided the Quaternary units based on sediment age and grain size (Tinsley and Fumal, 1985) was acquired. This map was intersected with the 'geologic mosaic map'. For reasons described earlier, the Saugus and Pico formations were preserved as distinct units, while the other Tertiary and older units unit remained undifferentiated. The map at this stage described the study area in terms of the geologic units of Table 4.

While this map was useful since it delineated surficial units with respect to grain size, it did not provide clay percentages. Detailed studies had been conducted at two sites (Malden St. and Wynne Ave.) within the fine-grained units (Holzer et al., 1996). Inasmuch as the sediment at these sites is believed to be typical of the fine-grained units throughout the valley (Tinsley, 1997), the average percent clay was calculated and then assigned to all of the fine-grained units (Qyf and Qof) mapped by Tinsley and Fumal (1985) (Table 4). To characterize the medium and coarse-grained units, detailed data from storm drain and California Department of Transportation (CALTRANS) highway investigations were examined. Storm drain data classified sediment according to the Unified Soil Classification System (Howard, 1984), while the highway investigations classified sediment using field-based estimates of particle size rather than mechanical

<table>
<thead>
<tr>
<th>Geologic Unit</th>
<th>Description</th>
<th>% clay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Af, Mx, Tv</td>
<td>artificial fill, crystalline rock, volcanic rock</td>
<td>0</td>
</tr>
<tr>
<td>Qoc, Qyc, Qyvc</td>
<td>coarse-grained alluvium</td>
<td>3</td>
</tr>
<tr>
<td>Qs</td>
<td>Saugus Formation (sandstone)</td>
<td>5</td>
</tr>
<tr>
<td>Qom, Qym</td>
<td>medium-grained alluvium</td>
<td>10</td>
</tr>
<tr>
<td>Tp</td>
<td>Pico Formation (sandy or clayey siltstone)</td>
<td>20</td>
</tr>
<tr>
<td>Qof, Qyf</td>
<td>fine-grained alluvium</td>
<td>26</td>
</tr>
<tr>
<td>Ts</td>
<td>Tertiary sediments (undifferentiated)</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 4. Percent clay assigned to the geologic units of the San Fernando Valley

$^7$ The term clay as it is used in this thesis refers to grains which are 4 microns or smaller in diameter.
Figure 18. Diagram showing the particle size classification system used by CALTRANS. Sediment is classified using field-based estimates of particle size rather than mechanical analysis. Medium-grained units are described as sands and silty sands. The average percent clay of these categories (10%) is used as an approximate clay content for all of the medium-grained units in the valley.
analyses (Figure 18). The medium-grained units were described as well-graded sands (SW), poorly graded sands (SP), silty sands (SM), or fine sands and silty/clayey fine sands (ML) by storm drain data and as sands and silty sands by CALTRANS data (Figure 18). The average percent clay for the sand and silty sand groups of Figure 18 was assigned to the medium-grained units (Table 4). Highway investigations revealed that the mapped coarse-grained units were dominated by coarse sands. These units were assigned a lower percent clay (3%) than the medium-grained units because very little clay was noted in borehole investigations.

The older units rimming the valley were assigned a clay percentage which is consistent with their unit description. For example, the Saugus Formation (Qs) was assigned 5% clay because it is a sandstone, whereas the Pico Formation (Tp) was assigned 20% clay because it is described as a sandy siltstone or a clayey siltstone (Yerkes and Campbell, 1993a). The clay percentages assigned to all map units are summarized in Table 4 and can be seen in Plate 2. While the stated values are approximations, they accurately represent the relative amount of clay within the deposits.

**Shear Wave Velocity**

Shear wave velocity, which is simply the rate at which seismic waves travel through soil or rock, is a useful factor in determining the amplification of ground motion at specific sites (Joyner and Fumal, 1985; Borcherdt et al., 1991). Borcherdt et al. (1979) found that average shear wave velocity inversely correlated to intensities of ground shaking generated by the 1906 San Francisco earthquake. In general, the slower the shear wave velocity, the greater the amplification of ground motion, and hence the greater the intensity of ground shaking and ground failure potential.

Shear wave velocity is dependent on the rigidity of the soil or rock mass through which seismic waves travel (Fumal and Tinsley, 1985). This rigidity can be measured using physical properties, such as grain size and standard penetration resistance for soil, and fracture spacing, hardness, and lithology for rock (Fumal and Tinsley, 1985). Therefore, one could group deposits into units, based on these physical properties and expect similar shear wave velocity throughout the unit. This is the basic process used to describe the shear wave velocity variable. Since the rigidity of the soil and bedrock units are measured differently, they will be discussed separately.
Soil Deposits

Shear wave velocity within soils varies depending on the grain size and age of the sediment (Fumal and Tinsley, 1985), so the very detailed age map of Tinsley et al. (1985) was combined with the grain size map of Tinsley and Fumal (1985). This derived map delineates units which should have similar average shear wave velocities.

The next step in characterizing these units was to examine their three-dimensional distribution, since the motion felt at the ground surface is in part a function of the sedimentary column at depth. The appropriate depth to use for shear wave velocity calculations depends on the kind of earthquake effects under consideration. If building damage is of interest, then shear wave velocity down to 30 m or one quarter wavelength of a 1 second shear wave (~100 m for alluvial deposits) (Fumal and Tinsley, 1985) is appropriate. If localized ground failure is of interest, as it is in this study, then shear wave velocity for shallower depths (15 m) should be considered (Tinsley, 1997).

With this in mind, shear wave velocity was calculated at 2.5 m intervals down to a 15 m depth from 28 downhole measurements from the San Fernando Valley (Fumal et al., 1981, 1982, 1984; Gibbs et al., 1996). The goal was to characterize each map unit with a shear wave velocity range which accurately reflects the upper 15 m of sediment and the three-dimensional geometry of the deposits. Unfortunately, not enough data were available to obtain meaningful shear wave velocity averages for the detailed units of the derived map.

Therefore, average shear wave velocities calculated by Fumal and Tinsley (1985) for map units delineated by Tinsley and Fumal (1985) were used (Table 5 and Plate 3). Inasmuch as these values incorporate downhole measurements from sites throughout the Los Angeles region, they are better than the detailed calculations. Admittedly, these average shear wave velocities are only approximations, and period dependencies have not been addressed. The surface unit is typically only 5 m to 8 m thick, but it is assumed that this unit is present to the depth that would influence shaking response. This is not considered a serious flaw, however, since this study is concerned with the shallow shear wave velocity that would cause ground failures. If anything, the shear wave velocity may be conservative; that is, it may overestimate the amplification hazard by assuming that the materials with slow shear wave velocities are thicker than they truly are.
Tertiary and Older Units

The geologic mosaic map described previously offered a starting point to characterize the shear wave velocity for the Tertiary and older units. Various bedrock units were grouped into eight engineering-geologic categories based on fracture spacing, hardness, and lithology by Wentworth et al. (1970). The rocks within a given category can be expected to have a similar shear wave velocity. The average shear wave velocity for each of the eight categories was determined by Fumal and Tinsley (1985). To create a shear wave velocity map, each of the mapped units from the geologic mosaic map was assigned to one of the eight categories and then given its appropriate shear wave velocity (Table 6 and Plate 3).

Slope

Slope is an important variable to include since it may have an effect on landslide initiation, liquefaction related failures, or topographic amplification of strong motion. The first step in generating a slope grid was to obtain a Digital Elevation Model (DEM). A DEM is comprised of regularly spaced grid cells which describe the ground surface elevation corresponding to the location of a given cell. Resolution and accuracy depend on the spacing of the grid cells. For this study, a composite 30 m resolution DEM for the Los Angeles region was acquired. This DEM was produced from 10 m resolution DEMs which

<table>
<thead>
<tr>
<th>Geologic Unit&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Description&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Average &lt;i&gt;V&lt;sub&gt;S&lt;/sub&gt;&lt;/i&gt; (m/s)&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Qyf</td>
<td>fine-grained Holocene alluvium</td>
<td>200</td>
</tr>
<tr>
<td>Qym</td>
<td>medium-grained Holocene alluvium</td>
<td>230</td>
</tr>
<tr>
<td>Qyc</td>
<td>coarse-grained Holocene alluvium</td>
<td>320</td>
</tr>
<tr>
<td>Qyvc</td>
<td>very coarse-grained Holocene alluvium</td>
<td>365</td>
</tr>
<tr>
<td>Qof</td>
<td>fine-grained Pleistocene alluvium</td>
<td>305</td>
</tr>
<tr>
<td>Qom</td>
<td>medium-grained Pleistocene alluvium</td>
<td>470</td>
</tr>
<tr>
<td>Qoc</td>
<td>coarse-grained Pleistocene alluvium</td>
<td>460</td>
</tr>
<tr>
<td>Qovc</td>
<td>very coarse-grained Pleistocene alluvium</td>
<td>650</td>
</tr>
</tbody>
</table>

Table 5. Average shear wave velocity (<i>V<sub>S</sub></i>) for alluvial deposits of the San Fernando Valley.

<sup>1</sup> From Tinsley and Fumal (1985)

<sup>2</sup> From Fumal and Tinsley (1985)
were mosaicked to eliminate edge effects, smoothed using a 4 x 4 focal mean filter to improve accuracy, and resampled at 30 m resolution. This composite DEM was clipped to include only the San Fernando Valley. The cell size was set to 100 m rather than 30 m to match the grids of the other predictor variables. Finally, the slope (in degrees) of each cell was calculated using an algorithm which fits a plane to the elevation values of a 3 x 3 cell neighborhood around the processing cell (see ESRI, 1996 for more details). The slope of the study area ranged from 0 to 42 degrees (Plate 4).

<table>
<thead>
<tr>
<th>Geologic Unit1</th>
<th>Description1</th>
<th>Group2</th>
<th>Average Vₛ (m/s)³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jsm, Jsms</td>
<td>Santa Monica Slate</td>
<td>IV</td>
<td>945</td>
</tr>
<tr>
<td>Kc, Ktc</td>
<td>Chatsworth or Chico Formation</td>
<td>VI</td>
<td>1265</td>
</tr>
<tr>
<td>Kgr</td>
<td>granitic rocks</td>
<td>IV</td>
<td>945</td>
</tr>
<tr>
<td>Tb</td>
<td>basalt</td>
<td>IV</td>
<td>945</td>
</tr>
<tr>
<td>Tc, Tcob</td>
<td>Conejo Volcanics</td>
<td>IV</td>
<td>945</td>
</tr>
<tr>
<td>Td</td>
<td>Domengine Formation</td>
<td>VI</td>
<td>1265</td>
</tr>
<tr>
<td>Ti</td>
<td>intrusive rocks</td>
<td>IV</td>
<td>945</td>
</tr>
<tr>
<td>Tm</td>
<td>Modelo Formation</td>
<td>II</td>
<td>550</td>
</tr>
<tr>
<td>Tp, QTp</td>
<td>Pico Formation</td>
<td>II</td>
<td>550</td>
</tr>
<tr>
<td>Tsc1-3</td>
<td>Simi Conglomerate</td>
<td>V</td>
<td>1080</td>
</tr>
<tr>
<td>Tt</td>
<td>Topanga Group (undivided)</td>
<td>IV</td>
<td>945</td>
</tr>
<tr>
<td>Tts, Tcb</td>
<td>members of Topanga Group</td>
<td>II</td>
<td>550</td>
</tr>
<tr>
<td>Ttc</td>
<td>Cold Creek Member of Topanga Group</td>
<td>V</td>
<td>1080</td>
</tr>
<tr>
<td>Tw</td>
<td>Towsley Formation</td>
<td>V</td>
<td>1080</td>
</tr>
<tr>
<td>Mx</td>
<td>intrusive and metamorphic rocks³</td>
<td>IV</td>
<td>945</td>
</tr>
<tr>
<td>Qls</td>
<td>landslide deposits</td>
<td>II</td>
<td>550</td>
</tr>
<tr>
<td>Qs, QTs</td>
<td>Saugus Formation</td>
<td>II</td>
<td>550</td>
</tr>
</tbody>
</table>

Table 6. Average shear wave velocity (Vₛ) of Tertiary and older units of the San Fernando Valley

2 Based on Wentworth et al. (1970) and Fumal and Tinsley (1985).
3 From Fumal and Tinsley (1985).
Depth to Ground Water

The depth to ground water is a very important determinant of ground failure susceptibility since sediment must be water saturated to liquefy or sustain soft clay failures. Liquefaction susceptibility decreases as depth to water increases because 1) the overburden pressure increases and 2) age, cementation, alteration, and compactness of sediment generally increase with depth (Tinsley et al., 1985). Soft clay failures are more likely with a shallow water table because clays which have never dewatered have low enough shear strength to fail if subjected to significantly high accelerations.

Depth to ground water has been used in liquefaction hazard studies by many workers (e.g. Youd and Perkins, 1978; Tinsley et al., 1985; Real et al., 1996). Youd and Perkins (1978) consider locations with a water table below 33 feet to have a low liquefaction susceptibility and locations with a shallower water table to be more susceptible. Similarly, Tinsley et al. (1985) attribute water table depths of less than 10 feet, 10 to 30 feet, 30 to 50 feet, and greater than 50 feet, with very high, high, low, and very low susceptibility rankings, respectively.

The present depth of the ground water in the San Fernando Valley is difficult to determine. Data describing ground water has been gathered throughout this century, but this data cannot simply be merged into one map since seasonal and historical fluctuations may change the depth to water significantly. For example, the water table will rise and fall in response to precipitation changes, and the water table has fallen as a consequence of pumping due to the increased water needs of the growing Los Angeles metropolis. Following the example of Tinsley et al. (1985), two possible states of the water table, representing high and low ground water conditions were examined.

The high ground water state was based on data gathered in 1944 (California Water Rights Board, 1962 modified by Tinsley et al., 1985). Water levels were high because 1941, 1943, and 1944 were wet years and since the valley was not heavily populated during this period. A contour map of water table elevation at this time is presented in Tinsley et al. (1985). This contour map was digitized and then converted to a raster image so that it could be compared to a DEM. The water table elevation was subtracted from the surface elevation given by the DEM to produce a depth to water map. This derived map contains some uncertainty associated with it because data points are generated only where DEM cells intersect water surface contours; thus, many areas of the map lack data. Even
so, the derived map conforms to expected water patterns. The derived map was contoured using 10, 30, and 50 foot contours to correspond to the susceptibility rankings of Tinsley et al. (1985). This final map (Figure 19) represents the highest ground water levels expected and thus delineates the maximum extent for high ground failure susceptibility.

To expedite analysis, it was necessary to specify a single value, rather than a range of values for each depth interval. Therefore, the midpoint of each depth range was taken as the depth to water for a given interval (Table 7). In cases where depth to water was greater than 50 feet, there was much variation in the data. Appropriate averages were determined using the descriptive statistics of the depth map; average depth was assigned differently depending on the geologic unit of the site (Table 7).

<table>
<thead>
<tr>
<th>Depth Interval</th>
<th>Average Depth to Water (ft)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;10 feet</td>
<td>5</td>
</tr>
<tr>
<td>10 to 30 feet</td>
<td>20</td>
</tr>
<tr>
<td>25 to 35 feet</td>
<td>30</td>
</tr>
<tr>
<td>30 to 50 feet</td>
<td>40</td>
</tr>
<tr>
<td>&gt;50 feet</td>
<td>80 for alluvium, Saugus, or Pico Formations</td>
</tr>
<tr>
<td></td>
<td>120 for other Tertiary units</td>
</tr>
</tbody>
</table>

Table 7. Average Depth to Ground Water assigned to Depth Intervals from Water Contour Maps.

1 Applies only to the more recent depth to water contour map.

A depth to water map describing the present, drier conditions was compiled by Tinsley et al. (1985). They used data from several thousand boreholes, drilled from 1950-1981, to construct this map. The depth to ground water, including perched ground water, is described by 10, 30, and 50 foot contours. This map was augmented with recently discovered shallow ground water in the Granada Hills area (Ponti, 1997). This map (Plate 5) shows shallow ground water in the same general areas as in the 1944 map, but the occurrences of shallow water are diminished in extent. The average values for depth intervals were assigned according to Table 7.
Figure 19. Depth to ground water map for the San Fernando Valley. This map represents high ground water conditions since it is based on data gathered after wet years and prior to pumping in the valley. Compiled from data from California Water Rights Board (1962) as modified by Tinsley et al. (1985).
Peak Ground Velocity

All of the independent variables discussed previously describe the susceptibility of a given location to ground failure, but susceptibility alone will not produce ground failure. Accelerations must surpass a critical level or velocities must be large enough to initiate ground failure mechanisms. Therefore, a parameter representing the variability of strong motion throughout the study area could be important. Three obvious variables which could be used include peak ground acceleration, peak ground velocity, or peak ground displacement. Peak ground velocity was chosen because it describes the directivity effects, which are a concern for underground pipelines, better than the other strong ground motion measures.

There are 10 strong motion stations in the study area which recorded peak ground velocity from the Northridge earthquake. These data offer only sparse coverage over large portions of the valley, hence modeled peak ground velocities (Wald et al., 1997) were used. Wald et al. (1997) produced these modeled velocities for 144 grid points in an approximate area of 3600 km² using a 5 km spacing, so they offer information where instruments are lacking. Wald et al. (1997) computed longer period ground motions (>1 second) using the finite fault methodology of Hartzell and Heaton (1983) and use actual recordings of shorter period ground motions (<1 second). Mathematical filters were used to remove short periods from the computed motions and to remove long periods from the actual recorded motions (Wald et al., 1997); finally, the two filtered records were summed to produce a ground motion estimate for each of the 144 points (Wald et al., 1997).

Of the 144 estimates of peak ground velocity, 22 fell within the study area. Triangulation was used to interpolate values between these points. Results were smoothed using a low pass filter, and the smoothed, triangulated map was converted into a grid with 1 hectare cells (Plate 6). Within the study area, peak ground velocity varies from 21 cm/sec to 136 cm/sec, with highs in the northern part of the valley. This is consistent with directivity effects expected from a thrust fault dipping to the south.

MODEL BUILDING

Once the variables were compiled, the ARC/INFO grids were exported as an ASCII file, which associated the location of each grid cell to all of the variables. This file could be read by LIMDEP (Greene, 1992), the computer statistical package chosen to perform the
analyses. Probit regressions were run using maximum likelihood estimation. The gradient, function, and parameters converged rapidly, usually after 6 or 7 iterations. Nearly one hundred regressions were performed before a suitable equation was found. Experimentation involved different combinations of variables, different transformations of the variables, and different study boundaries. The goal of the regression analysis was to find coefficients that made sense physically. Once realistic coefficients were determined, it was necessary to see if the spatial distribution of probabilities fit the natural environment. The calculated coefficients were multiplied by their corresponding variable grid to produce a grid of Z values. The standard normal distribution function was then used to convert the grid of Z values to probabilities. The final product is a map of the study area which gives the probability of ground failure for every 1 hectare grid cell.
CHAPTER FIVE: ANALYSIS

EXPERIMENTATION

Numerous choices about the study area extent, data transformations, and combinations of variables affect the probabilities obtained. Much experimentation was required to develop the best model, and nearly one hundred probit regressions were run. In choosing a study area, it is desirable to include all possible environments (geologic units, water table levels, etc.) so that the model is representative of the physical world and can be applied to broad regions. On the other hand, including large areas which are not very susceptible to the hazard of interest can reduce the model's power. As a compromise between these considerations, a 560 km$^2$ study area was chosen which includes Tertiary units of the mountains and Quaternary units of the valley, but excludes a large portion of the granitic and metamorphic rocks in the eastern valley (see plates for study boundary).

Descriptions of the variables, using this study area, are given in Table 8. The total number of grid cells is 56,111; ground failure occurred in 4,499 of these cells. All of the independent variables are on the same approximate scale, except for the shear wave velocity, which ranges from 200 m/s to 1265 m/s. The natural logarithm of the shear wave velocity was taken to make it comparable to the other variables. The natural logarithm is appropriate because it emphasizes differences in the damaging slow shear wave velocities, more than differences in the fast shear wave velocities. Other data transformations and variable manipulations were attempted, such as shear wave velocity divided by ten or clay multiplied by a dummy variable for coarse-grained deposits; however, none of these were used in the final models. In the end, the variables included average age, percent clay, ln (shear wave velocity), slope, depth to water for dry conditions, and peak ground velocity. The water table grid which represents the present, dry conditions of the valley was used because the Northridge earthquake occurred after a six week dry period. If another earthquake occurred after an unusually wet period, or if the metropolitan region reduced its rate of ground water pumping, then the grid representing high ground water conditions would be appropriate.
REGRESSION MODELS

Model One: Susceptibility Version

The initial goal was to develop a model using only site conditions as the independent variables; that is, all the independent variables except peak ground velocity were used. Following hazard map terminology, this first model will be referred to as the ground failure susceptibility model. The following equation was produced:

\[
\text{Probability} = \Phi \left( 1.9522 - 0.00673 \text{ water depth} + 0.0123 \text{ clay} - 0.549 \ln(\text{shear wave velocity}) \right). 
\]

The age and slope variables were dropped because they are not significant at the 99% confidence level (\( p = 0.24 \) for age; \( p = 0.52 \) for slope). All other variables are highly significant at the 99% confidence level (Table 9).

### Table 8. Descriptive Statistics of Dependent and Independent Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damage</td>
<td>.0802</td>
<td>.2716</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Average Age</td>
<td>4.301</td>
<td>2.142</td>
<td>1.5</td>
<td>7.057</td>
</tr>
<tr>
<td>Percent Clay</td>
<td>20.44</td>
<td>12.60</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Shear Wave Velocity</td>
<td>413.9</td>
<td>275.5</td>
<td>200</td>
<td>1265</td>
</tr>
<tr>
<td>ln (shear wave velocity)</td>
<td>5.853</td>
<td>.5579</td>
<td>5.298</td>
<td>7.143</td>
</tr>
<tr>
<td>Slope</td>
<td>4.146</td>
<td>5.643</td>
<td>0</td>
<td>41.52</td>
</tr>
<tr>
<td>Water depth</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>low</td>
<td>80.39</td>
<td>35.84</td>
<td>5</td>
<td>120</td>
</tr>
<tr>
<td>high</td>
<td>75.91</td>
<td>39.84</td>
<td>5</td>
<td>120</td>
</tr>
<tr>
<td>Peak Ground Velocity</td>
<td>58.31</td>
<td>25.84</td>
<td>20.72</td>
<td>135.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The signs of the coefficients are understandable in terms of the physical processes of ground failure. As depth to the water table increases, the probability of ground failure decreases because 1) sediment is more difficult to liquefy at depth and 2) a deep water table enables shallow clays to dewater, thereby increasing their shear resistance. As the percentage of clay within a deposit increases, the probability of ground failure increases; this reflects the large number of ground failures which occurred in fine-grained units of the valley. As shear wave velocity increases, the probability of ground failure decreases because the amplification of seismic energy would not be as great.

The equation was used to calculate probabilities for each cell of the study area, thereby generating a ground failure probability map (Plate 7). Probabilities range from 0.004 to 0.26\(^8\), with a mean of 0.08 and a standard deviation of 0.067. The spatial distribution of the probability categories conforms fairly well to expectations. High probabilities occur where ground water is shallow and geologic units are fine-grained. Two of the three damage pockets coincide with high probability categories (greater than 0.20), but the northern damage pocket is predicted to have less than 0.15 probability of ground failure.

To examine the probability distribution more rigorously, relative histograms of the map probability categories\(^9\) for failed (y = 1) and non-failed (y = 0) cells were produced (Figure 20). For cells where y = 1, there is a general increase in the number of cells in higher probability categories, with a notable lack of cells in the 0.15 to 0.20 probability

\(^8\) The low probabilities are a function of the fact that only 8% of the total number of cells sustained failure. A way to obtain higher probabilities would be to limit the study area so that damage cells are a greater proportion of the total.

\(^9\) The 0.20 to 0.25 and 0.25 to 0.26 probability categories were combined on the histogram.
Figure 20. Relative Histogram of Probabilities for Model One

Figure 21. Comparison of Actual Failures with Predicted Probabilities for Model One
range. For cells where $y = 0$, there is a decrease in the number of cells as probabilities increase.

While these trends follow expectations, it is difficult to make definitive conclusions based on these histograms. Therefore, a new quantity, termed the failure ratio, is calculated. The failure ratio is the number of failed cells in each probability category divided by the total number of cells in that category. The failure ratio for each probability category is plotted against probability on the x axis (Figure 21). Of the cells predicted to have a 0 to 0.05 probability of ground failure, one would expect 0 to 5% of these cells (a failure ratio of 0 to 0.05) to have failed during Northridge. Indeed, 3% (failure ratio = 0.03) of the cells in this probability category failed during Northridge. Failure ratios in the other probability categories match what is expected, except for a slight discrepancy for the 0.15 to 0.20 range. This confirms that the modeled probabilities accurately describe the observed failures.

**Model Two: Earthquake-Specific Version**

The first model identifies significant variables and predicts probabilities which match the failure ratios quite well, but the spatial fit of the model could be improved. Specifically, the northern damage pocket is predicted to have a relatively low probability (less than 0.15). It is postulated that ground failure in this northern pocket was the result of directivity effects (i.e. large ground velocities) in this region. Therefore, a variable representing peak ground velocities generated by the Northridge earthquake was added to the model, and the following equation was produced:

$$\text{Probability} = \Phi \{- 0.00858 \text{ water depth} + 0.0169 \text{ clay} - 0.354 \ln(\text{shear wave velocity}) + 0.0133 \text{ ground velocity}\}.$$ 

The constant and the age and slope variables were dropped because they were not significant at the 99% confidence level; p-values are 0.03, 0.24, and 0.21, respectively. All other variables are highly significant (Table 10). As with the susceptibility equation, all of the signs of the coefficients make sense physically. Following hazard map terminology, this model is a ground failure potential model because it combines susceptibility and earthquake opportunity factors.
Plate 8 shows the probability map generated using this equation. This map has a greater range of probabilities (0.001 to 0.532) than the previous map, but the mean and the standard deviation are similar (Table 11). For ease of comparison, the same probability categories are used in both maps, with only the highest, red category differing. The second map predicts higher probabilities for the northern and central damage pockets than the susceptibility map; however, the southernmost damage pocket is predicted with lower probabilities on the earthquake-specific map. In general, the actual failure cells fall into high probability categories. On the other hand, some regions are predicted with high probabilities but have few failed cells. Some of these regions can be explained. For example, severe damage in the 1971 San Fernando earthquake and more modest damage in the Northridge earthquake affected the Van Norman Dam and associated complex. This area is correctly identified by the model as being at high risk but shows few failed cells due to the lack of pipes in the region.

<table>
<thead>
<tr>
<th>Variable</th>
<th>coefficient</th>
<th>std. error</th>
<th>t-ratio</th>
<th>prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>waterdepth</td>
<td>-0.00858</td>
<td>0.000259</td>
<td>-34.462</td>
<td>0.0000</td>
</tr>
<tr>
<td>clay</td>
<td>0.0169</td>
<td>0.000830</td>
<td>20.303</td>
<td>0.0000</td>
</tr>
<tr>
<td>ln (shear wave velocity)</td>
<td>-0.354</td>
<td>0.00634</td>
<td>-55.943</td>
<td>0.0000</td>
</tr>
<tr>
<td>ground velocity</td>
<td>0.133</td>
<td>0.000330</td>
<td>40.413</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 10. Results of the Earthquake-Specific Model

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model One</td>
<td>0.080</td>
<td>0.067</td>
<td>0.004</td>
<td>0.261</td>
</tr>
<tr>
<td>Model Two</td>
<td>0.087</td>
<td>0.087</td>
<td>0.001</td>
<td>0.532</td>
</tr>
</tbody>
</table>

Table 11. Descriptive Statistics of Probability Models

The nature of the probability distribution was examined using another relative histogram (Figure 22). As in the first model, where \( y = 1 \), there is an increase in the number of cells present in higher probability categories; where \( y = 0 \), there is a decrease in the number of cells in higher probability categories. Figure 23 compares the probability categories to the failure ratio within those categories. In general, the model predicts the actual failure ratio accurately. As mentioned previously, there are too few failure cells in the highest category; the failure ratio was 0.24 in the 0.25 to 0.53 probability range. This
Figure 22. Relative Histogram of Probabilities for Model Two

Figure 23. Comparison of Actual Failures with Predicted Probabilities for Model Two
is not a serious cause for concern, however, because all of the other failure ratios fall within the appropriate probability ranges.

WHAT DO THE MODELS MEAN?

The maps presented here identify three or four variables which are significant in predicting ground failure occurrences, and combine them to produce probabilistic estimates of hazard. The damage variable has an expected value of 0.08 (Table 8). Thus, a simple estimate of the probability of ground failure would be 0.08 for the entire study area, if no predictor variables were used. This probability estimate is refined through the use of relevant variables, such as depth to water, percent clay, ln(shear wave velocity), and peak ground velocity. The new probability estimates identify areas of higher and lower risk of ground failure than 0.08 (Plates 7 and 8). The use of the models improves the estimates of hazard throughout the study area.

The maps are probabilistic and hence differ from traditional hazard maps. Traditional maps often identify qualitative zones of high, medium, and low hazard. For these types of maps, a good fit to the data would mean that nearly all failures fall in the high hazard regions. This measure of fit does not apply to probabilistic maps. If a probability of 0.25 is predicted (a relatively high probability of ground failure in this study), then only 1 out of every 4 cells assigned this probability would be expected to fail. Therefore, it is unreasonable to expect failure everywhere in the highest probability zones. Likewise, it is unreasonable to expect actual failures to only occur in the highest probability zones. Instead, a distribution of actual failures across probability categories can be expected. Overall, the fit between the failure ratios and the predicted probability ranges is extremely good (Figures 21 and 23). Minor deviations can be attributed to uncertainties in the data and the fact that only 3 or 4 variables are being used to explain the complex mechanisms of ground failure occurrence.

WHICH MODEL IS BETTER?

Both models are valid as guides for locating areas where ground failure could be a problem in the San Fernando Valley. All the variables are highly significant and are consistent with physical processes. Both maps predict probabilities which are spatially plausible, and failure ratios match probability predictions. In order to determine which
model is better at explaining the observed damage pattern, the mean squared error (mse) was calculated.

\[
mse = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n},
\]

where \( y_i \) is the observed probability and can equal 0 or 1, \( \hat{y}_i \) is the predicted probability, and \( n \) is the total number of cells. The overall mean squared error showed no difference between the models (Table 12). Therefore, the mean squared error was calculated separately for non-failed (\( y = 0 \)) and failed (\( y = 1 \)) sites. For \( y = 0 \), the first model has a the smaller mean squared error; for \( y = 1 \), however, the second model has a reduced error. Thus, the mean squared error offers no way of distinguishing between the models.

<table>
<thead>
<tr>
<th></th>
<th>Mean Squared Error</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( y = 0 ) or 1</td>
<td>( y = 0 )</td>
<td>( y = 1 )</td>
</tr>
<tr>
<td>Model One</td>
<td>0.069</td>
<td>0.010</td>
<td>0.749</td>
</tr>
<tr>
<td>Model Two</td>
<td>0.068</td>
<td>0.013</td>
<td>0.695</td>
</tr>
</tbody>
</table>

Table 12. Mean Squared Error for Both Models

Despite ambiguous mean squared error results, there are reasons to favor the second model. First, the earthquake-specific model incorporates an additional, highly significant variable. Second, it has a wider range of probabilities than the susceptibility version, so it offers more discrimination between different locations in the valley. As a final way of comparing the models, a likelihood ratio test was performed. The likelihood ratio test statistic is defined as :

\[
-2\ln\left( \frac{L_R}{L_U} \right),
\]

where \( L_R \) is the likelihood function evaluated using parameter constraints, and \( L_U \) is the likelihood function evaluated without using the constraints (Greene, 1990). This statistic follows a chi-squared distribution. In this case, \( \chi^2 = 1484 \), where the susceptibility model gives the restricted likelihood function (\( L_R \)) and the earthquake-specific model gives the unrestricted likelihood function (\( L_U \))\(^{10}\). This \( \chi^2 \) statistic is highly significant (\( \chi^2_{(0.005)} = 7.88; df = 1 \)). From this it may be concluded that the two models are statistically different and that the unrestricted likelihood function of the second model is superior to the restricted

\(^{10}\) Actually, a version of the earthquake-specific model which included the constant was used for comparison with the susceptibility model. The likelihood function was the same for both versions of the earthquake-specific model, but the one with the constant had to be used to avoid the degrees of freedom for the likelihood ratio test statistic being equal to zero.
likelihood function of the first model. All of the above reasons indicate that the second model is slightly better at explaining the observed damage pattern of the Northridge earthquake.
CHAPTER SIX: DISCUSSION

COMPARISON WITH OTHER HAZARD STUDIES

Having accomplished the goal of this thesis, to produce regional estimates of hazard due to earthquake-induced ground failures in the San Fernando Valley, it is instructive to put the models in the proper context of previous hazard studies. The best way to establish the hazard potential of a site is to conduct detailed geotechnical investigations (e.g. Seed and Idriss, 1971). Because it is impossible to study every site in detail, regional models of hazard must be employed. Regional hazard studies typically produce qualitative rankings of hazard based on expert opinion (Tinsley et al., 1985; Youd and J.B. Perkins, 1987), use algorithms which require experts to assign weights to the variables (Mejia-Navarro et al., 1994; King and Kiremidjian, 1994), or use statistical methods to produce quantitative hazard estimates (Mark, 1992; Pike et al., 1994).

Statistical methods in general are favored because they produce quantitative results which attempt to account for uncertainty in the analysis. As emphasized throughout this thesis, quantitative results are desired because they offer more information than qualitative estimates. For example, a particular location can be said to have a 20% higher chance of ground failure than another location. This offers much more information than the difference between a 'high' hazard zone and a 'medium' hazard zone. Some hazard studies do produce quantitative results through the use of algorithms which assign weights to predictor variables, but the logic behind the weights of the different variables is often unclear. The probabilistic maps presented here are more robust because the weights of the predictor variables are estimated through regression analysis. In addition, stochastic methods account for and quantify the uncertainty of the hazard estimates. For these reasons, statistical methods are favored over traditional methods.

Besides offering more information than traditional studies, the models produced here are more sophisticated than their statistical counterparts which employ the same method (Bernknopf et al., 1988, 1993; Mark, 1992; Bernknopf and Soller, 1994; Pike et al., 1994). It is difficult to compare all these studies directly to the present study because only Bernknopf and Soller (1994) and Pike et al. (1994) have considered seismic hazards, and none of these studies have studied all ground failure effects. Even so, some general comparisons can be made. The present models were based on extremely detailed data which pinpoint the locations of actual ground failures. That alone is an improvement over
the model of Pike et al. (1994) which used the susceptibility decisions of an expert as the dependent variable. The detailed dependent variable also allowed for the use of a study area which is at least twice as large as those used in Bernknopf et al. (1988, 1993), Bernknopf and Soller (1994), and Pike et al. (1994). In addition, this study incorporates some important variables omitted from previous studies, such as depth to ground water, which was unavailable for use in the lateral spread study of Pike et al. (1994). The incorporation of a seismic variable is another improvement. Pike et al. (1994) estimated susceptibility to lateral spreading and so did not incorporate a seismic component. Bernknopf and Soller (1994) included probabilistic estimates of earthquake recurrence in their analysis of earthquake-triggered landslides, but such recurrence models cannot offer the precision of a seismic variable such as peak ground velocity. These examples demonstrate that the models developed in this thesis are not only an application of an existing method, but they are a refinement and an expansion of a hazard mapping tool.

**APPLICABILITY OF MODELS**

Two versions of ground failure models have been developed: a susceptibility version and an earthquake-specific version. The susceptibility version may be easier to apply in the future because earthquake-specific information is not required. This model could be used to predict ground failure resulting from any earthquake with an epicenter in the study area. In contrast, the second model only gives the probability of ground failure during another Northridge event; this stems from the incorporation of the peak ground velocity variable which describes the Northridge ground motions. To generate a similar probability map for another earthquake would require knowledge about the future event. The model could be run based on likely earthquake scenarios, but this introduces considerable uncertainty into the model. The probability estimates would be dependent on the occurrence of the earthquake scenario.

The two versions of the model are applicable to different circumstances. To predict ground failures in a future earthquake, the susceptibility model could be used in its present form, but the earthquake-specific model would have to be modified using an estimated peak ground velocity for a potential scenario earthquake. The addition of a seismic variable could improve the fit of the second model, but if inaccurate, it introduces much uncertainty.

It will be interesting to see if the actual equations can be transported to other alluvial basins. Bernknopf et al. (1993) transported equations estimated for landslides in Ohio
(Bernknopf et al., 1988) to a site in Virginia; these authors also estimated the probability of contamination of land surrounding a waste disposal facility in New Jersey and applied the equation to a site in Virginia. The validity of such extrapolations has not been proven. Furthermore, it is likely that the seismic phenomena being modeled are far too complex to expect extrapolations to be meaningful. Much more study is required to determine whether the equations developed in this thesis are portable.

**USES OF THE HAZARD MAPS**

Probabilistic maps which delineate seismic hazard are extremely useful products. Much of their power stems from their ability to communicate hazard to different groups of people. The models translate what a geologist knows is important about ground failure into a quantitative measure of hazard which is displayed in map form. These maps can be easily understood by scientists and nonscientists alike. They enable informed decisions for land-use zonation, emergency response planning, mitigation strategies, and the setting of earthquake insurance rates, to name a few examples.

Along with the potential power of such maps comes a responsibility to ensure they are explained in an appropriate manner. First and foremost, it should be recognized that the maps delineate hazard on a regional level. It is tempting to believe that the maps are extremely detailed, since a 1 hectare grid cell was used as a unit of analysis, but only the dependent variable warrants such detail. Other variables are approximations derived primarily from geologic maps. Facies changes within geologic units make the assignment of percent clay to an entire map unit only a rough approximation. Ground water levels fluctuate dramatically with precipitation or pumping changes. Because the end users of the hazard maps (nonscientist policy makers) may take the hazard categories as fact, it is the responsibility of the scientist to explain that the models are only as good as their constituent parts, and that the variables have uncertainty associated with them. The probabilistic maps are a first step in the quantitative delineation of hazard, but more refinements in the model are required. Accordingly, the maps can act as guides for the regional specification of hazard, but site-specific investigations should be conducted to definitively determine hazard potential at a given building site.
CHAPTER SEVEN: CONCLUSIONS

History has shown that ground failures can lead directly and indirectly to damage and fatalities. Geologists understand factors which influence ground failure occurrence, but this knowledge is often not in a form that can be used quantitatively in seismic safety planning. This thesis develops a model and refines a method for translating earth science information into quantitative hazard estimates which can be incorporated into seismic policies to prevent damage and casualties from ground failure in the future.

A model which estimates ground failure potential in the San Fernando Valley, CA has been developed using GIS and regression analysis. Predictor variables, chosen for their demonstrated influence on ground failure occurrence, include: age of the surficial geologic unit, average percent clay in the surficial geologic unit, average shear wave velocity, slope, depth to shallow ground water, and peak ground velocity. The dependent variable is described using the locations of sewer pipe, water pipe, and street damages caused by the 1994 Northridge earthquake. Information for all of the variables was compiled in a GIS, which was essential for the georeferencing, management, and synthesis of diverse datasets. Maps of the variables were converted into grids with one hectare cells. Each cell, containing information about all the variables specific to the cell location, is an observation for the regression analysis. The independent variables are continuous quantities, but the dependent variable could take only two values: \( y = 0 \) (no ground failure occurred) or \( y = 1 \) (ground failure occurred).

A probit regression model was used to 1) test the hypothesis that variables which influence ground failure location could be found, 2) identify the most significant variables for ground failure prediction, 3) determine the weights of those variables in ground failure prediction, and 4) provide quantitative, probabilistic estimates of ground failure hazard. After much experimentation, two equations were determined:

\[
\text{probability} = \Phi \{ 1.9522 - 0.00673 \text{ water depth} + 0.0123 \text{ clay} - 0.549 \ln(\text{shear wave velocity}) \};
\]

\[
\text{probability} = \Phi \{ -0.00858 \text{ water depth} + 0.0169 \text{ clay} - 0.354 \ln(\text{shear wave velocity}) + 0.0133 \text{ ground velocity} \}.
\]
The first equation describes the susceptibility of the San Fernando Valley to ground failure since only site parameters are included, while the second equation describes the potential of ground failure since an earthquake opportunity factor is incorporated. All of the variables are significant at the 99% confidence level. The signs of the coefficients make sense in terms of physical processes. As depth to ground water increases, the probability of ground failure decreases because 1) sediment is more difficult to liquefy at depth and 2) a deep water table enables shallow clays to dewater, thereby increasing their shear resistance. As percent clay increases, the probability of failure increases, reflecting the large number of failures which occurred in fine-grained units. As ln (shear wave velocity) increases, the hazard decreases because materials with faster shear wave velocities do not amplify seismic energy as much as those with slow shear wave velocities. As the peak ground velocity increases, the hazard increases because the ground is subjected to higher levels of strong motion. When the probabilities are mapped, they follow expected patterns. For example, the high probabilities in the susceptibility version occur where sediment is fine-grained and the water saturated; the high probabilities in the earthquake-specific version occur where high ground velocities coincide with susceptible units. Further evidence that the models provide reasonable estimates of hazard is found in the correspondence between observed failure ratios and predicted probability categories.

Given that southern California is overdue for other major earthquakes which are certain to induce ground failure, the maps presented here offer useful tools to guide decisions about seismic safety planning. The key to their power is their quantitative nature. Quantitative results facilitate communication between scientists who understand the hazard and the nonscientists who make decisions about public safety; the confusion inherent in qualitative hazard ratings is eliminated. In addition, these maps provide much more information than traditional, qualitative maps. This is essential when working in the decision framework of seismic safety. Resources are limited and a risk-free existence is unattainable, so choices about earthquake hazards must be made. These maps enable informed decisions. They can act as guides for such issues as land-use zonation, emergency response planning, mitigation strategies, and earthquake insurance rate determination. Besides being extremely useful, the maps were relatively quick and easy to create. Available data for the variables was used, so the time and cost of additional site-specific investigations was avoided. The models also can be updated to accommodate new information readily. Given the nature and the ease of production of these maps, they could prove to be quite useful in the regional delineation of hazard zones.
The models and method developed in this thesis have exciting implications for future work. The probability maps can expressed in economic terms. For each cell, the probability of ground failure can be multiplied by the property value within that cell. Such a derived map would represent the economic risk due to ground failure in the San Fernando Valley. This map would be directly applicable in the decision-making process since cost-effective measures are always favored. In addition to making derivative maps, it would be interesting to test the limits of the models. For example, the earthquake-specific model could be rerun using the peak ground velocity from the 1971 San Fernando earthquake. The predicted probabilities could be compared to the observed ground failures from that earthquake.

The models could also be used to determine the optimum scale for similar hazard studies. More detailed information could be gathered, and then the models could be reestimated incorporating the new information (e.g. Bernknopf et al., 1993). If the new information makes a large difference in the result, then one would conclude that more detailed data than were used in this study is required to accurately estimate the ground failure hazard. Alternatively, if there is not a large difference in the results, then one would be reassured that the models are reliable representations of hazard levels.

The most significant future task will be to apply the models and methodology elsewhere. Perhaps the models could be applied to other alluvial basins in southern California which are similar to the San Fernando Valley. While the portability of the exact equations remains to be seen, the methodology presented here can definitely be applied to seismically active regions worldwide. The method provides a vital translation of scientific information about earthquake hazards to quantitative estimates of hazard which can be incorporated into seismic safety decisions.
REFERENCES


McCarthy, R., 1997, Remarks made by the executive director of the CA Seismic Safety Commission, while giving a lecture on the Stanford University campus.


Wald, D.J. and Heaton, T.H., 1994, A multidisciplinary source analysis of the 1994 (Mw 6.7) Northridge earthquake using strong motion, teleseismic, and geodetic data: Seismological Society of America (abstract), Program for Northridge Abstracts, Pasadena, California.


Yerkes, R.F., 1996b, Preliminary geologic map of the Sunland 7.5' quadrangle, southern California: USGS Open-file report 96-87, scale 1:24,000.

Yerkes, R.F., 1996c, Preliminary geologic map of the San Fernando 7.5' quadrangle, southern California: USGS Open-file report 96-88, scale 1:24,000.


Youd, T.L., 1978, Major cause of earthquake damage is ground failure: Civil Eng., v. 48, no. 4, p.47-51.


