EFFICACY OF MATHEMATICAL MODELS USING PHYSICAL SOIL FACTORS IN DETERMINING LANDSLIDE HAZARD

A thesis submitted to the faculty of
San Francisco State University
In partial fulfillment of
The degree

Master of Science
In
Geographic Information Science

by

William Goedecke

San Francisco, California

May 2013

CERTIFICATION OF APPROVAL

I certify that I have read *Efficacy of Mathematical Models Using Physical Soil Factors in Determining Landslide Hazard* by William Goedecke, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirements for the degree: Master of Science in Geographic Information Systems at San Francisco State University.

Leonhard Blesius
Assistant Professor of Geography

Jerry Davis
Professor of Geography

EFFICACY OF MATHEMATICAL MODELS USING PHYSICAL SOIL FACTORS IN DETERMINING LANDSLIDE HAZARD

William Goedecke
San Francisco, California
2013

This research evaluates and compares the use of a terrain stability model with a statistical method in developing landslide susceptibility maps. The conceptual approach was to model slope stability in a geographical information system for a wide area. The study area is a 75.4 km^2 area due east of Tomales Bay in Marin County (California) characterized by rugged terrain underlain by competent bedrock to areas of hilly terrain with chaotic bedrock. The primary methods were a mathematical model (SINMAP – Stability Index Mapping) based on the infinite slope equation and a steady state hydrology and statistical models based on binomial variance (using logistic regression). The terrain stability model worked well in rugged terrain with well-defined drainages where SINMAP could model threshold saturation and resulting instability but did not work well in areas with less-rugged terrain and immature drainages. In such areas the statistical models offered greater detail regarding areas of potential hazard given the variance of the sample.

I certify that the Abstract is a correct representation of the content of this thesis.

_____ _____

Leonhard Blesius Date

ACKNOWLEDGEMENTS

I would like to acknowledge the work of my thesis advisor, Leonard Blesius, in helping me understand slope mechanics and physically-based mathematical models in evaluating hazard. He helped me comprehend the power of equations in explaining complex conditions without the need of lengthy explanations. Jerry Davis introduced me to the complexities of the ArcGIS software and the use of this software in developing models and in working with raster datasets. Seth Hiatt provided many enjoyable hours of discussion regarding spatial statistics, GIS and computer programming, plus just shooting the breeze. He was very helpful to me in providing a platform in which we could bring forth and share ideas. Marilyn Smulyan helped me get through the final lap with her feedback and encouragement. And I want to thank all the students that I encountered during my time at SFSU and to thank my other professors for the educational experience.

TABLE OF CONTENTS

List of Tables vii

List of Figures viii

I. Introduction 1

II. The Study Area 4

Geomorphic Setting 4
Geology 5
Soils 7
Climate 7
Landslides 9

III. Methods 12

Data Sources 12
Data Variance 15
Terrains 19
One-Dimensional Physically-Based Models 20
SINMAP Theory and Application 23
Logistic Regression 27
Comparison between SINMAP and Logistic Regression Models .. 29

IV. Results 30

Logistic Regression Terrain Based Models 30
Analysis of Subareas Using SINMAP and Logistic Regression 31

V. Analysis 42

VI. Conclusions 48

VII. References 51

LIST OF TABLES

Table Page

1. Descriptive statistics for input variables 17
2. Stability class definitions 25
3. Soft terrain logistic regression model coefficient values and probability output ... 30
4. Hard terrain logistic regression model coefficient values and probability output .. 31
5. SINMAP data for Black Mountain subarea 34
6. Confusion matrix, SINMAP results for Black Mountain subarea 34
7. Logistic regression model outputs for Black Mountain 35
8. SINMAP data for Tomales Bay east shore subarea 39
9. Confusion matrix, SINMAP results for Tomales Bay east shore subarea 39
10. Logistic regression model outputs for Tomales east shore subarea 40

LIST OF FIGURES

Figure Page

1. Research study area.....5
2. Average monthly rainfall 8
3. Rainfall totals in reference to years where landslide activity was significant.....8
4. Landslide habitats.....10
5. Scatter plot of all landslide point samples 16
6. Histograms of all landslide and random point samples 18
7. Study area mapped terrains, study subareas and sampling/validation areas 20
8. Slope area plot 26
9. Black Mountain, west Marin County.....33
10. SINMAP stability analysis for Black Mountain 34
11. SA plot for Black Mountain.....35
12. Logistic regression terrain models probability outputs Black Mountain.....36
13. Watershed subarea in Tomales Bay east shore watershed 38
14. SINMAP stability analysis for Tomales Bay East Shore subarea 39
15. SA plot for Tomales Bay east shore subarea.....40
16. Logistic regression terrain models probability outputs Tomales East Shore subarea 41
17. Model comparison for soft terrain logistic regression model and SINMAP, Black Mountain.....42

18. Hard terrain logistic regression model output, Black Mountain 43

19. Saturation values and slope instability Black Mountain 44

20. Slope Values and hard terrain model logistic regression Black Mountain..... 45

21. Slope Values and soft terrain model logistic regression Tomales Bay East 46

22. Slope Values and Hard Terrain Model Logistic Regression Tomales Bay east shore 47

23. Saturation values and slope instability Tomales Bay east shore 47

I. INTRODUCTION

Landslide hazard is important for Marin County, California, given the amount of housing and infrastructure built upon the steeper slopes and uplands of a geologically young landscape. The USGS conservatively estimated that in the county of Marin there were 442 damaging landslides over a 40 year period with an estimated loss of $71.35 million dollars (Crovelli & Coe, 2009). One landslide inventory recording impacts in eastern Marin County from a single storm (January 3rd through 5th, 1982) listed 124 landslides with 14 destroyed structures, 17 structures damaged, three deaths and three serious injuries (Davenport, 1984). During that storm it was estimated that there were roughly 18,000 debris flows for the entire San Francisco Bay Area resulting in 15 deaths (out of a total of 25 deaths from landslides and of 33 deaths overall) and damage to at least 100 homes (Ellen et al., 1988). These events clearly show that the hazard exists to both property and life and that debris flows are particularly dangerous given the rapidity of the flow and the potential force the flow may carry. Knowledge of areas susceptible to landslides provides the basis for mitigation of the hazard.

Maps of areas susceptible to landslides have been made through heuristic, statistical and physically-based methods. Statistical models provide a way to incorporate past landslide history comprehensively in an analysis of a landscape. Methods include multiple regression (Campbell et al., 1996), discriminant analysis (Dhakal, et al. 1999) and logistic regression (Dai et al., 2001; Ohlmacher & Davis, 2003). Statistical probability of a landslide event is the probability of an event occurrence given

the input parameters and, therefore, the output is constrained by the sampled landslide data. The completeness and accuracy of landslide inventory data is therefore essential for a valid statistical study (Van Westen et al., 2008). For this research logistic regression is used, given that the method provides a (constrained) range of values indicating landslide probability from 0 (not probable) to 1 (probable) which is a more useful output than linear regression methods which tend to be unconstrained (Dai et al., 2001; Ohlmacher & Davis, 2003) or discriminate analysis which does not state probability (Ohlmacher & Davis, 2003). Physically-based models focus on a measure of the factor of safety given soil shear strength parameters coupled with a hydrologic model (Santini et al., 2009). These approaches typically employ an infinite slope analysis with either a simplified hydrological model (Montgomery & Dietrich, 1994; Pack, et al., 1998) or a transient hydrological model (Baum et al., 2002; Wu & Sidle, 1995). For this research a model based on a simplified hydrological model is used (SINMAP - Stability Index Mapping) given that a transient model requires detailed rainfall data with high spatial and temporal resolution and more complete hydrological data for study area soils than is available (Godt et al., 2008).

In the San Francisco Bay Area landslide susceptibility maps have been developed for hazard awareness. Geologic field mapping of lands in and near towns and cities of Marin County including the location of relict landslides, was completed by the California Department of Conservation, Division of Mines and Geology, in the 1970s (Wagner, 1976). Such documents were created to inform land-use planning and engineering design (Wagner, 1976) because geology (Van Westen et al., 2008) and relict landsides (Ellen et al., 1988b) are considered factors in landslide susceptibility.

These documents are referenced in a relatively recent county planning document (Marin County, 2002). Bay Area regional work in identifying landslide susceptibility was done by the USGS in preparation for the 1997-1998 El Niño winter season (USGS, 1997). Maps created included areas deemed susceptible to landslides due to the presence of relict landslides (Wentworth et al., 1997), areas where debris flows could initiate given slope (20° or greater), and curvature (+0.01 and less) (Ellen et al., 1997) and maps of rainfall thresholds for debris flows to occur (Wilson & Jayko, 1997). These maps were to be used by local government along with the National Weather Service rain gauge data in monitoring more extreme storm events and resulting local landslide occurrence in areas considered prone to such phenomena (Wilson & Jayko, 1997).

Even though these maps provide general information regarding landslide hazard, they are not comprehensive in their analysis nor do they take advantage of current computing technologies. The intent of this research is to explore methods that lead to the development more effective susceptibility maps. The questions addressed here are: Would logistic regression and physically-based models provide greater and more useful specificity in mapping areas susceptible to landslides in Marin County than what has already been completed? How does the surface morphology of Marin County affect the output of these models? What are the limitations and strengths of each of these approaches and when is it more appropriate to use one or the other?

II. STUDY AREA

Geomorphic Setting

The study area (Figure 1) is located in western Marin County, California, due east of Tomales Bay and southwest of the city of Petaluma. It is 75.4 km^2 (13 km north to south and 5.8 km east to west) in area and is part of the coastal range of mountains that run from Humboldt County to Santa Barbara County. Elevations are near sea level in the southeast corner of the study area, where Lagunitas Creek enters a flood plain before draining into Tomales Bay to a maximum height of 421.5 m in the northwestern part. About half of the terrain is characterized by gently sloping grassy hills, with most of that land in the southern to southwestern part. Over 35% of the area, mostly in the northern part, has steeper and more mountainous morphology. The mountains trend west/northwest to east/southeast, forming narrow valleys often with the drainage following the structural grain of faults or folds (Norris & Webb, 1976). In the southeast corner is the 400 m tall Black Mountain which is characterized by distinctive rounded ridges and attached spurs. Much of the remaining land consists of alluvial basins.

The entire study area is part of the Tomales Bay watershed. There are three major divisions of this watershed that intersect with the area. Half of the area, located north, is a sub-watershed that drains into Tomales Bay through Walker Creek. The Walker Creek drainage is 194 km^2 in size of which the upper 20% resides in the study area. In the south to southeast there is a portion of the Lagunitas Creek sub-watershed and in the central-southwest there is an area that drains directly into Tomales Bay.

Figure 1: Research study area

Much of the lower hills and adjacent valley bottoms are predominantly grasslands. The upper reaches of south-facing slopes also tend to be grassy with the north-facing slopes tending towards sclerophyllous woodlands. Often the higher ridges have areas that are relatively bare or hard surfaces. Sclerophyllous chaparral may be found in the upper drainages of the steeper hillsides with oak woodland on the lower slopes. Riparian woodland can be found at the bottom of the steeper drainages and

along creeks. Outside of protected areas, land use has been and is currently agricultural with the primary activity being raising cattle for beef (MCCDA, 2004).

Geology

The entire study area is underlain by the Franciscan Complex, an accreted terrain of heterogeneous clastic sedimentary and volcanic rocks. The most common lithological component is greywacke sandstone which may be interbedded with shale. Other rocks found are reddish radiolarian chert, limestone, and mafic volcanic rock (pillow basalt and greenstone) (Norris & Webb, 1976; Sloan, 2006). The Franciscan complex is a remnant of the Franciscan trench which was formed as the ancient Farallon plate was being subducted under the North American plate. The rocks of the Franciscan complex are estimated to be 100 to 80 million years old. It is theorized that in this process semi-coherent blocks were episodically scraped off and accreted onto the North American plate. This material was eventually thrust up to form the Coast Range 26 to 20 million years ago as the Farallon plate was being completely subducted. Material stacked in this way resulted in the structurally highest rocks in the eastern side of the Coastal Range being the oldest. The Franciscan Complex can have either unsheared bedrock or bedrock of sheared and more highly fragmented mélange (Rice et al., 1976; Alt & Hyndman, 2000; Elder, 2001).

In the study area, the steep ridges are underlain by coherent bedrock of greywacke interbedded with shale, except for Black Mountain, which is underlain by bedrock of greenstone with some outcroppings of pillow basalt. These hard terrains are characterized by regular steep, sharp crests alternating with flutes. The area characterized by more gentle terrain is underlain by Franciscan mélange (Blake et al.,

2000). A mélange is composed of competent blocks of rock within weakly bonded and intensely sheared matrix rocks which can be somewhat soil like. These blocks of rock have a variety of sizes and placements and can be common and somewhat massive in the local area (Medley & Rehermann, 2004).

Soils

Soils in the steeper uplands tend to be thinner, to have low available water capacity and moderate permeability and to be subject to rapid runoff. These soils are often more gravelly near the ridge, thickening on side slopes and in the drainage. They are derived from sandstone and sometimes sandstone and shale with relatively low clay content and generally low plasticity. The Felton Variant-Soulajule complex soils on Black Mountain are somewhat different, being deeper with low to moderate clay content, low to greater plasticity at depth and moderate shrink-swell capacity. These soils have moderate to very high water capacity and are subject to moderate runoff. The expansive soils found in the more gently sloped areas underlain by Franciscan mélange tend to have greater clay content with low to high plasticity at depth. Available water capacity is generally higher but runoff may also be rapid if transmissivity is very slow (Kashiwagi et al., 1984).

Climate

Marin County is situated along the Pacific Ocean and is subject to the influence of the south-flowing cold California current whose cooling effect brings in coastal fog in the cool summers, in a general Mediterranean precipitation regime of wet winters and dry summers. Rainfall averages are higher on the coast and on higher elevations whereas temperature averages are higher inland closer to the San Francisco Bay than

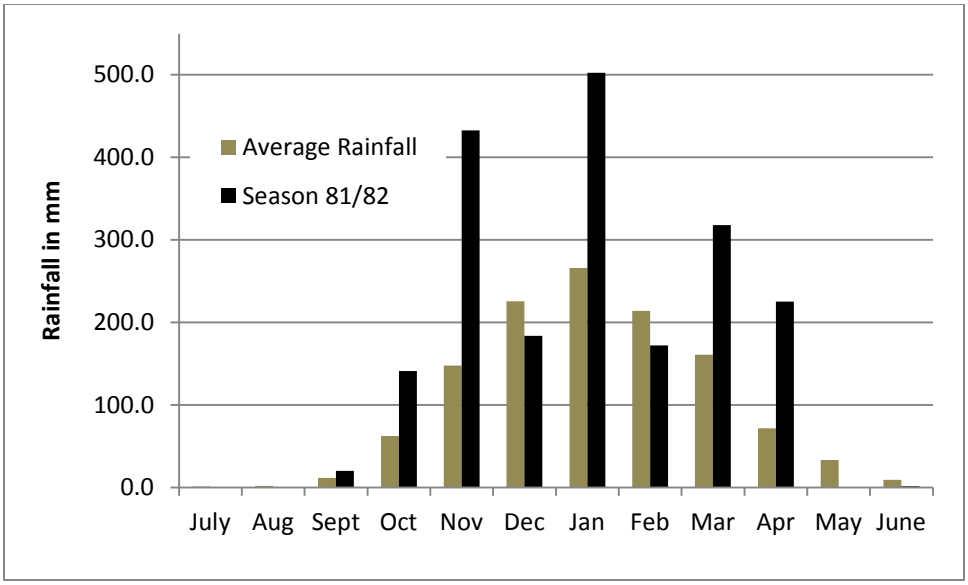

Figure2: Average monthly rainfall from the Kentfield Weather Station, Marin County compared to rainfall totals during the 1982 season (NWS, ND).

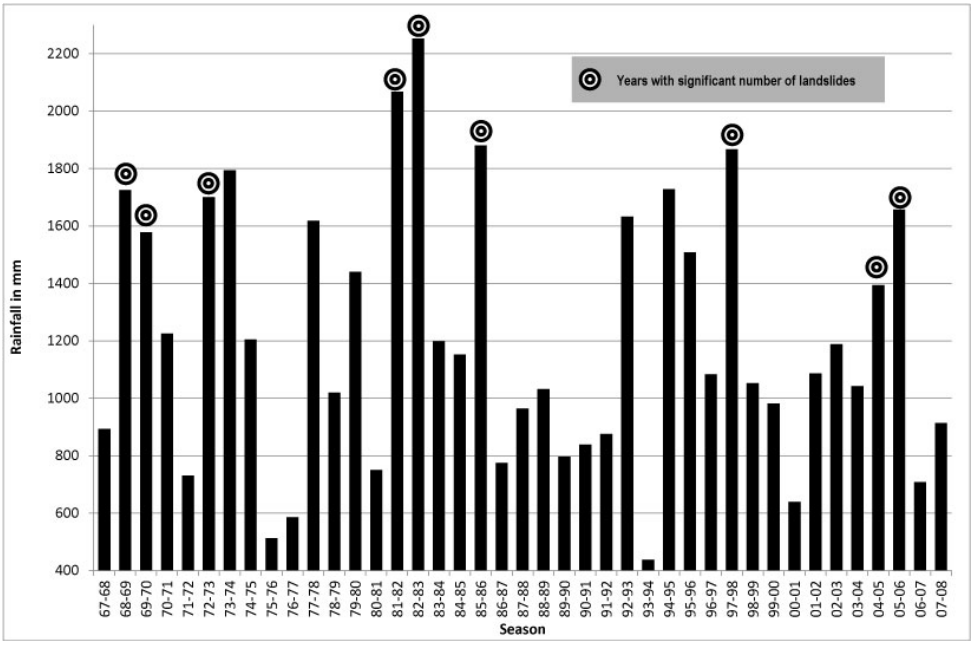

Figure 3: Rainfall totals in reference to years where landslides were deemed significant according to the Marin County Sheriff's Office of Emergency Services (Marin County, 2005).

on the coast. The average high temperature for the (inland) Kentfield weather station is 28.5°C (July) with the low at 13.1°C (January). The average high rainfall amount is 265.9 mm in January with an average low of 1.3 mm in July (Figure 2) (NWS, ND). Rainfall totals from the Kentfield weather station for a 40 year period are found in Figure 3, with periods of significant landslide activity marked.

Landslides

Landslides in the study area tend to be either fast-moving slides and debris flows or slow-moving slumps and earthflows (Ellen et al., 1988b). Slow-moving slumps and earthflows tend to occur in highly plastic (high clay content) and expansive soils and occur after prolonged periods of saturation when pore water pressure increases through accretion (Sidle & Ochiai, 2006). Areas where they occur are characterized by gently rolling hills and immature drainages. They were considered the principle hazard until the 1982 storm (Ellen et al, 1988). Debris slides and flows typically occur in steep terrain (>25°) during the height of storm intensity in a period of prolonged rainfall (Sidle & Ochiai, 2006); in the 1982 storm debris flows were found most often in slopes between 27.5° to 37.5° (Ellen et al., 1988b).

Ellen et al. (1988b) classified landslides inventoried on Hicks Mountain (a mountain in western Marin County) according to the natural condition, or habitat, in which the landslide would occur (Figure 4). Almost half of the landslide scars were found at the top of first-order drainages on the hillsides above the flutes in the steeper uplands (habitat 1). Another 37% of the landslide scars on Hicks Mountain were found in areas without well-defined erosional surfaces that slope down to nonalluviated drainages (habitat 2). About 13% of the landslide scars were found in areas without

well-defined erosional surfaces that abut alluviated surfaces (habitat 3). A sampling of other landslide scars found in the steeper uplands from the January 1982 storm found an even greater percentage (71%) of landslides occurred in habitat 1 (Ellen et al., 1988b). This phenomenon may reflect the presence of colluvial hollows that sit on top of the drainage and may be filled scars from relict landslides. The subsurface morphology would tend to concentrate the water in the hollow (Dietrich et al., 1982). Landslide scars were almost never found in the drainages, although debris flow trails would be found there.

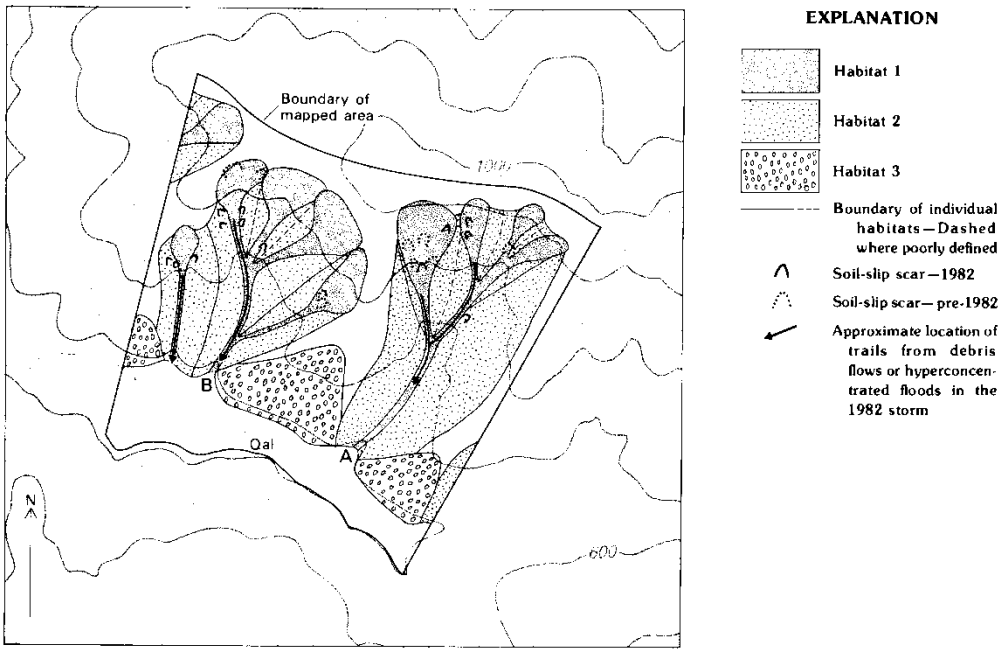

Figure 4: Landslide habitats (Ellen et al., 1988b).

For debris flow activity in the San Francisco Bay area, rainfall has to be of substantial duration and intensity (Cannon & Ellen, 1985; Cannon, 1988). Slow and steady rainfall does not produce debris flows, as the amount of water passing through the established drainage does not create positive pressure heads (Campbell, 1975).

Mature drainages develop in accordance to the historical climate and thus establish patterns of drainage that are sufficient for normal throughput (Carlston, 1963). However, when soils near or at saturation are subject to intense bursts of rainfall, areas where water concentrates may accumulate a perched water table in which water entering will be greater than the amount that is exiting (Campbell, 1975; Ellen et al., 1988c). With continuing intense rainfall positive pressure heads in the perched water table will develop and a debris slide and flow will follow (Ellen et al., 1988).

An analysis of the storm rainfall patterns during the January 1982 storm in Marin County found that the density of debris flows in the entire San Francisco Bay Area increased substantially with rainfall totals above 250 mm and where storm rainfall totals normalized to mean annual precipitation was more than 30 percent. Pre-storm totals above 300 mm and below 400 mm (not normalized to historical patterns) were found to be correlated to landslide density (Mark et al, 1988).

III. METHODS

The initial approach taken was to look at data sources and then at the distribution of phenomena using samples taken from debris flow and random points in order to understand the difference between what is typical for the study area to what would be found at a debris flow initiation point.

Data Sources

Data for geomorphic surfaces such as elevation, slope and curvature were derived from digital elevation model (DEM) datasets acquired from the Golden Gate LiDAR Project. This project consisted of an aerial survey of over 2162.6 km^2 of Marin and San Francisco counties and parts of San Mateo and Sonoma counties between April 23rd, 2010 and July 14th, 2010. Data collected were aerial photography, LiDAR data and hyperspectral data. LiDAR data was collected at an average density of 2 points per square meter. A 32-bit bare earth surface DEM with a grid resolution of 1.0 meter was derived from the processed LiDAR data (Hines, 2011).

Landslide inventory data is essential for both the calibration and the evaluation of the SINMAP and statistical models. Landslide inventory data was acquired from a study completed by the USGS in 1988 (Professional Paper 1434 - Landslides, Floods, and Marine Effects of the Storm of January 3-5, 1982, in the San Francisco Bay Region, California). Paper maps were completed for the USGS study, with debris slide/flow initiation points and debris flow trails marked. These points and trails were drawn from black and white stereoscopic pairs of vertical aerial photographs of the study area taken midday, January 6th and 7th, 1982 (two days after the historic 1982 storm). The drawn

maps were scaled at 1:24,000 for the entire 10-county Bay Area and at 1:12,000 for the study area (Ellen et al, 1988b). The maps, aerial photographs and high resolution (1 meter) orthorectified images of the land cover, dated 2004, available from a county government website (Marin Map, 2004) were used to digitize the locations of debris slide scarps (source locations for the debris flows). A total of 808 landslides were digitized in the study area.

Due to the shadows created by the low angle of the midday winter sun it was difficult for the USGS to identify landslides in north-facing slopes from the aerial photographs. The USGS concluded in their 1988 report that landslides on north-facing slopes were underrepresented in the resulting inventory (Ellen et al, 1988b). Of the 808 digitized slides, only 33 were found to have a northern aspect. After the initial exploration of the data, these slides were removed from the sample and random sample points were only generated for slopes without a northern aspect.

For the analysis certain parameters had to be developed. Before any datasets were derived the original DEM received from the Golden Gate LiDAR Project was run through a low-pass filter in order to reduce potential noise in the data. The drainage surface referred to in this study is a raster dataset that represents the Euclidean distance to hillside drainage in its base 10 logarithm derivative. Hillside drainage is a polyline shapefile which represents a potential line of drainage given water flow accumulation. This polyline shapefile was derived using ArcGIS 10.x hydrological modeling tools. Slope and curvature were derived from the DEM using the slope and curvature ArcGIS Spatial Analysis tools.

In order to understand the general distribution of land cover over the study area and how land cover differs at landslide initiation points from what is typical for the area, the 1 meter land cover imagery from Marin Map (Marin Map, 2004) was first classified in ArcGIS 10.x using an ISO cluster unsupervised classification with a specification of 50 classes. Observation of the classification showed that areas in the higher classification bins included grassy to bare land cover whereas areas with low numbered classification bins included dark forest land cover, water and shadow. After this initial exploration, land cover was approached as categorical data and classified with a specification of 5 classes (greener and thinner forests, chaparral and brown grasses, green grass, hard surfaces, bare ground or bright surfaces and not significant (water and dark forests were found to be not significant)).

Slope, curvature, classification data and the Euclidean distance drainage raster were sampled using digitized landslide initiation points and generated random points. An equal amount of random points was used to landslide points, given that an unequal random to landslide sampling would produce bias to the more common event (Lobo et al., 2007). Slope and curvature data was sampled using bilinear interpolation in order to take into account the surrounding cells.

Other available data included elevation, geology and soil classification. Geology and soil data was acquired from the county government GIS website (Marin Map) (MarinMap, ND) and elevation data was derived from the Golden Gate LiDAR dataset. Using the generalized linear modeling (glm) program in R (open-source statistical programming language and environment (http://www.r-project.org/)) with binomial

variance in a logistic regression analysis, based on landslide and random point sampling, it was found that these variables did not have significance.

Data Variance

The pairs chart (Figure 5) does not show any obvious correlation between the different parameters other than some correlation that would naturally exist between curvature and drainage. The histograms (Figure 6a) and sample statistics (Table 1) from slope data for landslide and random points both have the mean value close to equal the median value, indicating a tendency towards normality. The kurtosis value (as calculated in Excel 2010) for landslide points is slightly greater than 3, indicating normality, whereas the kurtosis value for random point slope values is negative, indicating a broad distribution of values. Landslide point values are negatively skewed, whereas random point data is positively skewed, as it takes in values from flat areas. Slope values for landslide points are more tightly clustered around the mean (35°), with a lower standard deviation and a smaller range than that for random points (mean of 19°). Landslide point data for curvature has similar mean and median values and a negative kurtosis value, indicating a somewhat uniform and broad distribution. This contrasts to the very peaked distribution of values for random points (Figure 6b, Table

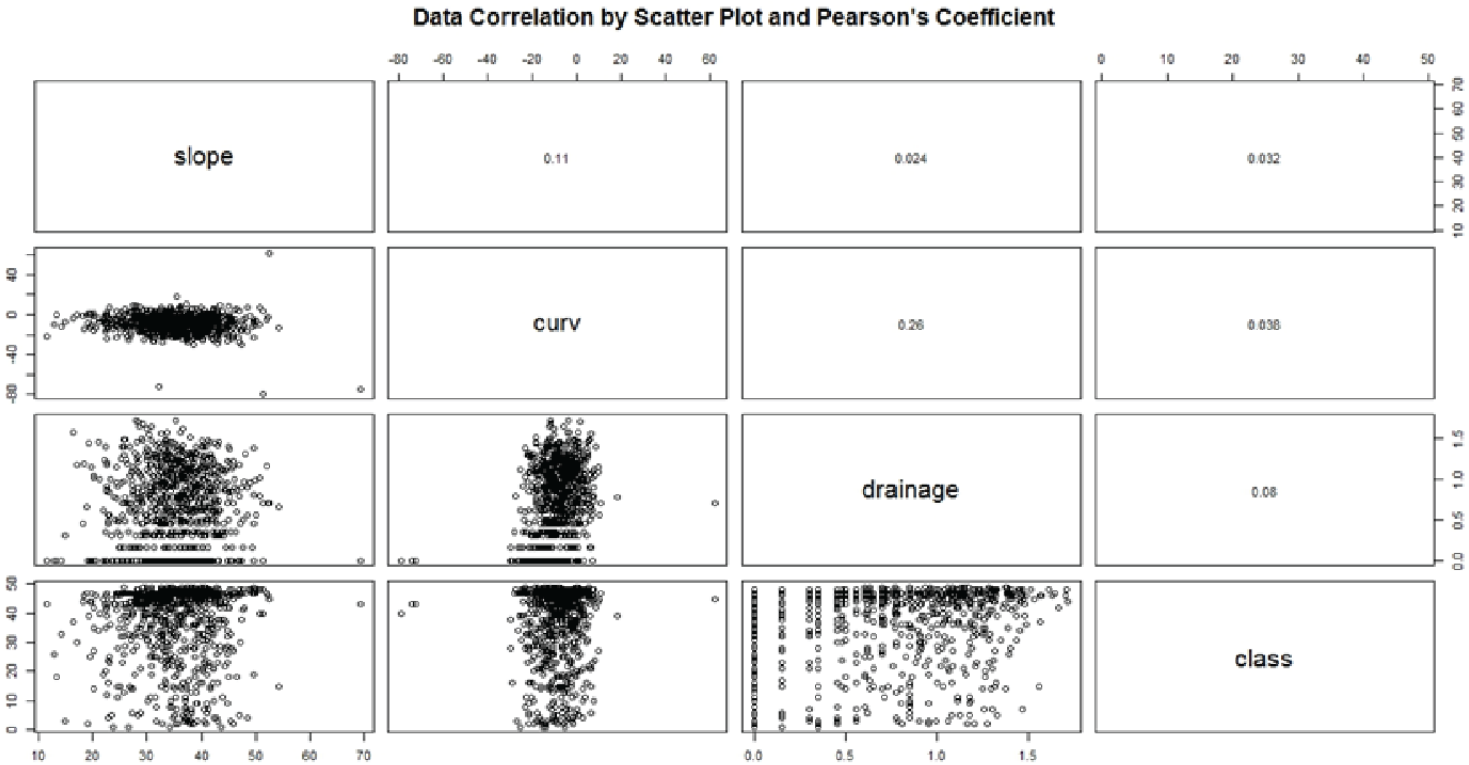

Figure 5: Scatter plot on all landslide point samples

	slope		*curvature*	
	landslide pts	*random pts*	*landslide pts*	*random pts*
Mean	35.1850074	19.4375131	-8.537791781	0.125146048
Median	35.541401	18.42675	-8.43903	0.2044825
Std Dev	6.694399501	10.95227199	7.690553454	4.821758269
Kurtosis	0.311124707	-0.792715947	-0.315173806	8.343496196
Skewness	-0.285149227	0.204984461	-0.118433332	-1.504534895
Range	42.675898	47.928683	40.408199	44.864901
Minimum	11.5589	0.017919	-29.858299	-28.319901
Maximum	54.234798	47.946602	10.5499	16.545
	drainage		*class*	
	landslide pts	*random pts*	*landslide pts*	*random pts*
Mean	8.553996744	16.09545146	37.13242574	26.96144279
Median	5.724465	12.18285	43	30
Std Dev	8.418908493	13.84130384	13.0512682	14.62055585
Kurtosis	4.028833851	1.089737772	0.633139818	-1.167650419
Skewness	1.75048423	1.146772064	-1.325705342	-0.369689231
Range	53.1119	69.708504	48	48
Minimum	0	0	1	1
Maximum	53.1119	69.708504	49	49

Table 1: Descriptive statistics for slope, curvature, drainage and land cover classification

1). The mean value for landslide points is -8.5, indicating concave curvature, whereas the mean value for random points is .13, indicating a slightly convex surface. Drainage (Figure 6c) has an overall positive skew for both landslide and random point data, with landslide point data having a larger skew and a smaller range (Table 1). Values representing land cover classification from the landslide point sample has a high negative skew and a positive kurtosis (Figure 6d, Table 1). This contrasts with the random point land cover classification sample which has a low, negative kurtosis indicating a flat, somewhat uniform, distribution. Both the mean and median bin values

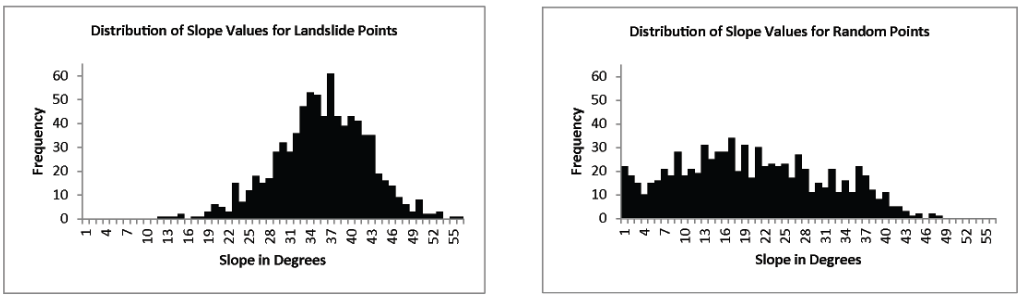

Figure 6a: Comparison of data distribution for slope degree, landslide and random point samples.

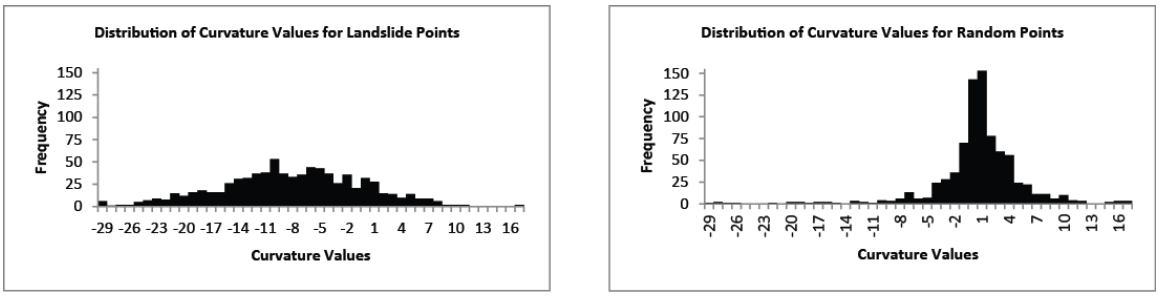

Figure 6b: Comparison of data distribution for curvature, landslide and random point samples.

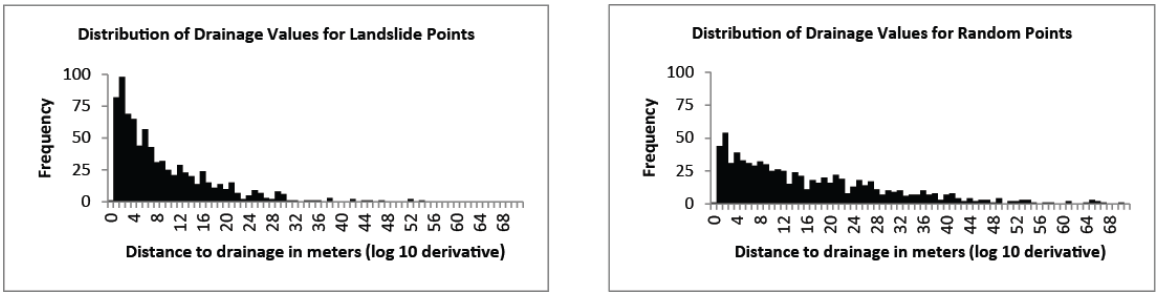

Figure 6c: Comparison of data distribution for distance to drainage values, landslide and random point samples.

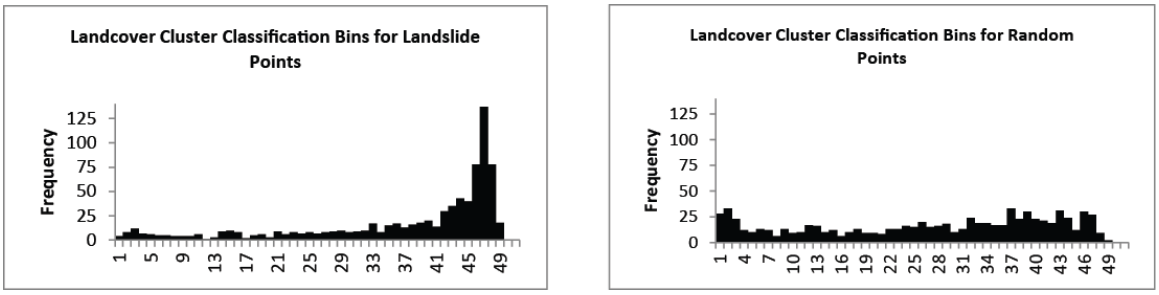

Figure 6d: Comparison of data distribution for vegetation classification bin values, landslide and random point samples.

for landslide point samples are higher than that for the random sample. The sampling indicates that the landslides found in the inventory tend to occur in grassy and bare ground areas.

Terrains

In order to understand how the statistical model works given the different input parameters and to have a common terminology in describing the land surface for both the statistical and the SINMAP models, the concept of terrains, based on surface morphology, is used. A hard terrain is steep topography with regularly placed ridges and flutes underlain by unsheared bedrock (often interbedded sandstone and shale) and covered by a granular soil. These areas have well-defined drainages. A soft terrain has gently rolling topography underlain by a highly sheared mix of bedrock materials within a matrix of crushed rocks covered by a clayey soil. These areas are characterized as having immature drainages. Intermediate terrain has features of both hard and soft terrain. In these areas, there are pockets of well-defined drainages interspersed with rolling hills (Ellen et al., 1988b). Terrain mapping based on these qualifications of the study area was done by the USGS (Reneau et al., 1988) and they were hand drawn on a paper map. The terrain markings were digitized for the study. Afterwards, parts of the study area classified as hard to hard intermediate terrains were reclassified to hard terrain and soft to soft intermediate terrains were reclassified to soft terrain. Figure 7 shows the hard and soft terrains with landslide points. Looking at the figure 7 suggests that debris flows tend to occur in predominately hard terrain.

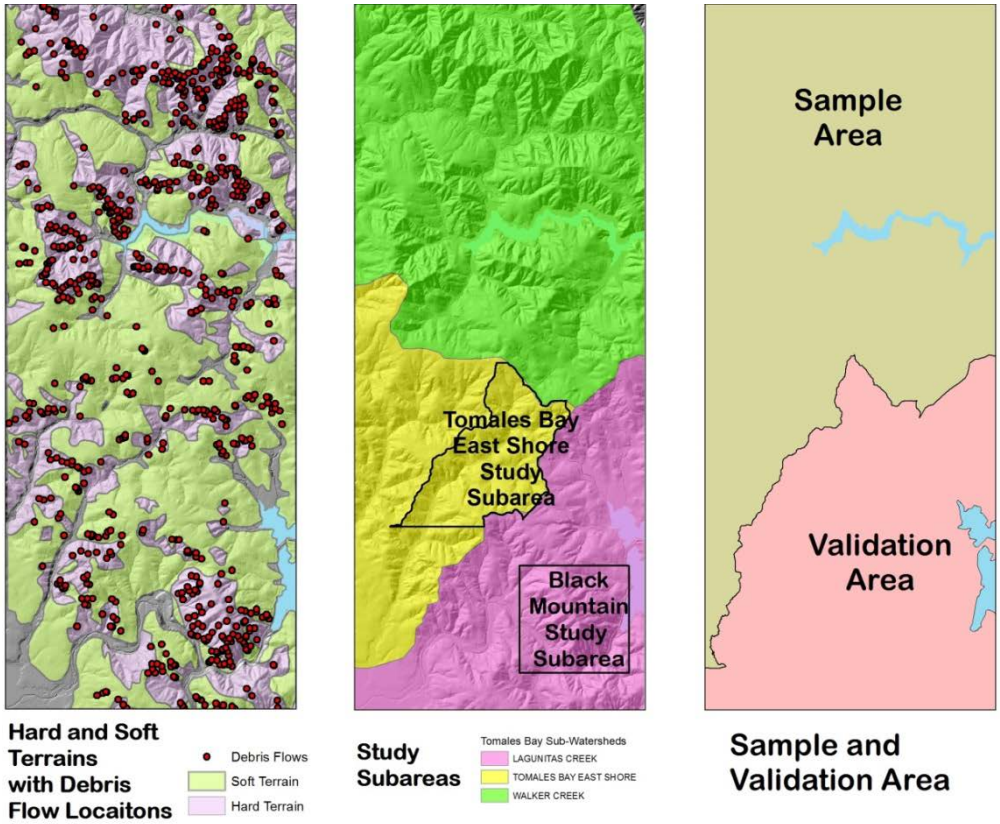

Figure 7: Study area mapped terrains, study subareas and sampling/validation areas

One-Dimensional Physically-Based GIS Models

In order to determine the Factor of Safety (FS) a limit equilibrium analysis was done. Limit equilibrium analysis determines the equilibrium between the shear stress or disturbing forces and the shear strength or the stabilizing forces. This approach uses a one-dimensional simplification of slope hydrology. Specific events in time are not considered (such as a day of heavy rainfall) other than to consider a worst-case scenario. The hydrology of the slope is assumed to be both isotropic and homogeneous in that the saturated hydraulic conductivity of the vertical component matches that of the horizontal component (only saturated flow is considered). The heterogeneity of slope

hydrology represented by wedge-shaped unsaturated zones and the resulting variation of different pore-water pressure heads is not accounted for (Rulon & Freeze, 1985).

One way to measure slope stability based on limit equilibrium principles is to use the infinite slope equation. The infinite slope equation assumes a sliding surface parallel to the ground surface. The term 'infinite slope' is applied to uniform slopes with a shallow soil mantle in relation to the length of the slope and a potential slip surface parallel to the soil surface (Sidle et al., 1985). The equation for the infinite slope equation is defined as the FS ratio (not accounting for groundwater):

$$FS = \frac{c' + (P_s\ z\ cos^2\ \theta - u)tan\emptyset'}{P_s\ z\ sin\theta cos\theta} \quad (1)$$

where P_s is the bulk density of soil, *z* is the soil thickness, *θ* is the slope angle, Ø is the soil friction angle, *u* is the pore water pressure and *c'* accounts for effective cohesion. The nominator accounts for effective normal stress whereas the denominator accounts for shear force. For fully saturated cohesionless soils the FS can be simplified as:

$$FS = \frac{(P_s - P_w)}{P_s}\frac{tan\emptyset}{tan\theta} \quad (2)$$

where P_w is the bulk density of water. This is due to the assumption that $tan\emptyset$ is the principle strength component and $tan\theta$ represents shear gravitational force (Graham, 1984).

Hydrology is simplified for the one-dimensional mathematical models used in GIS for slope stability studies as:

$$W = {}^{RA}/_{bT} \sin\theta \quad (3)$$

Here, saturation, or wetness (*W*), is determined by taking into account the area upslope (*A*) from a point on the slope (*b*) given recharge (*R*), soil transmissivity (*T*) and the steepness of the slope (*sin θ*) (Montgomery & Dietrich, 1994). Note that soil transmissivity is hydraulic conductivity multiplied by soil thickness (Pack, *et al*., 1998).

Montgomery and Dietrich (1994) employed the infinite slope equation with a simplified hydrological model in the development of a process-based model named SHALSTAB (Shallow Landsliding Stability Model). SHALSTAB is a relatively simplistic model that is used as a way to isolate topographic controls on slope stability (Montgomery & Dietrich, 1994). SHALSTAB provides a critical effective rainfall value in which the slope becomes unconditionally unstable (FS equals 1) (Montgomery & Dietrich, 1994; Meisina & Scarabelli, 2007) and thus can be considered deterministic. Pack et al. (1998) created the SINMAP model (Stability Index Mapping) which is similar to SHALSTAB in that it is based on directional flow and flow accumulation (Sidle & Ochiai, 2006) and it also employs an infinite slope equation assuming a steady state hydrology (Pack, et al., 2005). SINMAP provides a stability index with each breakpoint representing lower and upper thresholds of a qualitative ‘predicted state’. Probabilities are distributed in a uniform manner and range from ‘Stable slope zone’ to ‘Defended slope zone’ (Pack, et al., 1998; Meisina & Scarabelli, 2007).

SINMAP Theory and Application

SINMAP was created to map slope stability in broad areas of forests in British Columbia (Pack, et al., 1998; Pack et al., 2005). The function of SINMAP is to identify areas prone to instability where slides and flows may be initiated. Like SHALSTAB, SINMAP was created to be used in areas where only sparse information is available and to be used beyond the local scale. The model output is not meant to be taken as a precise quantitative measure of stability. Instead it is meant to show relative hazard and should be used in conjunction with other terrain stability mapping methods (Pack et al., 1998).

SINMAP incorporates topographic patterns, saturation and physical soil properties in order to primarily derive a stability index map, as well as other maps, including a map showing saturation given input parameters. The tool also provides a way of modeling a landslide inventory given the derived stability index summary statistics and therefore can be calibrated by adjusting parameters for a 'best fit' with the mapped landslide inventory points.

The stability index (SI) for the minimum FS, worst-case value is found through the following equation:

$$SI = FS_{min} = \frac{C_1 + cos\theta\,[1 - \min\left({}^{R}/_{T_2}\frac{a}{sin\theta}, 1\right)\frac{P_w}{P_s}]\,tan\emptyset_1}{sin\theta} \qquad (4)$$

For the worst-case scenario, C_1 is the minimum cohesion value, ${}^{R}/_{T_2}$ (water recharge over soil transmissivity) is the highest value, *a* is the specific catchment area and $\emptyset_1$ is

the smallest soil friction angle. Recharge or *R* is in m/hr. The value for $\frac{P_w}{P_s}$ is set to a constant (.05). After saturation $\left({R}/{T_2}\frac{a}{sin\theta}, 1\right)$ is reached, overland flow occurs.

The SI for the maximum FS, best-case value as found through the following equation:

$$SI = FS_{max} = \frac{C_2 + cos\theta\,[1 - \min\left({R}/{T_1}\frac{a}{sin\theta}, 1\right)\frac{P_w}{P_s}]\,tan\emptyset_2}{sin\theta} \quad (5)$$

For the best-case scenario, C_2 is the maximum cohesion value, ${R}/{T_1}$ (water recharge over soil transmissivity) is the lowest value and $\emptyset_2$ is the largest soil friction angle (Pack *et al*., 1998).

SI values represent a range of FS values from 1.5 (unconditionally stable) to 0 (unconditionally unstable) (Table 2). Areas labeled ‘stable’ are areas that would not fail given the conservative end of the input parameter range. The terms ‘lower threshold’ and “upper threshold’ represent a qualification of instability, either less than 50% potentially unstable or greater than 50% potentially unstable respectively. Areas labeled ‘defended’ are areas where SINMAP cannot model stability. These rankings are qualitative, as they can be calibrated given different input parameters (Pack et al., 2005).

After the stability index map is created SINMAP has a function to create a slope area (SA) plot. The SA plot tool generates random points (2,000 is the default) and displays these points along with the landslide points. The graph plots the tangent of the slope on the X axis and the area of moisture accumulation on the Y axis (Figure 8).

Condition	Class	Predicted State	Parameter Range	Possible Influence of Factors Not Modeled
SI > 1.5	1	Stable slope zone	Range cannot model instability	Significant destabilizing factors are required for instability
1.5 > SI > 1.25	2	Moderately stable zone	Range cannot model instability	Moderate destabilizing factors are required for instability
1.25 > SI > 1.0	3	Quasi-stable slope zone	Range cannot model instability	Minor destabilizing factors could lead to instability
1.0 > SI > 0.5	4	Lower threshold slope zone	Pessimistic half of range required for instability	Destabilizing factors are not required for instability
0.5 > SI > 0.0	5	Upper threshold slope zone	Optimistic half of range required for instability	Stabilizing factors may be responsible for stability
0.0 > SI	6	Defended slope zone	Range cannot model stability	Stabilizing factors are required for stability

Table 2: Stability Class Definitions (Pack et al., 1998)

Different regions represent the threshold values for the FS and points are plotted given the slope, area of accumulation and region. Additional boundary lines indicate whether or not the area is saturated given slope and catchment area. The FS value is derived by:

$$FS = \frac{C + \cos\theta(1 - \frac{z_w}{z}\frac{P_w}{P_s})\tan\emptyset}{\sin\theta} \tag{6}$$

where $\frac{z_w}{z}$ is the relative wetness (height of saturated zone divided by soil thickness).

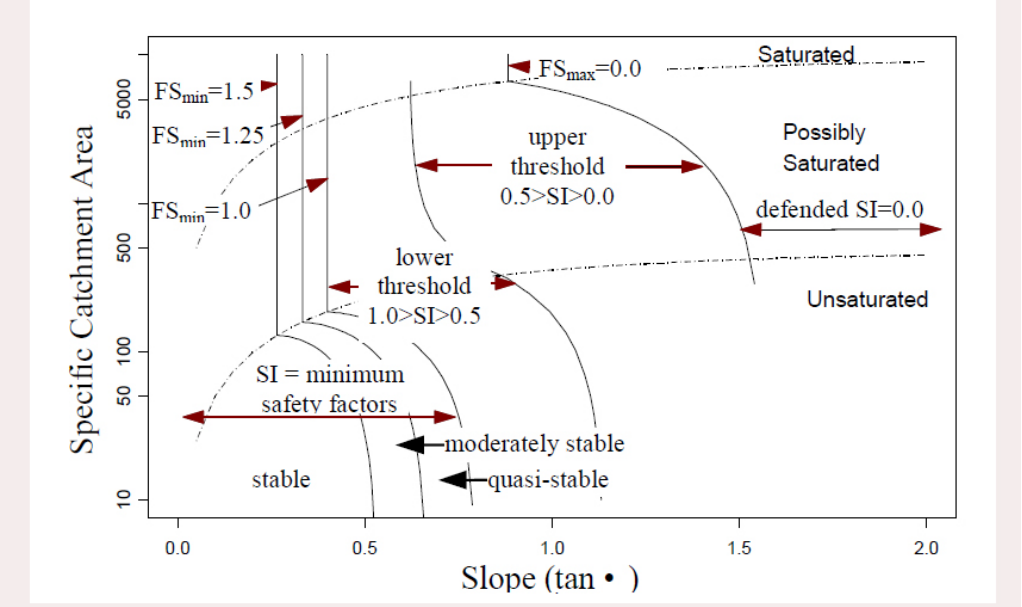

Figure 8: Slope Area Plot (Pack *et al*., 1998)

Grid size processing capacity for SINMAP is limited to 7000 x 7000 cells (Pack et al., ND). Lower cell counts resulted in faster SINMAP completion rates. Therefore, for the purpose of this study, SINMAP was run on two subsets of the study area (Figure 7). Sub-setting of the study area was based on a similarity of terrain. An area that is composed of a northern part of a valley in the Tomales Bay east shore watershed was selected as representative of soft terrain as it is mostly composed of soft terrain with most soils classified in the soil survey map as containing swelling clays. Black Mountain was the second area selected as the mountain is mostly a hard terrain with well-defined strike-ridges interspersed with incised flutes and a soil mantle composed mainly of sandy clay with low plasticity (NRCS, ND).

SINMAP was created as a plugin for Environmental Systems Research Institute, Inc. (ESRI) GIS application ArcView and ArcGIS 9.x (Pack et al., 2005). There is no

SINMAP plugin for the ArcGIS versions above 9.x. For this research, SINMAP, version 2.0, was utilized in ArcGIS 9.3.1. All other GIS processes employed were run in ArcGIS 10.x.

Logistic Regression

Another approach to determining susceptibility to debris slides and flows is to look at what characterizes those slopes that are prone to failure. Conditions that lead to slope instability may be inherent to the landform and may make it more likely for an area to be subject to episodic mass-wasting when a threshold is crossed (Schumm, 1979). These conditions can be sampled from a landform and then analyzed statistically given the historical record in order to determine susceptibility.

Logistic regression is a statistical method that provides a (constrained) range of values indicating landslide susceptibility from 0 (not susceptible) to 1 (susceptible) (Dai et al., 2001). The log-likelihood (probability) is expressed as:

$$\Pr(event) = \frac{1}{1 + e^{-z}} \tag{7}$$

where $z = b_0 + b_1x_1 + b_2x_2 \ldots + b_nx_n$ (with b_0 being the intercept, b_n the slope coefficient for the nth independent variable and x_n the independent variable). The probability being defined in this research is the likelihood of the slope being susceptible to a landslide given the input parameters. The parameters to be used as coefficients in the logistic regression model are those most likely to produce observed results (maximum likelihood method) (Dai, *et al.*, 2001; Ohlmacher & Davis, 2003;

Pradhan, 2010). The input parameters are considered to be independent to the resulting dependent variable (Can *et al.*, 2005).

In this study the dependent probability of instability is determined by slope, curvature, drainage and land cover classification. Slope, curvature and drainage are continuous variables, whereas land cover classification is categorical. The logistic regression model using these variables is stated as:

$$\frac{1}{1 + e^{-(\beta_0 + \beta_1 slope + \beta_2 curvature + \beta_3 drainage + \sum \beta_k \, classification_k)}} \tag{7}$$

where $\beta_k classification_k$ is the coefficient value for land cover classification *k.* Since each cell value will have only one land cover classification, the indicator value for the present land cover classification is set to 1, whereas for all other land cover classifications the indicator value is set to 0. Slope, curvature and drainage are variable, whereas land cover classification is constant for the specific classification (Ohlmacher & Davis).

A generalized linear modeling program (glm) in R was employed to derive coefficients for the logistic regression model based on sampled data from equal numbers of random points and landslide points in the sample area (Figure 7). The model output provides estimated coefficients for each parameter and the intercept, with associated probability values. The intercept is the log –odds of a landslide given the input data, converted into odds by using the exponential function, with each coefficient a measure of the contribution of the associated parameter to the probability of slope instability (Can et al., 2005).

Hard and soft terrains were sampled separately in the northern part of the study area in order to create a statistical model for each. The logistic regression model was then run in ArcGIS 10.x as described in equation 8 above with a resulting probability raster. This probability raster was then sampled from an equal number of landslide and random points in the validation area (Figure 7) for the purpose of validating the result.

Validation on the probability output was done by examining the sensitivity and specificity of the model. Taking the cutoff point at 50% (Lobo et al., 2007), sensitivity is indicated by the percentage of landslide points with values equal or above the 50% cutoff point (true positives). Specificity is 1 minus the number of random point sample data with values above that cutoff point (false positives) (Das et al., 2010; Zweig & Campbell, 1993). A polygon feature file was derived from the probability raster based on the 50% cutoff point in order to visually examine the results of different models.

Comparison between SINMAP and Logistic Regression Models

Confusion matrixes were used to compare SINMAP and Logistic Regression models in order to capture and compare true and false positives and true and false negatives, thus showing sensitivity in relation to specificity. Polygon features derived from the logistic regression models showing greater probability for landslides were then overlaid onto the SINMAP SI raster output in order to identify patterns of similarity and difference.

IV. RESULTS

Logistic Regression Terrain Based Models

The glm function output for the soft terrain model indicated that drainage and land classification for hard surfaces were insignificant (p-value above .05) (with brown grasses/shrubs, green grass and bare ground or bright surfaces significant). This may have to do with the lack of well-defined drainages and less rock outcroppings and other hard surfaces in the hilly areas. Slope, curvature and areas of grass and small shrubs were most significant. The derived coefficients for significant values were used in the logistic regression model to derive a probability surface. Probability values for both landslide and random points were then sampled from the validation area (Figure 7) to determine model sensitivity and specificity. As indicated in the results (Table 3), the model output has slightly greater specificity than sensitivity (with fewer false positives than false negatives). The model values decreased deviance by 35% below what would be found using just the intercept (null deviance).

Coefficient Values		
	Estimate	Pr(>|z|)
(Intercept)	-8.78689	6E-08
Curvature	-0.26901	6.09E-07
Slope	0.26536	2.5E-10
Hard surfaces	2.31864	0.0175
Brown grasses, shrubs	1.81878	0.0145
Green grass	1.64579	0.0221
Null deviance:	329.94 on 237 degrees of freedom	
Residual deviance:	115.97 on 230 degrees of freedom	

Classification			
0: absence 1: presence		**prediction**	
	p=0.5	0	1
data	0	228	28
	1	35	221
		Sensitivity:	0.8633
		Specificity:	0.8906
		≥ 0.5:	0.877

Table 3: Soft terrain logistic regression model coefficient values and probability output

The glm function output for the hard terrain model found slope, curvature, drainage and 4 land surface classifications all significant. As with the soft terrain model, probability values for both landslide and random points were then sampled from the validation area (Figure 7). The results (Table 4) show the model to be more specific than sensitive. The model values decreased deviance by 46% below what would be found using just the intercept (null deviance).

Coefficient Values		
	Estimate	Pr(>|z|)
(Intercept)	-6.04059	< 2e-16
Curvature	-0.12476	< 2e-16
Slope	0.14463	< 2e-16
Drainage	-0.94971	9.56E-05
Hard surfaces	1.65313	4.51E-07
Brown grasses, shrubs	1.63949	4.35E-08
Green grass	2.40895	4.14E-14
Bare ground or bright	3.18658	1.26E-12
Null deviance:	1067.4 on 769 degrees of freedom	
Residual deviance:	577.9 on 762 degrees of freedom	

Classification			
0: absence		**prediction**	
1: presence	*p=0.5*	0	1
data	0	252	4
	1	82	174
		Sensitivity:	0.6797
		Specificity:	0.9844
		≥ 0.5:	0.832

Table 4: Hard terrain logistic regression model coefficient values and probability output

Analysis of Subareas Using SINMAP and Logistic Regression

The first study area subset completed was Black Mountain (Figure 9). This area is principally composed of a hard terrain of well-defined and rounded strike ridges interspersed with incised flutes. Lagunitas Creek borders the mountain to the south. According to the SSURGO soil database, the soil is mainly classified as Felton Variant-Soulajulie Complex (NRCS, ND). Using the Unified Soil Classification System (USCS) textural classification, the soil's classification is mainly CL (sandy clay of low plasticity).

Values for the bulk density of soil, soil cohesion and soil friction angle were chosen based on reported constitution of the soil as well as consideration that the model tends to over-predict (Messina & Scarabelli, 2006). Bulk density was set at 1500 kg/m3 which is the highest value in the range given for moist bulk density in the SSURGO database. The range of soil friction angle was set from 20° to 40° given the range of soil granules (clay to some gravel) (Das, 2008) and the range of values for cohesion was set from 0 to 0.2 to account for clay content and vegetation (Kenney, 1984). The maximum recharge value (144 mm day) was acquired from the USGS rainfall threshold map for landslide initiation (Wilson & Jayko, 1997) and the minimum threshold value was acquired from the National Weather Service station records in Kentfield, with the value (8.6 mm) being an average daily value for the rainiest month (NWS, ND). The maximum saturated hydraulic conductivity value from the SSURGO database for the soil classification was used along with a soil depth of 2 meters in order to derive the range of values for transmissivity over recharge (104m to 1688m).

Even though there are measures of soil depth in the SSURGO database, these values are very general and may not be accurate nor represent variations in the surface morphology (Tesfe et al., 2009). For example, in Tennessee Valley along the Marin County coast, Montgomery and Dietrich (1994) measured soil depths of 0.1 to 0.5 m in convex areas to a maximum of 4 m in topographic hollows (Montgomery & Dietrich, 1994). The SSURGO database listed soils in this area with up to ~1.4 meters in depth (NRCS, ND). Given this uncertainty, soil depth is held constant at 2 meters for the entire study area.

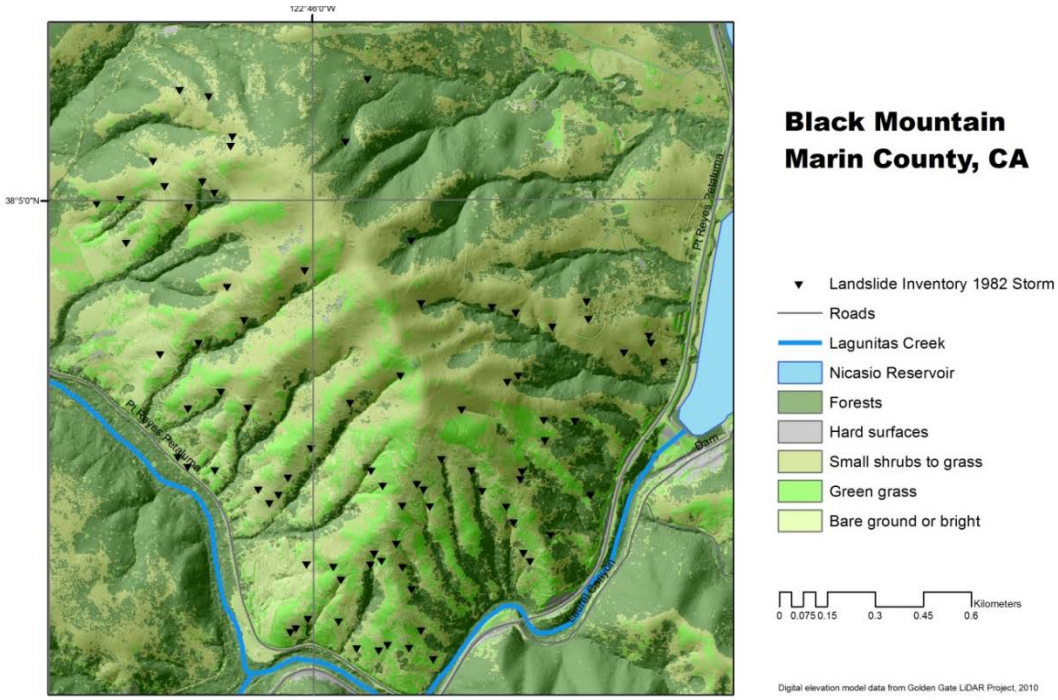

Figure 9: Black Mountain, west Marin County. Landslide inventory points marked.

The statistics from the model (Table 5) run show ~24% of the Black Mountain subarea more likely than not to be unstable given the input values (areas found in the upper threshold or defended regions as described in the SINMAP tool) (Figure 10). This captures 84% of the inventoried landslide points. Looking at the resulting confusion matrix (Table 6) based on landslide and random point sampling of SINMAP stability index (SI) data, the model has greater sensitivity to the landslide hazard than specificity, given that the false positive value is less than the false negative value.

The output for the hard terrain logistic regression model on the mostly hard terrain landform of Black Mountain showed the same high sensitivity and specificity and an overall correct prediction rate of 88% (Table 7). This contrasts with the output for the soft terrain logistic regression model which shows very high sensitivity but low specificity

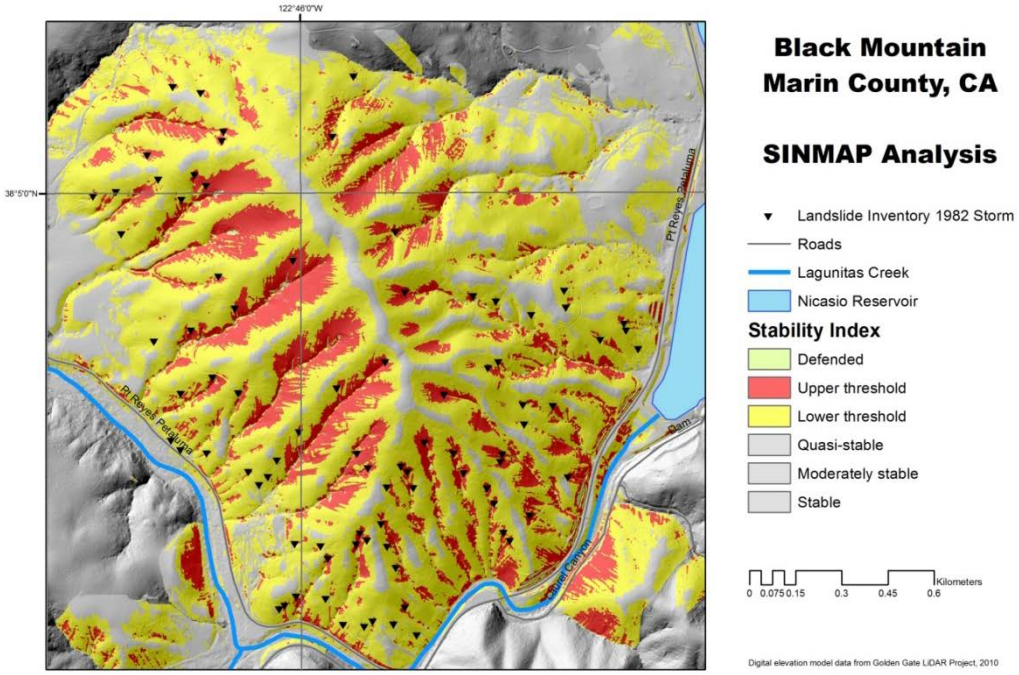

Figure 10: SINMAP stability analysis for Black Mountain

	Stable	Mod Stable	Quasi-Stable	Lower Threshold	Upper Threshold	Defended	Total
Area(km^2)	0.40825	0.116271	0.232532	2.115941	0.865449	0.020063	3.7585
% of Region	10.8621	3.09354	6.186813	56.297336	23.026385	0.533802	100
#Landslides	0	0	0	12	69	4	85
% of Slides	0	0	0	13.953488	80.232558	4.651163	98.837
LS Density (#/km^2)	0	0	0	3.192755	18.358339	1.064252	22.615

Table 5: SINMAP data for Black Mountain subarea

		Classification	
		Prediction	
	p=0.5	**less likely**	**more likely**
data	less likely	66	19
	more likely	12	73
		Sensitivity:	0.8588
		Specificity:	0.7765
		%Correct:	0.8176

Table 6: Confusion matrix, SINMAP results for Black Mountain subarea. Points where upper threshold values are found are areas more likely to fail.

and a lower prediction rate (79%). About 10% of the Black Mountain subarea is more likely than not to be unstable using the hard terrain model, whereas about 33% of the subarea was found tending towards instability using the soft terrain model. The greater sensitivity of the soft terrain logistic regression model and greater specificity of the hard terrain model can be seen in the resulting probability maps (Figure 12).

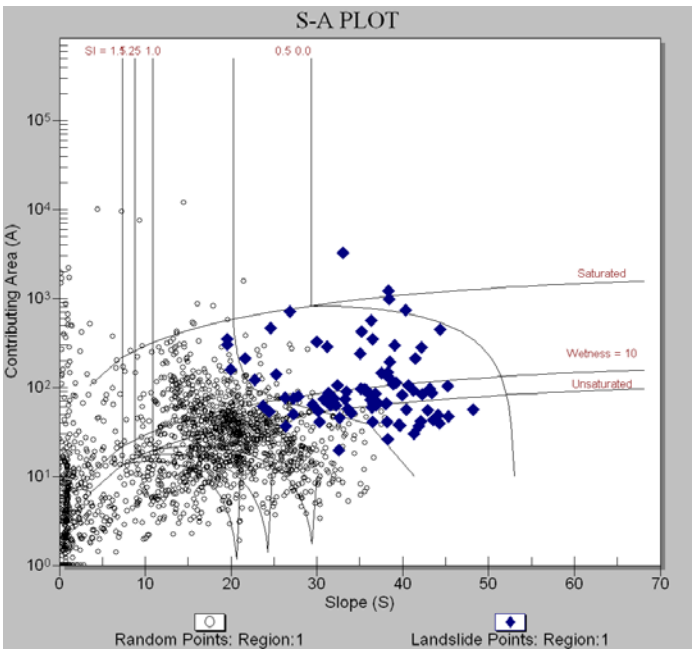

Figure 11: SINMAP output SA plot for Black Mountain. Plot clearly shows landslide points mostly plotted in the upper threshold. The two points found in the defended area are located in the drainage. Points in areas deemed partially wet (not saturated) tend to be on the higher slopes.

Hard Terrain Model Classification				Soft Terrain Model Classification			
0: absence		prediction				prediction	
1: presence	*p=0.5*	0	1		*p=0.5*	0	1
data	0	75	10	data	0	56	29
	1	10	75		1	7	78
		Sensitivity:	0.8824			*Sensitivity:*	0.9176
		Specificity:	0.8824			*Specificity:*	0.6588
		%Correct:	0.8824			*%Correct:*	0.7882

Table 7: Logistic regression model outputs for Black Mountain

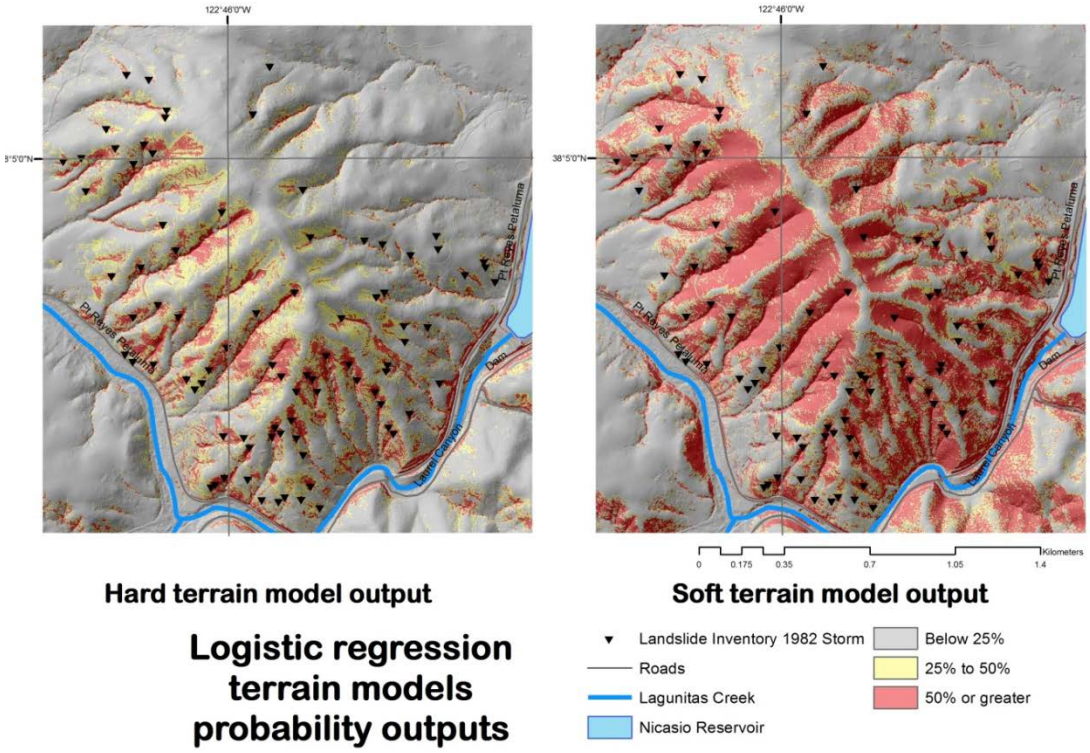

Figure 12: Logistic regression terrain models probability outputs Black Mountain

The second area evaluated using the SINMAP tool was the northern 5.2 km^2 of a valley that is part of the Tomales Bay east shore sub-watershed (Figure 13). Much of the valley is composed of soft terrain and is underlain by Franciscan Mélange. There are areas of harder terrain and these are the areas where debris flows are more likely to occur (Figure 13, inset). The major classified soil for the valley (75% of the spatial extent of three mapped soils) is the Los Osos-Bonnydoon Complex which a fine loamy claypan (Figure 13). This soil has some swelling clay content and is described in the SSURGO database as being a meter to half a meter in depth with medium drainage. Two other classified soils are mapped for the southeast facing slopes above the drainage. The Yorkville clay loam (16% of the area) is a deeper soil with greater

swelling clay content and moderately low saturated hydraulic conductivity. North of the mapped Yorkville clay loam is the Tocaloma-Saurin Association (8% of the area). This soil is a somewhat thin and sandy soil that is gravelly at depth and is described as having a high saturated hydraulic conductivity (NRCS, ND). Of the 36 debris flows in this area, two-thirds were in the area classified as Los Osos-Bonnydoon with all but 2 of the rest in areas classified as the Yorkville clay loam. Given the predominance of the Los Osos-Bonnydoon soil and given that the majority of the landslides occurred in this soil, parameters selected were based on this soil's constituents. The parameters chosen are very similar as those for the Black Mountain subarea, as the soils are similar, with the Los-Osos-Bonnydoon soil described in the SSURGO database as having slightly more clay content and not as deep (NRCS, ND). To account for the higher clay content a higher cohesion values is set (the cohesion value range was set at 0 to .25) (Kenney, 1984) and a lower soil friction angle was set from 20° to 35°) (Das, 2008).

The statistics from the model (Table 8) run show ~18% of the Tomales Bay east shore subarea more likely than not to be unstable given the input values (areas found in the upper threshold or defended regions as described in the SINMAP tool) (Figure 14). This captures ~93% of the inventoried landslide points (one landslide point fell outside of the area analyzed by the SINMAP tool). Looking at the resulting confusion matrix (Table 9) based on landslide and random point sampling of SINMAP stability index (SI) data, the model has greater sensitivity to the landslide hazard than specificity, given that the false positive value is higher than the false negative value.

The output for both hard and soft terrain logistic regression models on the subarea in the Tomales Bay east shore sub-watershed showed high specificity given

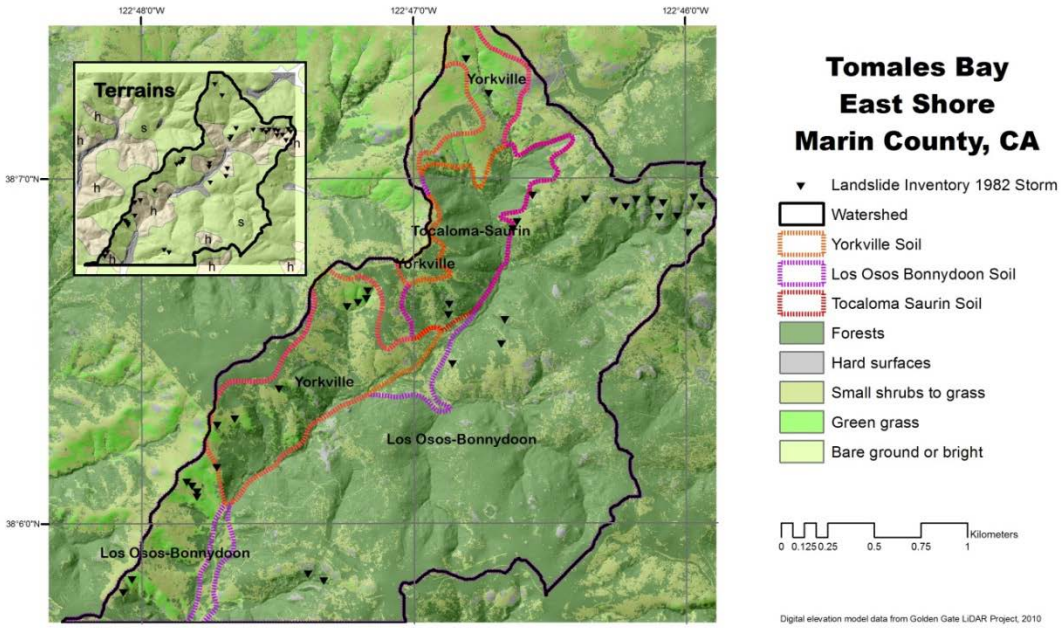

Figure 13: Watershed subarea in Tomales Bay east shore watershed

there was only one false positive for each model (Table 10). The hard terrain model was less sensitive as it only captured 70% of the landslide sample. For the soft terrain model, the area with probability values 50% or greater constituted ~8% of the subarea (Figure 16). For the hard terrain model, the area in this range constituted only ~1.5% of the subarea.

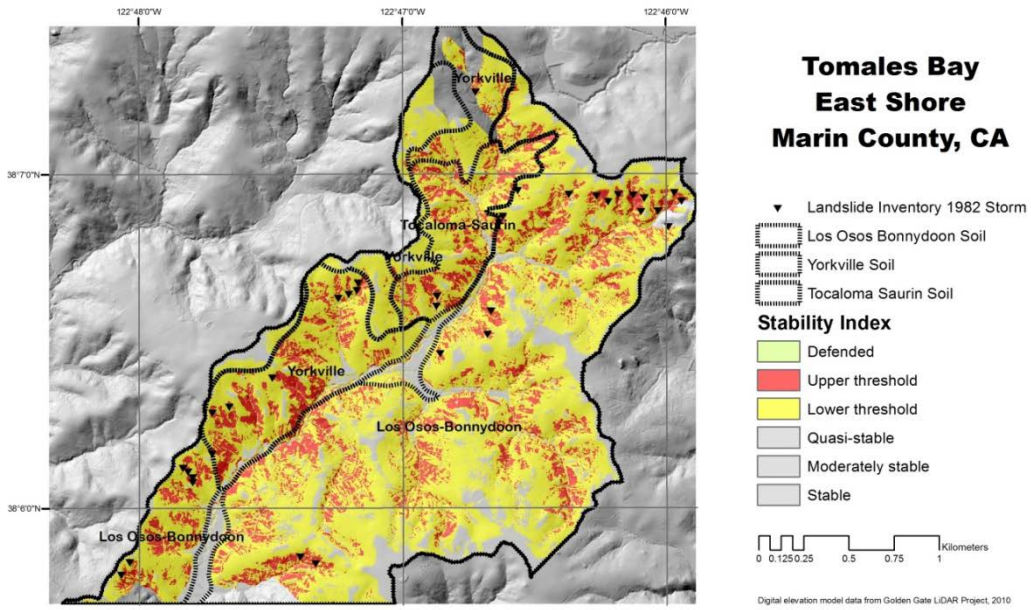

Figure 14: SINMAP stability analysis for Tomales Bay East Shore subarea

	Stable	Mod Stable	Quasi-Stable	Lower Threshold	Upper Threshold	Defended	Total
Area(km^2)	0.404524	0.124326	0.234134	3.390613	0.93947	0.007297	5.100364
% of Region	7.931277	2.437591	4.590535	66.477863	18.419666	0.143068	100
#Landslides	0	0	0	2	33	1	36
% of Slides	0	0	0	5.405405	89.189189	2.702703	97.2973
LS Density (#/km^2)	0	0	0	0.392129	6.470126	0.196064	7.058319

Table 8: SINMAP data for Tomales Bay East Shore subarea

		Classification	
		prediction	
	p=0.5	less likely	more likely
data	less likely	29	7
	more likely	2	34
		Sensitivity:	0.9444
		Specificity:	0.8056
		%Correct:	0.8750

Table 9: Confusion matrix, SINMAP results for Tomales Bay East Shore subarea. Points where upper threshold values are found are areas more likely to fail.

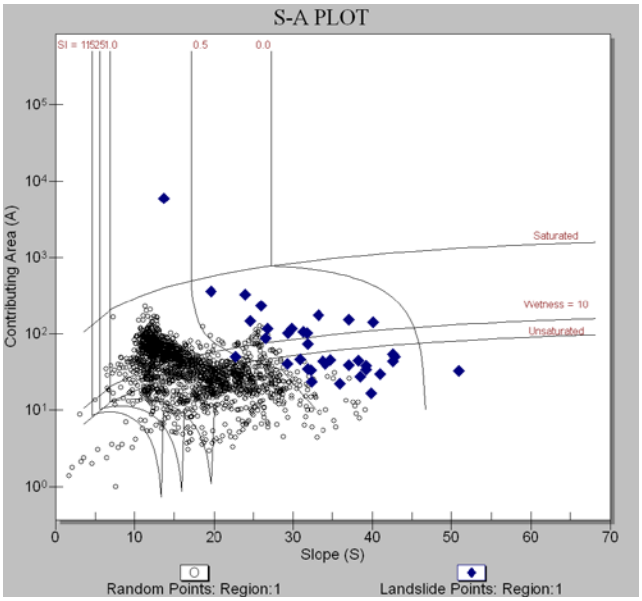

Figure 15: SINMAP output SA plot for Tomales Bay east shore subarea. Plot clearly shows landslide points mostly plotted in the upper threshold. There are a greater number of landslides in areas determined to be not fully saturated than for Black Mountain.

Hard Terrain Model Classification				Soft Terrain Model Classification			
0: absence		prediction				prediction	
1: presence	*p=0.5*	0	1		*p=0.5*	0	1
data	0	36	1	data	0	36	1
	1	11	26		1	5	32
		Sensitivity:	0.7027			*Sensitivity:*	0.8649
		Specificity:	0.973			*Specificity:*	0.973
		%Correct:	0.8378			*%Correct:*	0.9189

Table 10: Logistic regression model outputs for Tomales east shore subarea

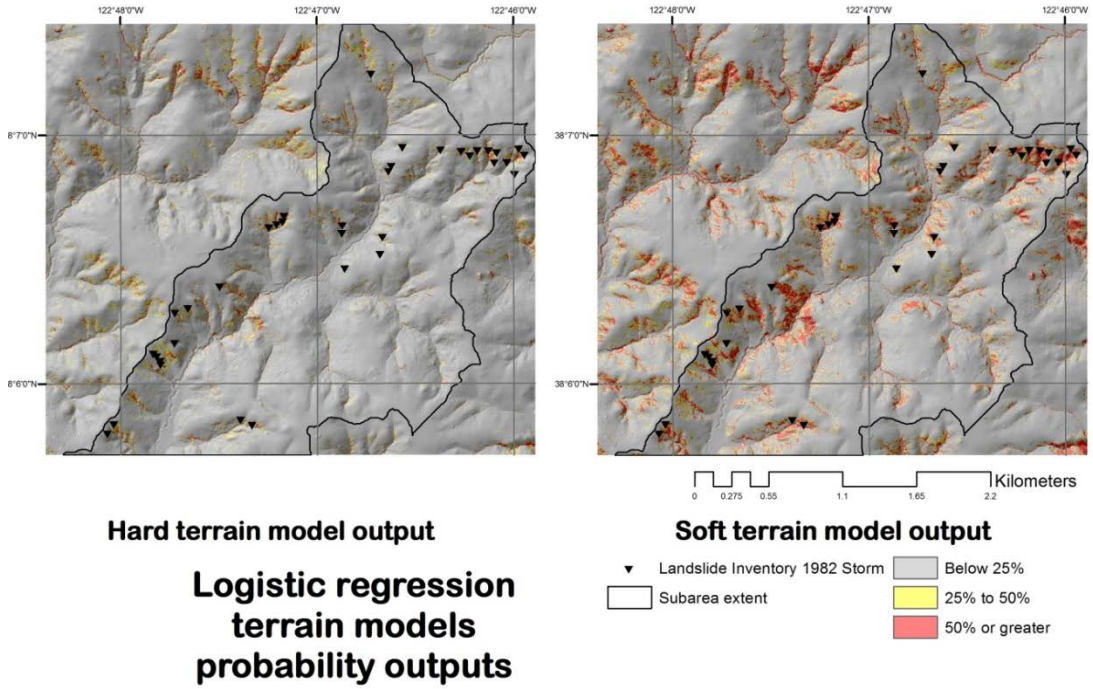

Figure 16: Logistic regression terrain models probability outputs Tomales East Shore subarea

V. ANALYSIS

In observing the results of the different models on Black Mountain, it can be seen how the models match up. The output of the soft terrain logistic regression model of 50% probability or greater more or less matches the extent of the area deemed prone to instability as indicated by the SINMAP tool (Figure 17). This result indicates how much slope influences both the soft terrain model and the SINMAP model. Running the glm in R as a function of slope only decreased deviance by 44% than what would be found using just the intercept.

The output of the hard terrain logistic regression model shows the greater

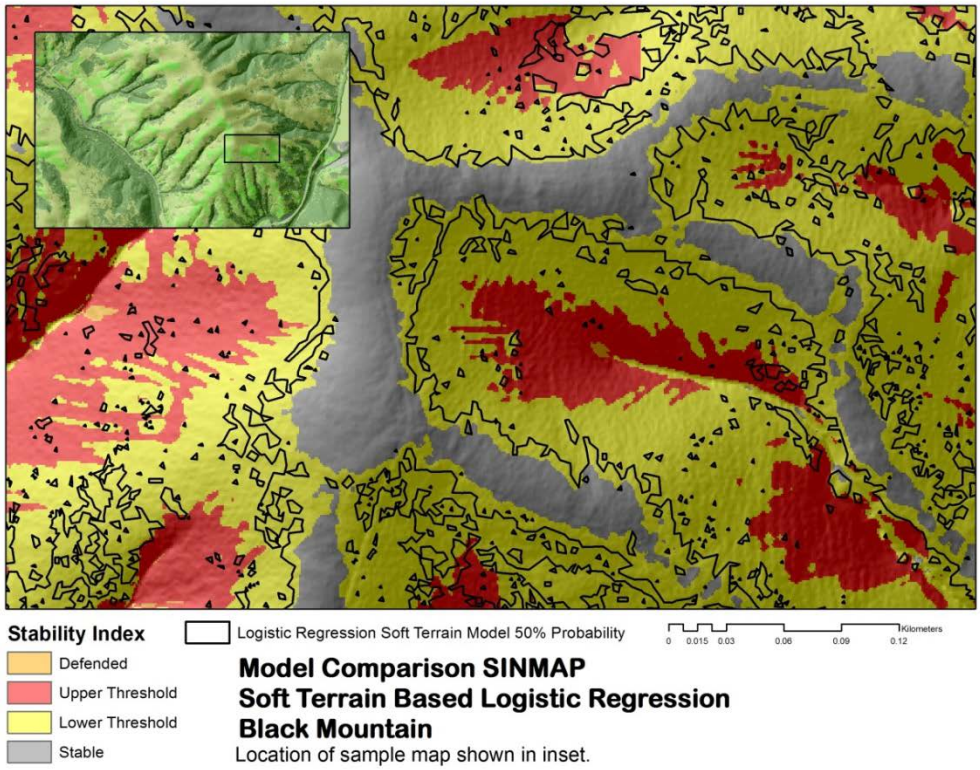

Figure 17: Model comparison for soft terrain logistic regression model and SINMAP, Black Mountain

influence of curvature on the output. Running the glm in R as a function of curvature only decreased deviance in about 25%. Slope had less influence, decreasing deviance about 17%. Although the inventoried debris flows tended not to occur directly in the drainage, they do occur in proximity to the drainage in areas which tend towards convex curvature along with significant slope. Additionally, the hard terrain model reflects the greater significance of land surface cover (Figure 18). As the probability extent indicates, debris flows tend not to occur in forested areas. This terrain model captured some of the extent of the upper colluvial hollows in very steep areas, and in other areas tended to follow the drainage, especially when the area is not forested.

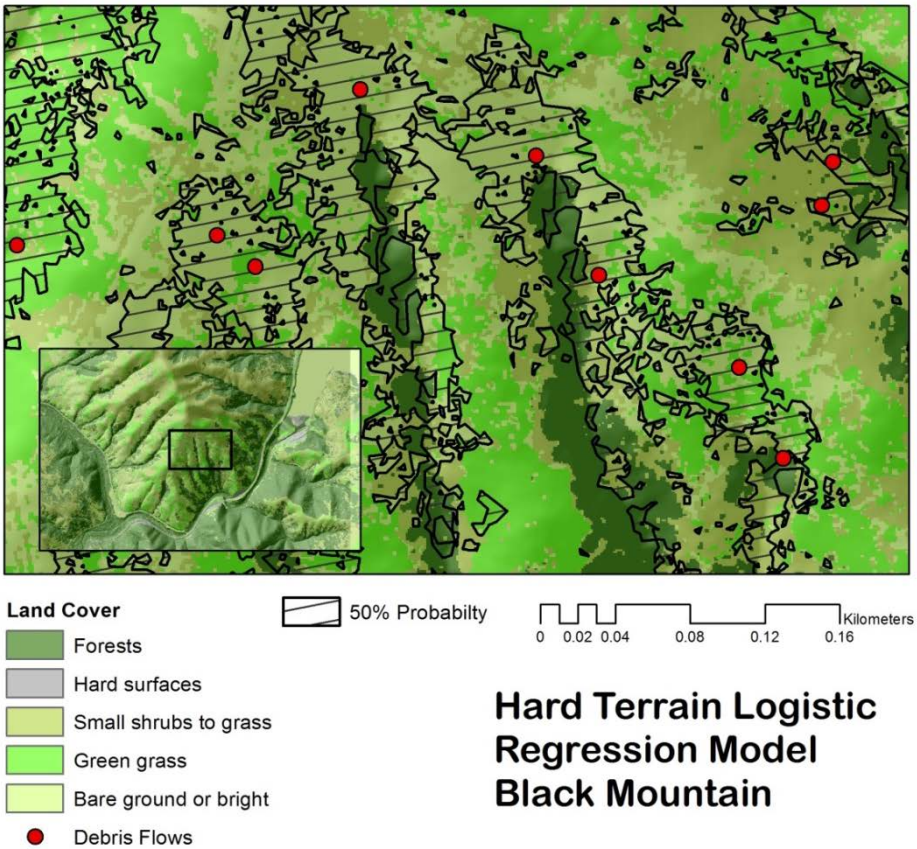

Figure 18: Hard terrain logistic regression model output, Black Mountain

The SINMAP tool output has lower specificity than the hard terrain logistic regression model. However it does provide information regarding where the slope will become saturated. Figure 19 shows the area in the upper threshold range of instability (indicated by dark green). The upper threshold is within the area of threshold saturation. Threshold saturation is one output of the SINMAP model, as it models shallow subsurface flow convergence given the input physical soil parameters and recharge rate (Pack et al., 1998).

This compares with the output of the hard terrain logistic regression model where curvature and slope have a more specific effect (Figure 20) on predicted instability. The

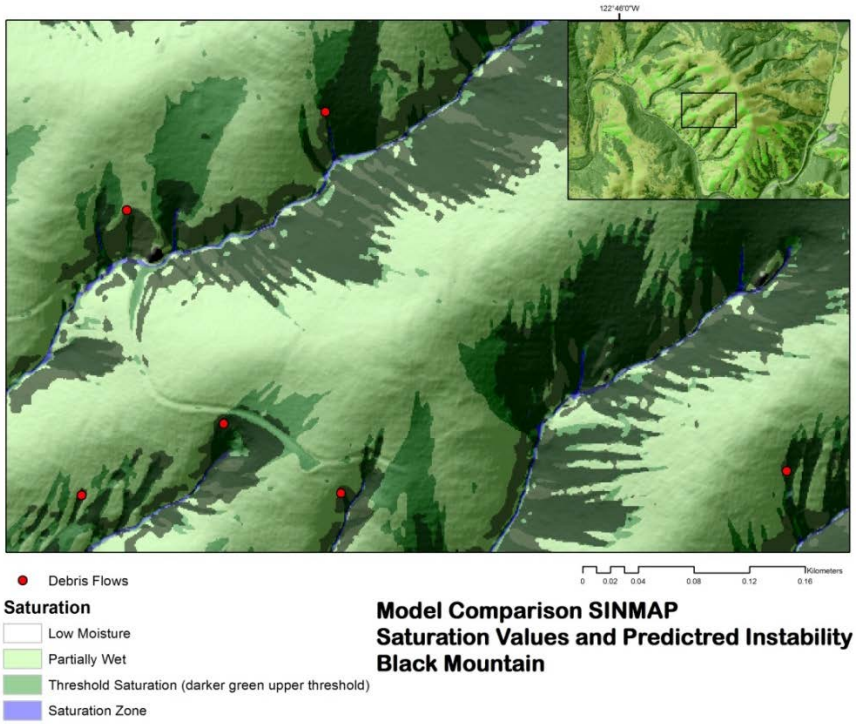

Figure 19: Saturation values and slope instability Black Mountain

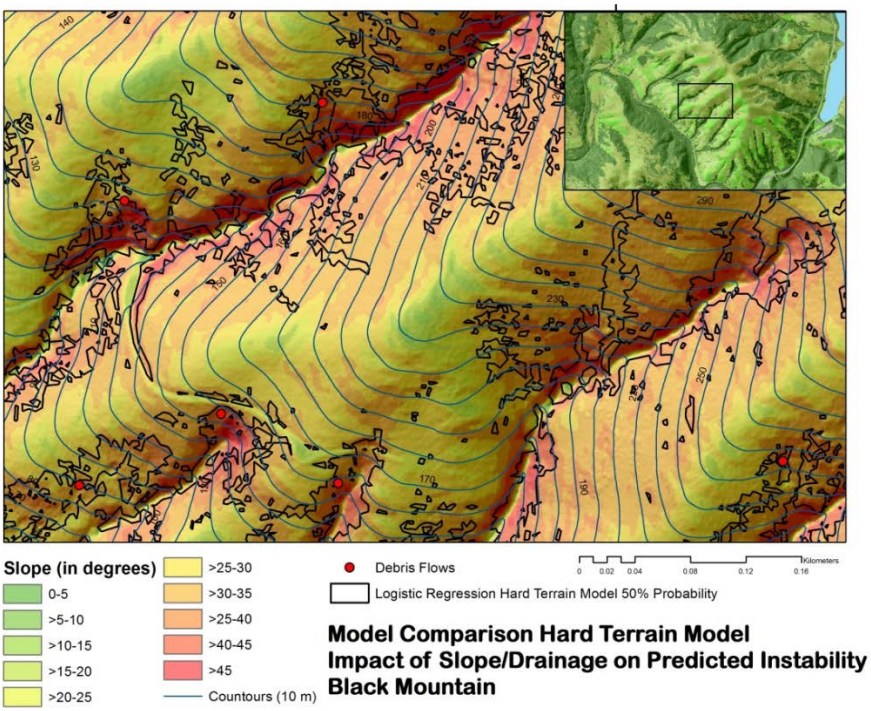

Figure 20: Slope Values and Hard Terrain Model Logistic Regression Black Mountain

question is does saturation as measured through SINMAP a more exact predictor of slope instability than what can be found through a statistical analysis of surface morphology? The hard terrain model is more sensitive and specific to the debris flow inventory than the SINMAP tool. This is most likely the result of the model coefficients being derived from sampling similar, but more rugged terrain as the validation area. The soft terrain model is more sensitive but lass specific as it overstates the influence of slope.

In looking at the output of the soft terrain model in the Tomales Bay East Shore subarea (Figure 21) we can see why slop is the primary determinant in the soft terrain logistical regression model. It is only where slope is significant that we have probability values equal or greater than 50%.

The hard terrain model output (Figure 22) has the same specificity as the soft terrain model (Table 10), yet the extent of the area in the 50% probability range is only 19% of that for the soft terrain model. It may be that curvature, which has greater weight in the hard terrain model, gives more definition to the output. The area in the northeast on that map is forested (not shown) and therefore weighted against greater probability.

The SINMAP model does not work well in these immature drainages. SINMAP tends to overstate instability as seen in the mapped output of upper threshold polygons overlaying saturation values (Figure 23). The SINMAP model indicated ~18% of the Tomales Bay East Shore subarea as more likely to be unstable than not whereas the soft terrain logistic regression model indicated ~8% of the subarea with probability values equal or greater than 50% and the hard terrain model ~2%. It is not clear from the output why a larger area would be prone to instability.

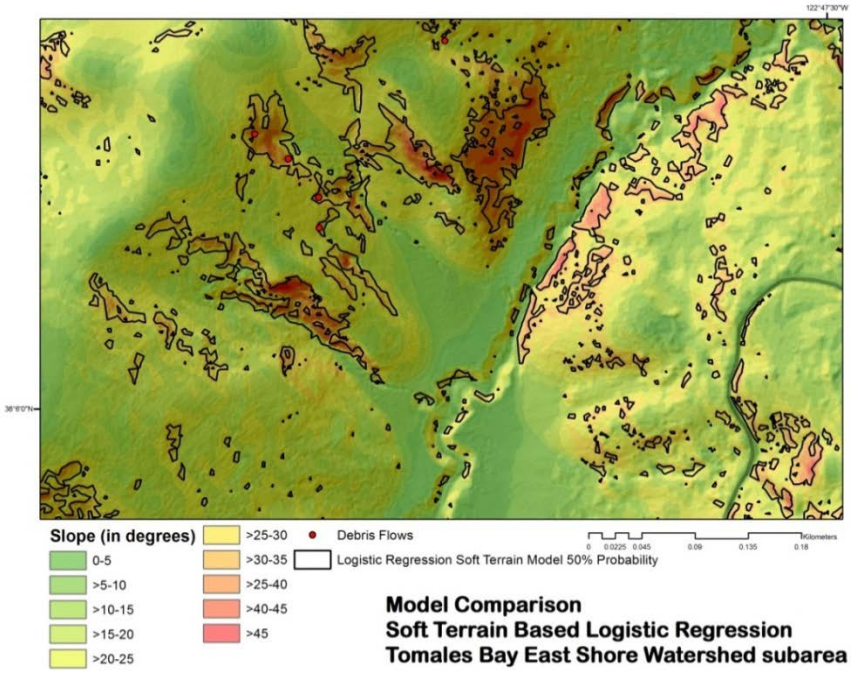

Figure 21: Slope Values and Soft Terrain Model Logistic Regression Tomales Bay East

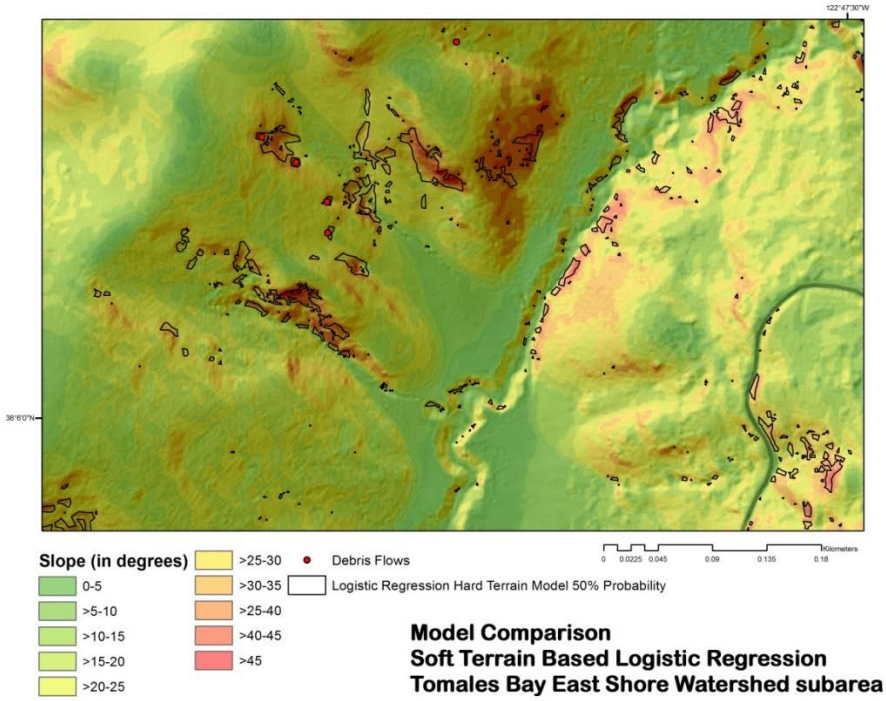

Figure 22: Slope Values and Hard Terrain Model Logistic Regression Tomales Bay east shore

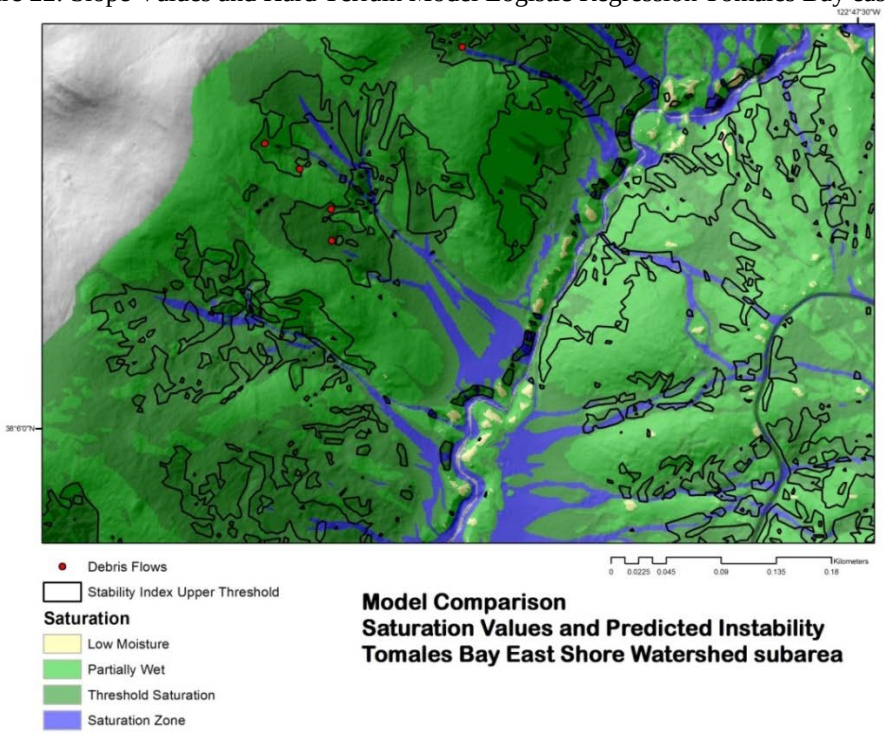

Figure 23: Saturation values and slope instability Tomales Bay east shore

VI. CONCLUSIONS

Both SINMAP and logistic regression bring greater specificity to the delineation of areas prone to landslide susceptibility than what was developed previously. They both provide detailed information that can inform decision makers as to why an area in susceptible to landslide hazard.

The SINMAP model requires a simplification of the subsurface and subsurface flow. It is therefore unable to accommodate heterogeneous environments, account for antecedent moisture or variations of rainfall intensity. However, for areas of hard terrain, the SINMAP tool does provide information on saturation given slope and soil parameters. Using this tool, areas where threshold saturation is reached are often found to be unstable. Such information can be used on the steep hillside housing communities in towns like Mill Valley in Marin County. The high resolution of publicly available digital elevation models (DEMs) provides good information on slope and curvature. For Marin county a sandy soil with some clay is the norm for the steeper slopes. Given this generalization, the SINMAP model could be provided with soil parameters with some justification. Model output could be used with other GIS datasets to determine areas and housing at risk. However to understand if the areas indicated by the tool are truly prone to elevated risk would require a longer historical record than the landslide inventory used for this research.

For areas of soft terrain, the SINMAP tool is limited. The tool performed adequately, given the confusion matrix. However the sample size was small for the soft

terrain area and it most likely does not represent how well the model works in such areas. Pack et al. noted that SINMAP was developed as a study of shallow debris flows that arise due to groundwater flow convergence (Pack et al., 1998). These types of slides are more likely to occur in hard terrains given the morphology and the underlying coherent bedrock (Ellen et al., 1988b). Areas underlain by Franciscan Mélange tend to have irregular bedrock configurations (Medley & Rehermann, 2004) which would lead to more localized hydrology and, thus, could not be addressed with an infinite slope equation. SINMAP tended to overstate hazard in such areas, given the landslide inventory and the results of the statistical models.

A logistic regression model in GIS does provide a way to take in a landslide inventory and create probability surfaces from that inventory. If the data are highly accurate and there are a number of different parameters being used, then the output probability surface can provide good detail as to areas of potential instability. The models cannot be considered independent of the areas where the samples were taken and coefficients derived since the output of the model is bound by the variance of the sample. Therefore, it is not a systematic approach outside of the area where the sample was taken, since the variance indicated by the coefficients is only relative to that area. Valuable information can be gained when applying the tool outside of the sample area if one understands the reasons for that variance. For example the hard terrain model worked well in an area similar to the sample area and in order to understand why the model worked, it is important to understand how that model was constructed and what variance it represents.

If there is a comprehensive landslide inventory for an area that takes into account multiple meteorological events over time and there is high resolution DEM data then a logistic regression model could be very useful for delineating hazard in the area in which the model was developed. The probability surface from the model would be more detailed if areas in which landslides would never occur were removed from the sampling area. Detailed information and history about human activities possibly causing or accentuating landslide hazard could be detailed statistically and information derived could be used to modify such activity thus mitigate against hazard. Such modifications to the landslide susceptibility model would not be possible with a SINMAP approach.

However, finding adequate landslide inventory data is problematic. In this research acquired landslide inventory data proved to be inaccurate. Data had to be digitized from paper maps due to the paucity of information. If there is no recorded and detailed history of landslides for an area, a physical model may be the only means to identifying hazard, outside of generalizing the hazard given the surface morphology.

VII. REFERENCE

Alt, D. & Hyndman, D.W. (2000).*Geology of Northern and Central California.* Missoula, MT: Mountain Press Publishing Company.

Baum, R.L., Savage, W.Z. & Godt, J.T. (2002). *TRIGRS—A Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis (Open-file Report 02-424).*Washington, D.C.: USGS.

Blake, M.C., Graymer, R.W. & Jones, D.L. (2000). *Geologic Map and Map Database of Parts of Marin, San Francisco, Alameda, Contra Costa, and Sonoma Counties, California.* Retrieved January 17th, 2012 from http://pubs.usgs.gov/mf/2000/2337/

Campbell, R. H., 1975, *Soil Slips, Debris Flows, and Rainstorms in the Santa Monica Mountains and Vicinity, Southern California*. U. S. Geological Survey Professional Paper no. 851, 51 p.

Campbell, R.H., Bernknopf, R.L. & Soller, D.R. (1996). *Mapping Time-Dependent Changes in Soil-Slip Debris Flow Probability (USGS IMAP: 2586)*. Retrieved December 23rd, 2012 from http://pubs.er.usgs.gov/publication/I2586.

Can, T., Nefeslioglu, H.A., Sonmez, H. & Duman, T.Y. (2005). Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression. *Geomorphology*, Vol. 72, 250-271.

Cannon, S. H. (1988). Regional rainfall-threshold conditions for abundant debris-flow activity. In Ellen, S.D. & Wieczorek, G.F. (Ed.) *Landslides, Floods, and Marine Effects of the Storm of January 3-5, 1982, in the San Francisco Bay Region, California*. Washington D.C.: United States Government Printing Office.

Cannon, S. H., and Ellen, S. D., 1985, Rainfall conditions for abundant debris avalanches, San Francisco Bay region, California. *California Geology*, Vol. 38 (12), 267-272.

Carlston, C. W., 1963, *Drainage Density and Streamflow*. U. S. Geological Survey Professional Paper 422-C, p. c1-c8.

Crovelli, R.A. & Coe, J.A. (2009). Probabilistic estimation of numbers and costs of future landslides in the San Francisco Bay region. *Georisk*, Vol. 3 (4), 206-223.

Crozier, M. J. (1984). Field assessment of slope instability. In Brunsden, D. & Prior, D.B. (Ed.) *Slope Instability*. New York.: John Wiley & Sons.

Dai, F.C., Lee, C.F., Li, J. & Xu, Z.W. (2001). Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmental Geology*, Vol. 40 (3).

Das, B.M. (2008). *Advanced Soil Mechanics, 3rd Ed*. New York: Taylor & Francis.

Das, I., Sahoo, S., van Westen, C., Stein, A. & Hack, R. (2010). Landslide susceptibility assessment using logistic regression and its comparison with a rock mass classification system, along a road section in the northern Himalayas (India). *Geomorphology*, Vol. 114, 627-637.

Dhakal, A.S., Amada, T. & Aniya, M. (1999). Landslide hazard mapping and the application of GIS in the Kulekhani Watershed, Nepal. *Mountain Research and Development*, Vol. 19 (1), 3-16.

Davenport, C.W. (1984). *An Analysis of Slope Failures in Eastern Marin County, California, Resulting from the January 3rd and 4th, 1982 Storm (CDMG Open-File Report 84-22SF).* Pleasant Hill: California Division of Mines and Geology.

Dietrich, W. E., Dunne, T., Humphrey, N.F. & Reid, L.M. (1982). Construction of sediment budgets for drainage basins. In Swanson, F.J., Janda, R.J. , Dunne, T. & Swanston, D.N. (Eds.) *Sediment Budgets and Routing in Forested Drainage Basins.* Portland, Oregon: U.S.D.A. Forest Service General Technical Report PNW-141, Pacific Northwest Forest and Range Experiment Station, p. 5-23.

Elder, W. (2001). *Geology of the Golden Gate Headlands*. Retrieved from January 17th, 2012 from http://pubs.usgs.gov/bul/b2188/b2188ch3.pdf

Ellen, S.D., Wieczorek, G.F., Brown III, W.M. & Herd, D.G.(1988). Landslides, floods, and marine effects of the storm of January 3-5, 1982, in the San Francisco Bay Region, California - Introduction. In Ellen, S.D. & Wieczorek, G.F. (Ed.) *Landslides, Floods, and Marine Effects of the Storm of January 3-5, 1982, in the San Francisco Bay Region, California*. Washington D.C.:United States Government Printing Office.

Ellen, S.D., Cannon, S.H. & Reneau, S.L. (1988b). Distribution of debris flows in Marin County. In Ellen, S.D. & Wieczorek, G.F. (Ed.) *Landslides, Floods, and Marine Effects of the Storm of January 3-5, 1982, in the San Francisco Bay Region, California*. Washington D.C.: United States Government Printing Office.

Ellen, S.D., Mark, R.K., Wieczorek, G.F., Wentworth, C.M., Ramsey, D.W. & May, T.E. (1997). *Map Showing Principal Debris-Flow Source Areas in the San Francisco Bay Region, California.* U.S. Geological Survey, Open-File Report 97-745 E. Menlo Park, CA: San Francisco Bay Region Landslide Folio. http://pubs.usgs.gov/of/1997/of97-745/.

Godt, J.W., Baum, R.L., Savage, W.Z., Salciarini, D., Schulz, W.H. & Harp, E.L. (2008). Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS framework. *Engineering Geology*, Vol. 102, 214-226.

Graham, J. (1984). Methods of stability analysis.In Brunsden, D. & Prior, D.B. (Ed.) *Slope Instability*. New York.: John Wiley & Sons.

Hines, E. (2011). *Final Report: Golden Gate Lidar Project*. San Francisco State University: Award #G10AC00122.

Hosmer, D. W., & Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics-Theory and Methods*, Vol. 9 (10), 1043-1069.

Kashiwagi, J.H.(1985). *Soil Survey of Marin County, California.* Washington D.C.: United States Department of Agriculture-Soil Conservation Service.

Kenney, C. (1984). Properties and behaviors of soils relevant to slope instability. In Brunsden, D. & Prior, D.B. (Ed.) *Slope Instability*. New York.: John Wiley & Sons.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2007). AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, Vol. 17 (2), 145-151.

Marin County (2002). *Marin Countywide Plan, Geology, Mineral Resources and hazardous Materials Technical Background Report.* San Rafael: Marin County Community Development Agency, Planning Division.

MarinMap, ND. *marinmap_dbo_geo_soil (soils shape file), marinmap_dbo_geo_soil (geology shape file),* Retrieved January 17th, 2012 from http://www.marinmap.org/PublicRecords/Default.asp?MyPhysPath=G:\PublicRecords\data\VectorData\MarinCounty

MarinMap (2004). *2004 Ortho Download.* Retrieved July 2nd, 2012 from: http://www.marinmap.org/dnn/Home.aspx.

Marin County (2005). *Marin County Operational Area Hazard Mitigation Plan*. Available online at http://www.marinsheriff.org/uploads/documents/Marin%20County%20Local%20Hazard%20Mitigation%20Plan.pdf. Accessed December 28, 2012.

Mark, K. M & Newman, E.B. (1988). Rainfall totals before and during the storm: Distribution and correlation with damaging landslides. In Ellen, S.D. & Wieczorek, G.F. (Ed.) *Landslides, Floods, and Marine Effects of the Storm of January 3-5, 1982, in the San Francisco Bay Region, California*. Washington D.C.: United States Government Printing Office.

MCCDA (2004). *Marin County Watershed Management Plan*. Retrieved November 2nd, 2012 from: http://www.co.marin.ca.us/depts/CD/main/comdev/Watershed/WMP_Pt1.pdf.

Medley, E.W. & Sanz Rehermann, P.F. (2004). *Characterization of Bimrocks (Rock/Soil Mixtures) with Application to Slope Stability Problems.* Paper presented at the Eurock 2004 & 53rd Geomechanics Colloquium, Salzburg, Austria, October, 2004.

Meisina, C. & Scarabelli, S. (2007). A comparative analysis of terrain stability models for predicting shallow landslides in colluvial soils. *Geomorphology*, Vol. 87, 207-223.

Montgomery, D.R. & Dietrich, W.E. (1994). A physically based model for the topographic control on shallow landsliding. *Water Resources Research*, Vol. 30 (4), 1153-1171.

Norris, R. M. & Webb, R.W. (1976).*Geology of California.* New York, NY: John Wiley & Sons, Inc.

NRCS, Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture (ND). *Soil Survey Geographic (SSURGO) Database for Marin County, CA*. Available online at http://soildatamart.nrcs.usda.gov. Accessed November 1, 2012.

NWS (ND). *National Weather Service Cooperative Observer Program (COOP) ID: 044500 (Kentfield).* Available online at http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?ca4500. Accessed December 1, 2012.

Ohlmacher, G.C. & Davis, J.C. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology,* Vol. 69, 331-343.

Pack, R.T., Tarboton, D.G., Goodwin, C.N.. & Prasad, A. (ND). *The SINMAP Approach to Terrain Stability Mapping* (website). Available online at http://hydrology.usu.edu/sinmap2/. Accessed February 10, 2013.

Pack, R.T., Tarboton, D.G. & Goodwin, C.N. (1998). *The SINMAP Approach to Terrain Stability Mapping.* Paper presented at the 8th Congress of the International Association of Engineering Geology, Vancouver, British Columbia, Canada 21-25 September 1998.

Pack, R.T., Tarboton, D.G., Goodwin, C.N.. & Prasad, A. (2005). *SINMAP, User's Manual*. Retrieved March 17th, 2012 from: http://hydrology.usu.edu/sinmap2/

Pradhan, B. (2010). Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. *Advances in Space Research,* Vol. 45, 1244-1256.

Reneau, S.L., Cannon, S.H. & Ellen, S.D. (1988). *Plate 6: Debris Flows Triggered by the Storm of January 3-5, 1982, and their Relation to Slope and Terrain Units in the Hicks Mountain Area, Marin County, California*. In Ellen, S.D. & Wieczorek, G.F. (Ed.) *Landslides, Floods, and Marine Effects of the Storm of January 3-5, 1982, in the San Francisco Bay Region, California*. Washington D.C.: United States Government Printing Office.

Rice, S.J., Smith, T.C. & Strand, R.G. (1976). *Geology for Planning: Central and Southeastern Marin County, California (DMG OFR 76-2)*. Sacramento: California Division of Mines and Geology.

Rulon, J.J. & Freeze, R.A. (1985). Multiple seepage faces on layered slopes and their implications for slope-stability analysis. *Canadian Geotechnical Journal*, Vol. 22 (3), 347-356.

Santini, M., Grimaldi, S., Nardi, F., Petroselli, A., & Rulli, M. C. (2009). Pre-processing algorithms and landslide modeling on remotely sensed DEMs. *Geomorphology*, Vol. 113 (1), 110-125.

Schumm, S. A. (1979). Geomorphic thresholds: the concept and its applications. *Transactions of the Institute of British Geographers*, 485-515.

Sidle, R.C., Pearce, A.J. &O'Loughlin, C.L. (1985). *Hillslope Stability and Land Use*. Washington D.C..: American Geophysical Union.

Sidle, R.C. & Ochiai, H. (2006). *Landslides, Processes, Prediction, and Land Use.* Washington, D.C.: American Geophysical Union.

Sloan, D. (2006). *Geology of the San Francisco Bay Region.* Berkeley: University of California Press.

Tesfa, T. K., Tarboton, D. G., Chandler, D. G., & McNamara, J. P. (2009). Modeling soil depth from topographic and land cover attributes. *Water Resources Research*, Vol. 45 (10).

U.S. Census Bureau (2010). *Profile of General Population and Housing Characteristics: 2010, 2010 Demographic Profile Data, Geography: Point Reyes Station CDP, California.* Retrieved February 9, 2013, from http://www.co.marin.ca.us/depts/CD/Main/pdf/planning/Census/CDP_PointReyesStation_DP1.pdf.

USGS (1997). *Introduction to the San Francisco Bay Region, California, Landslide Folio.* U.S. Geological Survey, Open-File Report 97-745 A. Menlo Park, CA: San Francisco Bay Region Landslide Folio. http://pubs.usgs.gov/of/1997/of97-745/.

Van Westen, C.J., Castellanos, E. & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology* Vol. 102, 112–131.

Wagner, D.L. (1976). *Geology for Planning: Western Marin County, California (DMG OFR 77-15)*. Sacramento: California Division of Mines and Geology.

Wentworth, C.M., Graham, S.E., Pike, R.J., Beukelman, G.S., Ramsey, D.W. & Barron, A.D. (1997). *Summary Distribution of Slides and Earth Flows in the San Francisco Bay Region, California.* U.S. Geological Survey, Open-File Report 97-745 C. Menlo Park, CA: San Francisco Bay Region Landslide Folio. http://pubs.usgs.gov/of/1997/of97-745/.

Wilson, R.C. & Jayko, A.S. (1997). *Preliminary Maps Showing Rainfall Thresholds for Debris-Flow Activity, San Francisco Bay Region, California.* U.S. Geological Survey, Open-File Report 97-745 F. Menlo Park, CA: San Francisco Bay Region Landslide Folio. http://pubs.usgs.gov/of/1997/of97-745/.

Wu, W., & Sidle, R. C. (1995). A distributed slope stability model for steep forested basins. *Water Resources Research*, Vol. 31 (8), 2097-2110.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry*, Vol. 39 (4), 561-577.