

DISCRIMINATING BETWEEN LANDSLIDE SITES AND POTENTIALLY
UNSTABLE TERRAIN USING TOPOGRAPHIC INDICES

By
Jeremy Appt

A THESIS
Submitted to
Oregon State University

In partial fulfillment of
The requirements for the
Degree of

Master of Science

Presented July 15, 2002
Commencement September 2002

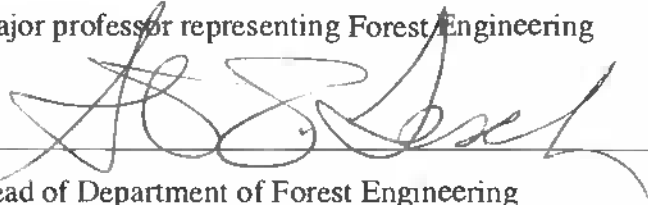
APPROVAL PAGE

Jeremy S. Appt for the degree of Master of Science in Forest Engineering presented
on July 15, 2002.

Abstract approved:



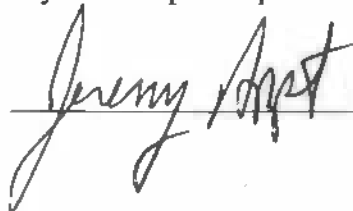
Major professor representing Forest Engineering



Head of Department of Forest Engineering

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon
State University Libraries. My signature below authorizes release of my thesis to
any reader upon request.



Jeremy S. Appt, author

AN ABSTRACT OF THE THESIS OF

Jeremy S. Appt for the degree of Master of Science in Forest Engineering presented on July 15, 2002.

Title: Discriminating between Landslide Sites and Potentially Unstable Terrain using Topographic Indices

Abstract approved: _____

Arne Skaugset III

A landslide inventory, statistical analyses and a Geographic Information System (GIS) are used to analyze landslide sites and potentially unstable terrain in the Oregon Coast Range. The objectives are to evaluate the efficacy of locating landslide sites with topographic variables and discriminate the difference between sites where landslides have and have not occurred. The population of known landslides are characterized as up-slope, non-road related, and associated with 1996 storm events. Topographic variables are derived from a Digital Elevation Model (DEM) for index construction forming six groups; i) slopes, ii) contributing areas, iii) ratios of slope and contributing area, iv) curvature v) infinite slope models, and vi) functions of slope and contributing area based on statistical models. Index groups employ different algorithms. Index performance is measured with landslide and aerial densities. Cumulative landslide occurrence is plotted against cumulative area on a continuous domain of the index to locate a maximum landslide density on equal size areas. Indices are used to generate model definitions of potentially unstable terrain based on similarity to the landslide population. Aerial densities of potentially unstable terrain based on index definitions are determined but no common metric is achieved. Statistical analyses on spatially stratified data suggest a significant ($\alpha < 0.05$) difference between landslides sites and adjoined terrain.

The minimum resolution at which a significant difference is achieved based on spatial stratification is a three cell radius surrounding the slide population. Curvature and area discriminate better than simple slope and topographic ratios. The relative performance is mostly a function of DEM error and resolution, and spatial correlation. Hydrologic geomorphic models perform about as well as the topographic ratios, and much less than the simple area index. There is no statistical evidence to suggest that the hydrologic geomorphic models accurately describe a threshold in the Mapleton slide population. The lack of significance is likely due to limitations on the available parameter sets. Logistic regression produced an index with the highest discrimination performance due to a maximum likelihood algorithm. Regression models have a physical basis in and parallel the behavior of linked hydrologic geomorphic and slope stability models. The measured differences in performance are a useful assessment of the DEM – index combination.

TABLE OF CONTENTS

| | |
|---|-----|
| 1.0 Introduction | 1 |
| 2.0 Objectives | 6 |
| 3.0 Hypothesis | 7 |
| 4.0 Literature Review | 8 |
| 4.1 Introduction to Potentially Unstable Terrain | 10 |
| 4.2 Landslides Defined in the 1996 Storm Impacts Inventory | 12 |
| 4.3 Site Similarity | 13 |
| 4.4 Landslides Viewed from Physical Geography | 15 |
| 4.5 Hazard Mapping and Types of Analysis | 19 |
| 4.6 Topographic Indices Related to Hydrology | 22 |
| 4.7 Dimensionless Analysis of Topography | 28 |
| 5.0 Materials and Methods | 33 |
| 5.1 Digital Elevation Model | 33 |
| 5.2 Preprocessing of Elevation Data | 34 |
| 5.3 Computation of Low Order Topographic Indices | 36 |
| 5.4 Higher Order Indices | 41 |
| 6.0 Results | 44 |
| 6.1 Empirical Modeling of the Slide Population | 44 |
| 6.2 Statistical Analysis of Index Distributions | 48 |
| 6.3 Index Performance | 54 |
| 6.4 Logistic Regression on Topographic Indices | 64 |
| 6.5 Discrimination by Definition | 78 |
| 7.0 Conclusions | 95 |
| Bibliography | 99 |
| Appendix A Statistical Report from Consultant | 102 |
| Appendix B Assumptions in Statistical Models and Logistic Regressions | 109 |
| Appendix C Regression Analysis of Soil Depth at Slide Sites | 113 |

LIST OF FIGURES

Figure

1. A conceptual landslide function showing a topographic variable plotted versus factor of safety. Bars indicate sensitivity to changes in other variables.4
2. Illustration shows the evolution of a bedrock hollow. A) After a landslide, bedrock is exposed and the risk of additional landsliding is reduced. B) Over a period of centuries to thousands of years, soil from the surrounding hillsides slowly fills the hollow. C) As the sediment thickens in the hollow and approaches a depth of 1 to 2 metres, the likelihood of landsliding increases. Recurrent landsliding in the hollow slowly erodes bedrock, maintaining the form of the hollow. (Adapted from Benda et. al., 1997).....11
3. Form shows the headwall hazard rating for the Mapleton Ranger District (After Swanson and Roach, 1987).....14
4. First order drainage basin morphology. a) Distribution of first order basin types, b) Basin type description, and c) First order sub-basin morphology. (Modified from Marcus, 1980).....17
5. Illustration shows idealized relationships between hollow source areas and first order basins. After Montgomery and Dietrich (1989).18
6. Matrix illustrates different types of hazard mapping. Map scale listed as Small (S), Medium (M), and Large (L), Detailed (D) (After the B.C. Earth Science Task Force, 1994).20
7. Illustration of conceptual drainage area with notations for local geometry.....23
8. Illustrations of converging, parallel, and diverging flow regimes. Shaded zones are areas draining across a section X_d from the stream. After O'Loughlin (1981).28
9. Graphs show sensitivity of the dimensionless area functions. Area function $g(x)$ is plotted against arbitrary values of dimensionless factors a) X and b) R29
10. Graph shows examples of dimensionless topographic ratios different flow regimes. After O'loughlin, 198131

LIST OF FIGURES CONTINUED

Figure

11. Grid extent constructed from four USGS grids. The solid shaded area classified as NODATA surrounds Mapleton Watershed boundary. The study area boundary is inside the watershed.34
12. Graph shows conceptual plots of a threshold function $A = C(1 - \tan \theta) \sin \theta$ marked on right axis with threshold coefficients; $C = 1$ and 10 . The sine and tangent slope transformations are shown on the left axis.45
13. Plot shows slope versus area for the Mapleton slope slide population. Threshold coefficients; $C = 1.6$ and 3.0 represent the 95% confidence interval, based on the function $A = C \sin \theta (1 - \tan \theta)$ fitted to the data set. Explanatory variables are slope (θ_1), and area (A_3)46
14. Graphic shows spatial stratification defined by three, five, and ten cell radii surrounding slides found in Mapleton. Slope stratification is defined by the 95% confidence interval on the mean of the slide population. Smaller scale section shows the intersection of spatial and slope stratification.....49
15. Box plot show the slope θ_1 distributions. Spatial stratification based on cells surrounding slide sites for the Mapleton slide inventory; a) slide population b) three cell radius, c) five cell radius, and d) ten cell radius.....50
16. Box plot of shows log transformations of the area index A_1 . Spatial stratification based on cells surrounding the Mapleton slide inventory; a) landslide population b) three cell radius, c) five cell radius, and d) ten cell radius.50
17. Sample area with spatial stratification of non-slide sites defined by a three-cell radius surrounding the slides in the Mapleton inventory with the down stream cells removed.52
18. Scatter plot showing slope, θ_1 , versus area, A_1 , for spatially stratified regions around landslides in the Mapleton inventory. Slides sites are shown as open circles.53
19. Scatter plots showing a) slope, θ_1 , and b) area, A_1 , versus curvature C_1 . Slides are shown as open circles. Stratification based on cells surrounding landslides in the Mapleton slide inventory.54

LIST OF FIGURES CONTINUED

Figure

21. Matrix plot shows the landslide association with slope index θ_1 in the Mapleton study area. Gradients on each curve represent; a) terrain character, b) index correlation with slides, and c) landslide density. Stippled area is an example definition for potentially unstable terrain in the θ_1 domain. The width of the stippled area measured on the horizontal axis of c) defines the aerial density over a continuous range of the index.57
22. Matrix plot shows landslide association with area index A_1 in Mapleton study area. Gradients on each curve represent; a) terrain character, b) index correlation with slides, and c) landslide density. Stippled area is an example definition for potentially unstable terrain for the A_1 index. The width of the stippled area measured on the horizontal axis of c) defines the areal density over the continuous range of the index.1
23. Graphic shows a contrast of the maximum landslide density measured from three different normalized areas. Densities are derived from topographic indices in the Mapleton area. The topographic indices are factored into five distinct types ranked from lowest to highest order (left to right).....59
24. Box plot shows three different slope index distributions for the Mapleton slide population. The white zones represent the 95% confidence intervals of the mean.60
25. Box plot shows three different area index distributions for the Mapleton slide population. The white zones represent the 95% confidence intervals of the mean.61
26. Plot shows slope index versus partial logit {1} with area terms only. Contours show lines of equal area ranging from 3.5 to 5.0 that represent the slide population in Mapleton. Arrow indicates direction of increasing area.66
27. Plot shows a family of slope index θ_1 versus probability curves derived from the full logistic {1} model in Mapleton. Contours show lines of equal area. Arrows show direction of increase in the area. Quadrants are marked Q1, Q2, Q3, and Q4.68

LIST OF FIGURES CONTINUED

Figure

28. Plots of slope versus probability derived from expected values of the logit{1} model from spatially stratified regions in the Mapleton area. Figure a) has slides shown as circles and non-slides shown as small crosses. Figure b) has slides only, circles are scaled in size and color proportional to the value of area. Lighter and smaller symbols are smaller values.69
29. Regression surface for landslide probability derived from the expected values of the logit{1} model in spatially stratified regions of the Mapleton area surrounding the slide population. Slides are shown as solid circles.70
30. Thematic map shows slide probability classified into quadrants based on the roots of the logit {1} model in Mapleton. A three-cell radius surrounding Mapleton slide inventory with down stream cells removed defines spatial stratification.1
31. Matrix of scatter plots show the explanatory variables versus the Logit{2} model derived from spatially stratified regions in the Mapleton area. Slides are shown as circles and non-slides as triangles. Symbols are colored on a gray scale from low (light) to high (dark) for a) slope, b) area c) area, d) curvature, e)slope, f) curvature. Indices are A_1 , θ_1 and C_174
32. Scatter plot shows area versus full and partial logit{2} derived from spatially stratified regions surrounding the Mapleton slide inventory, a) full logit with symbols colored on a gray scale for curvature from low (light) to high (dark) and, b) partial logit without curvature terms and no gray scale.75
33. Box plot shows distributions of the Logit{2} model factored by response in the Mapleton area, a) slides and b) non-slide sites.76
34. Box plots show the explanatory variable distributions based on systematic range limits of the logit{2} model, a) slope, b) area, and c) curvature. Box plots on left are coincident (in) with a 95% C.I. on the logit{2} model defined by the Mapleton upslope slide population. Box plots on the right are not coincident (out). There are 4,343 cells classified as “in” and 19,511 cells classified as “out”. Indices are A_1 , θ_1 , and C_183

LIST OF FIGURES CONTINUED

Figure

35. Box plots show explanatory variable distributions systematically defined by a simple slope model, a) slope, b) area, and c) curvature. Plots on left are coincident (in) with a 95% C.I. on the slope index θ_1 model defined by the Mapleton slide population. Box plots on the right are not coincident (out). There are 2,941 cells classified as “in” and 20,913 cells classified as “out”. Indices are A_1 , θ_1 , and C_185
36. Box plot shows a contrast of Curvature index C_1 distributions for the logit{2} and slope θ_1 models in the Mapleton area.86
37. Thematic map for a systematic definition of the logit{2} model. Theme indicates expected logit{2} values within the 95% C.I. of the Mapleton slide population. The similar sites are sub-classified based on curvature. Planar terrain is defined as +/- 1 ft. The extent of convergence is defined by the data minimum and divergence by the data maximum.88
38. Thematic map for an intersection of the logit{2} and simple slope models. Logit{2} and slope models show arbitrary definitions of site similarity based on the 95% C.I. of the slide population in the Mapleton area. The similar sites are sub-classified based on model type and their aerial intersection.89
39. Performance data inside spatially stratified areas. Slide and non-slide sites in spatially stratified areas surrounding the Mapleton slide population. Systematic definitions of domain similarity based on a 95% CI of the slide population. a) areas sub stratified by domain, b) Proportions only and c) Proportions sub stratified by domain.93

LIST OF TABLES

Table

| | |
|---|----|
| 1. Expressions for dimensionless area | 29 |
| 2. Summary of drainage area algorithms | 37 |
| 3. Logistic Regression Model for the EVI Index | 43 |
| 4. Wilcoxon-Rank-Sum test of area distributions on spatially stratified regions. | 51 |
| 5. Logit {1}-model output for a spatially stratified data set | 65 |
| 6. Logit {2} model output for spatially stratified data set | 73 |
| 7. Slide and areal densities using domain stratification | 80 |
| 8. Slide and areal densities with spatial and domain stratification relative to the slide population | 92 |

Discriminating between Landslide Sites and Potentially Unstable Terrain using Topographic Indices

1.0 Introduction

Mass wasting is the dominant process accounting for most of the total sediment moving through Coast Range watersheds [Dietrich and Dunne, 1978; Swanson et al. 1982]. Landslides are a natural phenomenon resulting from the dynamic equilibrium between slope forming process and the surrounding environment. The effects of changes in the forest environment are likely to impact landslide occurrence. The physical causes are changes in vegetation and hydrology above and below ground, with interaction between the two. It is generally perceived that landslides occur due to rainfall events that cause a reduction in the soil strength. This loss of soil strength produces the largest change in the overall strength equation. Tree roots are usually considered to add cohesive strength to the soil. Loss of soil strength due to a rainfall is much larger than the contribution by tree roots. A portion of the landscape exists near failure during a rainfall and is characterized as potentially unstable. Therefore, an important question that arises is when and where will forestry impact the subsurface hydrology and soil strength enough to tip the balance toward a landslide? Further, when and where are forest management activities likely to cause change that exceeds natural forest dynamics? This thesis makes use of an inventory of non-road related landslides. Of the main forestry activities, roads are generally considered to have the greatest ability to impact landslide occurrence and vegetation the least impact. Thus, the landslide inventory used here contains the subtlest of forestry impacts. The subject of this thesis is to develop a process for locating potentially unstable terrain and until this

terrain can be delineated and mapped satisfactorily, impacts due to forestry related activities cannot be discriminated.

Many attempts have been made to document the effects of forest management on the incidence of landslide occurrence and erosion rates using landslide inventories [Sidle et. al, 1985]. The majority of these studies show that forestry increases the incidences of landslides. However, interpretation of results is confounded by the scale of the project, measurement error, and uncertainty in the application of the results of aerial photos to heavily timbered terrain [Pyles and Froelich, 1987]. The Oregon Department of Forestry (ODF) carried out a recent landslide inventory in the Oregon Coast Range. ODF carried out a systematic, ground-based landslide inventory after two landslide-producing storms. The data indicates that there is an increase in landslide density compared with 100+-year-old forests for the first 10 years after timber harvest. This increase is followed by a reduction in landslide density compared with 100+ year old forests between 10 and 100 years after harvest.”[ODF, 1998] The erosion rate is estimated to increase from 5.2 to 7.4 landslides per square kilometer for clear cut and mature forest respectively [Skaugset and Keim 2002 in press].

Forest management activities can impact mass wasting processes resulting in an environmental damage, thus managers need to identify and locate potentially unstable terrain to make decisions about management-induced risks. Potentially unstable terrain is often equated with areas that have a high likelihood of mass wasting or alternately referred to as *high hazard* areas. Such qualitative statements demonstrate the need for tractable definitions and standards to be included in the management plan.

Forest managers try to meet environmental impact goals through the implementation of Best Management Practices or BMP's. The very nature of some

BMPs highlights the scientific uncertainty involved with defining potentially unstable terrain and its identification on the ground. This uncertainty confounds and introduces variability in the application of BMPs to forest operations. Forest managers may be lacking time, information and resources to meet a desired level of certainty in the application of BMPs. The design of BMPs for larger-scale ecosystem management also contains uncertainty. The later application may have more time and resources available on a unit area basis but the science surrounding definitions of potentially unstable terrain and the interaction with forest operations and ecosystem management is still highly uncertain. Therefore, good definitions of potentially unstable terrain contain i) known degrees of uncertainty and, ii) low relative uncertainty. Further, potentially unstable terrain should be identifiable on the ground, both locally and on a landscape scale.

To discriminate means to perceive the distinguishing features of objects, in this case slide versus non-slide sites. The distinguishing feature of slide sites is their location on some continuum of stability. There is a perceived systematic continuum of stability that varies with location relative to existing slide sites. For example, this continuum may be described with functional relationships between Factor of Safety (FS) and topographic variables, see Figure [1].

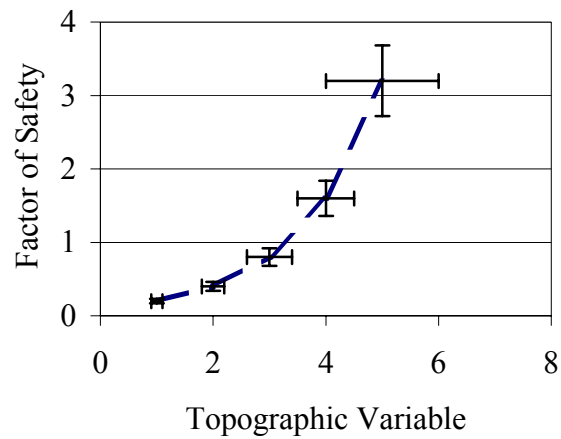


Figure1. A conceptual landslide function showing a topographic variable plotted versus factor of safety. Bars indicate sensitivity to changes in other variables.

The FS in Figure [1] is a ratio representing the forces resisting failure divided by forces causing failure. This is referred to as a deterministic approach. FS is determined from model parameters that vary across the landscape. Therefore, parameters are treated here as variables. If probability is introduced the relationship becomes probabilistic. FS may be a function of topography or just correlated to topography. A traditional sensitivity analysis is used to produce a functional relationship between FS and changes in the variable of interest, given that all other variables are held constant. However, the range in FS becomes obscure if the other model variables are not held constant.

Landslide inventories may be used to analyze functional relationships between FS and topographic variables on a landscape scale. Figure [1] demonstrates the use of a conceptual landslide function to contrast sites that failed ($FS < 1.0$) and sites that did not fail ($FS > 1.0$) for a generic topographic variable. The bars in Figure [1] acknowledge that all other variables in the model are not constant. Some of the variance is due to non-topographic factors. The over-lap in the range of a

topographic variable between slide and non-slide sites makes it difficult to use as a discriminator.

2.0 Objectives

The objectives of this thesis are to determine the correlation between topographic variables and the occurrence of landslides. Further, define potentially unstable terrain by discriminating against sites with recent slide activity, and place those sites without landslides in the context of that definition. The objectives imply a calibration of the available data set meaning; determine the uncertainty surrounding the definition of potentially unstable terrain, and the efficacy of locating existing landslides. The analyses are limited to topographic variables due to time constraints and the data set available.

3.0 Hypothesis

The hypothesis tested in this thesis may be stated as;

H₁: Slide and non-slide sites possess different probability distributions on topographic variables; slope, area, and curvature

H₂: Non-slide sites that are topographically similar to slide sites can be identified

These hypotheses are simple statements that require a scope to have any real meaning. The scope may vary as a function of

- area used to define slide and non-slide sites
- variables used to define topography
- data source and method of measuring topography

A large amount of this thesis effort is spent defining scope.

4.0 Literature Review

There is variability in the definitions of some common terms applied to mass wasting phenomena. It is constructive to review some of these terms. For example, in reference to potentially unstable terrain, definition of the term potential may be problematic since it can be used as an adjective, or a noun. Use of the term, potentially in the context of mass wasting phenomena, implies a high possibility of failure.

A generic reference to potentially unstable terrain implies the following;

- Mass wasting phenomenon is defined
- probability of occurrence for a mass wasting phenomenon
- physical characteristics of the site are definable based on a map, visual inspection in the field, or possibly in a lab specimen
- cause and consequence of the mass wasting phenomenon is known
- spatial and temporal scales are defined

For example, the B.C. Forest Practices Code [B.C. MoF and MoE 1999] defines terrain stability on a reconnaissance level map as “the likelihood that landslides will occur after logging and/or road construction”. Further, detailed stability mapping breaks the classes down into sub categories based on specific cause and/or consequence of that failure (e.g., a road related failure that places debris into a stream).

Sometimes a synonym that is preferable to potential is susceptibility.

The type of hazard event, i.e., shallow landslides, defines the nature of unstable terrain. The terrain analyzed in this thesis is limited by study area boundaries, and slope position of the landslide (see introductory section on landslides).

The scope of any definition for potentially unstable terrain is dependant on the time and space scales chosen. For example;

| Scale | Definition |
|-------|---|
| Time | On the order of soil profile development on steep slopes in the coast range \cong 100 to 1,000 years \cong return period defined by the 1996 flood event, and |
| Space | One the order of a gully headwall in size. The minimum resolution restricted by the topographic maps available is 30m x 30m. |

Therefore, a concise definition of potentially unstable terrain in this thesis is, “upslope ground areas with a high probability of shallow landslide occurrence, given a landslide producing storm (1996)”.

The terms hazard, risk and consequence are commonly used in discussions on landslides. However, they are not uniquely defined in the literature. The Terrain Stability Mapping Guidelines of British Columbia [B.C. R.I.C., 1998] offer the following discussion on the term hazard;

The word 'hazard' is derived from the Arabic word for 'a die' (singular of dice) and is often related to 'chance or probability', as in the phrase 'to hazard a guess'. This definition is reflected in the United Nations definition of natural hazard: "the probability of occurrence of a potentially damaging natural phenomenon" (Varnes1984)[4]. In reference to landslides, Fell (1994)[5] defines 'hazard' as "the magnitude of the event times the probability of its occurrence". In British Columbia, however, 'hazard' is also often used to describe the damaging phenomenon, as in 'natural hazard',

'geological hazard', 'landslide hazard', or a specific type of landslide hazard, such as, a 'debris flow hazard'.

Hazard and risk are sometimes treated as synonyms, and usually related to a consequence.

4.1 Introduction to Potentially Unstable Terrain

Many shallow landslides originate in a landform known as the bedrock hollow that is also referred to as a swale or zero order basins. Hollows are narrow depressions formed in bedrock and can occur on hillsides of varying steepness. Hollows create conditions conducive to shallow sliding [Benda, 1997]. During storms, hollow topography concentrates water flowing down slope through the soil into the center of the hollow causing a higher ratio of saturation depth to soil depth saturation than adjacent hill slopes. A qualitative model exists for hollow development [Dietrich, 1991]. Hollows are sites that experience infrequent, but rapid, evacuation of colluvium, primarily by landsliding, followed by periods of slow colluvium accumulation. This is illustrated in Figure [2].

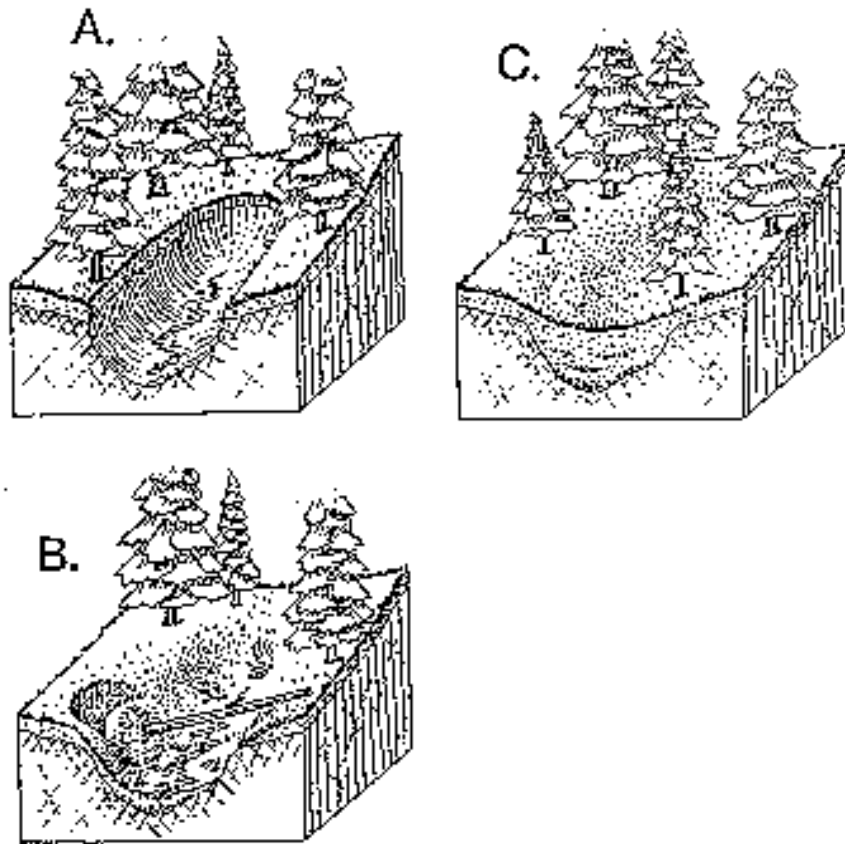


Figure 2. Illustration shows the evolution of a bedrock hollow. A) After a landslide, bedrock is exposed and the risk of additional landsliding is reduced. B) Over a period of centuries to thousands of years, soil from the surrounding hillsides slowly fills the hollow. C) As the sediment thickens in the hollow and approaches a depth of 1 to 2 metres, the likelihood of landsliding increases. Recurrent landsliding in the hollow slowly erodes bedrock, maintaining the form of the hollow. (Adapted from Benda et. al., 1997^[a11])

Each bedrock hollow has a unique failure history because of i) variability in soil properties (e.g., soil strength) that resist failure, and ii) probabilistic nature of forces causing failure (e.g., catastrophic rainfall event). This unique failure history has an influence on landslide occurrence during a landslide-producing storm. For example, during any landslide producing storm only a small proportion (e.g., 3%) of hollows will fail.

This discussion is a simplified view of slope forming processes in bedrock hollows. Many of the landslides that occurred during the 1996 storms were not associated with topographic features indicative of bedrock hollows. In fact, only 50% of the up-slope slides were marked as occurring on concave terrain.

4.2 Landslides Defined in the 1996 Storm Impacts Inventory

Mass wasting is divided into several categories based on landslide processes that are governed by environmental controls [Benda et al., 1997]. This thesis deals with *shallow landslides* that occur in thin soils, less than 2 metres deep, overlying steep bedrock or compacted glacial till. The depth of the landslide is small compared to its length. Other names for shallow landslides include *debris flows*, *debris avalanches*, and *planar failures*. The first two terms refer to the water and debris content of the failed material after movement has started. The later term refers to the shape of the failure surface. Debris slide – avalanche flow combinations are the most common type of soil mass movement on steep forested terrain and on other steep ($>25^\circ$) natural slopes [Sidle et al., 1985].

This thesis uses landslide data from the 1996 Oregon Department of Forestry storm impacts study [ODF, 1997]. Landslides are defined by a Channel Impact Field

Protocol (CHIP) intended to identify slides that impact streams. Landslides defined by the CHIP are therefore a reference to;

1. an initiation point for a shallow landslide with a discrete failure surface,
2. a resultant debris flow impacting a well defined channel downslope, and
3. located in an upslope position,
4. in response to a catastrophic rainfall (i.e., 1996 flood)
5. non-road related.

This thesis uses only the upslope slides in order to limit the scope of the complex environmental processes and forest management treatment effects involved.

4.3 Site Similarity

There is a common perception that sites with the same surficial data, will behave the same. Thus, if a non-landslide site is surficially the same as a landslide site, then the non-landslide site will possess a susceptibility to landslides similar to the landslide site. This assumes that all other factors that cannot be seen are the same for both sites. The first step in facilitating this discussion is to explore what features define similarity.

Terrain is assessed starting with maps in the office and then in the field. Most field assessments investigate geomorphology and other terrain attributes. An example of such a field assessment is shown in Figure [3]






| HEADWALL RATING SYSTEM | | | | |
|--|--|--|---|--|
| SALE _____ UNIT _____ | | SCORE <input type="checkbox"/> | | |
| HEADWALL ID. _____ DATE _____ | | | | |
|  ZONE OF INTEREST | 1 SLOPE - STEEPEST PORTION ON FALL LINE, LOWER 2/3 HEADWALL | | | |
| | (16) <input type="checkbox"/> 90%+ | (8) <input type="checkbox"/> 80%+ | (4) <input type="checkbox"/> 70%+ | (2) <input type="checkbox"/> <70% |
| 2 VEGETATION | (8) <input type="checkbox"/> FEW TREES SALMONBERRY UNDERSTORY PISTOL BUTTING OF UPSLOPE CONIFER | (4) <input type="checkbox"/> PATCHY CONIFER SOME ALDER SOME PISTOL-BUTTING PRIMARILY SALMONBERRY UNDERSTORY | (4) <input type="checkbox"/> EVENLY DISTRIBUTED HARDWOOD SALMONBERRY/ SWORDFERN UNDERSTORY | (2) <input type="checkbox"/> EVENLY DISTRIBUTED CONIFER-STRAIGHT SWORDFERN UNDERSTORY |
| 3 SIDE SLOPE | (8) <input type="checkbox"/>  | (6) <input type="checkbox"/>  | (4) <input type="checkbox"/>  | (2) <input type="checkbox"/>  |
| 4 SOIL DEPTH | (6) <input type="checkbox"/> SHALLOW, <4' | (4) <input type="checkbox"/> NO DATA OR INDICATORS | (2) <input type="checkbox"/> DEEP, >4' | |
| 5 HEADWALL CONFIG | (8) <input type="checkbox"/> MULTIPLE CONVERGENT DEPRESSIONS | | (4) <input type="checkbox"/> SINGLE DEPRESSION | |
| 6 SLOPE ASPECT | (9) <input type="checkbox"/> NORTH 270 - 090 DEGREES | | (3) <input type="checkbox"/> SOUTH 090 - 270 DEGREES | |
| 7 MICROTOPOGRAPHY | (12) <input type="checkbox"/> TENSION CRACKS, ISLANDS HUMMOCKY MICRO-RELIEF BLOWDOWN OFTEN COMMON | (8) <input type="checkbox"/> SCARPS, BENCHES, BULGES OFTEN SCATTERED BLOWDOWN | (4) <input type="checkbox"/> SMOOTH, GENERALLY EVEN SLOPE | |
| * RECENT SLIDING (0 - 10 YEARS) ADD 5 PTS. | | | | |
| COMMENTS BOOMER HOLES LINEATIONS/FAULTS ON PHOTOS SEEPS, UNUSUAL AMOUNTS OF WATER PRESENT IGNEOUS BANDS (DIKES-SILLS) EXPOSED OR RESISTANT BEDROCK BANDS/FACES HIGH POTENTIAL LEAVE AREA BLOWDOWN | | OTHER COMMENTS _____ _____ _____ _____ | | |

Figure 2.3. Headwall rating form used on Mapleton Ranger District. (After Swanson and Roach, 1987).

Figure 3. Form shows the headwall hazard rating for the Mapleton Ranger District (After Swanson and Roach, 1987)

This field assessment method is based on expert opinion. Site similarity is achieved when two sites look the same and receive the same score. However, landslides are complex. Thus, it is highly unlikely that any two sites are the same even if they look the same after a field assessment.

This thesis is limited to an analysis of the topographic variables, local slope (*slope*), upslope contributing area (*area*), and local plan curvature (*curvature*).

Hence, similarity is achieved in this thesis when sites have the same topographic variables or some combination of those variables expressed as an index.

4.4 Landslides Viewed from Physical Geography

A view from physical geography provides some useful terms of reference and context for topographic variables.

The evolution of a bedrock hollow is complex. The *system* encompasses an energy and landscape development state. The system may be classified as *open*, which assumes there is an exchange of both mass and energy with the surroundings. The bedrock hollow system may be classified as a *process-response* system [Chorely, 1971] consisting of *morphological* and *cascading* sub-systems. The morphological system is a structural relationship between local topographic variables. Their connectivity is revealed by correlation. The morphological system includes the size and geometry of the hollow itself and surrounding soil. A systematic path of energy and mass defines cascading systems. The cascade system is characterized by thresholds having both spatial magnitude and geographic location that are dynamically linked by a cascade of mass and energy. The bedrock hollow is assumed to be a function of a hydrologic and debris cascade.

Marcus (1980) investigated the morphology and distribution of first-order drainage basins within the Hubbard Brook Experimental Forest (New Hampshire). He suggests that first-order drainage basin morphology is a basis for first-order watershed dynamics. A loose correlation can be drawn between first order drainage basins and potentially unstable terrain. First order basin morphology is illustrated in Figure [4] and consists of two complimentary sub-basins, a headwater region or valley head (VHD) and a stream region or channelway (CHW). The

morphology of each sub-basin is represented by a set of factors that includes area, length, slope, relief, elongation, and plan curvature.

The CHW region is a morphological unit dominated by a *size-shape* factor that indicates an organized flow system and the presence of a permanent channel. The size-shape factor is made up of length, area, relief, and elongation variables. The magnitude of the size-shape factor is responsible for providing a sufficient volume of erosive water and sediment to maintain a channel. The interpretation is supported by the presence of a pronounced plan curvature and permanent channel that dominate first order basin morphology. The VHD region has insufficient volumes of water and sediment available to form a permanent channel. The VHD and CHW regions converge at the *channel head*.

The VHD and CHW definitions are used to classify first-order basins into different morphological types; *length*, *side*, and *head* (see Figure. 4b). The morphological type is related to location in the larger drainage network, and this relationship helps explain *sub-basin* morphology. The channel-way's morphology is influenced by the stream junction location on the network, and the valley head morphology varies as a function of basin divide location (Figure.4c).

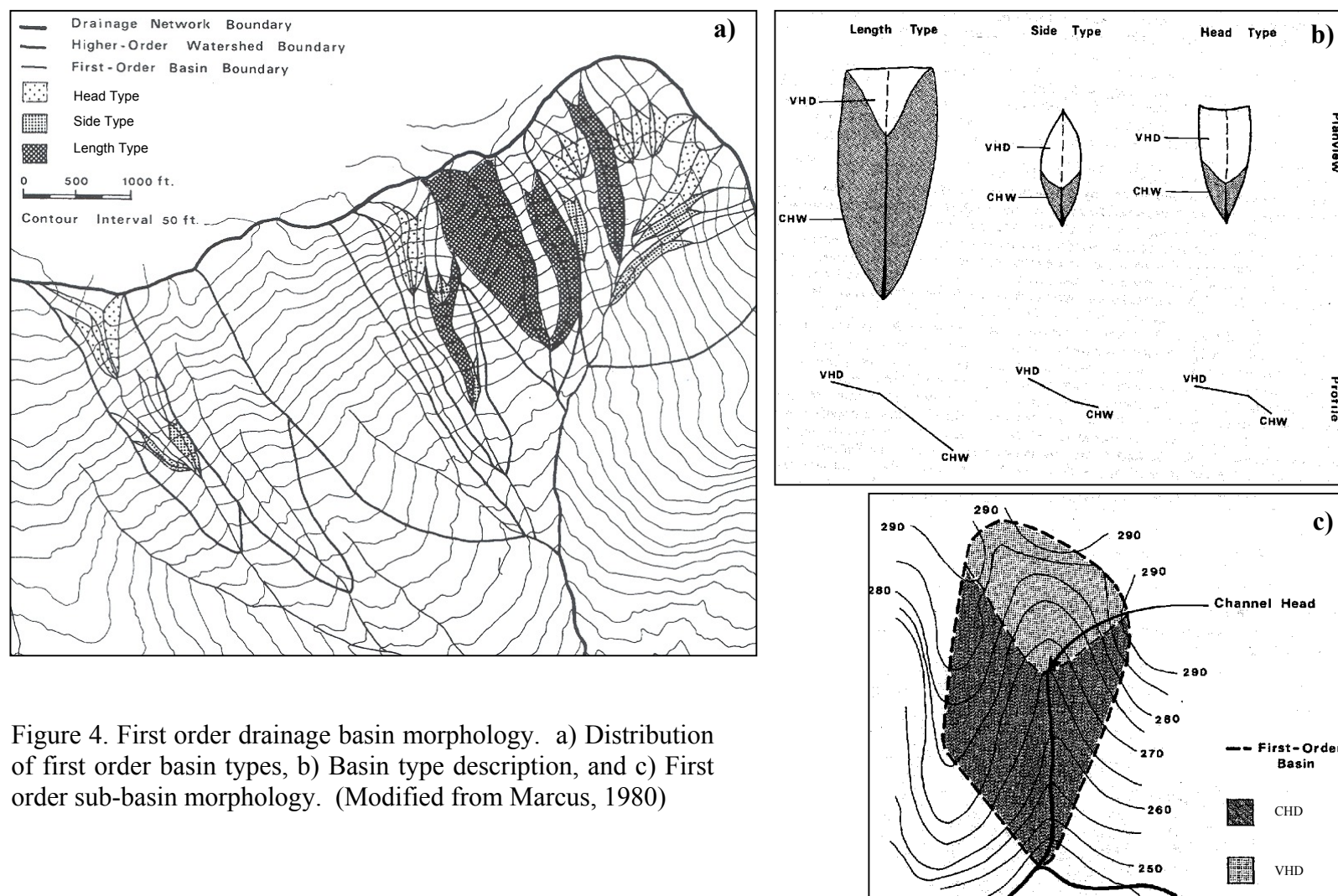


Figure 4. First order drainage basin morphology. a) Distribution of first order basin types, b) Basin type description, and c) First order sub-basin morphology. (Modified from Marcus, 1980)

Marcus (1980) highlighted a complex set of geomorphic variables associated with first order basins, i.e., length, area, slope, relief, plan curvature and elongation. Further, there are distinct sub-types of first order basins at a relatively fine scale of resolution ($\approx 100\text{m}^2$). These variables are all significant at the first order watershed scale. Marcus does not say how the significance of these variables changes with direction up or down the drainage network. Some variables may be significant at one scale and not another. The geomorphic variables are analog in nature. This thesis uses Digital Elevation Models (DEMs) that discretize all subsequent derivatives at a pre-determined resolution. Thus a relatively coarse resolution of digital data may not be able to pick up the significance of some variables or differences in basin types. An important question that arises is, what geomorphic variables are significant at a sub-first order watershed scale, and can a DEM provide enough accuracy, precision and resolution to measure them?

Montgomery and Dietrich (1989) suggests that a bedrock hollow lies in the valley head area as shown in Figure [5]

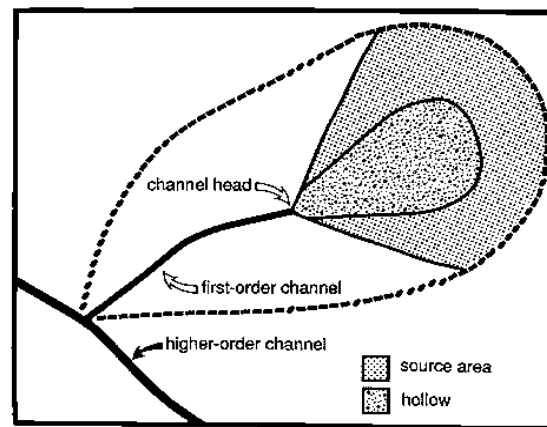


Figure 5. Illustration shows idealized relationships between hollow source areas and first order basins. After Montgomery and Dietrich (1989).

Montgomery and Dietrich (1989) concluded that channel head locations might be controlled by landsliding on steeper slopes and by seepage erosion and saturation overland flow on gentler slopes. The channel head is difficult to distinguish in the field, and is sometimes located up inside the hollow itself. Models on channel head initiation demonstrate a variable source $Area = A(t, R)$, where t = time, and R = discharge rate. Many channels do not connect to the drainage network; therefore, channel head initiation in humid environments must be due to hill-slope process, rather than headward extension of the network.

O'Loughlin (1981) determined that zones of saturation and stability of lower area boundaries are inversely proportional to the degree of convergence.

A morphological system may be defined by the slope, area and curvature variables measured from the ground surface. The present landslide inventory occupies sites somewhere in the valley head and channel-way of the first order basin. The landslide inventory may also be defining the channel heads, the location of which is highly variable over time, and partially a function of the area morphology.

4.5 Hazard Mapping and Types of Analysis

Many types of hazard mapping have been proposed. They differ by the type of information that is collected and method that is used to translate the information into a hazard rating. The strong points, weak points, and applicable scales of nine different methods of hazard mapping are referenced in Figure [6].

| Method of Analysis | Map Scale* | Strong Points | Weak Points |
|-----------------------------|------------|---|--|
| A-Landslide distribution | S-D | - objective - useful data base | - qualitative - no prediction |
| B-Landslide activity | M-D | - objective - useful data base | - no prediction |
| C-Landslide density | S-L | - objective - quantitative | - no prediction |
| D-Subjective geomorphic | S-D | - flexible - makes use of expert skills - useful data collection | - qualitative - difficult to review - requires high skills |
| E-Subjective rating | S-L | - flexible - makes use of expert skills - work can be delegated and checked | - qualitative - danger of oversimplification |
| F-Univariate susceptibility | M-L | - objective - shows effect of factors | - qualitative - relies on data quality |
| G-Univariate probability | M-L | - quantitative - flexible | - danger of wrong factor selection - relies on data quality |
| H-Multi-variate probability | M-L | - quantitative - precise | - danger of wrong factor selection - high reliance on data quality - difficult |
| I-Stability | M-D | - quantitative - can be reviewed - shows influence of some factors | - danger of oversimplification - may conceal lack of knowledge |

Figure 6. Matrix illustrates different types of hazard mapping. Map scale listed as Small (S), Medium (M), and Large (L), Detailed (D) (After the B.C. Earth Science Task Force, 1994).

This thesis will focus on the quantitative methods F, G, H, and I. Note, the Mapleton Headwall Rating System falls under category E (see section on Site Similarity).

The following are excerpts from the *Landslide Hazard Mapping Guidelines for British Columbia* (B.C. Earth Science Task Force, 1994);

Statistical methods produce a quantitative relationship between hazard ratings and the performance of slopes. Two types of correlation exist;

1. A relative correlation (*susceptibility*) is based on the assumption that units of land surface, which are similar in certain critical attributes to areas, which failed in the past, are most likely to fail in the future.
2. An absolute correlation (*probability*) goes one step further in assuming that future landslide frequency in similar units can be predicted, based on known frequency of occurrence in failed units, over a given time period.

The nature of the term probability will on occasion be characterized as *retrogressive*, since it is calculated based on landslide events that have occurred previously. Therefore, susceptibility has the same connotation as retrogressive probability.

Performance data may take two forms;

1. Landslide density, measured in the number of events per unit area.
2. Areal density, expressed as a percentage of the susceptible area to the total area.

Univariate probability analysis is a direct statistical correlation between the probability of landslide occurrence and a single variable (e.g., slope, and area). The statistical models are commonly factored with indicator variables (e.g., vegetation age class, and sub-basin). The probability of occurrence is usually a spatial distribution and may also be a temporal distribution.

Multivariate probability analysis uses multiple regression techniques to establish a correlation between landslide frequency and a group of attributes. The method can be applied to points or polygons. “The main disadvantage of the multivariate approach is that it removes any possibility of participation of the mapper’s experience and judgement in producing correlations. Thus, the results are totally dependant on the quality of input data”.

Stability analysis methods may be based on the infinite slope equation, which determines the limit equilibrium of a long shallow slope segment of uniform description. The *Factor of Safety* is used in the engineering sense as the ratio by which shear strength of the slope material exceeds the value needed to maintain equilibrium. A ratio of 1 indicates that failure is imminent. The Factor of Safety is spatially distributed, and may be temporally distributed if the ground water table is modeled over time.

There are two types of Factor of Safety;

1. Deterministic – single factor of safety given a single set of input parameters
2. Probabilistic- a probability distribution of factors of safety given probability distributions on the input parameters.

4.6 Topographic Indices Related to Hydrology

Hewlett and Hibbert (1967) introduced the concept of the variable source area that coupled runoff to watershed morphology. Beven and Kirkby (1979) developed the concept of area density, or Specific Area, a , defined as the area, A , divided by the contour width, b (Figure [7]).

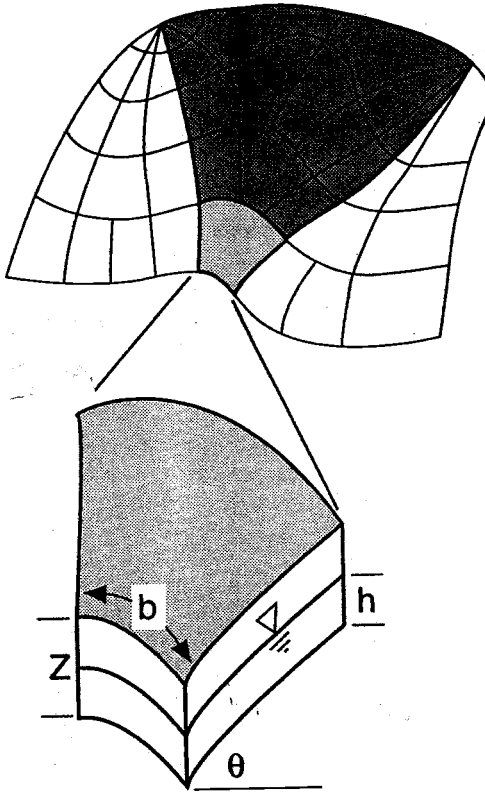


Figure 7. Illustration of conceptual drainage area with notations for local geometry

Their modeling efforts produced an index called the topographic ratio; $\ln(a/\tan\theta)$ where $\tan \theta$ is the local ground slope used to approximate the hydraulic gradient. The index provides an indication of the relative wetness across a landscape that refers to the saturated portion of the soil profile above a layer with highly contrasting permeability. Relative wetness is proportional to the area and inversely proportional to the site's ability to drain water.

O'loughlin (1981) showed that a simplified physical analysis of uniform lateral flow in idealized hill slopes leads to predictions of surface saturation using a relationship between *capacity* and *drainage flux*. Hillslope capacity, C , is defined as the maximum flow that can occur in the profile determined as;

$$C = \sin \theta \int_0^l K(z) dz = \sin \theta T = \sin \theta K z \cos \theta$$

K is the hydraulic conductivity and $z \cos \theta$ is the perpendicular depth of the soil profile. T is the transmissivity and θ is the local angle of slope.

The drainage flux (discharge per unit width) is defined as $q(x) = AQ/b$. The contributing area, A , drains through an element with an x-axis measured perpendicular to the contour, Q is the discharge per unit area, and b is the unit contour width.

The capacity is equated with drainage flux and re-arranged to separate the topographic and hydrologic variables. The resulting equality is;

$$\frac{A}{\sin \theta b} = \frac{T}{Q}$$

A criterion for saturation is established where flux exceeds capacity, and notated as;

$$\frac{A}{\sin \theta b} > \frac{T}{Q}$$

The result is an explicit expression for the topographic effect on the initiation of saturated areas. The right hand term is known as the *hydrologic ratio*. The left hand term is the *topographic ratio* and consists of two important factors;

$$\text{Factor \#1} \quad \frac{1}{\sin \theta} \quad \text{and} \quad \text{Factor \#2} \quad \frac{A}{b}$$

Factor #1 is a vector characterized by a direction perpendicular to the contours. Factor #2 is a scalar known as the specific surface area discussed earlier. Some useful inferences on saturation behavior as it relates to these factors are;

- As local angle increases, factor #1 decreases, thus topographic ratio decreases. In other words, the ability to achieve profile saturation is inversely proportional to the element's local angle
- As the area increases, factor #2 increases, thus topographic ratio increases. In other words, the elements ability to achieve profile saturation is proportional to the area.
- It follows that Factor #1 and Factor #2 are inversely proportional to one another, and their combined effect on saturation ability could be an increase or decrease.

The combined effect of Factor #1 and Factor #2 on the analysis of hillslope hydrology is referred to as an equi-finality problem. Different values of the numerator and denominator can result in the same quotient. Thus, analytical results that depend on unique values of the quotient will be confounded.

Dietrich et. al (1986) derived a relationship between a steady state hydrologic model and the infinite slope stability model. The major hydrologic assumptions are conservation of mass, steady state rainfall, and uniform lateral flow. A criterion for

the accumulation of flow is determined by equating capacity with rainfall. The resulting equality is $qA = T \sin \theta b$ where q is the effective rainfall after losses. A partial saturation ratio is defined as;

$$\frac{h}{z} = \frac{(qA)_h}{(qA)_z} = \frac{(qA)_h}{T_z \sin \theta b}$$

Where h is the perpendicular depth of the saturated soil profile. Note, the uniform flow assumption requires this ratio to be less than or equal to 1.0.

Next, a simplified version of the infinite slope model on a partially saturated slope is expressed as;

$$\frac{h}{z} = \frac{\gamma_s}{\gamma_w} \left(1 - \frac{\tan \theta}{\tan \phi'} FS \right) \text{ where FS is Factor of Safety}$$

Where γ_s is the unit weight of saturated soil, γ_w is the unit weight of water, and ϕ' is the internal friction angle. This is the Mohr-Coulomb failure criterion for a cohesionless soil when FS is equal to one. The moist unit weight is assumed equal to the saturated unit weight. When the left side quantity becomes larger than a known set of parameters on the right, a failure will theoretically occur. There is no assumption that the ratio be less than or equal to 1.0 like the uniform flow approximation.

Finally the partial saturation ratio is equated with the infinite slope equation to yield the following expression;

$$\frac{A}{b} = \frac{\gamma_s}{\gamma_w} \left(1 - \frac{\tan \theta}{\tan \phi'} \right) \frac{T_z}{q} \sin \theta$$

When the quantity on the left exceeds the quantity on the right a failure is supposed to occur.

The linked hydrologic-geomorphic model is a useful tool to examine the effects of topography on slope stability. However, the physical parameters are presently impossible to measure on a landscape scale. Effective values may be derived on a landscape scale through calibration with a landslide inventory. Variations on the linked hydrologic-geomorphic model exist and include such modifications as cohesion terms, and probabilistic distributions for physical parameters [R.T. Pack et.al, 1988].

4.7 Dimensionless Analysis of Topography

Consider an area draining across a finite element width b shown in Figure [8]

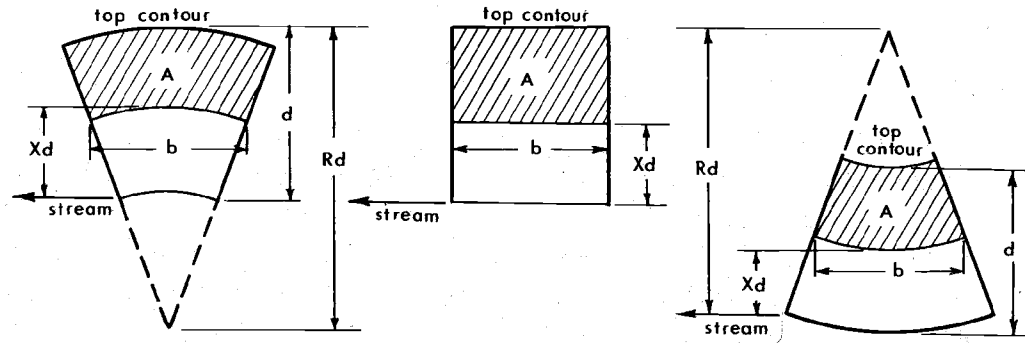


Figure 8. Illustrations of converging, parallel, and diverging flow regimes. Shaded zones are areas draining across a section X_d from the stream. After O'Loughlin (1981).

Figure [8] shows three types of area in plan view with the y-axis perpendicular to the flow gradient, and tangent to the contour. The x-axis is shown parallel to the flow gradient. A specific area function $g(x)$ may be examined entirely in a dimensionless coordinate system $X = x/d$, and $R = r/d$, where x is the coordinate, and r is the radius of curvature. Dimensionless terms shift the focus of the analysis onto shape by removing scale. Analysis of idealized terrain can assist in the understanding the shape effects on $g(x)$. The $g(x)$ functions are listed in Table [1].

Table 1. Expressions for dimensionless area

| Flow Regime | g(x) | dX/dQ | Approximate dX/dQ for X <<1 | | | |
|-------------|------------------------------------|------------------------------------|-----------------------------|--------------------------|----------------------------|----------------------------|
| | | | R=1 | R=2 | R=∞ | |
| Converging | $\frac{R^2-(R-1+X)^2}{2(R-1+X)}$ | $\frac{2(R-1+X)^2}{R^2+(R-1+X)^2}$ | $\frac{\sin\theta}{T}$ | $2X^2$ | $\frac{2 \sin\theta}{T}$ | $\frac{0.4 \sin\theta}{T}$ |
| | | | Q^2d | Q^2d | Q^2d | Q^2d |
| Parallel | $1-X$ | | $\frac{\sin\theta}{T}$ | | | $\frac{\sin\theta}{T}$ |
| | | | Q^2d | - | - | Q^2d |
| Diverging | $\frac{(R-X)^2 - (R-1)^2}{2(R-X)}$ | $\frac{2(R-X)^2}{(R-X)^2+(R-1)^2}$ | $\frac{\sin\theta}{T}$ | $\frac{2 \sin\theta}{T}$ | $\frac{1.6 \sin\theta}{T}$ | $\frac{\sin\theta}{T}$ |
| | | | Q^2d | Q^2d | Q^2d | Q^2d |

A sensitivity analysis of the g(x) functions is shown in Figure [9].

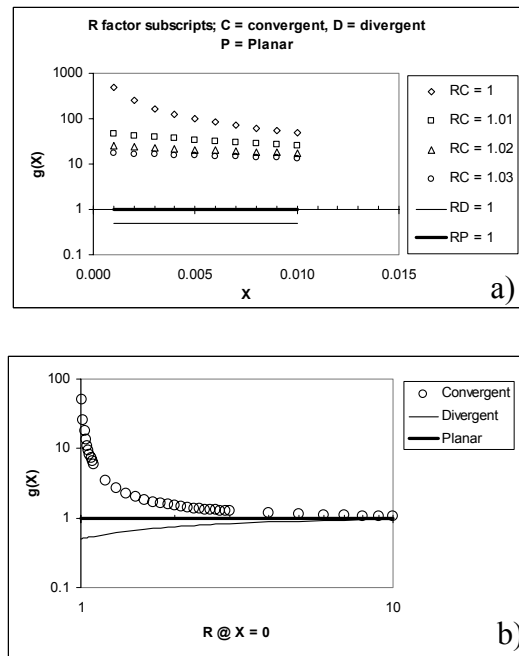


Figure 9. Graphs show sensitivity of the dimensionless area functions. Area function $g(x)$ is plotted against arbitrary values of dimensionless factors a) X and b) R .

Figure [9] shows that the X factor is inversely proportional to $g(X)$ in all types of terrain, however, the effect of a change in X is negligible for values of R marginally greater than 1. The $g(X)$ function changes rapidly as X goes to 0, i.e., small marginal zone located at the bottom of a drainage. Thus, curvature measured at the bottom of the drainage should be suitable index or surrogate for a true radius

measured across the entire contributing area of non-idealized terrain. In convergent terrain the value of $g(X)$ gets very large as X goes to zero, and R goes to 1 (i.e., increase in convergence), therefore saturation is likely. The $g(X)$ function is the same for all types of terrain at high values of R , since they all look planar and the value goes to unity.

The hydrologic and topographic ratios may be further normalized with the basin length d measured along the x -axis. The result is a function $g(X) = A / bd$ which is a dimensionless ratio. At the lower boundary of the drainage element the saturation criterion becomes $g(X) = \sin\theta T / Qd$. The derivative of X with respect to Q can be examined for each flow regime. The derivatives are also listed in Table[1] along with their approximations as X goes to zero. The value of $\partial X / \partial Q$ describes how a wet area expands or contracts as the drainage flux changes. Note the value of $\partial X / \partial Q$ is X^2 smaller in convergent zones than divergent zones. This means the location of the saturation boundary is much more stable in convergent than divergent zones.

Next, consider the dimensionless elevation ratio as it relates to the slope position; $Z = z/H$, where z is the vertical coordinate, and H = total height of the slope. A function $Z(X)$ can be used to construct a 2 dimensional slope profile. Further the slope for any X may be estimated by $\sin\theta = \frac{dz}{dx} = \frac{H}{d} f(X)$. Where $f(X)$ is a slope function not an elevation profile. The expression may be solved for $f(X)$ and placed back into the saturation criterion. The result is expressed as;

$$h(X) > \frac{g(X)}{f(X)} > \frac{TH}{Qd^2} \text{ or in words this becomes;}$$

$$\frac{\text{Normalized area}}{\text{Normalized gradient}} > \text{hydrologic ratio} * \text{scalar} \quad \text{or} \quad \text{flux} > \text{capacity}$$

A seepage boundary will occur at a distance X from the bottom of the area wherever the inequality is satisfied. The function $h(X)$ is plotted for an arbitrary $Z(X)$ profile against the dimensionless distance X in Figure [10] below;

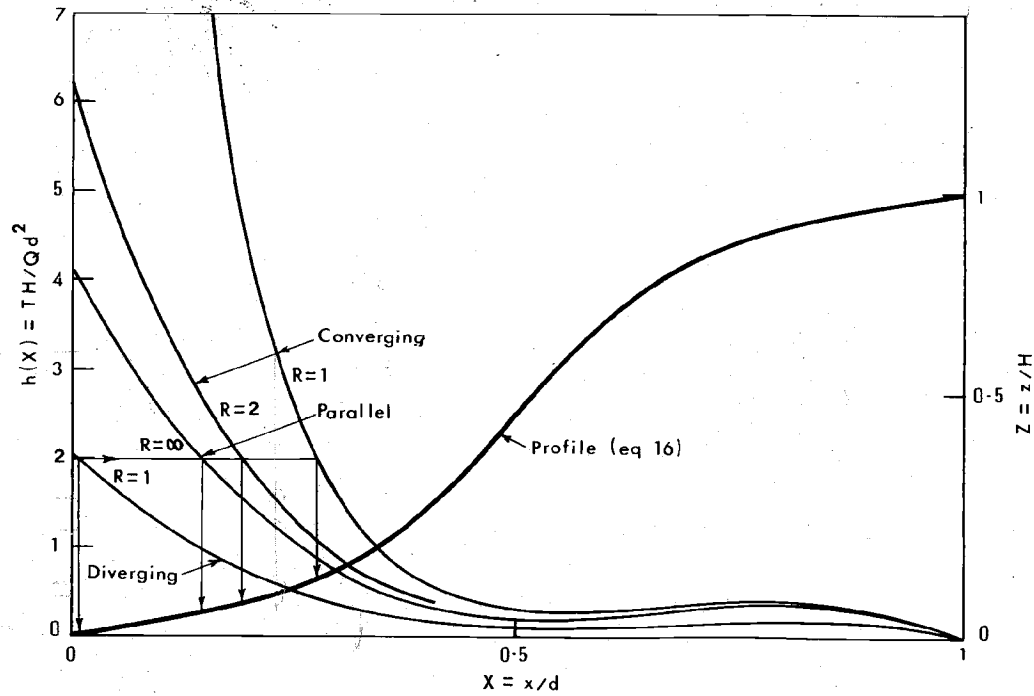


Figure 10. Graph shows examples of dimensionless topographic ratios different flow regimes. After O'loughlin, 1981

The $h(X)$ curves in Figure [10] show the location of saturated boundaries in convergent zones are insensitive to various drainage conditions (i.e., hydrologic ratio). The converse statement is also worth making, i.e., for a given hydrologic ratio, the location of a saturation boundary is highly sensitive to the degree of convergence.

Non-dimensional measurements in empirical model of landslide occurrence are unlikely since so much depends on scale and threshold values of dependent variables. However, finite estimates on the degree of convergence measured with variables r , and d , may prove useful. Another example may include weighting the area variable with $g(x)$ to produce an *effective area*. The added significance (i.e. statistical power) of a convergence variable will be proportional to the measured degree. Therefore, convergence should be insignificant in planar and divergent flow regimes. It follows that the area measured in the planar and divergent flow regimes will also experience a decrease in significance with decreasing convergence. Thus there is a physical basis for an Area:Curvature interaction. Simple measurements of area in any flow regime will always be highly significant in a statistical model due to its relative importance as an explanatory variable.

In retrospect, the practical significance of the Area:Curvature interaction is evident in a landslide attribute study in the Queen Charlotte Islands of British Columbia. (Rood, 1990) The slope variable was the dominant factor in open slope type (i.e., planar and divergent terrain) landslide frequency. In gullied terrain, the failure rates were more sensitive to slope form and position which are crude surrogates for an area term weighted with $g(x)$.

5.0 Materials and Methods

This thesis uses a landslide inventory because of the power offered by a large sample size related to one storm event in an area with homogeneous geology. Geographic Information System (GIS) are used to accommodate the large quantity of spatial data, and to perform the basic numerical operations. Topographic data are extracted from a Digital Elevation Model (DEM) because the data are freely available. The elevation data is used to directly determine lower order indices and then combined to higher order forms.

5.1 Digital Elevation Model

Topographic data are extracted from a Digital Elevation Model (DEM) with a 30-meter square cell resolution. The analysis begins with grids in a raw state that require pre-processing in order to delineate watersheds and remove spurious data. The raw grids have map extents equivalent to the common 1:24 000, 7.5 minute quad maps. The USGS publishes the raw grids that can be freely downloaded from the internet [O.G.D.C., 2001].

A combination of Geographic Information Systems (GIS) are used to process and extract topographic data from the DEMs. These include the full-featured system *ARC/INFO8.0* and a desktop variety *ArcView3.2*. Analyses are carried out inside the *ARC/INFO8.0* with the Arc and Grid modules, and inside the *ArcView3.2* environment with the *Spatial Analyst*, *SHALSTAB*, and *SINMAP* extensions. Some analyses may be carried out with different procedures inside different systems with the same intended result. Unfortunately, there may be subtle differences in the software algorithms that create differences in the analytical output. Therefore, particular algorithms are documented where appropriate in the follow sections.

5.2 Preprocessing of Elevation Data

Four grids are merged using the together to cover the entire Mapleton watershed area. Cells outside the watershed area are assigned *NODATA* values. Modification was completed inside the *ARCINFO8.0* Grid module with the MERGE, and *BASINS* command. An illustration of the Mapleton grid extent is shown in Figure [11].

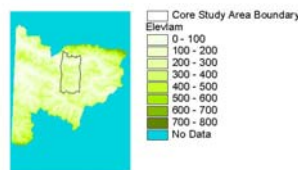


Figure 11. Grid extent constructed from four USGS grids. The solid shaded area classified as *NODATA* surrounds Mapleton Watershed boundary. The study area boundary is inside the watershed.

The large watershed extent shown in Figure [11] is intended to minimize edge effects with *NODATA*.

The raw grids are in a Universal Trans-Mercator (UTM) projection with X,Y, and Z units of metres. Grid processing and topographic data extraction are carried out in the UTM projection. The output grids are re-projected into the Oregon State Centered Lambert Conformal Conic (Lambert) using Grid commands to match the

projection used by the ODF landslide inventory. Statistical analyses are carried out on data in the Lambert projection.

Most grids contain sinks which are depressions in a DEM. Drainage area algorithms require that sinks be removed to define a flow direction. The *ARCINFO8.0* documentation provides a concise summary of sinks.

To create an accurate representation of flow direction and therefore accumulated flow, it is best to use a data set that is free of sinks. A digital elevation model that has been processed to remove all sinks is referred to as a depressionless DEM. Sinks in elevation data are most commonly due to errors in the data. These errors are often due to sampling effects and the rounding of elevations to integer numbers. Naturally occurring sinks in elevation data with a cell size of 10 meters or larger are rare except for glacial or karst areas, and generally can be considered errors. As the cell size increases, the number of sinks in a data set often also increases.

Sinks are removed using the *SHALSTAB*, and *SINMAP* extensions, or with Grid commands in the *ARCINFO8.0* environment. The software documentation for all three programs declares similar algorithms for sink removal [Jenson and Dominique, 1988] [On-line users Manual at Utah State University] [On-line users Manual at University of California, Berkeley]. Examination of the sink areas offers some insight to the elevation errors present and may serve as a partial quality control check on slope and area calculations. One standard deviation of the sink distribution is 0.034 to 0.225 m based on the *ARCINFO8.0* system. Most of the sinks are located on the valley bottom and none are coincident with the landslides population. Therefore, sinks are not likely to have any deleterious effects on the present analysis.

5.3 Computation of Low Order Topographic Indices

Indices of slope, area, and curvature are extracted directly from the DEM and notated by;

- slope θ_i [degrees]
- Up-slope contributing area A_i [L^2]
- Curvature C_i [L]

The subscript i refers to the particular algorithm.

The indices may be referred to as local, since they are based on local data surrounding a center cell. The algorithms extract elevations from raster cells forming a 3x3 window surrounding the j 'th center cell. Different algorithms use all or some of the data inside the 3x3 window in different ways.

Low order topographic indices and then combined to form higher order indices. Two geometric transformations of slope are chosen to help discriminate between processes dominated by hydraulic gradients (sine) and friction (tangent). Logarithm base ten transforms are carried out on the area indices for convenience and to produce normal distributions..

There are many ways to determine the up-slope contributing area. The purpose is to recognize that various assumptions influence the distribution of the contributing area variable across the landscape. Every contributing area algorithm makes an assumption on sink removal, flow directions, and flow portioning. Sink removal is discussed in section 5.1. Some slopes are intended to represent an area inside a cell and others are intended to define a flow direction out of the cell. Flow directions may be defined using different portions of the data inside the 3x3 window and will

be discussed individually in the following sections. Flow portioning is an attempt to define flow boundaries at a sub-cell scale. The importance of the flow portioning assumption occurs on divergent terrain where water could follow significantly different directions. Endreny, 2001, offers a thorough review and contrast of drainage area algorithms. A summary of the contributing area algorithms and their attendant flow direction algorithms used in this thesis in Table [2]

Table 2. Summary of drainage area algorithms

| Drainage Area Index | Slope Index | Flow Partitioning |
|---------------------|--|---|
| A_1 | 3x3 geometric mean (θ_1) | multiple-flow-direction; proportional to the local gradient out of the center cell to each adjacent and diagonal cell. <i>SHALSTAB</i> |
| A_2 | 3x3 maximum drop (θ_3) | single-flow-direction; area is accumulated in the direction of flow. <i>ARCINFO8.0</i> |
| A_3 | Dip slopes calculated from triangular facets | multiple-flow-direction; apportioned to two down-slope pixels according to flow angle (strike) separation from cardinal direction relative to the pixel center. <i>SINMAP</i> |

The algorithms used for calculating slope, area, and curvature are discussed below.

5.3.1 Slope Indices

Slope θ_1 is calculated as the geometric mean from the surrounding eight cells [Montgomery and Dietrich, 1988]. The two directions normal to the cell have a spacing of 2 times the grid size, whereas the slopes diagonally across the cell have a distance that is 2.83 times the grid size. Slope θ_1 is determined by default inside the *SHALSTAB* extension. Default units are $\text{Tan}\theta_1 \times 100$ (i.e., percent).

Slope θ_2 is calculated from a third-order finite difference estimator of the surrounding eight cells [Burroughs and McDonnell, 1988]. The formula is;

$$\text{Tan}\theta_i = \sqrt{\left(\frac{dz}{dx}\right)^2 + \left(\frac{dz}{dy}\right)^2}$$

The deltas dx, dy, and dz are calculated using a 3x3-roving window. The window is defined by a through i

| | | |
|---|---|---|
| a | b | c |
| d | e | f |
| g | h | i |

$(dz/dx) = ((a + 2d + g) - (c + 2f + i)) / (8 * x_mesh_spacing)$ (dz/dy)
 $= ((a + 2b + c) - (g + 2h + i)) / (8 * y_mesh_spacing)$

Slope θ_2 is determined inside the *ARCINFO8.0* environment using the Grid module *SLOPE* command. Default units are given as $\text{Tan}\theta_2$

Slope θ_3 is calculated relative to a center from the surrounding eight cells [Jenson and Dominique, 1988]. The maximum elevation difference from the eight adjacent

values is selected for the slope calculation. The flow direction out of a cell is assumed to follow this maximum drop. The slope is therefore representative of a segment located between the grid cells, but the value is assigned to the center cell. Index θ_3 is determined inside the *ARCINFO8.0* environment using the Grid module *FLOWDIRECTION* command. Default units are $\tan\theta_3$.

Slope θ_1 smoothes the landscape relative to the local maximum slope θ_3 . Montgomery and Dietrich, 1998 indicate that slope θ_1 is most desirable for minimizing grid artifacts and produces a superior index of landslide occurrence. Grid artifacts are the result of the relative orientation of the grid. Montgomery and Dietrich did not declare a basis for index superiority. They also report minor differences with slope θ_2 . Burroughs and McDonnell, 1988 tested slope θ_2 and determined it is best for accurately measuring slopes from rough surfaces. Slope θ_3 is biased towards steeper values that represent inter-cell line segments. Slope θ_3 is the least accurate on rough surfaces (Burroughs and McDonnell, 1988 [39]). The USGS also tested slope θ_3 for defining flow directions and found that it accurately matched planimeter estimates of contributing areas 97% of the time [Jenson and Dominique, 1988].

5.3.2 Area Indices

A_1 is a multiple-flow-direction algorithm [41]. The total area that a grid cell has to hand-off is divided in proportion to the local gradient out of the cell to each adjacent and diagonal cell. The gradient is determined using the slope variable θ_1 (See section 5.2.1). A_1 is determined by default inside the *SHALSTAB* extension. Default units are given as $\log_{10}(a_1/b)$ then converted over to A_1 .

A_2 is a single-flow-direction algorithm [42]. Water is assumed to follow the maximum gradient out of each cell. The flow direction is determined using the slope variable θ_3 (see section 5.2.3). The present cell is not included in the area accumulation of cells. A_2 is determined inside the *ARCINFO8.0* environment with the Grid module using the *FLOWACCUMULATION* command. Default units are given as the number of grid cells then converted over to A_2 .

A_3 is a multiple-flow-direction algorithm [Montgomery and Dietrich, 1998]. The area is defined using a slope variable not defined in the previous section on slopes. The steepest flow direction is represented as a continuous quantity between 0 and 2π and is determined from the eight triangular facets formed in a 3x3 window. Each triangular facet defines a plane. The dip slope is calculated from each plane with vectors. The maximum dip slope defines the flow direction. When the flow does not follow one of the cardinal or diagonal directions, area is calculated by portioning the flow between the two down-slope pixels according to how close the flow angle is to the direct angle formed with the window center. This flow-portioning algorithm is part of the *SINMAP* program. Default units are given as $\log_{10}(a_3/b)$, then converted over to A_3 .

5.3.3 Curvature

Curvature variable C is measured in plan-form from a DEM using a 4th order polynomial using data inside a 3x3 window surrounding each cell. The unit of measure is taken from a second derivative of the polynomial function yielding a range of numbers; negative corresponding to convergent, and positive corresponding to divergent. Thus values of curvature near zero are on planar

slopes. Note, this is a local measure of curvature and not necessarily a measure reflecting the shape of the whole upslope contributing area that likely extends beyond the upper half of a 3x3 window. Curvature is determined inside the *ARCINFO8.0* environment with the Grid module using the *CURVATURE* command. Default units are 1/100 z-units,, then converted over to z units.

5.4 Higher Order Indices

Several types of higher order indices are constructed from the basic topographic variables of slope and area.

The first higher order index is the topographic ratio and is defined as;

$$\log_{10}\left(\frac{A_1}{\tan \theta_1}\right) \quad , \quad \log_{10}\left(\frac{A_1}{\sin \theta_1}\right)$$

The tangent form comes from Beven and Kirkby, 1969 [44] and the sine form comes from Oloughlin 1981 [45]. Both topographic ratios were calculated inside the *ARCINFO8.0* environment with Grid algebra.

A second type of higher order index is the infinite slope type, which comes in two varieties, the hydrologic ratio and the SINDEX. Both are calculated using the linked hydrologic geomorphic model (see section 4.6). The hydrologic ratio is defined as;

$$\frac{q}{T} = \frac{\rho_s}{\rho_w} \left(1 - \frac{\tan \theta_1}{\tan \phi'} \right) \frac{b}{A_1} \sin \theta_1$$

Where T , ρ_s , ρ_w , ϕ' , and b are constants. The equation has three topographic terms that are defined by the DEM; drainage area a , outflow boundary length b , and slope angle θ . There are four parameters that need to be assigned to apply this model; soil bulk density ρ_s , angle of internal friction ϕ , soil transmissivity T , and the effective precipitation q . Default parameters are used in this analysis, specifically; $\phi = 45^\circ$, and $\gamma_s = 1700 \text{ kg/m}^3$. The hydrologic ratio is determined using the *SHALSATB* program. The *SHALSTAB* program is supported by an on-line publication of University of California, Berkely[]

The SINDEXT is similar to the hydrologic ratio, however, the model is re-arranged into a probabilistic form of the factor of safety that uses A_3 . The physical parameters are assumed to possess a uniform distribution within user-defined limits. The graphical user interface allows iteration of distribution limits for the purpose of maximizing the number of landslides in a SINDEXT bin class. The SINDEXT is determined using the *SINMAP* (Stability Index MAPping) program. The *SINMAP* program is supported by an on-line publication of Utah State University[46].

The highest order index is derived from regression analysis on slope and area. The product of the regression is an expected value function or Expected Value Index (EVI) defined as;

$$EVI = \beta_0 + \beta_1 \tan \theta_1 + \beta_2 A_1$$

The coefficients are estimated from a maximum likelihood algorithm used in a logistic regression with a binary response variable (slide site = 1, non-slide site = 0). The maximum likelihood algorithm iterates coefficients until it achieves the highest retrogressive probabilities for the slide population. The result is a maximum difference between the EVI distributions for the slide and non-slide

populations. The non-slide sites were randomly selected from the Mapleton study area. The regression output is shown in Table [3].

Table 3. Logistic Regression Model for the EVI Index

| Terms | Coefficient | Std. Error | Z statistic* |
|----------------------------|-------------|------------|--------------|
| β_0 * (Intercept) | -9.27 | 1.65 | -5.64 |
| β_1 * $\tan\theta_1$ | 7.88 | 1.77 | 4.46 |
| β_2 * A_1 | 2.18 | 0.38 | 5.77 |

*Z statistic is from a standard normal distribution. Absolute values greater than 2 are significant at approximately the $\alpha = 0.05$ level.

6.0 Results

Empirical modeling and regression analyses are used to examine the behavior of higher order indices and their ability to explain variability in the slope and area distributions for the landslide population. Statistical methods are used to analyze topographic data to sort out the variability in the data. Stratification design is used to sort out spatial issues while attempting to distinguish between slide and non-slide sites. Logistic regression models are developed on spatially stratified areas to determine the retrospective odds of landslide occurrence with parameter sensitivity. The logistic models are also used to establish topographic definitions of landslide similarity. Finally, inter and intra-index performance is measured with the landslide and aerial densities (see Section 4.5 for definitions)

6.1 Empirical Modeling of the Slide Population

The linked hydrologic geomorphic failure criterion derived in Section 4.6 is tested to explain variability in the distributions of slope and area of the landslide population. The model parameters are simplified and the equation rearranged to establish a threshold area as a function of slope;

$$\frac{A}{b} = \frac{\rho_s}{\rho_w} \left(1 - \frac{\tan \theta}{\tan \phi'} \right) \frac{T}{q} \sin \theta \quad \text{becomes} \quad A = C(1 - \tan \theta) \sin \theta$$

The ρ_s , ρ_w , T , and q parameters are lumped into a single coefficient C , and $\tan \phi'$ is assumed equal to one. The equality holds at $FS = 1$, thus the C term is a minimum threshold coefficient. The area function is plotted in Figure [12].

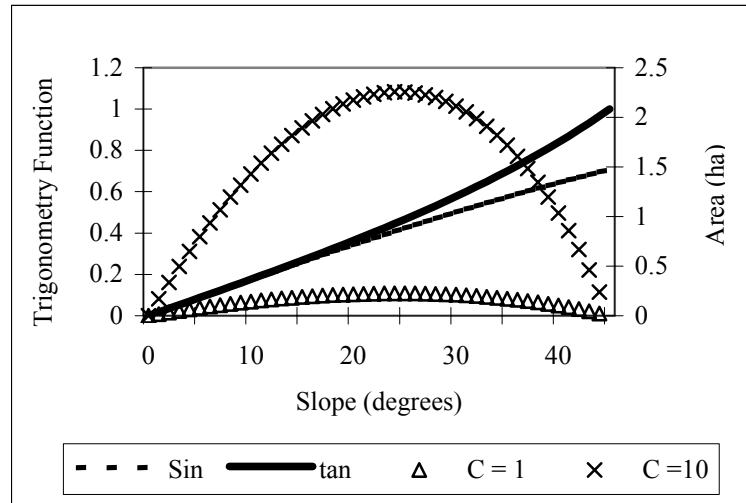


Figure 12. Graph shows conceptual plots of a threshold function $A = C(1 - \tan \theta) \sin \theta$ marked on right axis with threshold coefficients; $C = 1$ and 10. The sine and tangent slope transformations are shown on the left axis.

The purpose of the model simplification is to examine the behavior of the failure criterion purely as a function of topographic variables. The $\sin \theta$ variable is the hydraulic gradient and proportional to area. The $\tan \theta$ variable defines the failure plane and causes the $(1 - \tan \theta)$ term to be inversely proportional to area. The area function starts at zero with slope equal to zero which is hard to understand since landslides generally don't occur on gentle slopes with small contributing areas. An unknown variable must be missing to explain the area function behavior at the low end of the slope spectrum. The area function also equals zero with slope equal to 45 degrees. This makes sense since the friction angle is assumed to equal 45 degrees. Thus, if the site is already failing due to gravity, no drainage area is required. The first derivative of the area function is equal to zero at approximately 24.4 degrees. The existence of an area maximum is more important than the location on the slope scale.

The behavior of the threshold function beyond the maxima shown in Figure [12] is consistent with observations by O'Loughlin (1981) that local slope is inversely proportional to the ability of soil to become fully saturated. Thus, at lower slopes landslide occurrence should be associated with high soil profile saturation. There is a wide range in transition where slope takes over and less area is required to achieve a lower degree of soil profile saturation.

The area function is tested using real slope and area values from the Mapleton slide population. A value of C is determined for each slide to produce a distribution of values. The mean value of C equals 2.3 ha (95% C.I.; 1.6 to 3.0). The Mapleton landslide data is shown with threshold coefficients in Figure [13].

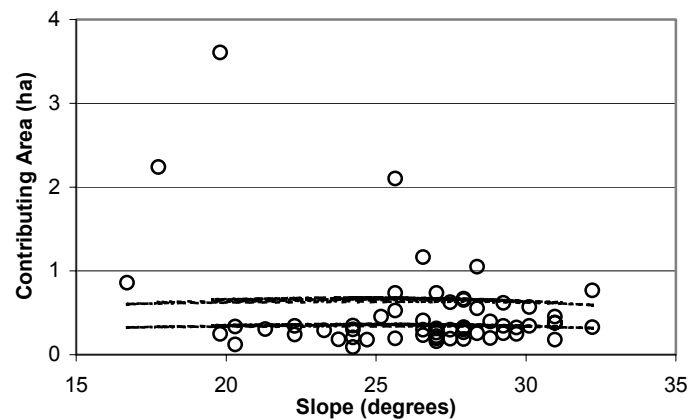


Figure 13. Plot shows slope versus area for the Mapleton slope slide population. Threshold coefficients; $C = 1.6$ and 3.0 represent the 95% confidence interval, based on the function $A = C \sin \theta (1 - \tan \theta)$ fitted to the data set. Explanatory variables are slope (θ_1), and area (A_3)

Most of the slides in Figure [13] are on slopes greater than the slope defined by the area maxima, therefore landslide occurrence should be dominated by the slope factor. There is significant deviance from the area function for slope values less

than the maxima. The deviance is most likely due to the model assumption that all factors are equal across the landscape. In reality, soil depth and transmissivity vary. Lower slopes should have deeper soils thus larger transmissivity thus larger area. It is also possible that some sites achieved a FS less than one given a constant rainfall. A FS less than one would result in an area larger than the threshold. The width of scatter in Figure [13] is inversely proportional to slope (i.e. possible narrowing). This may be a trend in soil depth and transmissivity but variability in the other parameters cannot be ruled out.

A least square's regression of the slope transform, $\sin \theta (1 - \tan \theta)$ against area is not significant (p-value = 0.3) with an R^2 of 0.02. Thus, there is no statistical evidence to support the application of a single threshold coefficient to the upslope landslides in Mapleton.

Another way to examine the effect of slope on a threshold value of area is to carry out a regression on $\tan \theta$ and $\sin \theta$. A full two-way regression model takes the form;

$$Area = \beta_0 + \beta_1 \tan \theta + \beta_2 \sin \theta + \beta_3 \tan \theta : \sin \theta$$

The coefficients are all significant (p values < 0.05). However, an R^2 value of 0.11 suggests that other factors account for most of the variability.

6.2 Statistical Analysis of Index Distributions

In this section, the differences in slope, area and curvature distributions between the slide and non-slide sites will be addressed. The first law of geography states that near attributes are more similar than attributes far away. In other words, topographic variables possess positive spatial correlation. The spatial contrast between slide and non-slide sites is proportional to the distance between them. The challenge is to determine where non-slide sites are no longer similar to slide sites. There is some minimum resolution at which a difference in variable distributions can be detected. The minimum resolution is determined using spatial stratification and statistical analyses for testing significant difference. The minimum resolution is a function of measurement error, sample size, and spatial rates of change in the variables. Stratification based on index is one way to examine questions regarding similarity.

Non-slide sites are spatially stratified, which limits the scope for sampling to specific topographic regions. A 3x3 window and three, five and ten cell radii surrounding a slide site defines the ranges of spatial stratification for the slide population. The spatially stratified regions are further stratified based on the slope variable using limits defined by the 95% confidence interval on the mean of the slide population. Spatial and slope stratification is illustrated in Figure [14].

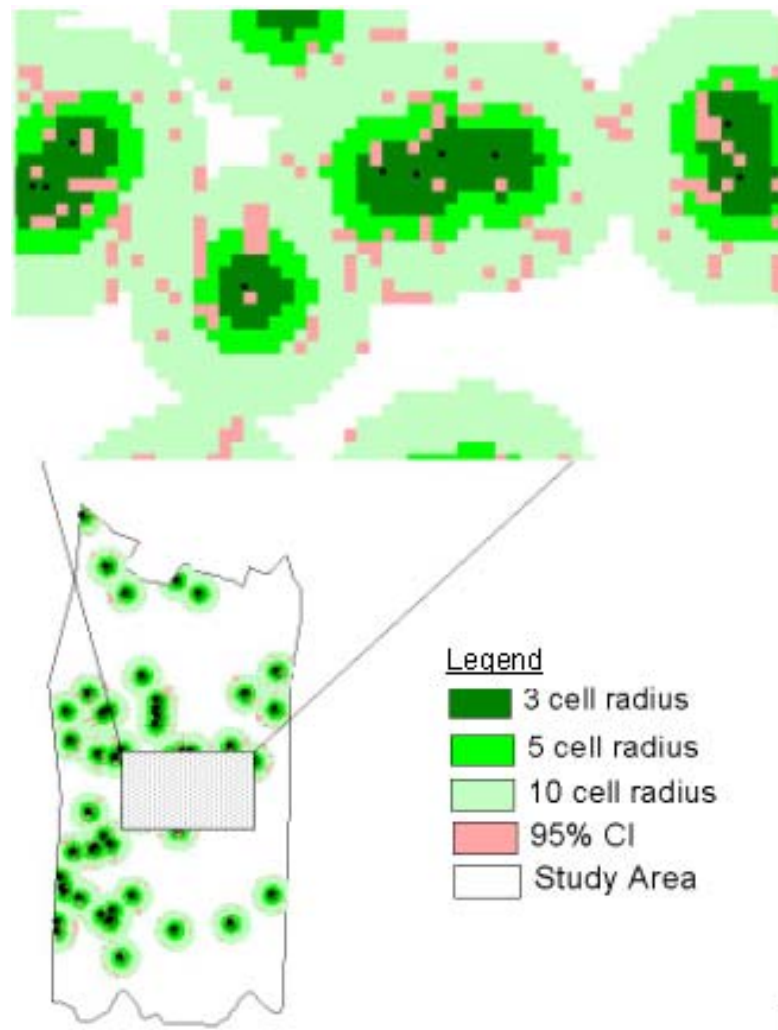


Figure 14. Graphic shows spatial stratification defined by three, five, and ten cell radii surrounding slides found in Mapleton. Slope stratification is defined by the 95% confidence interval on the mean of the slide population. Smaller scale section shows the intersection of spatial and slope stratification.

Slope and area distributions using only spatial stratification are displayed using box plots in Figures [15] and [16].

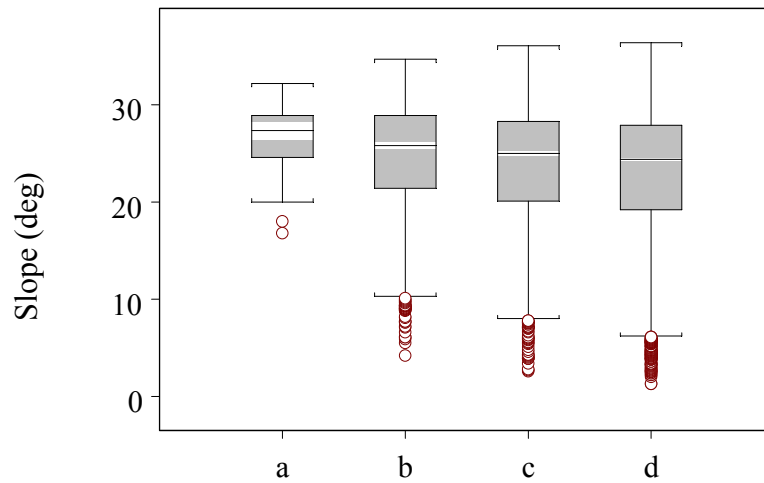


Figure 15. Box plot show the slope θ_1 distributions. Spatial stratification based on cells surrounding slide sites for the Mapleton slide inventory; a) slide population b) three cell radius, c) five cell radius, and d) ten cell radius.

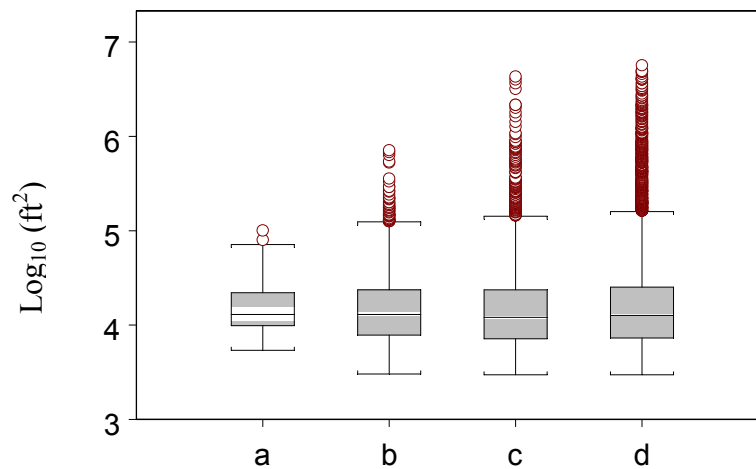


Figure 16. Box plot of shows log transformations of the area index A_1 . Spatial stratification based on cells surrounding the Mapleton slide inventory; a) landslide population b) three cell radius, c) five cell radius, and d) ten cell radius.

The “box” in a box plot is that shaded portion containing three horizontal lines indicating the 25, 50 and 75th percentiles respectively. The off colored portion of

the box, white or black, indicates the 95% confidence interval about the mean. Outliers are indicated by circles and defined as 1.5 times the inter quartile range. The extreme horizontal lines indicate values that lie within the inter-quartile range.

Figures 15 and 16 suggest that there is a significant difference between the distributions of slope, but no significant difference in the area. Distribution skew is proportional to the scope of spatial stratification for both slope and area. The mean value of slope is inversely proportional to the scope of spatial stratification, and the mean value of area is proportional to the scope of spatial stratification.

Logarithmic and inverse transforms do not remove the skew in the area distributions; therefore t-tools are limited for testing equal means. Further, as the scope expands it becomes more difficult to justify using t-tools, and larger scale variations complicate the search for a significant difference.

A rank transformation is performed on the area distributions. The results from a Wilcoxon-Rank-Sum test are shown in Table [4].

Table 4. Wilcoxon-Rank-Sum test of area distributions on spatially stratified regions.

| Population 1 | Population 2 | p-value |
|--------------|--------------|---------|
| Slides | 3cell | 0.43 |
| Slides | 5 cell | 0.19 |
| Slides | 10 cell | 0.38 |
| 3 cell | 5 cell | 0.09 |
| 5 cell | 10 cell | 0.01 |

Interpretation of the p-values suggests that a significant difference exists between the five-ten cell radius and three-five cell radius distributions. Unfortunately, this result has little practical significance since the topographic data can barely

discriminate basin form on a headwall scale (e.g., 5 to 6 acres). Thus, there is no evidence to suggest a difference in the area distributions for the slide and non-slide sites.

Distributions of area resulting from slope stratification are more similar than the distributions of area resulting from spatial stratification. Another set of random samples was taken with no spatial restriction, but with slope stratification still in effect and there was still no significant difference in the distributions.

Attempts to discriminate between distributions of area were unsuccessful.

Inspection of the distributions of area suggests a change in the design of spatial stratification. An area threshold at each slide site requires that all cells downstream belong to the slide population. Therefore data sets tested thus far contain response errors. Further, the down-stream cells are responsible for the extreme skew in the distributions of area. Spatial stratification using a three-cell radius with the down stream cells removed (*3cell-minus*) is shown in Figure [17].

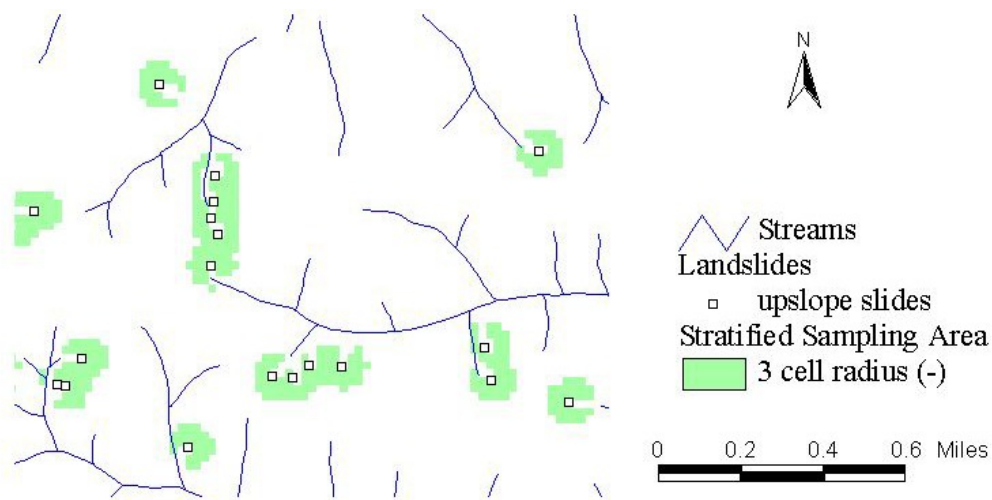


Figure 17. Sample area with spatial stratification of non-slide sites defined by a three-cell radius surrounding the slides in the Mapleton inventory with the down stream cells removed.

Removing the down-stream cells removes response errors and most of the skew in distributions of area. Therefore, t-tools can be used with the 3cell-minus data set. The test of equal means was used on the 3cell-minus data set. There is strong evidence of a difference between slide and non-slide sites for the distributions of slope (p-value = 0.013) and area (p-value = 0.006). A slope versus area plot for the 3cell-minus data set is shown in Figure [18].

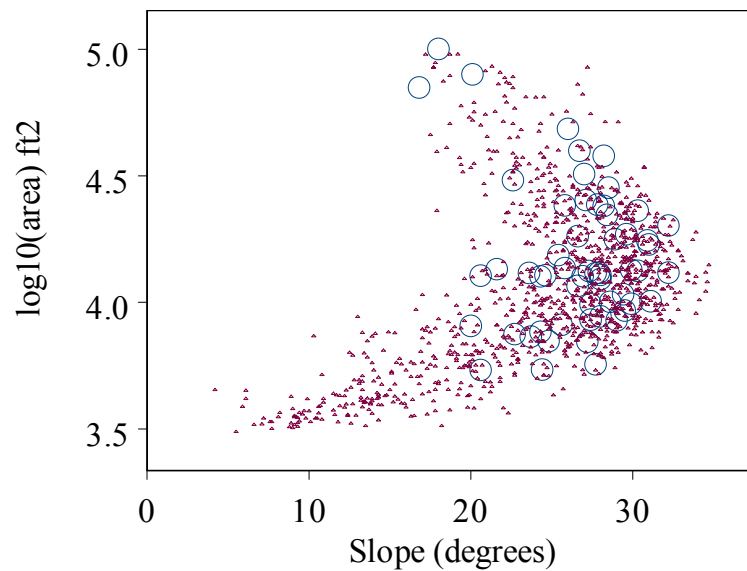


Figure 18. Scatter plot showing slope, θ_1 , versus area, A_1 , for spatially stratified regions around landslides in the Mapleton inventory. Slides sites are shown as open circles.

Figure [18] illustrates the difficulty in visually determining the difference in the slope and area distributions.

The distributions of curvature index are examined using the 3cell-minus data set. There is a significant difference ($p\text{-value} < 0.001$; rank sum test) in distributions of curvature between the slide and non-slide populations.

A scatter plot of slope and area versus curvature is shown in Figure [19].

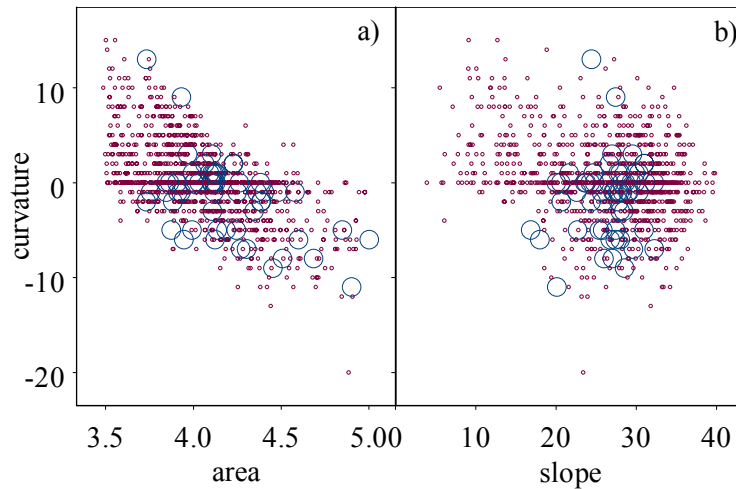


Figure 19. Scatter plots showing a) slope, θ_1 , and b) area, A_1 , versus curvature C_1 . Slides are shown as open circles. Stratification based on cells surrounding landslides in the Mapleton slide inventory.

A trend for curvature with area is shown in Figure [19], however, no trend is observed with curvature and slope.

6.3 Index Performance

Definitions that are accurate and result in the smallest area delineated as potentially unstable is the objective of most land managers. Indices are used here to define potentially unstable terrain. Indices that define the highest density of existing slides will be considered to have the best performance. Slide density is related to index correlation and may be the result of some role the index has in the physical

processes producing slides. However, slide density cannot distinguish between the role in physical process and simple correlation.

Topographic indices are derived from continuous variables with unimodal bell shaped distributions, see Figure [20]. A small variance in the index value for the slide population will result from a high correlation. The consequence of a small variance is a narrow distribution of index values compared to non-slide sites. Therefore, a small variance results in a high density of slides. Slide index distributions located on the extreme ends of the non-slide distribution will also produce high slide densities. The later comment describes the objective of logistic regression on topographic indices, see section 6.4. Non-slide distributions with positive spatial correlation and extreme skew possess high rates of change with distance across the landscape. The later type of index is a more precise tool for locating slides and also has higher densities. An ideal conceptual index is shown in Figure [20]

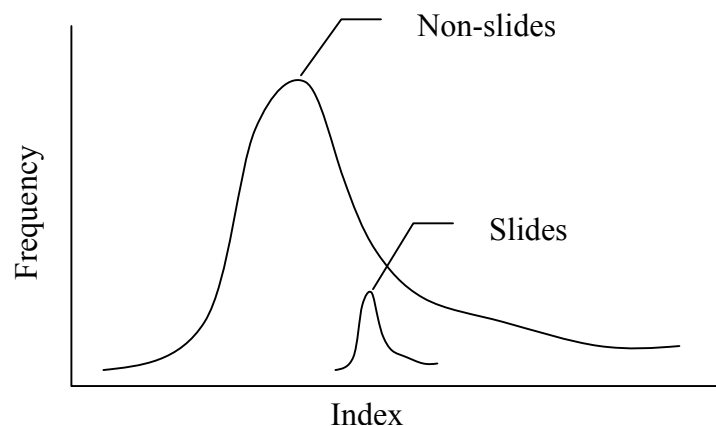


Figure [20] Graph shows a conceptual index distribution with ideal shape for high performance in locating landslides. The ideal index has a slide population with a small variance resulting in a narrow distribution located on the extreme end of a skewed non-slide distribution.

Both the slide and non-slide distributions in Figure [20] are assumed to be unimodal. Slide density increases as the slide population moves further to the right in Figure [20].

Two types of index performance are considered here.

1. Landslide density, measured in the number of events per unit area.
2. Areal density, expressed as a percentage of the unstable (susceptible) area to the total area.

Landslide associations with topographic variables are expressed as cumulative percent to allow easy reference between variable range, density, and sensitivity. The procedures for determining landslide density on cumulative scales are listed below.

1. Calculate the θ , A, and C etc.. for each grid cell within the study area. Associate these values with the presence or absence of landslide sites (i.e., slide = 1, non-slide = 0).
2. Rank the grid cells in order of least to greatest value of the variable.
3. Calculate the percentage of total area and percentage of landslide population to produce cumulative area (%) and cumulative landslides (%) functions.
4. Develop a matrix plot of all three variables, cumulative area (%), cumulative landslides (%) and topographic variable.

The performance statistic is a density ratio defined as;

$$DensityRatio = \frac{\Delta\% \text{ Cumulative Slides}}{\Delta\% \text{ Cumulative Area}}$$

Density varies with the range therefore each index has a maximum density or maximum intra-index performance. Alternatively, inter-index performance may be assessed for maximum values of landslide density measured across equal cumulative areas.

Intra-index performance data for slope and area indices are shown in Figure [21] and Figure [22] respectively.

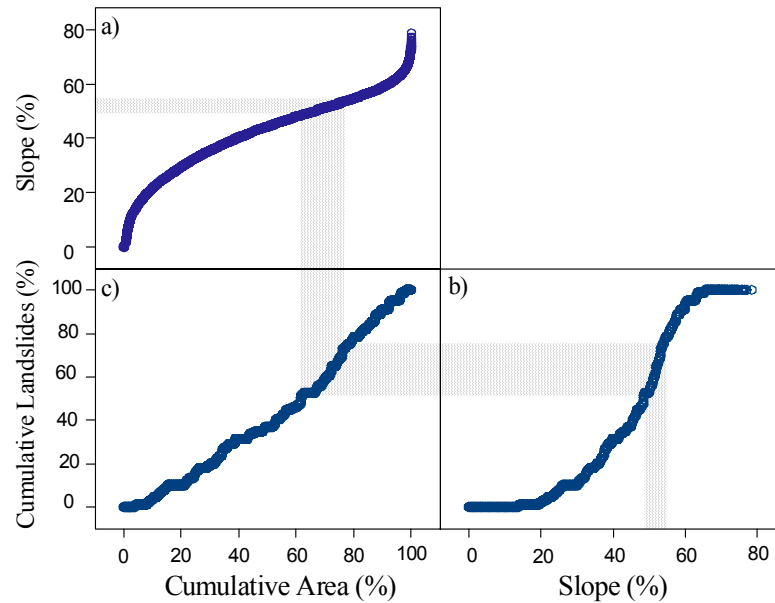


Figure 21. Matrix plot shows the landslide association with slope index θ_1 in the Mapleton study area. Gradients on each curve represent; a) terrain character, b) index correlation with slides, and c) landslide density. Stippled area is an example definition for potentially unstable terrain in the θ_1 domain. The width of the stippled area measured on the horizontal axis of c) defines the aerial density over a continuous range of the index.

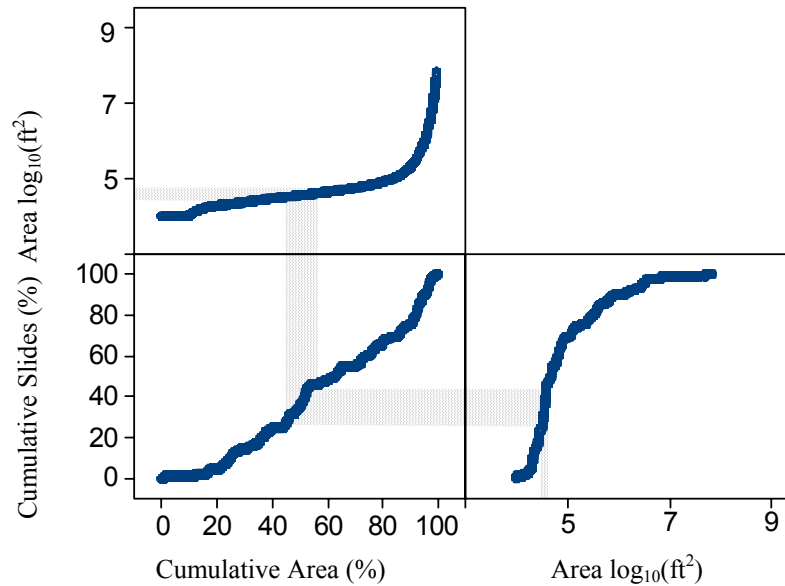


Figure 22. Matrix plot shows landslide association with area index A_1 in Mapleton study area. Gradients on each curve represent; a) terrain character, b) index correlation with slides, and c) landslide density. Stippled area is an example definition for potentially unstable terrain for the A_1 index. The width of the stippled area measured on the horizontal axis of c) defines the areal density over the continuous range of the index.

The matrix plots in Figures [21] and [22] provide a means to associate slide density with index performance. Slide density on these plots is defined across a continuous range of the index, and the smallest and largest values of the index define the study area. Indices are continuous variables discretized at a 30m scale, while slide accumulations create step plots due to their binary values. Good index qualities include a high sensitivity (i.e., steep average slope) to landslide occurrence and high landslide density.

A contrast of the density ratios from all indices is shown in Figure [23].

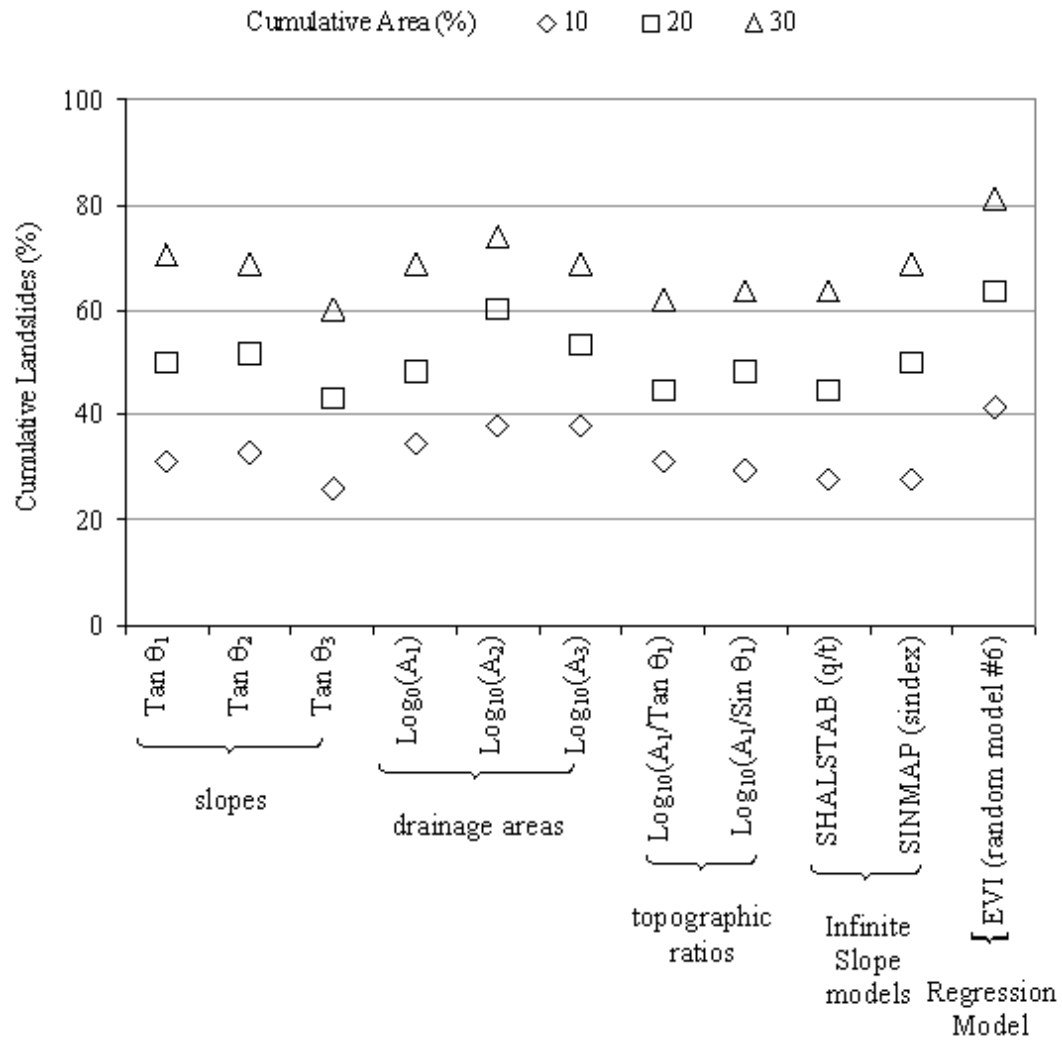


Figure 23. Graphic shows a contrast of the maximum landslide density measured from three different normalized areas. Densities are derived from topographic indices in the Mapleton area. The topographic indices are factored into five distinct types ranked from lowest to highest order (left to right).

The performance of an index is a function of the algorithm and the DEM accuracy and precision. Any bias caused by the algorithm should be manifest in the distributions.

6.3.1 Slope

The relative performance of the slope indices depends on the normalized area of measurement. Figure [23] shows that in general, slope θ_1 and θ_2 perform better than θ_3 . The slope index distributions for the slide population are shown in Figure [24]

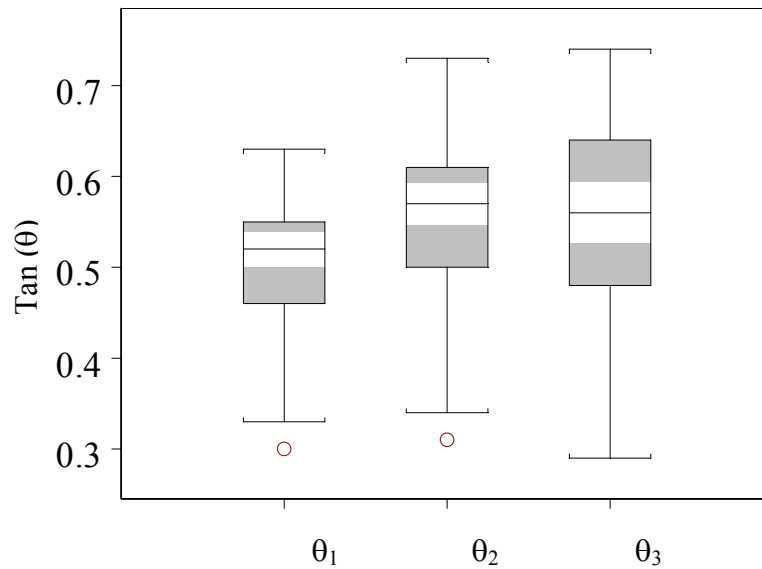


Figure 24. Box plot shows three different slope index distributions for the Mapleton slide population. The white zones represent the 95% confidence intervals of the mean.

Slope index θ_1 is significantly (p -values < 0.05) smaller than slope index θ_2 and slope index θ_3 . Given that they both use the same data and are centered inside a 3x3 window of grid cells, the increase in performance of slope index θ_1 over slope index θ_2 and slope index θ_3 may be due to a better model design and not just a shift in mean values of the distribution. The literature claims that the geometric mean of the slopes (θ_1) around a center cell make a better index of slope and avoids bias caused by orientation of the grid [Montgomery and Dietrich, 1998]. Further, any algorithm that has a bias towards steeper slopes that would counteract the

smoothing caused by the 30m resolution should increase performance. This is not the case shown in Figure [23] and Figure [24]. Slope index θ_3 has the steepest mean slope, but it represents a line segment in-between two grid cell centers, therefore introduces a location bias that may be undermining model performance, i.e., the slopes are not being correctly associated with the slides. It is difficult to say anything conclusive about why one index performs better than any other. This necessarily compounds the problem of interpretation since these slopes are propagated into higher order indices.

6.3.2 Area

The relative performance of the area indices depends on the normalized area used for comparison. Figure [23] shows that in general, A_2 performs better than A_1 and A_3 . The area distributions are shown in Figure [25].

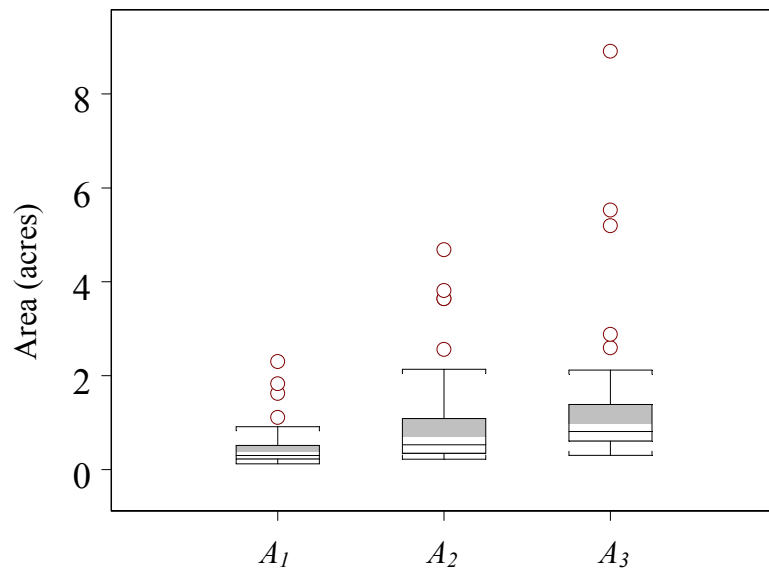


Figure 25. Box plot shows three different area index distributions for the Mapleton slide population. The white zones represent the 95% confidence intervals of the mean.

In spite of the skewed shape in Figure [25], the distributions are all significantly different (p-value < 0.007, Wilcoxon Rank Sum Test).

All three area indices use preconditioning to produce a depressionless DEM and follow the same sink filling procedure. Further, every algorithm for generating area must make assumptions on defining flow direction and the accumulation of flow. Remember that each of these area indices uses a different slope algorithm, therefore, the analysis thus far is not a comparison of the later assumptions with all other factors being equal. The bias in the area indices is proportional to the area. The major differences are that indices A_1 and A_3 make attempts at portioning flow on a sub grid scale. The effect should be larger values in small areas. The area index A_2 accumulates flow in the direction of the steepest drop only. Flow apportioning does not have an obvious effect on the bias above or below index A_2 .

A rigorous examination and testing of the code would be required to determine the exact nature of the bias, i.e., effects caused by relative grid orientation, preconditioning, slope indices used to define flow directions, and finally flow accumulation assumptions. Such testing is beyond the scope of this thesis.

6.3.3 Topographic Ratio

Figure [23] shows that in general the topographic ratio performances are approximately equal;

$$\log_{10}\left(\frac{A_1}{\tan \theta_1}\right) = \log_{10}\left(\frac{A_1}{\sin \theta_1}\right)$$

Given uncertainty in the DEM, differences in the performance cannot be distinguished.

6.3.4 Hydrologic-Geomorphic Models

Figure [23] shows that in general the SINMAP performs better than SHALSTAB;

$$SINMAP(\sin dex) > SHALSTAB\left(\frac{q_{critical}}{T}\right)$$

The *SINMAP* program produces a probability distribution on a factor of safety. The *SHALSTAB* program produces a hydrologic ratio that solves for a critical rainfall to reach a factor of safety of one given a constant transmissivity (T). A direct comparison of the distributions is not warranted. Further, a fair contrast against slope and area cannot be made since both have independent algorithms. The graphical user interface for the *SINMAP* program has an option for manual iteration of the parameters, that *SHALSTAB* does not have, and is likely the source of increased performance.

6.3.5 Expected Value Type

The statistical model was structured as;

$$EVI = \beta_0 + \beta_1 \tan \theta + \beta_2 A_1$$

The model used here sampled data from the entire area, not just from stratified regions surrounding the slide population. Section 6.4 on logistic regression draws comparison and contrast of several types of models. There is no direct comparison between the statistical performance criteria on regressions to densities.

6.3.6 Inter index Performance

Figure [23] shows the inter index performance based on equal-area-densities is greatest for the EVI index. The EVI distribution for the slide population is on the high side of the non-slide distribution resulting in higher densities. Next the area indices lead slope indices in performance due the greater amount of skew and rates of change across the landscape in the area distributions. The performance of the slope and area indices propagates through the higher order indices. Performance for the topographic ratio suffers from a division that allows equal outcomes given different values of the numerator and denominator. The hydrologic-geomorphic models also contain topographic ratios in their algorithms, and thus experience the same equi-finality problem. The hydrologic-geomorphic model algorithms are higher order in the sense that they have additional parameters but do not have an objective optimization function.

6.4 Logistic Regression on Topographic Indices

Logistic regression was used to determine the odds of landslide occurrence. These odds are used to define the similarity between slide and non-slide sites. Logistic regression was performed using a full-two-way model that included the slope and area terms. This regression will be referred to as the logit $\{I\}$ model. The

regression was performed on spatially stratified regions defined by a 3x3 window surrounding the population of slide sites. The slope and area terms are not significant. The 3x3 window is smaller than the minimum resolution that differences in the variable distributions can be detected. A regression was then carried out on spatially stratified regions defined by the 3cell-minus data set. The results are shown in Table [5].

Table 5. Logit {1}-model output for a spatially stratified data set

| Full Model | | | | | Odds = exp (Value) | | |
|--|--------|------------|--------|---------|--------------------|----------|------------|
| Term | Value | Std. Error | Z-stat | P-value | L. 95% C.I. | Mean | U 95% C.I. |
| (Intercept) | -24.1 | 9.65E+00 | -2.49 | 0.006 | 2.09E-19 | 3.42E-11 | 5.59E-03 |
| Slope | 0.744 | 0.417 | 1.78 | 0.075 | 0.929 | 2.10 | 4.77 |
| Area | 4.84 | 2.34 | 2.06 | 0.044 | 1.29 | 126 | 12400 |
| Slope: Area | -0.169 | 0.101 | -1.67 | 0.095 | 0.693 | 0.845 | 1.03 |
| Null Deviance: 452.72 on 1140 degrees of freedom | | | | | | | |
| Residual Deviance: 438.65 on 1137 degrees of freedom | | | | | | | |

A sensitivity analysis of the odds is performed using the coefficients of the model. Back transformation of the coefficients yields the multiplicative effect of a unit change in the variable on the odds of landslide occurrence, see Table [5]. The relative multiplicative effects on the odds are greatest for the area term and least for the slope:area interaction term.

All of the variables in the logit{1} model are significant and the full model explains the most variability. The linked hydrologic geomorphic model presented in section 4.6 supports the presence of the slope:area interaction term. The negative value of the slope:area coefficient plays an important role in the behavior of logit{1} model. A closer look at the full model equation is required to fully understand what it says

about the relative contributions of slope and area to the odds and variable thresholds that exist.

One way to examine the role of the slope:area interaction term in model behavior is with partial functions;

$$\text{Partial logit } (\pi) = 4.84 (\text{Area}) - 0.169 (\text{Slope:Area})$$

Solving for slope in the equality when the partial logit (π) is equal to zero yields

$$\text{Slope} = \frac{\log it(\pi) - 4.84 \text{Area}}{-0.169 \text{Area}} \quad \text{solve for the root} \quad \text{Slope} = 28.6^\circ$$

A graphical solution for the full logit {1} model is shown in Figure [26]

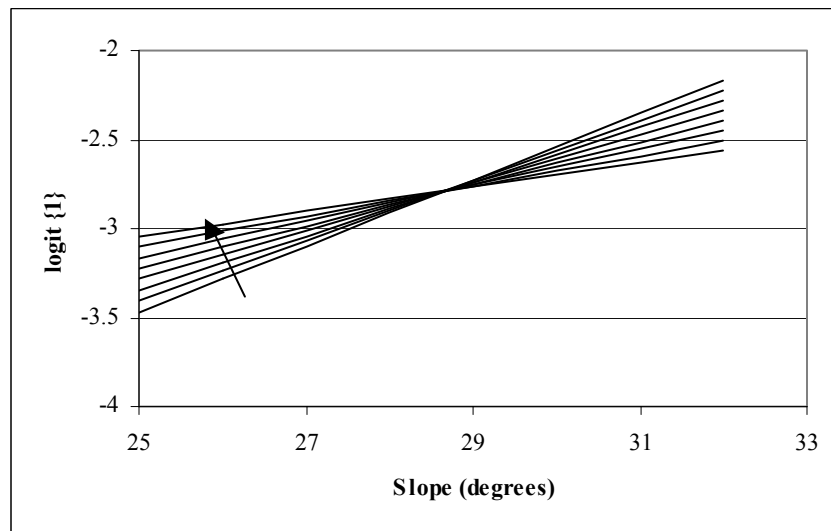


Figure 26. Plot shows slope index versus partial logit {1} with area terms only. Contours show lines of equal area ranging from 3.5 to 5.0 that represent the slide population in Mapleton. Arrow indicates direction of increasing area.

A root is located at 28.6 degrees where the partial logit(π) is equal to zero. The root is evidence that the logit {1} model behaves similar to the linked hydrologic geomorphic model in section 4.6. The arrow in Figure [26] indicates that on the lower side of the root (slopes less than 28.6 degrees), area is proportional to slope and on the right side of the root (slopes greater than 28.6 degrees), area is inversely proportional to slope. The size of the interaction term determines the size of the root. Remember the area threshold function, $A = C \sin \theta (1 - \tan \theta)$ described in section 6.1 has a maxima at 24.4 degrees. The presence of a root marking a change in the logit {1} model behavior that is proportional to the physical model is important.

The concept of dominant and submissive factors is introduced to help differentiate model behavior on either side of the root. A variable can dominate the logit. Dominance indicates that the sum contribution of a variable and its interaction term to the logit is relatively high when contrasted to another variable and its interaction term. For example, on the lower side of the root, the partial logit (π) = 4.84 (Area) – 0.169 (Slope:Area) is greater than partial logit(π) = 0.74 (Area) – 0.169 (Slope:Area). It follows that area is dominant on the lower side of the root and slope is dominant on the upper side of the root.

Conversion of the logit(π) function to a probability (π) does not change the interpretation, however, probabilities are easier to reason with. All values of probability must be within zero or one. Therefore the graphical display of Figure [26] is transformed into probabilities shown in Figure [27].

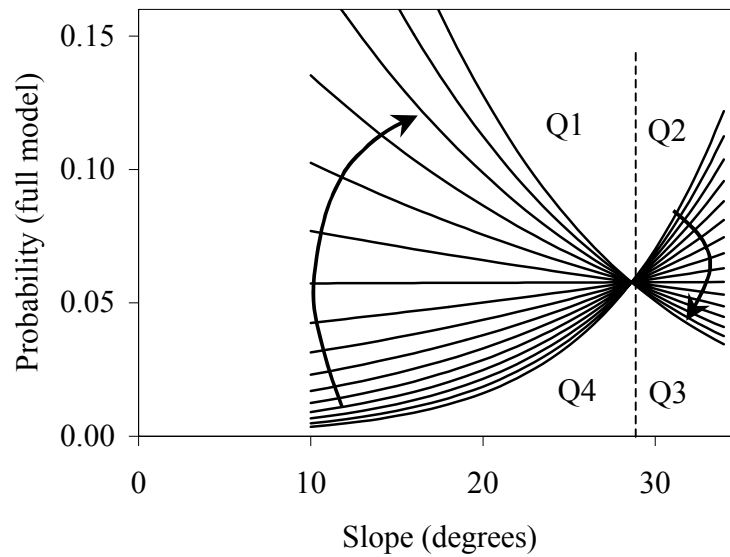


Figure 27. Plot shows a family of slope index θ_1 versus probability curves derived from the full logistic model in Mapleton. Contours show lines of equal area. Arrows show direction of increase in the area. Quadrants are marked Q1, Q2, Q3, and Q4.

The trends from the plot in Figure 27 suggest that given equal slopes on the right side of the root (quadrants Q2 and Q3) that area is inversely proportional to probability and it is the submissive factor. Conversely, on the left side of the root (quadrants Q1 and Q4) the area is proportional to probability and it is the dominant factor. Slope and area interaction are described by the gradient on the curves in Figure [27]. The trends from Figure [27] also suggest that given equal areas in quadrants Q2 and Q4, the slope is proportional to probability and interaction with area is inversely proportional to the magnitude of area. Conversely for quadrants Q1 and Q3 the slope is inversely proportional to probability and interaction with area is proportional to the magnitude of area.

O'Loughlin's observations suggest that for quadrants Q1 and Q4 a high saturation of the soil profile is required for landslide occurrence. Alternatively, for quadrants Q2 and Q3 a low saturation of the soil profile is required. A plot of the expected probabilities is shown in Figure [28].

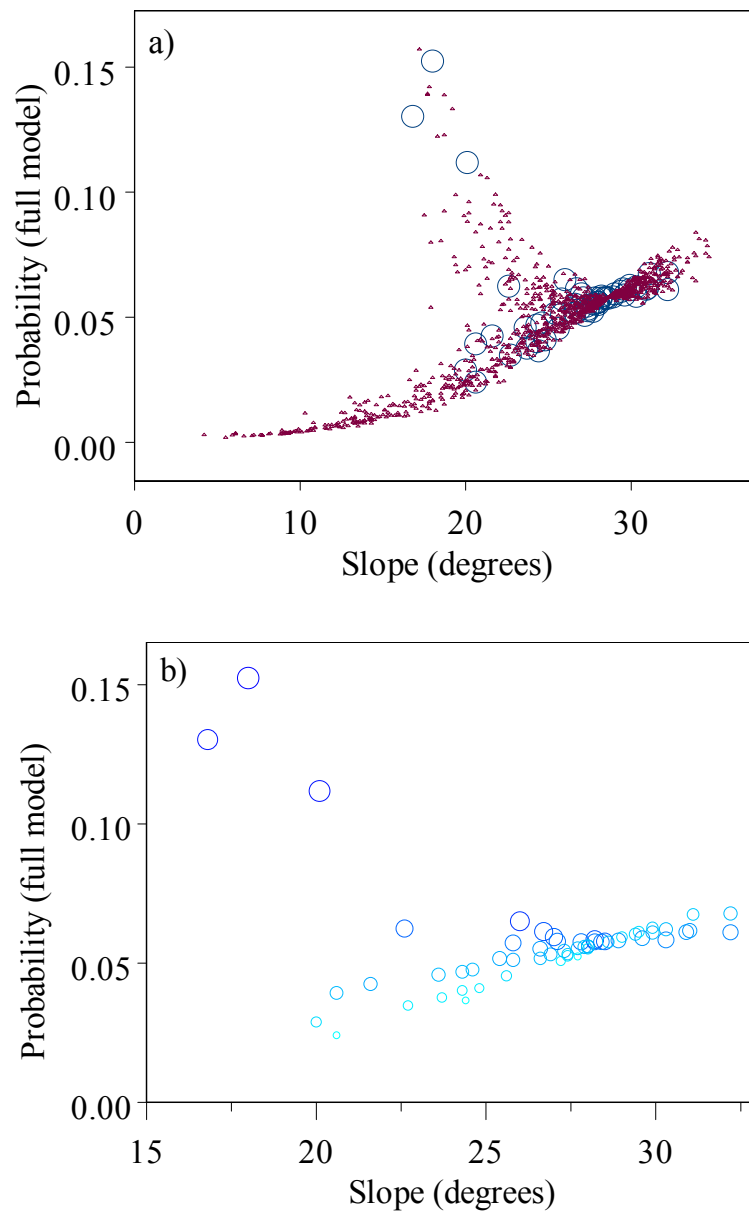


Figure 28. Plots of slope versus probability derived from expected values of the $\text{logit}\{1\}$ model from spatially stratified regions in the Mapleton area. Figure a) has slides shown as circles and non-slides shown as small crosses. Figure b) has slides only, circles are scaled in size and color proportional to the value of area. Lighter and smaller symbols are smaller values.

Figure 28 indicates there is no terrain present in quadrant Q3.

A three dimensional grid surface of expected probability values is shown in Figure [29].

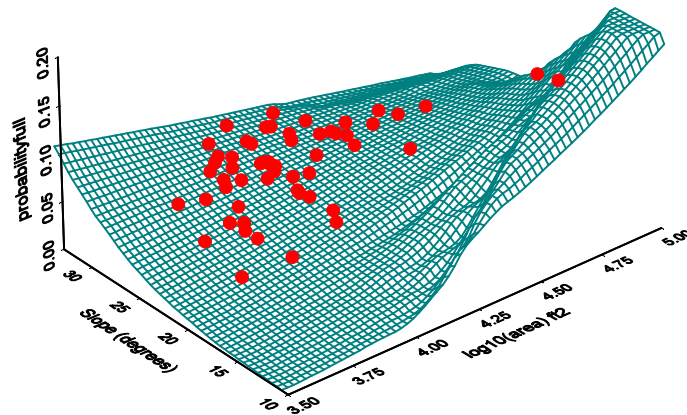


Figure 29. Regression surface for landslide probability derived from the expected values of the $\text{logit}\{1\}$ model in spatially stratified regions of the Mapleton area surrounding the slide population. Slides are shown as solid circles.

The regression surface in Figure [29] shows that probabilities gradients for slope (measured in degrees) are less than area (measured in \log_{10} transformed ha). This follows from the relative size of the coefficients in the full $\text{logit}\{1\}$ model and the sensitivity analysis on multiplicative effects. Thus landslide occurrence is more sensitive to the area than the slope.

These are subtle distinctions because it is difficult to say which variable is more important given their similar p-values. Area terms should have lower p-values relative to slope due to a smaller relative uncertainty in the DEM. This would put slope at a disadvantage in the regression. However, the multiplicative effect for slope seems to have a much smaller confidence interval than area.

Figure [30] shows a contrast of the logit $\{1\}$ model probabilities classified by quadrants in a thematic map.

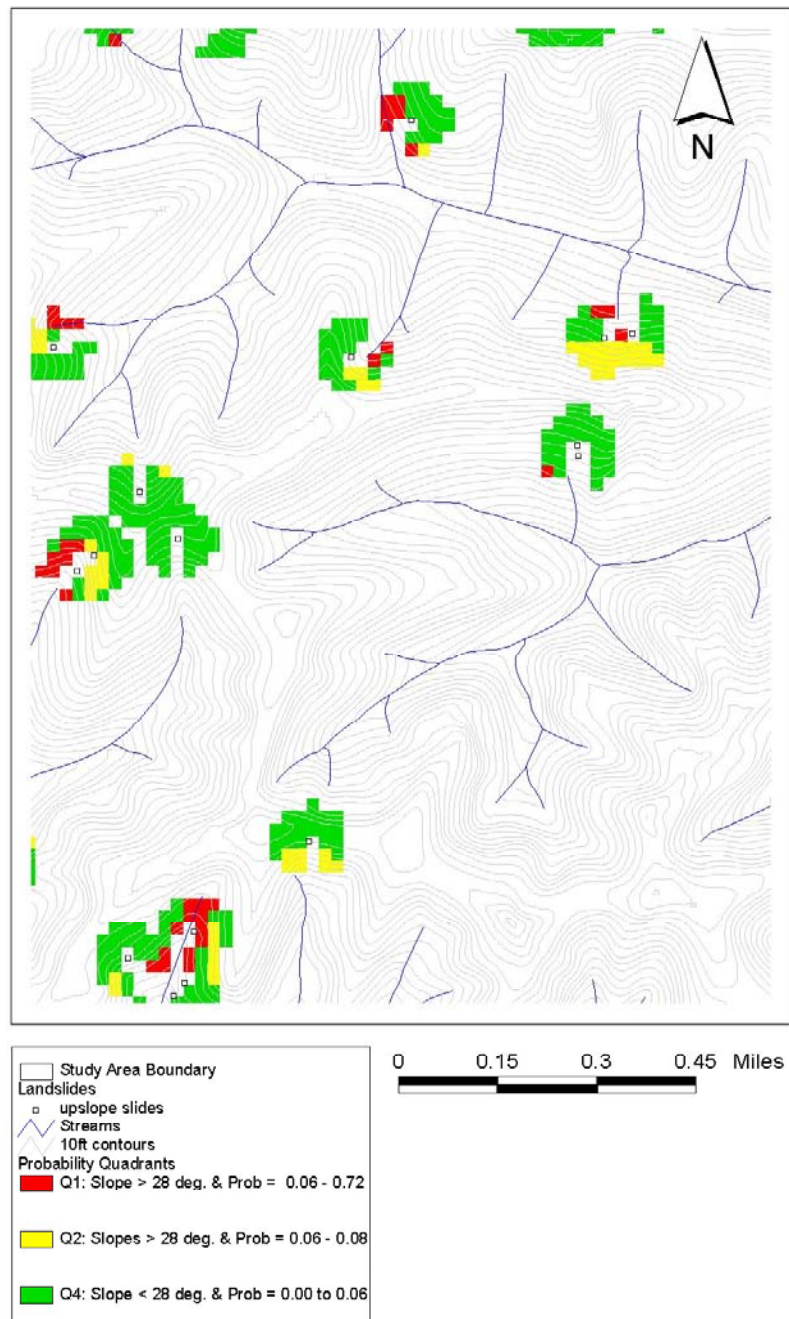


Figure 30. Thematic map shows slide probability classified into quadrants based on the roots of the logit {1} model in Mapleton. A three-cell radius surrounding Mapleton slide inventory with down stream cells removed defines spatial stratification.

Next, the curvature variable is added to the analysis. The logit {2} model is the result of a backward selection of significant terms using three-way saturation by slope, area and curvature. The results of the logit {2} model are shown in Table [6].

Table 6. Logit {2} model output for spatially stratified data set

| Term | Coefficient | Std. Error | Z-stat | P-value |
|----------------|-------------|------------|--------|---------|
| Intercept | -24.9 | 7.03 | -3.54 | 0.0002 |
| Slope | 0.828 | 0.317 | 2.61 | 0.0045 |
| Area | 5.50 | 1.72 | 3.19 | 0.0007 |
| Curvature | -0.727 | 0.444 | -1.64 | 0.0507 |
| Area:Curvature | -0.147 | 0.104 | 1.42 | 0.0774 |
| Slope:Area | -0.213 | 0.078 | -2.74 | 0.0031 |

The logit {2} model explains more variability than the logit {1} model. Addition of a curvature term produces a better model. However, little change took place in the magnitudes of the coefficients for the intercept, slope, area and slope:area. The curvature and area:curvature interaction are significant to the model but of lesser importance than the other terms. The area:curvature term is positive and the slope:area term is negative. The increase in convergence in the formation of hollows results in negative values of curvature and divergent terrain has positive values of curvature. The area:curvature term reduces the magnitude of the logit(π), which is similar to the behavior of the slope:area term. Both interaction terms work together to reduce the logit in convergent terrain and in opposition to one another in divergent terrain. Basically, the curvature terms says convergent terrain looks like a slide, and the area:curvature slightly reduces the strength of that argument.

The distributions of logit{2} with each main effect variable are shown in Figure [31].

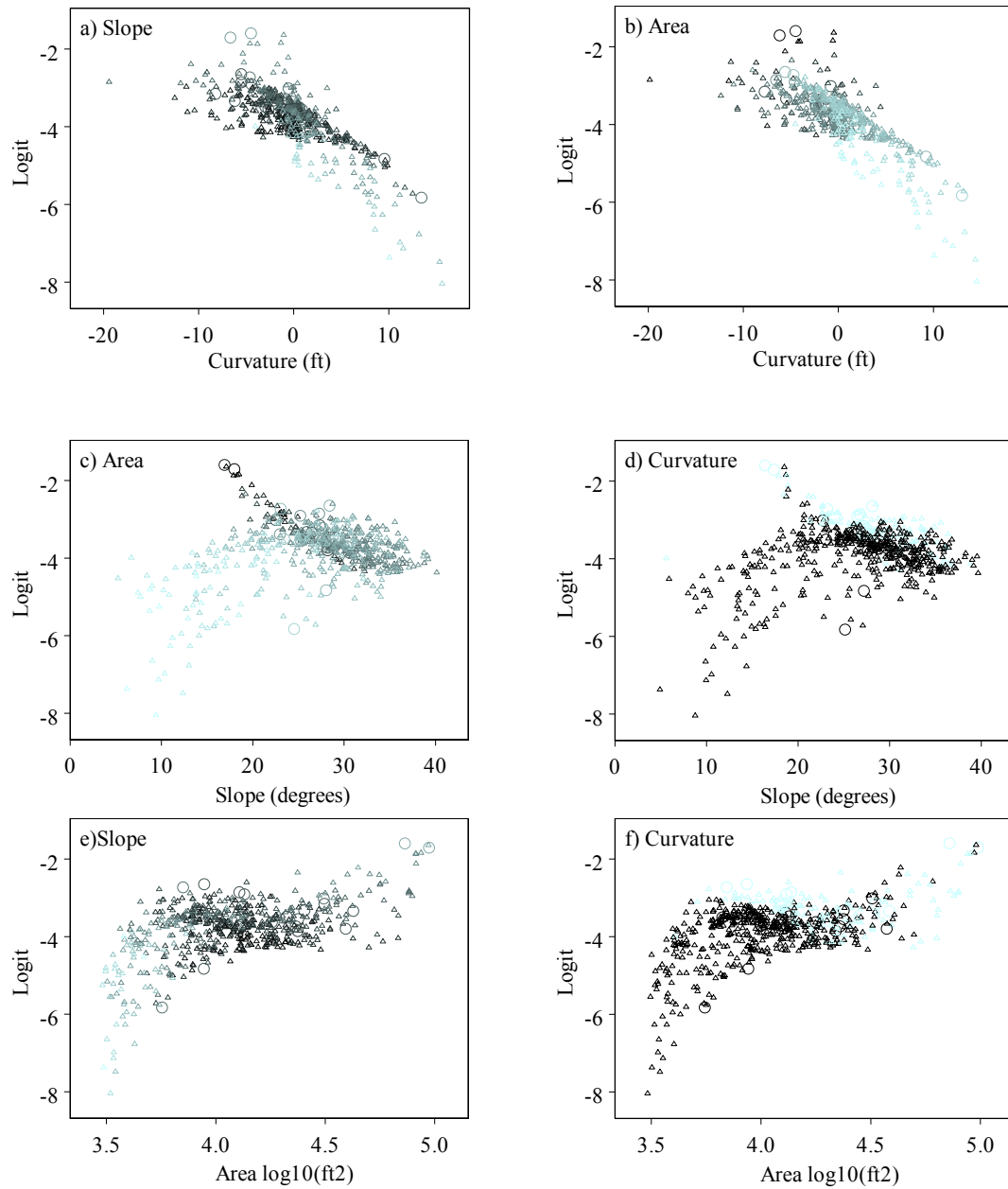


Figure 31. Matrix of scatter plots show the explanatory variables versus the Logit{2} model derived from spatially stratified regions in the Mapleton area. Slides are shown as circles and non-slides as triangles. Symbols are colored on a gray scale from low (light) to high (dark) for a) slope, b) area c) area, d) curvature, e) slope, f) curvature. Indices are A_1 , θ_1 and C_1 .

Figure [31a] and [31b] show the relationship between $\logit\{2\}$ and curvature is linear with a steep gradient and constant variance. The $\logit\{2\}$ versus slope plot in Figure [31c] and [31d] is non-linear. The $\logit\{2\}$ versus area plot in Figure [31e] and [31f] has non-linear tails separated by a potential threshold that is characterized by a segment with zero gradient.

Figure [31] suggests there is some distinctions on the gray scale, which indicate the ability of a variable to define a threshold. The curvature distinction is most visible on the slope and area plots. A contrast of a full and partial $\logit\{2\}$ model with and without curvature and area:curvature terms is shown in Figure [32]

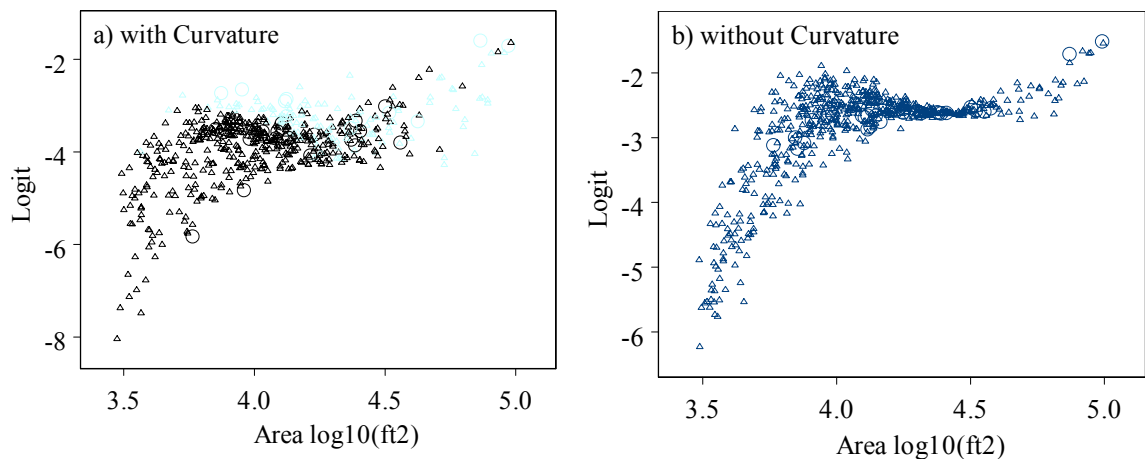


Figure 32. Scatter plot shows area versus full and partial $\logit\{2\}$ derived from spatially stratified regions surrounding the Mapleton slide inventory, a) full logit with symbols colored on a gray scale for curvature from low (light) to high (dark) and, b) partial logit without curvature terms and no gray scale.

A visual inspection of Figure [32] shows how the curvature terms account for a lot of variability in the upper range of the logit.

Similarity based on the $\text{logit}\{2\}$ between the slide and non-slide sites is shown in Figure [33].

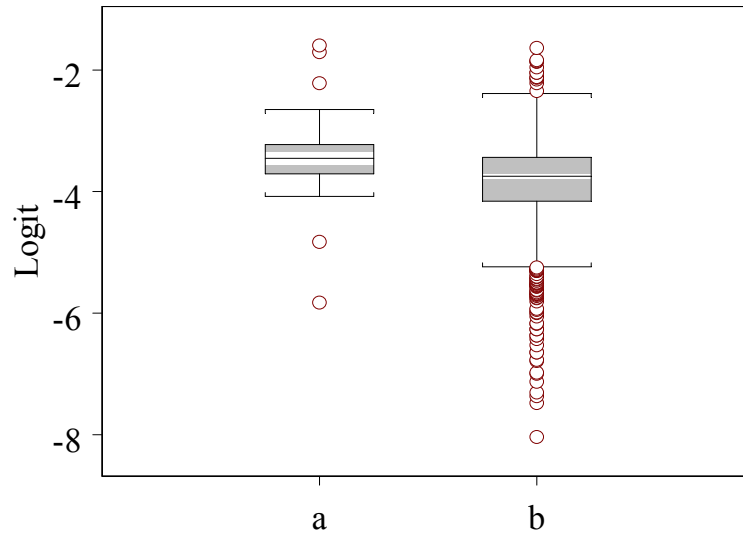


Figure 33. Box plot shows distributions of the $\text{Logit}\{2\}$ model factored by response in the Mapleton area, a) slides and b) non-slide sites.

A contrast of the logit in Figure [33] is used to make observations regarding slide and non-slide sites. Non-slide sites have a lower mean value of the $\text{logit}\{2\}$ than slides sites, thus indicating higher odds (lower probabilities) of slide occurrence. On average, slide sites are 1.5 times (95% C.I., 1.3 to 1.9) as likely to have a response of one as non-slide sites. Non-slide sites have a logit distribution with long tails and a lot of overlap with slides sites. Therefore the slope, area and curvature variables play a small but significant role in the response.

The interaction terms make it difficult to interpret the effects of a unit change in the explanatory variables. A sensitivity analysis was performed using calculated changes in the logit centered on mean values of the explanatory variables from the slide population. The multiplicative effect of a unit changes are slope 1.06, area 1.49 and, curvature 1.54.

Logistic regression indicates those variables of slope, area, and curvature, and the interactions between Slope:Area, and Area:Curvature are all significant terms in the odds function. In general, the slide sites are steeper, more convergent and have a larger area.

The scope of these statements is limited to topographic variables derived from a USGS 30m DEM using the particular algorithms for slope (θ_1), Area (A_1), and plan form curvature (C_1). The spatial scope is limited to spatially stratified areas within a three-cell radius adjoined to the slide sites in the Mapleton area with cells down stream of the slides removed. If the spatial scope defines the morphologic system, then the temporal scope is defined by the process response system (see section 4.4 on physical geography). Strictly speaking the odds are retrospective. However, the retrospective odds may be treated like prospective odds if the temporal scope encompasses the process response e.g., a geologic time scale.

In terms of assigned hazard ratings where life and property are at risk, the difference in odds are likely not practical for a land manager. In terms of research questions on where landslides occur, the difference in odds may be practical.

The odds may also be phrased in terms of site similarity. The similarity assumption means that similar odds are equivalent to similar topography. Sites that possess the same odds as the slide population but did not fail, likely have some other first order-confounding variable that was not measured. Non-slide sites cannot be interrogated for confounding variables without more information.

Logistic models may be extrapolated outside the spatial scope given restrictions on the range of the topographic variables inside the source area. This assumes that source areas are representative of the other terrain under similar conditions. The

logistic regression models presented here have not been tested outside the source areas.

6.5 Discrimination by Definition

Discrimination between slide and non-slide sites is facilitated by an examination of the definitions on topographic similarity with the slide population. The analysis presented here is a variation on the assessments of index performance in Section 6.3. The primary metric of interest is areal density calculated from definitions of site similarity extrapolated across the Mapleton study area. An example includes using a map with arbitrary boundaries on a slope index to define non-slide sites that are similar to the existing slide population. Arbitrary boundaries are set on each index using the 95% confidence interval about the mean calculated from the slide population. Application of these boundaries is index stratification.

A combination of indices used in a definition of similarity may involve logical operators such as an;

- IF statement used to define arbitrary boundaries set on an index. When statement is true, the non-slide site is similar to slide site
- AND statement used for variable intersection (i.e. common domain and cumulative area). The intersected area defines non-slide sites that are similar to slide sites.
- OR statement used for variable union (i.e., overlapping domain and cumulative area). The union of areas defines non-slide sites that are similar to slide sites.

Performance for a group of low order indices using domain stratification for the entire Mapleton study area is shown in Table [7].

Table 7. Slide and areal densities using domain stratification

| | <u>Slide Limits</u> ¹ | | Areal Density ² | Slide Population ³ | Density Ratio ⁴ |
|---|----------------------------------|-------|-------------------------------|-------------------------------|----------------------------|
| | LCL | UCL | (%) | (%) | (%) |
| <u>Single Variable</u> ⁵ | | | | | |
| Slope (degrees) | 25.7 | 27.6 | 11.7 | 22.4 | 2.1 |
| Area log ₁₀ (ft ²) | 4.10 | 4.25 | 13.7 | 32.8 | 2.4 |
| Curvature (ft) | -2.57 | -0.46 | 13.6 | 46.6 | 3.4 |
| <u>Variable Intersection</u> ⁶ | | | | | |
| Slope & Area | | | 2.4 | 3.7 | 1.5 |
| Slope & Curvature | | | 1.8 | 5.5 | 3.0 |
| Area & Curvature | | | 1.8 | 1.8 | 1.0 |
| Slope & Area & Curvature | | | 0.4 | 0.0 | 0.0 |
| <u>Variable Union</u> ⁷ | | | | | |
| Slope or Area | | | 23.1 | 43.1 | 1.9 |
| Slope or Curvature | | | 23.5 | 62.1 | 2.6 |
| Area or Curvature | | | 25.3 | 63.8 | 2.5 |
| Slope or Area or Curvature | | | 33.0 | 72.4 | 2.2 |
| <u>Optimization of Variables</u> ⁸ | | | | | |
| Logit {1} | -2.75 | -2.59 | 13.6 | 44.8 | 3.3 |
| Logit {2} | -3.59 | -3.26 | 18.2 | 39.7 | 2.2 |

Notes:

- 1) 95% confidence limits for each variable based on Mapleton upslope slide inventory
- 2) Percentage of Mapleton red zone that falls within the domain of the variable
- 3) Percentage of the upslope slide population that falls within the domain of the variable
- 4) Density ratio = percent slides / percent total area
- 5) Algorithms are slope = θ_1 , area = A_1 , and curvature = C_1 .
- 6) Intersections include common areas only. Same domain limits as the single variables.
- 7) Unions include intersected areas and non-intersected areas. Same domain limits as the single variables.
- 8) Regression models based on stratified areas surrounding Mapleton slides. Expected values extrapolated to entire Mapleton red zone, therefore models are valid within the domain limits of the explanatory variables only.
 $\text{logit } \{1\} = \beta_0 + \beta_1 \text{Slope} + \beta_2 \text{Area} + \beta_3 \text{Slope:Area}$
 $\text{logit } \{2\} = \beta_0 + \beta_1 \text{Slope} + \beta_2 \text{Area} + \beta_3 \text{Curvature} + \beta_4 \text{Slope:Area} + \beta_5 \text{Area:Curvature}$

The data in Table [7] have fixed domains and variable areas, where as the data in Figure [23] have fixed areas and variable domains. Therefore a direct comparison is possible. However, it highlights the need for an objective such as fixed area or proportion of the slide population. For example, a high-density ratio is worthless if the variable regime captures an unacceptable number of slides. Ideally, the domain limits could be iterated until an objective was achieved.

6.5.1 Single Variable Performance

Table [7] shows the relative performance of the single variable definitions based on slide densities is;

$$\text{Curvature} > \text{Area} > \text{Slope}$$

The domain limits were arbitrary, but all three had similar representative areas, therefore, this is a fair comparison. There are thresholds in all three indices facilitating failure and the order of arrangement highlights their importance in physical process.

6.5.2 Variable Intersections and Unions

Variable intersections based on the slide populations should define the most similar sites and produce the highest performance. However, the arbitrary limits on the definitions yielded such small representative areas that conclusions on the performance cannot be drawn.

Table [7] shows the performance of the variable unions based on slide densities is;

Slope or Curvature >Area or Curvature>Slope or Area or Curvature>Slope or Area

These definitions occupy the largest area and are therefore the most thoroughly tested of all the definitions. Variable unions have lower density ratios than the single variable definitions. Variable unions represent the most conservative definitions of site similarity.

6.5.3 Optimization of Variables

Iteration of the domain limits on variables for an objective is known as combinatorial optimization (pers. Comm. John Session, 2002). Combinatorial optimization would produce a set of logical rules (IF, AND, OR) with domain limits. Future research efforts using this type of analysis is within the scope of the thesis objectives.

A specific type of optimization is the maximum likelihood algorithms employed in logistic regression. The expected value is the log odds (logit) of landslide occurrence. The product is a distribution of the logit generated from a combination of variables with linear coefficients. This is distinctly different from a union or intersection of variable domains. The coefficients are generated such that the slide population has the largest and most extreme values of the logit. Given domain limits on the logit, non-slide sites that are similar to slide sites can be selected. However due to extrapolation, blatant use of logit values on the extreme ends of the slide population will pick up terrain that has explanatory values outside the domain of the model.

For example, the $\text{logit}\{2\}$ model was defined as;

$$\text{logit}\{2\} = \beta_0 + \beta_1 \text{Slope} + \beta_2 \text{Area} + \beta_3 \text{Curvature} + \beta_4 \text{Slope:Area} + \beta_5 \text{Area:Curvature}$$

A sample definition of site similarity is given by the 95% confidence interval about the mean of the $\text{logit}\{2\}$ from the slide population. Table [7] shows the performance stats for two logistic models. Next, the explanatory variable distributions from cells that fit the definition are presented in Figure [34].

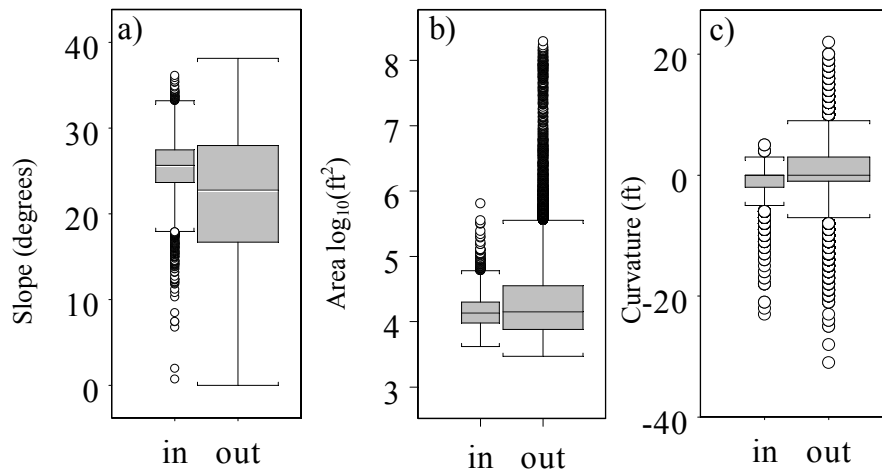


Figure 34. Box plots show the explanatory variable distributions based on systematic range limits of the $\text{logit}\{2\}$ model, a) slope, b) area, and c) curvature. Box plots on left are coincident (in) with a 95% C.I. on the $\text{logit}\{2\}$ model defined by the Mapleton upslope slide population. Box plots on the right are not coincident (out). There are 4,343 cells classified as “in” and 19,511 cells classified as “out”. Indices are A_1 , θ_1 , and C_1 .

Figure [34] shows that extrapolation of the logit outside the stratified areas (defined by the slide population) has violated one of the basic requirements, i.e. explanatory values must be inside the domain limits defined by the slide population. Next, the $\text{logit}\{2\}$ data are intersected with the domain limits defined by the explanatory values in the 3cell-minus data set. The result is 188 cells are removed from the

“in” tails in Figure [34]. This translates into 17.4% of the total area defined as being similar to slides.

The challenge here is to determine what makes a good definition on site similarity. Recall from Section 5.0 the reduced model $\text{logit}\{1\}$ was defined as;

$$\text{logit}\{1\} = \beta_0 + \beta_1\text{Slope} + \beta_2\text{Area} + \beta_3\text{Slope:Area}$$

The $\text{logit}\{1\}$ model demonstrates better performance (defined by the density ratio) in Table[7] than does $\text{logit}\{2\}$. The $\text{logit}\{1\}$ model also has similar density ratios as the single variables. Further, it is remarkable that the $\text{logit}\{1\}$ model has the same density ratio as the single curvature variable.

Drop in deviance testing suggests the $\text{logit}\{2\}$ is a better model than $\text{logit}\{1\}$. The $\text{logit}\{2\}$ was derived from stratified areas characterized as being mostly convergent. Thus terrain outside the stratified regions may be characterized as more planar and even divergent. The curvature distributions for slide headwalls and non-slide headwalls have not been examined because there is no definition of a headwall. The “in” and “out” terrain in Figure[34] contains a mix of headwalls with and without slides. Extrapolation of the $\text{logit}\{2\}$ to terrain with different curvature distributions may weaken the model.

Inspection of the $\text{logit}\{1\}$ and $\text{logit}\{2\}$ formulae highlights an essential behavioral difference between the two. $\text{Logit}\{2\}$ should pick up more convergent terrain and less divergent than the $\text{logit}\{1\}$ model. Basically, the $\text{logit}\{1\}$ model is looking for convergent terrain and cannot find it all, but the $\text{logit}\{2\}$ model can. The only evidence available is from drop in deviance testing, therefore, the later interpretation is accepted.

In spite of the low performance in the single variables the slope index is still the “gold standard” for comparison. Therefore, the explanatory variable distributions based on a slope index definition are shown in Figure[35].

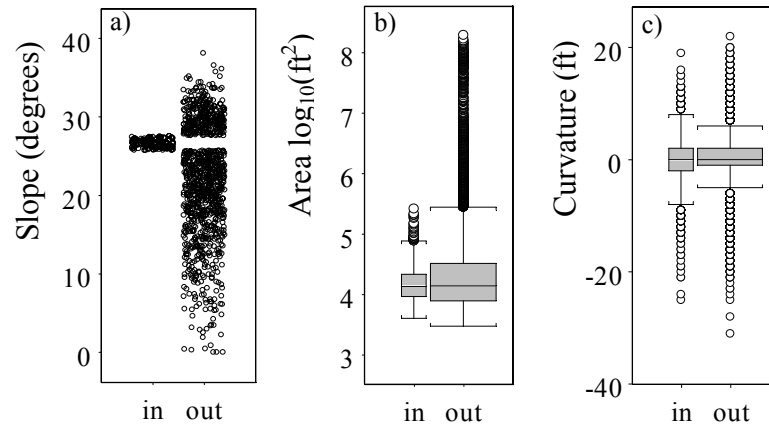


Figure 35. Box plots show explanatory variable distributions systematically defined by a simple slope model, a) slope, b) area, and c) curvature. Plots on left are coincident (in) with a 95% C.I. on the slope index θ_1 model defined by the Mapleton slide population. Box plots on the right are not coincident (out). There are 2,941 cells classified as “in” and 20,913 cells classified as “out”. Indices are A_1 , θ_1 , and C_1 .

A contrast of the $\text{logit}\{2\}$ model in Figure [34] with the simple slope model in Figure [35] indicates there is a significant shift in the curvature distribution. The slope index definition of similarity has a bias toward higher values of curvature (i.e., planar terrain). Conversely the $\text{logit}\{2\}$ model has a bias toward convergent terrain (negative curvature). A visual comparison is made in Figure [36].

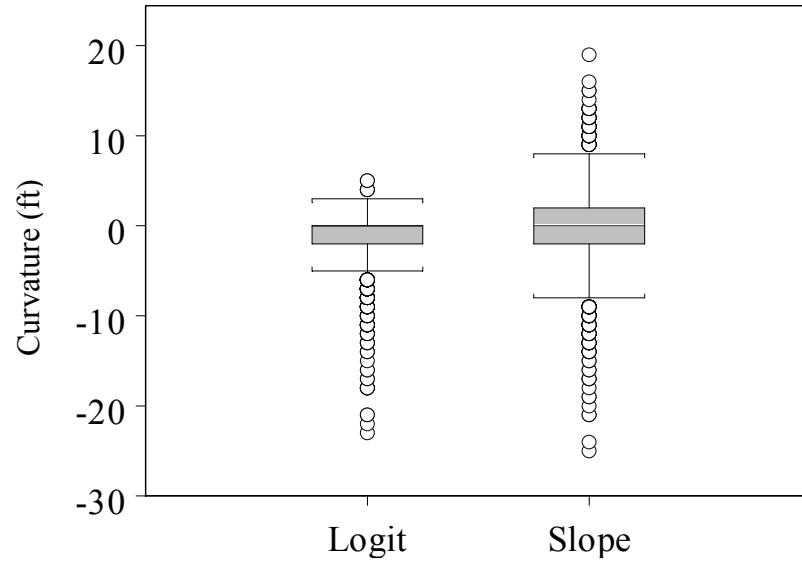


Figure 36. Box plot shows a contrast of Curvature index C_1 distributions for the $\text{logit}\{2\}$ and slope θ_1 models in the Mapleton area.

The mean difference in the curvature distributions of Figure [36] is 0.9ft (95% C.I. equal to 0.7 to 1.0 ft). The slope model has 36% of the cells classified as convergent (curvature < 0), and the $\text{logit}\{2\}$ model has 44% classified as convergent. This difference in classification accounts for 830 cells or 4% of the total area inside the Mapleton study boundaries.

Analysis of densities and index domains does not tell the whole story. The location of sites defined as similar to slide sites is the next item of interest.

The difference in the explanatory variable distributions hints at a shift toward a uniquely different set of cells. An aerial intersection on the $\text{logit}\{2\}$ and slope models is carried out to examine this shift. The proportions from an aerial intersection are shown below;

| | Number of cells |
|-------------------|-----------------|
| Slope θ_1 | 2941 |
| Logit{2} | 4155 |
| Difference | 1214 |
| Area intersection | 1060 |

The logit{2}-model definition of site similarity is mapped out in Figure [37]. A picture of the slope and logit{2} models with aerial intersection is mapped out in Figure [38].

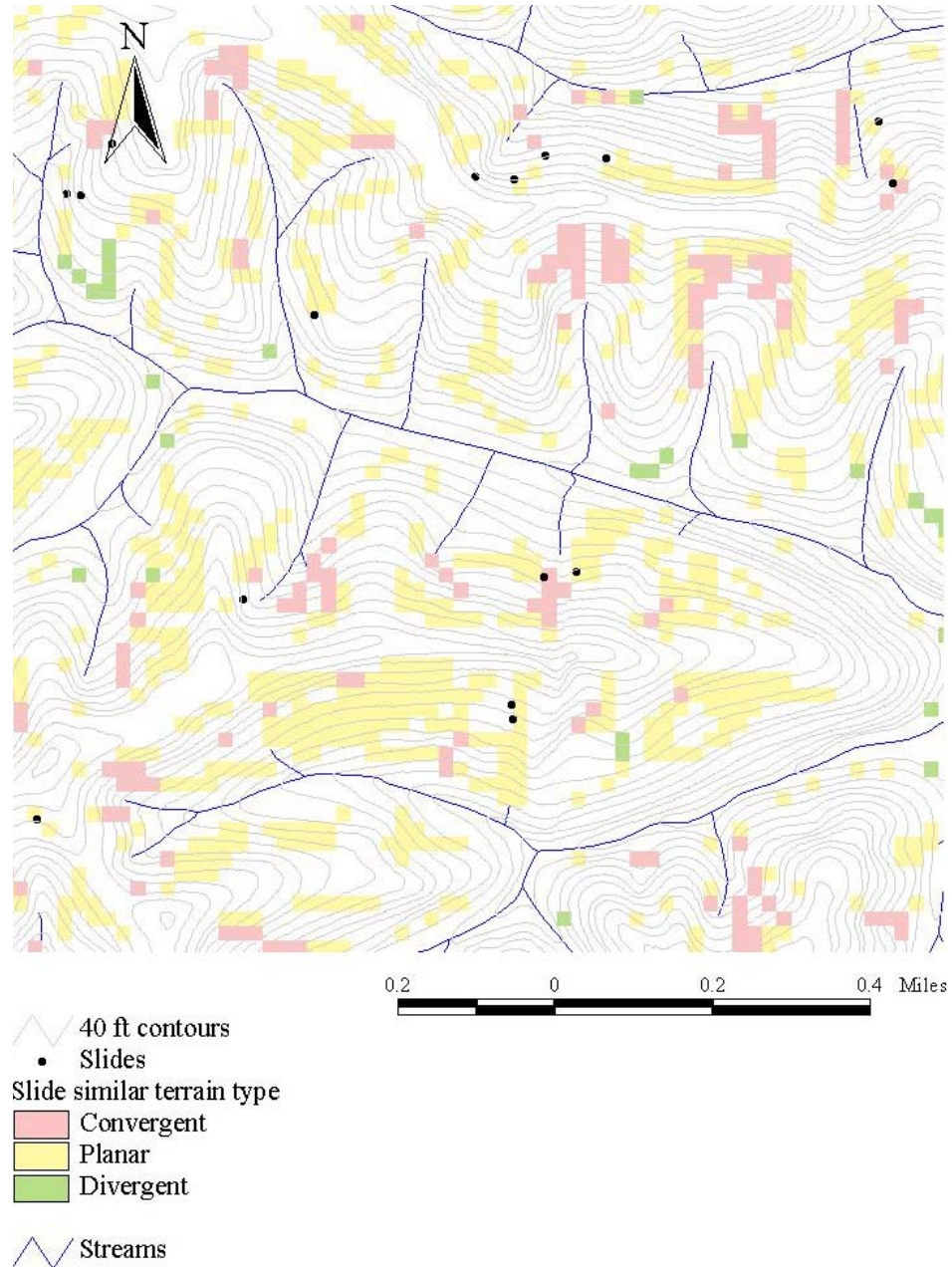


Figure 37. Thematic map for a systematic definition of the $\text{logit}\{2\}$ model. Theme indicates expected $\text{logit}\{2\}$ values within the 95% C.I. of the Mapleton slide population. The similar sites are sub-classified based on curvature. Planar terrain is defined as ± 1 ft. The extent of convergence is defined by the data minimum and divergence by the data maximum.

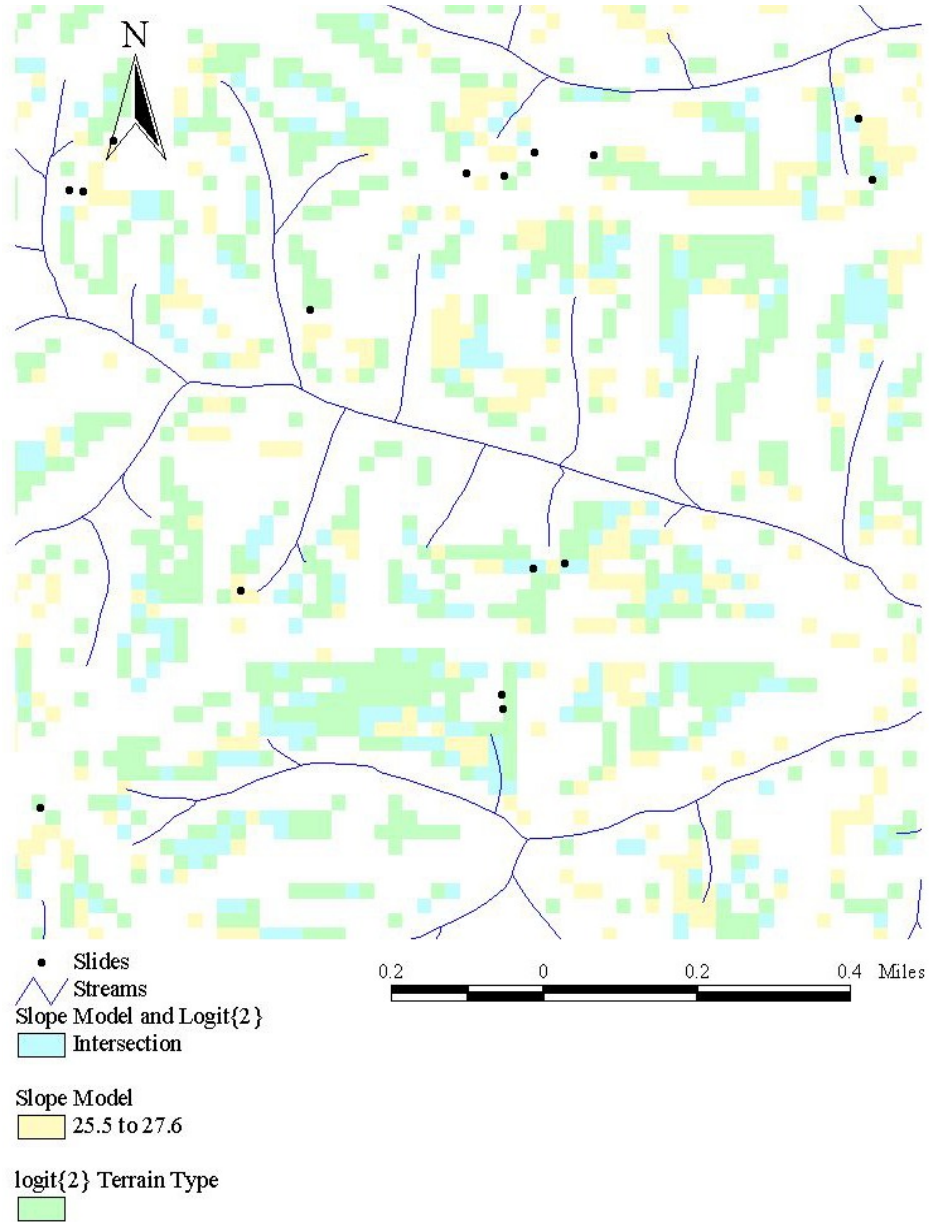


Figure 38. Thematic map for an intersection of the logit{2} and simple slope models. Logit{2} and slope models show arbitrary definitions of site similarity based on the 95% C.I. of the slide population in the Mapleton area. The similar sites are sub-classified based on model type and their aerial intersection.

A fair comparison is difficult to make between the logit{2} and slope models because the representative areas have not been normalized as they were for different model comparisons made in Figure[23]. The portion of the landslide population captured may best normalize aerial density performance. The discrepancy in coincidentally classified cells created by the aerial intersection of the logit{2} model and the slope θ_1 model of site similarity is disconcerting and is worthy of future research.

The data in Table [7] are derived from confidence intervals about the mean index value measured from the slide population. Land managers want prediction intervals that are necessarily wider than a confidence interval on the mean. Further, the expected values from logistic regressions are mean values and have standard errors attached to them. Thus there are actually two distributions of any given logit model that define the desired prediction interval, i.e. a lower and upper 95% bound of the mean logit. Extrapolation of the logit model is worthy of further research.

Given the following assumption;

- Topographic similarity is achieved by a 95% confidence prediction interval about the mean of the index for the slide population, i.e., some function of slope, area, and curvature.

The following inferences from the evidence in Table [7]

- The portion of the landscape that looks like a landslide ranges from 13.6 to 17.4 % depending on which regression model is examined, and
- The portion of the landscape that looks like a landslide ranges from 11.7 to 13.6 % depending on which single variable model is examined, and

- The portion of the landscape represented by the more conservative regression models exceeds the single variables models by 4%.

The later inference is important because a simplified view of slope stability hazard ratings (e.g. slope theme only) could systematically ignore potentially unstable terrain.

6.5.4 Performance in Spatially Stratified Areas

The proximity of non-slide sites to slide sites is yet another way to define similarity. The regression analyses in Section [6.0] defined stratified areas surrounding the Mapleton upslope slide population. Index performance for a select group of low order variables using domain and area stratification is summarized in Table [8] and displayed graphically in Figure [39].

Table 8. Slide and areal densities with spatial and domain stratification for the Mapleton slide population

| Fraction | Sample Size In Domain (n) ¹ | Population Proportions | | | Density Ratio ⁵ <u>Slide-Slide</u> Proximity -Total | Density ⁶ (slides/acre) | Site Stratification | |
|-----------|--|--------------------------------|--------------------------------------|------------------------------------|--|---------------------------------------|---|----------------------------------|
| | | Slide ² (n / 58) | Proximity ³ (n / 1377) | Domain ⁴ (n / Total) | | | Proximity ^{7,8} | 95% C.I. on Domain |
| Slide | 11 | 0.19 | 0.01 | 0.06 | 1.40 | 0.26 | Slide population (N = 58) | Slope ⁹ |
| Non-slide | 176 | 3.03 | 0.13 | 0.94 | | | | |
| Total | 187 | | 0.14 | 1.00 | | | | |
| Slide | 18 | 0.31 | 0.01 | 0.08 | 1.87 | 0.35 | Three cell radius of slides (N = 1377) | Area ¹⁰ |
| Non-slide | 211 | 3.64 | 0.15 | 0.92 | | | | |
| Total | 229 | | 0.17 | 1.00 | | | | |
| Slide | 3 | 0.05 | 0.00 | 0.08 | 1.83 | 0.35 | | - Intersection - Slope & Area |
| Non-slide | 36 | 0.62 | 0.03 | 0.92 | | | | |
| Total | 39 | | 0.03 | 1.00 | | | | |
| Slide | 26 | 0.45 | 0.02 | 0.07 | 1.64 | 0.31 | | - Union - Slope OR Area |
| Non-slide | 351 | 6.05 | 0.25 | 0.93 | | | | |
| Total | 377 | | 0.27 | 1.00 | | | | |

Notes:

1) 30m-USGS-DEM

2) Fraction of landslide population

3) Fraction of total area stratified by proximity to the slide population

4) Fraction of total area further stratified by 95% C.I. on variable domains of slide population

5) Density Ratio is a measure of efficiency

6) Density of slides per unit of similar terrain

7) The three cell radius is intended to be an approximation of total area in the associated zero order basin

8) Cells down stream of the landside population are included, representing an additional 290 cells

9) Slope variable θ_1

10) Area variable a_1

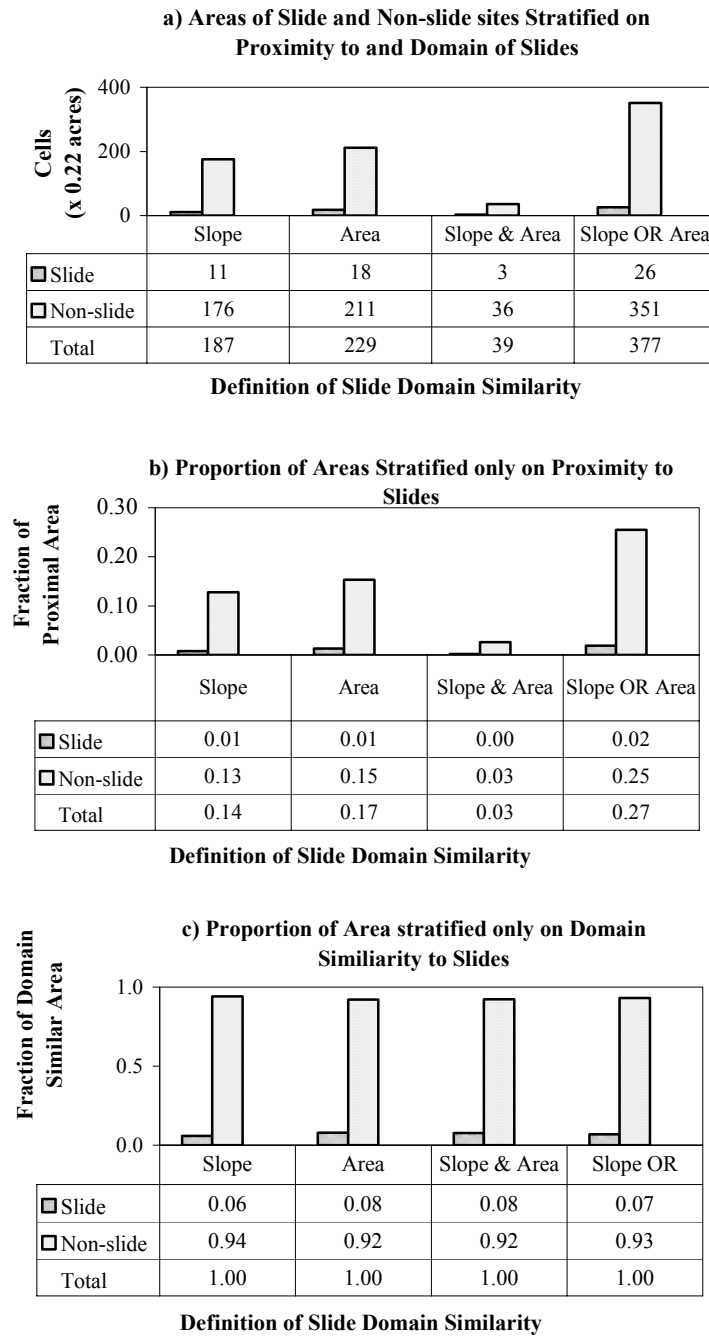


Figure 39. Performance data inside spatially stratified areas. Slide and non-slide sites in spatially stratified areas surrounding the Mapleton slide population. Systematic definitions of domain similarity based on a 95% CI of the slide population. a) areas sub stratified by domain, b) Proportions only and c) Proportions sub stratified by domain.

A couple of inferences from Table [8] and Figure [39] are;

- The most efficient definition of slide site similarity is given by the area variable. The least efficient is given by the slope variable.
- Efficiencies defined by intersections and unions of the slope and area variables fall in between these two extremes.

Given the high similarity of cells inside the stratified regions the performance of the indices is lower and less variable than for the total study area data shown in Table[8].

7.0 Conclusions

Single topographic variables and higher order indices have been contrasted for slide and aerial density (performance) with varying degrees of success depending on the metric of performance. Interpretation of performance of all types and indices is confounded by measurement uncertainty and GIS technique.

Inter index performance based on equal-area-landslide-densities measured across a continuous domain is

$$\text{EVI} > \text{Area} > \text{Slope} > \text{Infinite Slope} > \text{Topographic Ratio}$$

The EVI index performs best because of optimization routines. The area index performs better than slope because area has a high spatial correlation of high landslide probability sites caused by convergence. The hydrologic-geomorphic model indices worked marginally better than the topographic ratios due to a modest amount of manual calibration. The topographic ratio performed least of all due to an equi-finality problem brought on by division.

The intra-index performance based on equal-area-densities for the area and slope index groups is also different. The numerous assumptions made in each one of the algorithms and lack of rigorous testing present makes it difficult to interpret why these differences exist.

There are significant differences ($\alpha < 0.05$) between mean values of the slope, area, and curvature index distributions for slide and non-slide sites in spatially stratified areas adjoined to the existing Mapleton slide population. The minimum resolution on this statement is approximately a 90m radius surrounding the landslides (three

cells). The differences are small but increase proportional to the scope. There is no significant difference between mean values of the area distribution in areas further stratified by a 95% confidence interval on slope defined by the slide population. However, given the highly skewed nature of the area distribution there are sites with extreme values of the logit representing much higher odds of failure.

Empirical modeling was unsuccessful and provided no evidence (p-value = 0.3) of an area threshold on landslide occurrence. Regression analyses of slope against area for the slide population suggests there is mild evidence (p-value of 0.1) of an area threshold on slide occurrence in Mapleton, but it was not distinct. Logistic regression analyses match the behavior suggested by the hydrologic-geomorphic model. Subsequent experiments suggest that such a simple view of the area variable is not appropriate. Other factors such as curvature and interaction with slope have to be factored in. The logistic regressions are an improvement over the geomorphic-hydrologic model in that they take advantage of curvature in the landscape, and optimize the data set. The hydrologic-geomorphic model does not have a curvature variable nor does it have an effective area weighted for convergence as demonstrated by O'loughlin, 1981.

The logistic regressions suggest that the relative multiplicative effect of topographic variables on the logit is;

$$\text{Area (5.5)} > \text{Slope (0.83)} > \text{Curvature (-0.72)}$$

Further the variable interactions act;

Area:Curvature (1.5) acts positively > Slope:Area (-0.21) acts negatively

The biggest obstacle to discrimination between slide and non-slide sites is the definition of similarity. An examination of domain limits on single variables, domain intersections and unions, and area intersections produced significantly different results. Performance measured by areal densities across the Mapleton area suggest the relative performances of the single variable definitions normalized for the slope index are;

Curvature (1.6) > Area(1.1) > Slope (1.0)

The variable intersections and unions using domains are a confusing mixture, but crudely parallel the performance already stated above. Further analysis in spatially stratified regions surrounding the slide population suggests the same pattern. Performances measured by areal densities in spatially stratified regions normalized for the slope index are;

Area(1.3) > [Domain Intersections and Unions] > Slope (1.0)

The number in brackets is density ratio normalized for slope. The optimized definition of topographic variables have areal densities that are less than or equal to single variables. More research is required to determine if these definitions are conservative with respect to future susceptibility of landslides.

Optimized definitions of site similarity using logistic regression may be extrapolated spatially beyond the sampling area. These definitions may not be extrapolated beyond the domain of the sample distributions.

Non-slide sites similar to slide sites, may be predicted and located correctly with known confidence using statistical definitions of site similarity based on a slide inventory. However, location DEM derived variables cannot be confused with ground measurements, thus similar terrain may only be located via X and Y coordinates.

The logit definitions may be transformed into odds and probabilities and are strictly defined as retrogressive in nature. The logit definitions may be extrapolated into the future if a carefully designed scope is carried with it. The physical processes suggest that this scope should include a specified time scale such as the cycle time on bedrock hollow filling and evacuation and the return periods on the 1996 storms. For example, X percentage of the landscape is susceptible to landslide occurrence on a forest management time scale, or DEM derived slopes within X and Y domain limits are Z times as likely to be susceptible to landslide occurrence.

Finally, good definitions of topographic similarity could allow slope stability assessors to invest more time into studying non-topographic factors.

Bibliography

- Alan Marcus. 1980. First-order Drainage Basin Morphology – Definition and Distribution, *Earth Surface Processes*, Vol 5, 389-398
- Benda, Lee. 1997. Slope Instability and Forest Land managers – A Primer and Field Guide, Earth Systems Institute
- Beven, K. and M. J. Kirkby. 1979. A physically based, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, 24, 43-69
- D. Montgomery, W.E. Dietrich. 1989. Source Areas, Drainage Density, and Channel Initiation, *Water Resources Research*, Vol. 25, No.8, Pages 1907-1918, August
- E.M. O’loughlin. 1981. Saturation Regions in Catchments and Their Relations to Soil and Topographic Properties, *Journal of Hydrology*, 53, pg 229-246
- Endreny A., Wood E. 2001. Representing Uncertainty in Runoff Modeling and Flowpath Mapping, *Hydrol. Process.* 15, 2223-2236.
- Fell, R. 1994. Landslide risk assessment and acceptable risk. *Can. Geotech J.* 31:261-272.
- Forest Practices Code of British Columbia.1999. 2nd. Ed. Mapping and Assessing Terrain Stability Guide book, B.C. Ministry of Forests and B.C. Environment, 36p.
- J.D. Hewlett, and A.R. Hibbert. 1967. Factors Affecting the Response of Small Watersheds to the Precipitation in Humid Areas. In: *Interantional Symposium on Forest Hydrology*, W.E. Sopper and H.W. Lull (Editors). Proc., the Pennsylvania State University, August 29 – September 10, 1965. Pergamon, Oxford, pp. 275-290.
- Jenson S. K. and J. O. Domingue. 1988. Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis, *Photogrammetric Engineering and Remote Sensing*. Vol. 54, No. 11, November.pp. 1593-1600.

- Kenneth M. Rood. 1990. Site characteristics and landsliding in forested and clearcut terrain, Queen Charlotte Islands, Land Management Report No. 4, B.C. Ministry of Forests
- Montgomery, D. and Dietrich, W. 1998. A digital terrain model for mapping shallow landslide potential, to be published as a technical report by NCASI
- ODF. 1997. Storm Impacts and Landslides of 1996 – Final Report, Forest Practices Technical Report No.4
- OGDC. 2001. Oregon Geospatial Data Clearinghouse. 2001. Internet web site <http://www.sscgis.state.or.us/>
- Pack, R. T., D. G. Tarboton and C. N. Goodwin. 1998. "The SINMAP Approach to Terrain Stability Mapping," Paper Submitted to 8th Congress of the International Association of Engineering Geology, Vancouver, British Columbia, Canada 21-25 September.
- Peter A. Burrough and Racheal A. McDonnell. 1998. Principles of Geographic Information Systems – Spatial Information Systems and Geostatistics, Oxford University Press.
- Pyles, M.R. and H.A. Froelich. 1987. Discussion of "Rates of landsliding as Impacted by timber management Activities in North Western California" by M. Wolfe and J. Williams. Bulletin of the Association of Engineering Geologist. Vol XXIV, No.3, pp 425-431.
- Resource Inventory Committee. 1998. Terrain Stability Mapping in BC: A Review and Suggested Methods for Landslide Hazard and Risk Mapping - Final Draft, HTML Created: Mar 98, Copyright © 1997 Province of British Columbia, Published by the Resources Inventory Committee
- RIC. 1994. Landslide Hazard Mapping Guidelines for British Columbia, Report to the Earth Sciences Task Force – B.C. Government Resources Inventory Committee, Prepared by Thurber Engineering Ltd. In association with Van Dine Geological Engineering.
- Richard J. Chorley, and Barbara A. Kennedy. 1971. Physical Geography- A systems Approach, Prentice Hall International Inc., London.
- Roy Sidle, Andrew J. Pearce, and Colin Oloughlin. 1985. Hillslope stability and Land Use, Research Monograph, American Geophysical Union

- SHALSTAB. 1998. On-line users Manual at University of California, Berkly, Nov.29, on-line website, <http://socrates.berkeley.edu/~geomorph/shalstab/>
© 1998, William Dietrich and David Montgomery
- SINMAP. 2002. On-line users manual at Utah State University.
<http://moose.cee.usu.edu/sinmap/sinmap.htm>
- Skaugset Arne E., and Richard F. Keim. 2002. Chapter "Landslides, Surface Erosion, and Forest Operations in the Oregon Coast Range" in Hobbs Stephen D., John P. Hayes, Rebecca L. Johnson, Gordon H. Reeves, Tomas A. Spies, John C. Tappeiner II, and Gail E. Wells, eds 2002. Forest and Stream Management in the Oregon Coast Range. Oregon State University Press, Corvallis.
- Tarboton, D. G. 1997. "A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models," Water Resources Research, 33(2): 309-319
- Varnes, D.J. 1974. The logic of geological maps with reference to their interpretation and use for engineering purposes. U.S. Geol. Surv. Prof. Paper 837.
- W.E. Dietrich, and T. Dunne. 1978. Sediment budgets for a small catchment in mountainous terrain, Zeitschrift fur Geomorphologie, Vol.1, pp. 191-206.
- W.E. Dietrich, C.J. Wilson, and S. Reneau. 1986. Hollows, colluvium, and landslides in soil mantled landscapes, Hillslopes processes, pp. 361-388
- William Dietrich. 1991. Hollows, Colluvium, and Landslides in Soil-mantled Landscapes, Proceedings of the 16th annual Geomorphology Symposium, Hillslope Processes, Dunne T., pp 361-388

Appendix A

Report from Statistical Consultant

Equations to predict the occurrence of landslides in the Oregon Coast Range

Project #: C0102-62
Client: Jeremy Appt, MS student, Forest Engineering (Maj. Prof.: Arne Skaugset)
Statistics classes: ST511, ST512, ST513.
Consultants: Vicente Monleon, Cynthia Cooper
Statistical consulting: This is the fourth time that Jeremy uses the Consulting Services.

Background and objectives:

Jeremy is trying to predict the occurrence of landslides in an area around Mapleton using slope and drainage area measurements derived from maps. The ultimate objective is to produce a map of risk of landslides, and identify areas of higher risk

Sampling / response design:

Following two large landslide-producing rainfall events in 1996, the Oregon Department of Forestry identified as many landslides as they could find in four pilot sites in the Coast Range. The main objective of that study was to determine if there was a link between specific forest management practices and landslide frequency. Personnel from ODF visited each site and determined the coordinates of the point of origin of the landslide using GPS. Jeremy is using data from one of those sites (Mapleton, an area about $3 \times 10 \text{ km}^2$). Only landslides that were not started by roads were included in the sample (total 92). There are 3 types of landslides – upslope, in stream channel, and adjacent to stream channel. Jeremy is only interested in upslope landslides – total 56.

The two explanatory variables, slope and drainage area, were derived using the 30m USGS Digital Elevation Model. A DEM is an array of x, y and z coordinates in a 30m grid. He used this grid to divide the study area into $30 \times 30 \text{ m}^2$ cells (about 24,000). Then, he computed the slope of each cell based on the elevation of that cell and that of the eight adjacent cells. He computed the drainage area of each cell by determining the cells that ‘flowed’ into it. The errors associated with the DEM and the derived slope and drainage area were discussed at length in previous consulting sessions.

In summary, for each of the 24,000 cells that comprise the whole study area, Jeremy has an indicator variable of whether that cell was the point of origin of a landslide, and an estimate of its slope and drainage area.

Modeling

Jeremy has used logistic regression to estimate the probability of presence of a landslide in a cell. He transformed the explanatory variables to the tangent of the slope angle and the log of the drainage area. He has used a retrospective sampling design to select the cells used to develop the models: he included all the cells with landslides, and randomly chose 3 times as many cells without landslides. In addition to sampling non-landslide cells from the whole population, he also restricted the areas from which he sampled. Domain restrictions were based on values of the slope and drainage area, selected from the range in which the landslides in the dataset occurred. Location restrictions limited the cells available for selection to a 5- or 10-grid cell radius around a cell with a landslide. He then used a likelihood ratio test of whether the coefficients for slope and drainage area are both 0. When using a location restriction, he did not account for the ‘pairing’.

Assistance needed:

1. Slope stability theory predicts that a landslide will occur when:

$$\frac{qA}{Tb \sin \theta} > \frac{\gamma_s}{\gamma_w} \left(1 - \frac{\tan \theta}{\tan \phi} \right), \quad \text{where:}$$

| | |
|------------|--|
| A | = upslope drainage area |
| θ | = slope angle |
| b | = unit contour width (known constant) |
| γ_w | = unit weight of water (known constant) |
| q | = effective rainfall after losses |
| T | = depth integral of the saturated hydraulic conductivity |
| γ_s | = unit weight of saturated soil |
| ϕ | = internal friction angle |

Of those variables, Jeremy obtained surrogates of A and θ from the DEM, and b and γ_w are known constants. The rest are parameters of “great theoretical importance, but impossible to measure in the field”. Jeremy re-arranged the terms of the equation to get:

$$A > \frac{Tb\gamma_s}{q\gamma_w} \sin \theta - \frac{Tb\gamma_s}{q\gamma_w \tan \phi} \sin \theta \tan \theta = \beta_1 \sin \theta + \beta_2 \sin \theta \tan \theta$$

He wants to know if he can use this information to improve his models.

2. Jeremy wants general assistance with the logistic regression and model development. He has spent a lot of time calculating NPP for the explanatory variables, trying to find a normalizing transformation; he wants to find a transformation to stabilize the variance of the response...He wants to know how to deal with the possible spatial dependence in the data, and how to estimate the uncertainty in the prediction.
3. Measurement error problem: Since the elevations from the DEM are not exact, he has performed a stochastic simulation of the elevation error by adding 'spatially correlated random error fields with a $N(0,15m)$ '. Somehow, he added noise to each cell, and re-calculated the slope and drainage area for the whole dataset. He repeated this process 30 times. He wants to know what to do next.

Recommendations

I. Theoretical model:

We do not know how you could use the theoretical model to improve your regression equation. The model identifies both slope and drainage area as important, and you are already using them.

We recommend that you try some data visualizations rather than focus on modeling – this way you may be able to explore your theoretical model. For example, you may want to draw scatterplots of $\log(\text{area})$ vs. $\tan(\theta)$ and use two symbols, one for landslide and the other for the rest (if there is too much clutter, you may want to use a random sample of the points without landslides). Then, you can choose some values of β_1 and β_2 and display curves of the form $A = \beta_1 \sin(\theta) + \beta_2 \sin(\theta) \tan(\theta)$ on those graphs. You can check whether the landslide points group together on one side of those curves.

II. Model fit and assumptions

1. The constant variance assumption only applies to normal regression. When the response is binary and you use logistic regression, there is no variance assumption to worry about.
2. You do not need to worry about the distribution of the explanatory variables – no need to check NPP or to try to attain normality. In regression, the explanatory variables are assumed to be known constants. You transform

the explanatory variables so that the response meets the assumptions of the kind of regression that you are using.

3. You need to worry about additivity and linearity. You are assuming that the log of the odds of landslide is linear and additive in the $\tan(\theta)$ and $\log(\text{area})$ scales. You may want to do some visualization of the data so find out if it is true. You may want to look at books that deal with discriminant analysis to find out useful plots (such as those described in Johnson and Wicherin, chapter 11), or to the book by Cleveland (1993). Since the data is binary, you may want to use binning or smoothing to be able find out whether the assumptions are met. Another option is to include squared terms (test for linearity) or interactions (additivity) in the regression, and test whether they are significant.
4. We do not know why you would want to use the location restrictions in this retrospective sample. You will not be able to use the resulting equation for prediction, since it is only valid for the domain restriction where they were developed. In any case, you should treat the paired cells with and without landslide as a block.
5. You may have to do some kind of restriction of the study area, since you are not including landslides that originated in roads or streams. For example, creeks and areas adjacent to creeks should not be included, same with roads.
6. Restricting by the range of slopes and drainage areas in which you observed landslides does not seem very appropriate. Are you saying that there cannot be landslides when the slope is greater than 65 degrees? You should decide domain restrictions before looking at the data.
7. There is some concern with the eastern boundary of the study area. It looks like it is a straight line – how did you calculate drainage area for those plots in the boundary? It may be a source of error.

III. Interpretation of the model

You are using logistic regression to get a measure of similarity to the locations where landslides occurred in 1995-1996. The measure of similarity is based only in the slope and drainage area, and assumes linearity and additivity in whatever scales you choose. When you get your prediction, a number closer to one would indicate a greater similarity to the landslide sites. The study is totally observational, so you can only find associations, not probabilities or risks of landslides. From a statistical point of view, we cannot say that areas more similar according to the measure that

you developed are more likely to have a landslide. So, any reference to probability or risk of landslide cannot be based on statistical principles. The observational nature of the study makes it such it could only be applied to the study area and to conditions similar to those of the 1995-1996 storms.

You are using a retrospective sampling scheme to develop your models. Therefore, you can only describe odds ratios, not probabilities. You may want to read McCullagh and Nelder (1989), p. 111, for some comments on this problem and possible remedial measures.

IV. Model validation

There is a general consensus that you should try an external validation of the model, specially given the limitations of an observational study. This seems specially feasible here, since there are other sites available. You could find out if the model developed in Mapelton could predict the locations of landslides at the other sites. Then, you could have an estimate of the false alarm and detection failure rates that is independent of the data used to develop the model and, therefore, more credible.

V. Measurement error

We can see problems with the approach that you are using. First of all, if your aim is to create a map of probabilities of landslide occurrence, using the same DEM data that you used to develop the logistic regression, there is no measurement error problem. There may not be any pattern either, but that is not going to improve by adding noise.

If you want to explore the relationship between some ‘true’ measurement of slope and drainage area, instead of those derived from the DEM, I would recommend that you go to the field and takes better measurements of those variables. Adding noise to the elevation data from the DEM according to some assumed model only accounts for one source of the many possible errors in the explanatory variables. For example, even if the DEM was known exactly, the algorithm that you used to calculate slopes tends to flatten the stepper areas significantly. Then, there is the problem of scale: given how variable is the topography in the Coast Range, averaging the slope over 30m does not seem reasonable. Also, drainage area is calculated in increments of 900 m² – quite a large area.

If you decide to pursue his approach, I guess that you would develop regression equations for each of the 30 DEM that you generated, carry the process to the end, and find a final ‘envelop’ for the CI.

References:

Cleveland, W.S. 1993. Visualizing Data. Murray Hill, NJ.

Johnson, R.A. and Wichern. 1998. Applied Multivariate Statistical Analysis, 4th edition. Prentice Hall, NJ.

McCullagh, P. and J.A. Nelder. 1989. Generalized Linear Models, 2nd edition. Cambridge University Press.

Appendix B

Assumptions in Statistical Models and Logistic Regressions

1. The natural link for a binary variable (i.e., slide = 1, and non-slide = 0) is the logit or log *Odds* function; $\text{logit}(\pi) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$, where π is the population mean of responses equal to one. This leads to logistic regression. The logit may be back transformed into the odds (ϖ) where $\varpi = e^{\text{logit}(\pi)}$. The odds may be back transformed into the probabilities(π) where $\pi = \frac{\varpi}{1 + \varpi}$. The range of possible probabilities is 0 to 1. The coefficients are determined by a maximum likelihood algorithm that maximizes the probabilities for the sub-population with response values equal to one.
2. Non-constant variance. Logistic regression is non-linear in the β 's however, all of the non-linearity is contained in the link function. The regression structure may be used like normal regression
3. Linearity in explanatory variables. The statistical model should match assumptions based on physical process. The hydrologic-geomorphic model does not suggest any non-linear explanatory variables.
4. Additivity in the expected values based on interactions between explanatory variables. The statistical model should match assumptions based on the physical process. The hydrologic-geomorphic model does suggest Slope:Area and Area:Convergence interactions.
5. Explanatory variables are constants. Therefore transformations like Sine, Tangent and Log_{10} are just matters of convenience.
6. Independence of Response

Response values equal to one in cells down stream of one another are not independent. If I remove response errors down stream of the landslides then I also further limit the scope of the analysis.

Response values equal to zero in the *Area* (converse of above) of a slide cell are not 100% independent of one another. The level of independence increases with the distance of separation.

Response values of zero that are not in the *Area* of the slide are by assumption of the DEM calculations assumed to be independent of the response in the slide cell, but still partially dependent on one another. Lack of independence causes no bias in least squares estimates of the coefficient, but standard errors are seriously affected (Sleuth, pg.203). Regression analyses require that after accounting for the effects of the explanatory variables, the responses be independent (Sleuth, pg.427). Is one slide causing another? This answer is not available. The severity of the consequences is a function of the severity of the violation.

7. Representative Sampling

Response values of zero that are not in the *Area* are by virtue of the GIS algorithm assumed independent of the landslide, therefore cannot be playing a role in the response. The selection of a sampling radius in lieu of a contributing area significantly changes the scope of analysis. The primary purpose of this analysis is to carefully distinguish between slide and non-slide sites within some useful scope of application. This does not necessarily include testing theory on *Area* thresholds in landslide occurrence. Including values outside the *Area*, confounds interpretation of regression output for potential thresholds. The data, which are non-

representative of threshold theory, may mask or create patterns in the fitted data that are unrelated.

Appendix C

Regression Analysis of Soil Depth at Slide Sites in Mapleton

Regression Analyses

Source: 1996 ODF Landslide report

Location: Mapleton core study area

Data: up-slope landslide subset population

The proposed regression model includes;

Response: ls_maxd (landslide maximum depth, units are in feet)

Explanatory variables: slp_blw, ls_width, drain_area (slope below the landslide headscarp in percent, landslide width in feet, and landslide drainage area in acres)

Factors: the data is factored by sub-basin (five in total)

Methods: A fully saturated two way interaction model was analyzed using least squares regression techniques. Backward selection was employed to simplify the model. Tabulated results are presented below;

Mapleton Core Study Area Up slope Landslide data Backward Variable Selection

| | | | Absolute value of t value | | | | | | | | | | | | | | | | |
|-----------------------|-------------------|---------------|---------------------------|---------|---------|---------|---------|---------|---------|---------|---------|----------|----------|----------|----------|----------|----------|----------|----------|
| Explanatory variables | | | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 | Round 6 | Round 7 | Round 8 | Round 9 | Round 10 | Round 11 | Round 12 | Round 13 | Round 14 | Round 15 | Round 16 | Round 17 |
| Intercept | | | 0.04594 | 0.0451 | 0.04466 | 0.07655 | 0.07294 | 0.05261 | 0.14572 | 0.51541 | 0.59241 | 0.75938 | 0.35585 | 0.99379 | 0.73576 | 3.61235 | 2.96584 | 3.28021 | 0.07344 |
| Main Effect | Slope below | | 0.409 | 0.42275 | 0.42829 | 0.54284 | 0.54753 | 0.56634 | 0.61709 | 0.96591 | 1.03722 | 1.13617 | 0.85501 | 1.56197 | 1.36377 | 2.44808 | 1.55869 | 1.6059 | 0.45091 |
| Main Effect | Drainage Area | | 0.1732 | 0.17826 | 0.1815 | 0.43929 | 0.48667 | 0.49469 | 0.43075 | 0.25775 | 0.11966 | 1.22037 | 0.97964 | 0.98423 | 1.031 | 0.73746 | 0.41168 | 0.02702 | 0.46748 |
| Main Effect | Slide Width | | 1.5547 | 1.62384 | 1.66089 | 1.7606 | 1.78255 | 1.87331 | 2.07203 | 2.96395 | 3.4627 | 3.41684 | 2.9689 | 3.18124 | 2.85338 | 2.31134 | 2.05065 | 2.65606 | 1.15424 |
| Indicator 1 | Bridge | | 1.42475 | 1.45601 | 1.50562 | 1.52677 | 1.54645 | 1.60111 | 1.60347 | 1.48178 | 1.58348 | 1.8303 | 0.6159 | 0.67689 | 1.08183 | 1.88242 | 1.90356 | | |
| Indicator 2 | Brush | | 1.57548 | 2.18302 | 2.21724 | 2.303 | 2.34966 | 2.38016 | 2.38456 | 2.47396 | 2.84823 | 2.66397 | 2.6287 | 2.31513 | 2.25696 | 1.66823 | 1.70722 | | |
| Indicator 3 | Sluslaw | | 1.14118 | 1.1576 | 1.1831 | 1.20729 | 1.22396 | 1.2371 | 1.17036 | 1.4408 | 1.77111 | 1.27887 | 1.11494 | 2.4589 | 2.47359 | 2.69531 | 2.48996 | | |
| Indicator 4 | Turner | | 0.2708 | 0.27466 | 0.27795 | 0.27983 | 2.31678 | 2.3702 | 2.70275 | 2.66254 | 2.83445 | 2.99288 | 2.52995 | 2.34119 | 1.78642 | 1.54747 | 0.49598 | | |
| Two Way Interaction | Drainage Area | Slope below | 0.03682 | 0.0386 | 0.03888 | | | | | | | | | | | | | | 0.40795 |
| Two Way Interaction | Slide Width | Slope below | 1.44463 | 1.50999 | 1.54015 | 1.64567 | 1.66754 | 1.79426 | 1.90385 | 2.58139 | 3.05365 | 2.9952 | 2.6215 | 2.85666 | 2.58713 | | | | 0.94825 |
| Two Way Interaction | Bridge | Slope below | 1.11416 | 1.13818 | 1.18564 | 1.20527 | 1.22012 | 1.25471 | 1.2218 | 1.07932 | | | | | | | | | |
| Two Way Interaction | Brush | Slope below | 1.85856 | 2.25139 | 2.28483 | 2.42338 | 2.46366 | 2.79717 | 2.8067 | 2.91622 | 3.35106 | 3.1874 | 3.12849 | 2.7962 | 2.72831 | 2.06345 | | | |
| Two Way Interaction | Sluslaw | Slope below | 1.33641 | 1.35533 | 1.38393 | 1.43937 | 1.45874 | 1.48075 | 1.52762 | 1.7856 | 2.12197 | 1.7359 | 1.5739 | | | | | | |
| Two Way Interaction | Turner | Slope below | 0.06127 | 0.06218 | 0.06184 | 0.06034 | | | | | | | | | | | | | |
| Two Way Interaction | Slide Width | Drainage Area | 0.56456 | 0.5791 | 0.61729 | 0.62521 | 0.68731 | 0.69976 | 0.64148 | | | | | | | | | | 0.3734 |
| Two Way Interaction | Bridge | Drainage Area | 0.01711 | 0.01688 | | | | | | | | | | | | | | | |
| Two Way Interaction | Brush | Drainage Area | 0.0116 | | | | | | | | | | | | | | | | |
| Two Way Interaction | Sluslaw | Drainage Area | 0.71168 | 0.72373 | 0.75468 | 0.76378 | 0.77144 | 0.77783 | 0.83559 | 1.25158 | 1.36477 | | | | | | | | |
| Two Way Interaction | Turner | Drainage Area | 1.72989 | 1.75444 | 1.78687 | 1.81082 | 2.24885 | 2.28744 | 2.31294 | 2.24496 | 2.40407 | 2.70815 | 2.44646 | 2.37726 | 3.10301 | 2.00587 | | | |
| Two Way Interaction | Bridge | Slide Width | 1.05816 | 1.0731 | 1.2382 | 1.25956 | 1.28156 | 1.33793 | 1.70517 | 1.63537 | 1.5723 | 1.73532 | | | | | | | |
| Two Way Interaction | Brush | Slide Width | 0.05859 | 0.06262 | 0.06332 | 0.07179 | 0.07636 | | | | | | | | | | | | |
| Two Way Interaction | Sluslaw | Slide Width | 0.42218 | 0.42808 | 0.4348 | 0.46201 | 0.47027 | 0.47071 | | | | | | | | | | | |
| Two Way Interaction | Turner | Slide Width | 1.44908 | 1.47023 | 1.49079 | 1.5136 | 1.66438 | 1.71357 | 2.09775 | 2.09013 | 2.24871 | 2.25453 | 1.72961 | 1.61946 | | | | | |
| Additional stats | | | | | | | | | | | | | | | | | | | |
| | min 2 way t value | | 0.0116 | 0.01688 | 0.03888 | 0.06034 | 0.07636 | 0.47071 | 0.64148 | 1.07932 | 1.36477 | 1.73532 | 1.5739 | 1.61946 | 2.58713 | 0.73746 | 0.41168 | 0.02702 | 0.07344 |
| | RSD | | 1.3594 | 1.3403 | 1.3221 | 1.3406 | 1.2879 | 1.2717 | 1.2596 | 1.2508 | 1.2532 | 1.6254 | 1.2934 | 1.314 | 1.3364 | 1.4135 | 1.5001 | 1.5695 | 1.5998 |
| | R ² | | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | 0.59 | 0.58 | 0.57 | 0.55 | 0.52 | 0.5 | 0.47 | 0.39 | 0.29 | 0.16 | 0.17 |
| | d.f. | | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 50 | 54 | 51 |

Notes:

- 1 Backwards selection routine using MLR on Mapleton upslope landslides
- 2 The minimum t value from the set of two way interaction variables was selected for subsequent deletion

We eliminated all two way interactions involving lswidth, but, the lswidth main effect still has t-value greater than 2, therefore we can keep.

Arbitrary cutoff. All two way interactions have t value greater than 2. Minimum t value search extended to all variables

Dropping two way interaction variables by t value no longer makes sense since some of the main effects are lower.

The fully saturated model has an R^2 of 0.59. Backward selection was carried out by dropping the interaction with the lowest absolute value of the t-statistic. The R^2 term rapidly decreases after approximately 9 rounds of backward selection. Round nine defines the simplest accurate model. The model still has some significant interaction with sub-basins, drainage area, landslide width, and slopes below landslides. This may cause speculation as to physical significance of some explanatory variables in some sub-basins and not others. Conversely, there may be confounding variables. Treating the sub-basins purely as main effects (i.e., no two way interaction in round 15), the R^2 drops to 0.29. Two-way interaction of the main effects (i.e., round 17), the R^2 drops to 0.17. Without any sub-basin interaction, the R^2 drops to 0.16.