QUANTITATIVE LANDSLIDE SUSCEPTIBILITY ASSESSMENT OF THE WAIKATO REGION USING GIS

A thesis

submitted in partial fulfilment

of the requirements for the degree

of

Master of Science

in Earth Sciences

at

The University of Waikato

by

Renée Deborah Schicker

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2010

Abstract

Areas that have experienced landslide events in the past and the conditioning factors present at these sites can be used to identify areas of the same or similar susceptibility. This can be achieved through a landslide susceptibility assessment using a landslide inventory, a set of predictor variables and specialised computer software.

A quantitative landslide susceptibility assessment was conducted for the Waikato Region using two statistical approaches and eleven predictor variables. A landslide inventory map was constructed from the GNS QMap landslide spatial data and GeoNet landslide catalogue. Parameter maps for slope, elevation, aspect, lithology, land cover, soil order, mean monthly rainfall, maximum monthly rainfall, distance from roads, distance from faults and distance from rivers, were constructed and compiled into a database with the landslide inventory.

The compiled data underwent both bivariate (weights of evidence) and multivariate (logistic regression) statistical analysis, and a landslide susceptibility map was derived for each. In the weights of evidence approach, the presence and absence of each class in relation to landslide occurrence and non-occurrence was individually assessed for each predictive factor. Logistic regression involves fitting a generalised non-linear model to the data based on a binary predictor (presence or absence of a past landslide event). For each method, a landslide susceptibility map was derived, and the model fit assessed using the landslide inventory.

Both susceptibility maps underwent an evaluation to determine the better predictive model. An independent landslide data set was compiled from observations made in Google Earth, and used to establish a set of prediction rate curves and cumulative area curves. Both susceptibility maps resulted in very similar prediction rate curves. Weights of evidence gave a better prediction rate in the 10, 20 and 30% most susceptible pixels, but not in the 40% most susceptible pixels. Neither susceptibility map could be justified as being better than the other based on the prediction rate curves alone. The cumulative area curves for each susceptibility map resulted in very different outcomes. Logistic regression gave the best result with a large proportion of the landslide area within a small proportion of the total area in the high susceptibility classes. Weights of evidence had a larger proportion of the landslide area in high susceptibility classes than logistic regression, but this was associated with a large proportion of total area. Based on the evaluation, the susceptibility map derived using logistic regression was determined to be superior.

Acknowledgements

This study would not have been achievable without the help and support from various people I have had the fortune to encounter. Thanks to Mark Rattenbury for the QMap landslide polygon spatial data for the Waikato Topo 260 Tile, and also for putting me in touch with Grant Dellow. Grant, I remain indebted for your generosity, without the GeoNet landslide catalogue this study would not have gone so far. Thank you once again. Also thank you to Peter Kamp for lending me both the Auckland and Waikato QMap vector data CDs.

A very special thanks to both my supervisors Dr. Vicki Moon and Dr. Megan Balks for their time and effort, proof reading, advice, and humour. I really appreciate what you have done for me. Thanks Vicki for the help with logistic regression and lithologies, and for the many other things I may have forgotten to mention.

I am sincerely grateful to have received the Broad Memorial Fund, and am thankful for the financial support towards this project. Thanks should go to Ray Littler for his attempts to help me with statistics, and Darryl Gillgren for the odd piece of GIS advice.

I am incredibly grateful to Aunty Kathryn and Uncle John for all their advice and support; and for really helping me out when things were a bit crazy. I should also extend my gratitude to Tania and Elliot Billings for taking me in for the first half of the year, and for all the good talks about anything and everything.

Also to two of my best friends, Sarah and Sez, thanks for always looking out for me, and just being the awesome, reliable friends that you are. Thanks also to Fi and Bashirah for making the office environment entertaining, it's been quite an enjoyable and eventful year. Thanks to my fellow EOS colleagues for making lunch times so much more interesting.

To my parents, thank you for your love and support, and for everything you have done for me. To my brothers Karl and Brendon cheers for keeping me on my toes all these years. To all my Hamilton and Northland friends and whanau, I look forward to catching up with you all again. To anyone I may have left out, I apologise, but I thank you all the same.

Table of Contents

Digital Appendices

List of Figures

Chapter Five

Chapter Six

List of Maps

Chapter Three

Chapter Four

Chapter Five

List of Tables

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Chapter Six

CHAPTER ONE

Introduction

1.1 Introduction

Landslides are a hazard to people and property which could potentially be avoided with forecasting. Landslide susceptibility assessments allow predictions to be made of where future landslide events are likely to occur. These follow a philosophy that the same or similar causative factors linked to past landslide occurrence can be used to predict where future events are likely to occur. By identifying likely causative factors and where they occur within a given area of interest, the landslide susceptibility can be modelled and mapped.

1.2 Aims and Objectives

This study has two aims:

(1) to derive landslide susceptibility maps for the Waikato Region using available data and applying various statistical techniques in conjunction with automated GIS; and

(2) to compare and contrast maps from each method chosen to assess which statistical model gives the best estimate of landslide susceptibility for this region.

In order to do this the following objectives will need to be met:

(1) collect data for a landslide database and a selection of causative (topographical, geological, environmental and anthropogenic) parameters;

(2) develop a set of scripts to execute automated GIS processes from which a set of parameter maps and a landslide inventory map can be produced;

(3) determine which bivariate and multivariate statistical analyses are most appropriate, and apply these to produce a landslide susceptibility map; and

(4) calibrate and evaluate the chosen model using an independent landslide dataset.

1.3 Landslides and Causative Factors

The simplest definition of a landslide is the movement of a mass of rock, debris or earth down a slope (Dai *et al.*, 2002; Ohlmacher and Davis, 2003; Selby, 1993; Wang and Sassa, 2006). Landslides are influenced by gravity without the direct aid of media such as water, air or ice (Selby, 1993), although these factors can have an indirect effect. Heavy rainfall, seismic shaking, rapid stream erosion, storm waves, volcanic eruptions and changes in water level can result in landslides as a natural hazard (Dai *et al.*, 2002; Wang and Sassa, 2006). Human (anthropogenic) activities which result in some modification to the land such as slope excavation and deforestation can also indirectly trigger a landslide (Dai *et al.*, 2002; He and Beighley, 2008; Remondo *et al.*, 2008). Landslides are remarkably complex, as there are a wide range of internal factors (slope, geology, land cover, aspect, elevation, soil, and many more) and external factors (natural and anthropogenic) which interact with each other that can lead to landsliding (Carrara *et al.*,1999). A range of factors such as slope, geology and land cover can be used as intrinsic or predictor variables to determine areas susceptible to landslides, while external factors are considered as extrinsic variables which tend to trigger landslides in areas susceptible to landslides (Dahal *et al.*, 2008a).

1.4 Costs

Extensive damage to both property and lives can result from a natural hazard such as landslides (Park and Chi, 2008). On a global scale the cost is huge, with Taiwan and Japan alone costing a gross domestic product (GDP) of about US$140 billion and US$120 billion respectively (United Nations, 2009). The cost of landslide related damage in New Zealand is significantly less than this. In the 2008-2009 period (Figure 1.1), there were 2,226 claims with a cost of

approximately $41 million relating to landslips, whereas there were 1,307 claims at a cost of $18 million the year before (EQC, 2008, 2009; Figure 1.1).

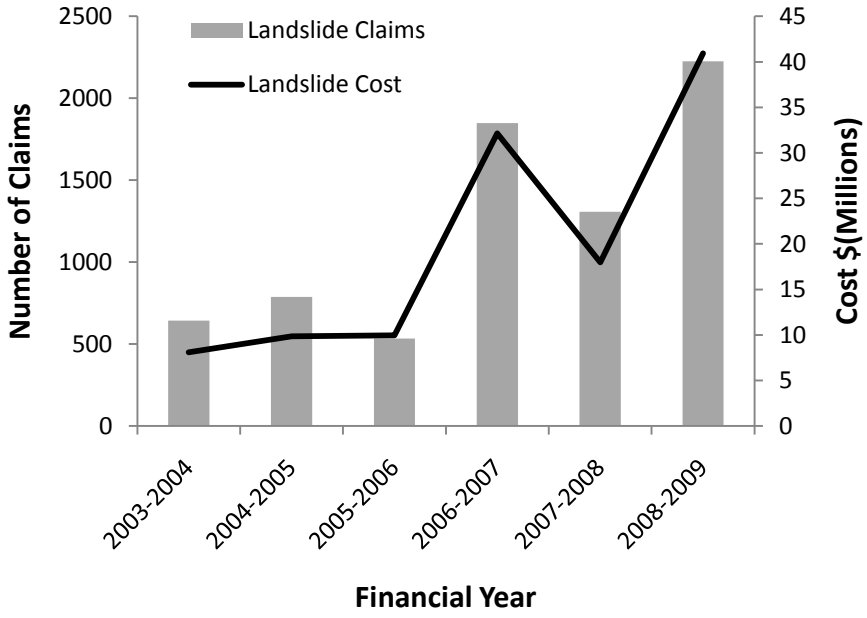


Figure 1.1 Plot of the number of claims per year and the resulting cost pertaining to landslides in New Zealand for the last six financial years (data from EQC, 2004, 2005, 2006, 2007, 2008, 2009).

The cost as a result of landslides becomes more of a problem with an ever increasing population, limited land, urban sprawl and the desire for some to have their own lifestyle block away from the city, which means more people could be affected, resulting in more claims and a higher financial cost. Added to this cost is the economic issue of inflation, as costs tend to increase with time. If building in areas susceptible to landslide is avoided, the hazard and cost in theory should be reduced.

1.5 Susceptibility Mapping

Many landslide susceptibility assessments have been carried out internationally, and these involve a wide range of both qualitative and quantitative techniques which have developed over time. Initially landslide susceptibility assessments relied on qualitative approaches, however semi-quantitative and quantitative approaches are now more commonly used. Quantitative methods can be implemented using a geographic information system (GIS) with or without the use

of an external statistical program. GIS programs and methods have also developed quite rapidly as more capabilities are now available, allowing for a greater range of spatial analyses.

Landslide susceptibility mapping and qualitative risk assessment has previously been carried out in the Waikato Region by implementing a GIS and assigning parameter class rankings for each parameter map (Smith, 1999). In her study, Smith (1999) applied both qualitative ranking and multiple regression, however the methods themselves were not outlined in detail, and so it was impossible to replicate the findings. A validation or evaluation of the resulting model, which is considered quite important by the current standards, was lacking in the 1999 Waikato susceptibility assessment. It is important however, to remember that this earlier assessment was done ten years ago and there have been many developments in the field of landslide susceptibility assessment. Developments include a wider knowledge base in regards to various statistical methods, as well as advances in computers and GIS among various other things. This study will take advantage of these advances to derive a more robust susceptibility assessment.

1.6 Waikato Region

The Waikato Region covers an area of some 25,000 km^2 (Environment Waikato, 1998) which is centrally located in the North Island of New Zealand (Figure 1.2). The Waikato Region stretches across both the western and eastern coasts and encompasses the Waikato River which travels from the central North Island volcanoes south of Lake Taupo (near the southern most part of the region) north to Port Waikato, near Auckland.

The topology of the Waikato Region ranges from extensive lowlands to relatively high elevation mountains (Leathwick *et al.,* 2003). The Regional Council (Environment Waikato) distinguishes four distinct topographical areas based on associated land and soil resource management issues: Taupo Volcanic Zone, western and central hill country, Waikato lowlands and Hauraki Plains, and the Eastern Ranges and coastline (Environment Waikato, 1998).

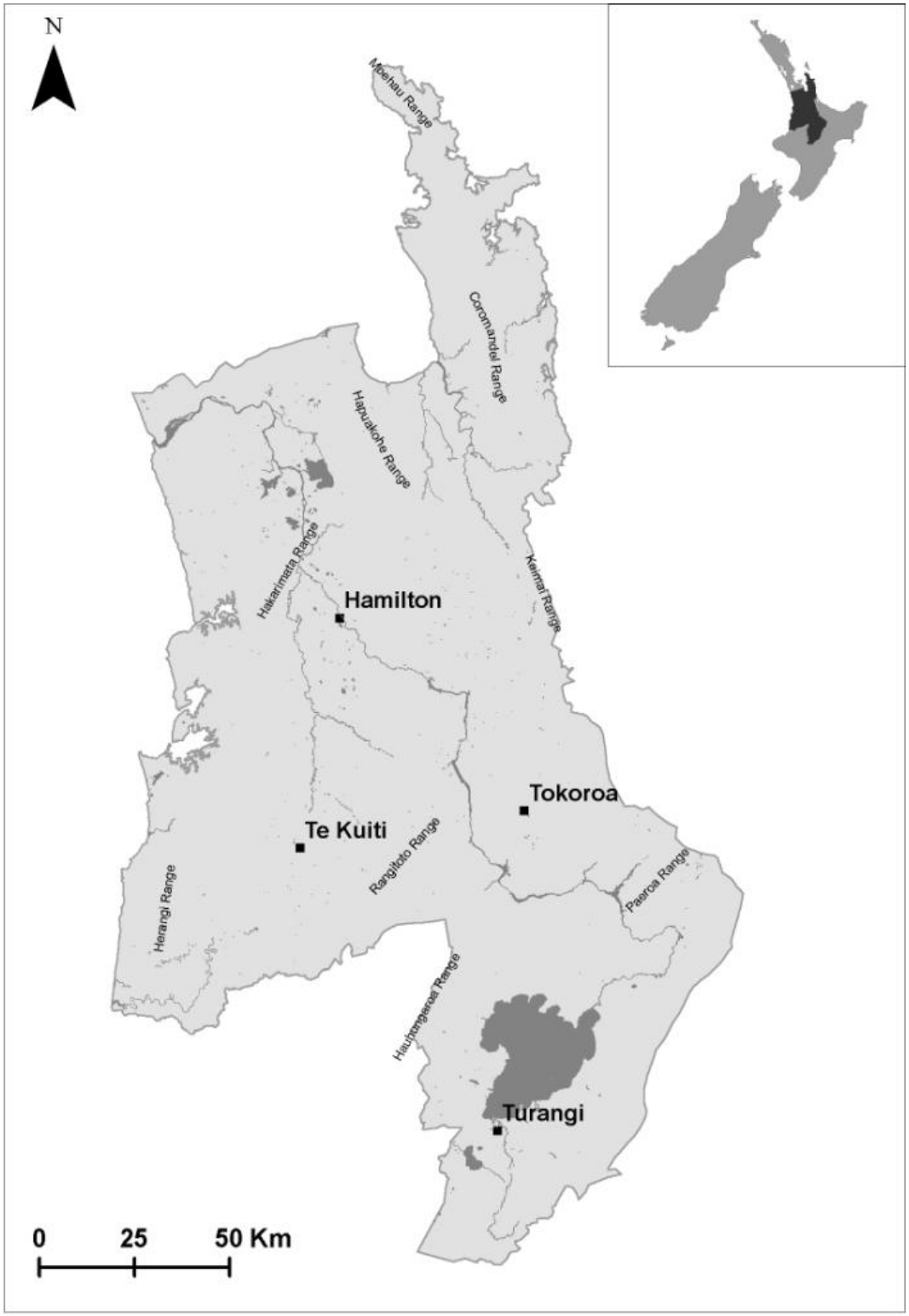


Figure.1.2 The Waikato Region and its location within New Zealand

The Taupo Volcanic Zone consists of pumice lands with high landscape values and geothermal features and the land use includes dairying and stock grazing, plantation and indigenous forest (Environment Waikato, 1998). The western and central hill country consists of steep sedimentary hills and widespread cave and karst systems which are covered by large expanses of indigenous forest with many areas used for plantation forestry and extensive grazing (Environment Waikato, 1998). The flat and gently rolling land of the Waikato lowlands and Hauraki plains comprises of the Hinuera Formation, large wetlands and peat soils and the land is generally used for urban and rural settlement and dairying with some horticulture and cropping (Environment Waikato, 1998). The Eastern Ranges and coastline are volcanic in origin and have good quality soils in low lying areas which are used for horticulture and dairying, with drystock farming, plantation forestry and indigenous forest elsewhere (Environment Waikato, 1998).

1.7 Structure of Thesis

A review of the literature outlining the differences between hazard and susceptibility assessments is given in chapter two. This includes a discussion of a wide range of variables considered in these assessments, as well as the various methods that can be applied to determine landslide susceptibility. Chapter three discusses the selection of variables used in this study and the GIS techniques employed to derive the landslide inventory map, parameter maps and compiled database. The selection and simplification of classes for each categorical spatial dataset, and their spatial distribution is also discussed. A bivariate statistical method, the weights of evidence approach, is explained and applied in chapter four. This involves the exclusion of parameters, in the search for a significant model, and deriving a landslide susceptibility map. Chapter five investigates the multivariate approach of logistic regression, random sampling, deriving a model, and validating its goodness of fit by applying the model to a second random sample. Both multivariate and bivariate statistically derived models are then evaluated using cumulative frequency curves and prediction rate curves in chapter six. Chapter seven concludes the findings of this landslide susceptibility assessment for the Waikato Region.

CHAPTER TWO

Literature Review

2.1 Introduction

This chapter reviews literature regarding landslide susceptibility assessment, in particular the use of geographic information systems (GIS) in these assessments. Both intrinsic (causative) and extrinsic (triggering) variables are explored, with some insight into which are most commonly considered to be influential and why. A wide range of qualitative (expert opinion based) and quantitative (statistical and physically based) methods are explored, along with consideration to the advantages and disadvantages in applying these to landslide susceptibility assessments. Finally, a discussion is given on the chosen methods for the landslide susceptibility assessment for the Waikato Region.

2.2 Landslide Susceptibility and Hazard

A landslide, being a naturally occurring event, is only considered a hazard if it jeopardises something or someone (Alexander, 2008). The most widely accepted definition for a natural hazard is that of Varnes (1984), who defines natural hazard as *"the probability of occurrence within a specified period of time and within a given area of a potentially damaging phenomenon"* (Ayalew *et al.,* 2005; Carrara *et al.,* 1991; Dai *et al.,* 2001; Guzzetti *et al.,* 1999; Thiery *et al.,* 2007; Wang and Sassa, 2006). As such, landslide hazard is understood to be based on the magnitude (size/speed) of the landslide and where and when it will occur as the probability of a landslide occurrence (Carrara and Pike, 2008; Dahal *et al.,* 2008a; Guzzetti *et al.,* 1999; Thiery *et al.,* 2007). Landslide hazard assessments include both spatial and temporal aspects of both causative (intrinsic) and triggering (extrinsic) factors in addition to some thresholds (Ayalew *et al.,* 2005; Dahal *et al.,* 2008a; Remondo *et al.,* 2003).

The implementation of landslide hazard analysis rarely complies with the above definitions (Ayalew *et al.,* 2005) and the analysis tends to be more consistent with landslide susceptibility assessment. As a result the terminology of hazard and susceptibility appears to be used interchangeably in the literature and the two have become somewhat synonymous (Carrara and Pike, 2008). This issue has likely arisen as, in reality, it is difficult to near impossible to accurately predict all three (location, magnitude, and time) attributes (Ayalew *et al.,* 2005; Carrara and Pike, 2008). In particular, temporal data such as past landslide occurrences is often difficult to obtain (Remondo *et al.,* 2003) and the time dimension generally defies most attempts at being incorporated into predictive models (Carrara and Pike, 2008).

An alternative is to look at landslide susceptibility which is the relative hazard in a given area without any reference to time and magnitude (Ermini *et al.,* 2005). Landslide susceptibility is an indication of the likelihood of future landslide occurrence in a given area based on a given set of environmental factors (section 2.4) (Ermini *et al.,* 2005; He and Beighley, 2008) which are related to past landslide occurrences (Fernandez *et al.,* 1999; Santacana *et al.,* 2003). This follows the general principle that "the past is the key to the future" in which future landslides are most likely to occur under the same conditions as past landslides (Can *et al.,* 2005; Carrara *et al.,* 1991; Dai and Lee, 2002; Ermini *et al.,* 2005; Guzzetti *et al.,* 1999; Lee and Talib, 2005; Varnes *et al.* 1984). It is important to both accurately detect past landslides, and to effectively determine the relationships between spatial data representing the physical conditions and landslide occurrences, in order to identify locations susceptible to future landslides (Park and Chi, 2008).

Traditional procedures for landslide susceptibility assessment were laborious and time-consuming as data was handled and processed manually (Dahal *et al.,* 2008a; Dai *et al.,* 2001). Due to developments in geographic information systems (GIS) and computer applications it is now easier to carry out landslide susceptibility analysis (Dahal *et al.,* 2008a). Landslide susceptibility assessment leads onto hazard mitigation (Can *et al.,* 2005; Cevik and Topal, 2003) and risk assessment. Hazard mitigation that can be implemented includes mitigation plans, land use restrictions and early warning systems (He and Beighley, 2008).

2.3 Geographic Information Systems (GIS)

A geographic information system (GIS) is a computer technology which enables the user to capture, acquire, validate, store, analyse, retrieve, transfer, manipulate, and display spatial data sets representing some aspect of the real world (Carrara *et al.*, 1999; Dai *et al.*, 2001; Lee and Choi, 2004; Pachauri and Pant, 1992). Advances in GIS technology, and more widespread use and awareness of GIS technology, have contributed to the advancement of hillslope mass-movement analyses (Alexander, 2008; Carrara *et al.*, 1999; Carrara and Pike, 2008).

Carrara *et al.* (1999) note that there are several popular misconceptions in the application of GIS based landslide susceptibility assessments. These include the belief that: a map in hand draft form is less accurate and credible than a similar one generated by computer; a map of landslide hazard obtained through GIS data manipulations is assumed to be more objective than a comparable hand-made product founded on the same conceptual model and derived from the same input data; and people with no expertise in GIS technology should no longer have a problem handling geographical data in a GIS environment. It is important to keep in mind that GIS can only do much the same as earlier methods (such as map overlay) only it can do it quicker, and with reduced manpower (Alexander, 2008). GIS and databases are useful technical tools which have greatly contributed to understanding landslides as a hazard, but not so in understanding the process itself in space and time (Alexander, 2008; Dikau *et al.*, 1996). Alexander (2008) reasons this may be due to the tendency of using GIS in an inductive mode, making generalisations based on specific instances.

2.4 Variables

Independent or explanatory variables are variables which are used to attempt to explain the dependent or response variable (Moore and McCabe, 2003). In landslide susceptibility assessments past landslide locations in the landslide inventory make up the dependent variable. Independent variables can be any factor that is suspected to have some causative influence in landslide occurrence. These broadly include, and are not limited to, a wide range of topographical, geological, hydrological, geomorphological, seismic, and anthropogenic factors.

While there are no universal criteria or guidelines for selecting independent variables there are some agreements in what a variable must have or be in order for it to be considered in analysis (Ayalew and Yamagishi, 2005). In order for an independent variable to be considered it must be operational, complete, heterogeneous, quantifiable, and non-redundant (Ayalew and Yamagishi, 2005). For the independent variable to be operational, it should on the outset, appear to potentially have some form of relationship with the dependent variable (Ayalew and Yamagishi, 2005) otherwise it would be irrelevant to consider it. A complete set of data is necessary in order to cover the whole area of interest (Ayalew and Yamagishi, 2005). If there are gaps, there is data unaccounted and unexplained, which will have a bearing on the end result. Non-uniformity in terms of spatial variation (Ayalew and Yamagishi, 2005) is important because if the data is homogenous, there is no spatial variation and there is no point including it. Heterogeneous spatial data should be used as it has spatial variation which allows for different possible outcomes to be drawn from it. The independent variable should be quantifiable, meaning that it should be possible to express it using any of the various types of measurement scales (Ayalew and Yamagishi, 2005). A redundant variable serves no additional purpose other than to explain what has already been explained by an existing variable (Ayalew and Yamagishi, 2005). Variable redundancy is undesirable as this would result in a double influence on the outcome (Ayalew and Yamagishi, 2005).

There are two categories of variables that are used to determine landslide hazard in an area: (1) intrinsic variables which contribute to, and determine landslide susceptibility; and (2) extrinsic variables which have a propensity to trigger landslides in a given area of susceptibility (Cevik and Topal, 2003; Dahal *et al.*, 2008a; Dai *et al.*, 2001; Wu and Sidle, 1995).

2.4.1 Extrinsic (Triggering) Variables

Extrinsic variables include both natural and human induced landslide triggers. Natural triggering factors include intense rainfall (Can *et al.*, 2005; Cevik and Topal, 2003; Dahal *et al.*, 2008a; Dai *et al.*, 2001; He and Beighley, 2008; Lee *et al.*, 2003a; Ohlmacher and Davis, 2003; Remondo *et al.*, 2003; Wang and Sassa, 2006; Wieczorek, 1996), earthquake shaking (Cevik and Topal, 2003; Dahal *et*

al., 2008a; Dai *et al.*, 2001; He and Beighley, 2008; Wang and Sassa, 2006), volcanic eruption (Cevik and Topal, 2003; Dahal *et al.*, 2008a; Wang and Sassa, 2006), water level change (Wieczorek, 1996), stream erosion of slope bases (Ohlmacher and Davis, 2003) and snowmelt (Wang and Sassa, 2006; Wieczorek, 1996). Anthropogenic triggers include deforestation (He and Beighley, 2008), improper construction or placement of fill, steeply cut slopes, and poorly controlled surface drainage (Ohlmacher and Davis, 2003).

Table 2.1 outlines the most commonly identified triggering factors in a range of literature. These have been itemised according to whether they have been discussed or briefly mentioned in the text. The rank identifies which triggering factor based on all observations is most commonly discussed in the literature.

The most commonly acknowledged triggering factor is heavy rainfall (Table 2.1) (Can *et al.*, 2005; Lee and Sambath, 2006; Santacana *et al.*, 2003; Wang and Sassa, 2006) which when infiltrated rapidly, increases pore water pressure resulting in a loss in shear strength which initiates the landslide (Wang and Sassa, 2006; Wieczorek, 1996; Wilson, 2004). In seismically active areas earthquakes can be a significant factor in triggering landslides (Yesilnacar and Topal, 2005). The strength of a slope can be reduced as a result of sesimic loading (Wieczorek, 1996). In loose, saturated, cohesionless soils ground shaking can raise the pore water strength and reduce the soil strength which can lead to landsliding by earthquake-induced liquefaction (Wieczorek, 1996).

Table 2.1 Count of extrinsic or triggering variables discussed (C_d), briefly mentioned (C_m), and the total count (C_T) in the literature (compiled from 128 journals articles; Appendix 2.1). A rank of those most commonly discussed in some context (R_T) is also given.

Extrinsic Variable		C_d	C_m	C_T	R_T
Natural	Heavy Rainfall	87	13	100	1
	Earthquake shaking	42	18	60	2
	Water level change	11	6	17	4
	Snowmelt	6	9	15	5
	Rapid stream erosion	5	6	11	6
	Volcanic eruption	6	1	7	8
	Storm waves	2	1	3	10
Anthropogenic	In general	17	6	23	3
	Slope excavation	7	4	11	6
	Deforestation	4	3	7	8

In addition to being site specific, extrinsic variables possess temporal distribution (Dahal *et al.,* 2008a) and as such, they are difficult to estimate as they may change over a short time span (Dai *et al.,* 2001) and often there is insufficient information regarding their spatial distribution (Dahal *et al.,* 2008a).

2.4.2 Intrinsic (Causative) Variables

Within a region, the spatial distribution of landslide susceptibility is determined by the spatial distribution of its intrinsic variables (Dai *et al.,* 2001). Influencing intrinsic variables (Table 2.2) can vary depending on the characteristics of the study area and there is currently no standard to limit the number of independent variables in a landslide susceptibility assessment, and nor should there be (Ayalew and Yamagishi, 2005).

Table 2.2 lists the frequency and rank of a series of identified variables commonly used and/or discussed in the literature as well as the total count and rank. Not all variables discussed in the literature are used, some are simply acknowledged in other studies, or may be applied but later rejected. A rank was taken for both those used to get an indication of what has successfully been used in susceptibility assessments, and total count of those used, discussed, and briefly mentioned to get some insight into what other considerations are important but may not have been applied for various reasons.

The variables selected in this assessment (in order of descending rank) are: slope angle, lithology (or geology), land cover (or land use), aspect, elevation, distance from waterways (drainage lines or streams), distance from fault lines, distance from roads, rainfall, and soil. Of the selected variables, the first nine ranked in the ten most used variables in the literature observed. Variables not included in this study were excluded on the basis of availability, quality, completeness, and scale of the data, having minor to no relevance for this particular study, or were too costly or time consuming to obtain data for. The chosen variables used will be discussed in turn.

Table 2.2 Count of intrinsic (or causative) variables used (C_u), discussed (C_d), and mentioned (C_m) in the literature (compiled from 146 journal articles; Appendix 2.1) and the total count (C_T) is given. The rank of those most commonly used (R_u) and those that most commonly discussed in some context (R_T) in other landslide susceptibility or hazard assessments is also given.

Intrinsic Variables		C_u	R_u	C_d	C_m	C_T	R_T
Topography	Slope Angle/ Gradient	98	**1**	16	13	127	**2**
	Slope Aspect	68	**4**	14	13	95	**4**
	Slope Elevation	48	**5**	10	6	64	**5**
	Slope Curvature	39	6	6	9	54	6
	Topography	18	11	13	12	43	9
	Relief/ Terrain	16	15	8	7	31	13
	Geomorphology	12	18	15	7	34	11
Geology	Lithology	81	**2**	21	29	131	**1**
Seismic	Distance to/from faults	28	8	9	5	42	10
	Peak Ground Acceleration	5	21	5	8	18	21
Soil	Soil Depth	18	11	2	5	25	15
	Soil Drainage	17	14	8	8	33	12
	Soil Type	13	16	4	4	21	18
	Soil Material	12	18	8	10	30	14
	Soil Texture	12	18	0	0	12	22
Land cover	Land Use	73	**3**	17	13	103	**3**
Forest	Tree/Timber specific	13	16	1	5	19	20
Hydrology	Distance to/from waterways	34	7	9	5	48	8
	Rainfall	19	10	20	13	52	7
	Ground water	4	22	11	5	20	19
Anthropogenic	Infrastructure	4	22	3	3	10	23
	Distance to/from roads	20	9	3	1	24	17
Other	Undetermined lineaments	18	11	2	5	25	15
	Various others	31	n/a	1	7	39	n/a

2.4.2.1 Slope Angle

Slope angle is considered to be the most significant causal variable related to landsliding, and as a result is the most important and commonly used factor in landslide susceptibility and hazard assessments (Ayalew and Yamagishi, 2005; Cevik and Topal, 2003; Dai *et al.*, 2001; Pachauri and Pant, 1992). The generalisation is that with increasing slope there is an increased probability of landslide occurrence (Dai and Lee, 2002; He and Beighley, 2008; Lee and Choi, 2004; Pachauri and Pant, 1992). With steeper slopes there is a greater vertical

component of gravity (Donati and Turrini, 2002) which results in increased gravitational-induced shear stress in the soil or other unconsolidated material in the slope (Dai *et al.,* 2001; He and Beighley, 2008; Lee and Choi, 2004; Lee and Sambath, 2006; Lee and Talib, 2005). Slopes of low gradients are generally associated with lower shear stresses and as such, gentle slopes are expected to have a low frequency of landslides (Lee and Sambath, 2006; Lee and Talib, 2005). There is an exception to this generalisation as some steeper natural slopes, such as those resulting from outcropping bedrock, may not be so susceptible to landslides (Lee and Sambath, 2006; Lee and Talib, 2005).

2.4.2.2 Lithology (or Geology)

Lithology is one of the most frequently used variables in the literature for landslide susceptibility and/or hazard analysis (Table 2.2). Lithology has a fundamental influence on a landscape's geomorphology (Dai *et al.,* 2001) and as lithological units vary, so do their susceptibilities to active geomorphological processes such as landslides (Cevik and Topal, 2003). Lithology affects both the rock mass shear strength and permeability (Donati and Turrini, 2002).

2.4.2.3 Land Cover (Land Use)

Land cover or land use is considered by some to be one of the main factors responsible for landslide occurrence (Dahal *et al.,* 2008a), and as such is utilised as an indirect indication of slope stability (Cevik and Topal, 2003). Land cover/land use data is most commonly classified from Landsat TM satellite imagery (Lee and Choi, 2004; Lee *et al.,* 2003a; Lee and Sambath, 2006; Lee and Talib, 2005).

Theoretically, there are a number of ways in which land cover can affect the susceptibility to slope failure (Donati and Turrini, 2002). The first is that the degree of land erosion may be more or less dependent on the extent and type of vegetation (Donati and Turrini, 2002). The general observation when comparing more vegetated areas (such as forests) to areas which are barren or sparsely vegetated, is that the less vegetated areas are more prone to landslides as they exhibit faster rates of erosion and greater instabilities (Carrara *et al.,* 1991; Cevik

and Topal, 2003; Dahal *et al.*, 2008a). Active slope failures or severe erosion related to mass movement are often the causative drivers responsible for barren land in several morphological environments (Carrara *et al.*, 1991). The second influence of land cover on landsliding is that plant roots increase the shear strength of the land for the depth of the roots (Donati and Turrini, 2002). Trees (especially those of a woody type) with large, strong root systems acting as a natural anchorage, have been found to help improve slope stability (Dahal *et al.*, 2008a; Dai and Lee, 2002). Thirdly, vegetation intercepts meteoric water, decreasing the infiltration rate (Donati and Turrini, 2002).

2.4.2.4 Aspect

Aspect refers to the direction of maximum slope of the ground surface (Dahal *et al.*, 2008a) or more basically, the direction the slope faces (Dai *et al.*, 2001). In some landslide susceptibility studies aspect is considered an important factor (Cevik and Topal, 2003). The influence of aspect could be as a result of the influence of aspect-related physical factors such as the number of sunshine hours (Park and Chi, 2008), exposure to drying winds and sunlight (Dai *et al.*, 2001; Cevik and Topal, 2003), and the degree of saturation as a result of rainfall (Cevik and Topal, 2003; Dai and Lee, 2002; Dai *et al.*, 2001). Aspect can influence vegetation and moisture retention which impact on soil strength and hence landslide susceptibility (Dai and Lee, 2002). Despite many investigations into the relationship between aspect and mass-movement there is still a lack of general agreement on the role of aspect and its importance (Carrarra *et al.*, 1991).

2.4.2.5 Elevation

Areas of relative relief are portrayed in an elevation map, and as landslides may develop in certain relief ranges, elevation is frequently used in landslide susceptibility studies (Cevik and Topal, 2003). A map of elevation can be obtained using GIS to create a digital elevation model (DEM) from a triangulated irregular network (TIN) derived from contour lines digitised from a topographic map (Dai *et al.*, 2001; Lee and Talib, 2005; Yesilnacar and Topal, 2005). The use of digital elevation models in landslide assessment has become common (Wang and Sassa, 2006) as DEMs have become increasingly available (Carrara and Pike,

2008) and act as a means to derive secondary geomorphological parameters for use in analyses (Ayalew and Yamagishi 2005). These secondary topographical parameters include: slope, aspect (Dai *et al.*, 2001; Ercanoglu *et al.*, 2008; Lee and Choi, 2004; Lee *et al.*, 2003a; Lee and Talib, 2005; Santacana *et al.*, 2003; Wang and Sassa, 2006; Yesilnacar and Topal, 2005), curvature (Lee and Choi, 2004; Lee *et al.*, 2003a; Lee and Talib, 2005; Santacana *et al.*, 2003), transverse and longitudinal curvatures (Santacana *et al.*, 2003), slope length, and surface area ratio (Yesilnacar and Topal, 2005), in addition to elevation (Dai *et al.*, 2001; Ercanoglu *et al.*, 2008; Wang and Sassa, 2006).

2.4.2.6 Distance from Waterways

The distance from drainage lines or streams is used to investigate the effects of regional geomorphology and localised processes on landslide occurrence (He and Beighley, 2008). Proximity to streams can be an important factor whether landslides occur next to streams (Cevik and Topal, 2003), or as a result of localised processes such as terrain modified by gully erosion (Dai and Lee, 2001, 2002; Dai *et al.*, 2001), stream flow undercutting the banks (Donati and Turrini, 2002; Saha *et al.*, 2002; van Westen *et al.*, 2003), or headward or backward stream channel erosion initiating slope failure (He and Beighley, 2008). The common observation is that with increased distance from drainage lines or streams there is generally a decrease in landslide frequency (Arora *et al.*, 2004; Dai and Lee, 2001, 2002; Dai *et al.*, 2001; Lee and Sambath, 2006; Lee and Talib, 2005). Drainage lines or streams are obtained from a topographic database or map, and buffer zones of selected distances created around them (Abdallah *et al.*, 2005; Dai *et al.*, 2001; Lee and Talib, 2005). Intervals of 1 m (Lee, 2007a), 50 m (Dai *et al.*, 2001) and 100 m (Lee and Lee, 2006; Lee and Talib, 2005) have been used in the literature for determination of buffer zones.

2.4.2.7 Distance from Fault Lines

Proximity to faults is considered in some landslide susceptibility and hazard assessments where seismic or tectonic activity may influence landsliding (Cevik and Topal, 2003; He and Beighley, 2008; Pachauri and Pant, 1992; Saha *et al.*, 2002). Faults reduce the strength of the rock mass by breaking or shearing the

rock (Donati and Turrini, 2002; Saha *et al.*, 2002). A generally accepted observation is that in seismically affected areas an inverse relationship exists between distance to faults and landslide distribution, where the number of landslides decreases with increased distance from a fault (Cevik and Topal, 2003; Pachauri and Pant, 1992; Saha *et al.*, 2002). To assess the effect of proximity to a fault line on landslide distribution, classes made up of a number of buffer zones of selected distances from the fault must first be created (Cevik and Topal, 2003). Cevik and Topal (2003) found that while previous studies indicate that within 250 – 1,000 m distance of a fault line more landslides occur, this did not apply to all areas as their investigation showed most landslides (about 63%) occurred beyond 1,000 m of a fault. They decided to exclude faults as (1) a meaningful relationship between landslides and faults could not be extracted as the study area was too small; and (2) some faults could not be detected due to burial by recent deposits.

2.4.2.8 Distance from Roads

Road cuts, road drainage, and road construction activity which alter the natural terrain and drainage system, can trigger slope instability (Ayalew and Yamagishi, 2005; Dahal *et al.*, 2008a; He and Beighley, 2008). Depending on the road segment and its location, it may act as a source of landslides, or alternatively, it could act as a barrier, or a passage for water flow (Ayalew and Yamagishi, 2005; Sharma and Kumar, 2008). Dahal *et al.* (2008a) considered landsliding may be more frequent along roads, and therefore it may be sensible to investigate the effects of roads in landslide susceptibility assessment (He and Beighley, 2008). While some slope failures may start above roads, the roads often intercept them (Ayalew and Yamagishi, 2005), thus preventing further downslope damage.

2.4.2.9 Rainfall

Precipitation can be a significant factor in slope instability, as areas of higher rainfall undergo greater soil saturation and have an increased likelihood of landsliding (He and Beighley, 2008). The amount of rainfall a slope receives can vary depending on the aspect (Garcia-Rodriguez *et al.*, 2008) and elevation of the slope (He and Beighley, 2008).

Spatial and temporal rainfall data, such as intensity, duration, and frequency of rainfall, can be collected and the relationship between rainfall and landslide occurrence investigated (Wang and Sassa, 2006). Including temporal data, in a strict sense, would transform a landslide susceptibility assessment into a landslide hazard assessment as a time dependent factor is being introduced. In landslide hazard assessment, the temporal aspect of rainfall-induced hazard tends to be unclear unless analyses considering frequency, intensity and duration of rainfall have been implemented (Wang and Sassa, 2006). While in landslide susceptibility assessments, mean annual rainfall (Can *et al.,* 2005; Fernandez *et al.,* 1999) and/or monthly average rainfall (Can *et al.,* 2005) have generally been included as part of the spatial investigation, without much consideration for time.

2.4.2.10 Soil

The soil factors chosen in landslide susceptibility or hazard assessment are somewhat dependent on the method taken. For site-specific studies employing physically based models which involve factors of safety, soil cohesion is a crucial component (Gorsevski *et al.,* 2006; Ohlmacher, 2007; Wu and Sidle, 1995). Other factors such as moisture content, bulk density (Gorsevski *et al.,* 2006; Ohlmacher, 2007; Wu and Sidle, 1995), unit weight (Ohlmacher, 2007; Wu and Sidle, 1995), and soil depth (Gorsevski *et al.,* 2006; Wu and Sidle, 1995) are also considered in physically based models. In regional assessments however, soil factors considered typically include: soil type (Lee, 2007a), material, texture (Lee and Choi, 2004; Lee *et al.,* 2002; Lee and Lee, 2006), drainage, depth or effective thickness (Lee and Choi, 2004; Lee *et al.,* 2002).

Soil material relates to the slope (Lee *et al.,* 2002), topography (Lee and Choi, 2004; Lee *et al.,* 2002; Lee and Lee, 2006), and geology (Lee and Choi, 2004; Lee and Lee, 2006). In some studies, colluvium was found to have a low probability of landslide occurrence as the colluvium had already collapsed (Lee and Choi, 2004; Lee *et al.,* 2002). Granite residuum (Lee and Choi, 2004; Lee and Lee, 2006) and acidic rock residuum (Lee *et al.,* 2002) however, were found to have a greater chance of landsliding. Soil texture is considered to be related to landsliding as larger grains have larger spaces between the grains which allows the soil to hold more water following heavy rainfall, which increases the probability of landslide

occurrence (Lee and Choi, 2004; Lee *et al.*, 2002, 2004a; Lee and Lee, 2006). The increased volume of water in some soils such as sandy soils leads to instability when pore water pressures develop (Abdallah *et al.*, 2005). Drainage impacts landslide susceptibility, but the relationship is complex and often site specific

2.5 Methods for Landslide Susceptibility Assessment

Methods for landslide susceptibility and hazard assessment can be divided into two groups (Figure 2.1): qualitative and quantitative (Guzzetti *et al.*, 1999; He and Beighley, 2008). Qualitative methods involve subjectivity and result in a susceptibility or hazard zoning map depicted in descriptive (qualitative) terms (Dai and Lee, 2002; Guzzetti *et al.*, 1999; Yesilnacar and Topal, 2005), whereas quantitative methods determine the numeric probability of landslide occurrence in any hazard area (Guzzetti *et al.*, 1999; Yesilnacar and Topal, 2005). Within each of these broad methods there are a wide range of approaches to landslide susceptibility assessment, which will be discussed in this section.

All methods of assessing landslide susceptibility or hazard have advantages and disadvantages (Carrara *et al.*, 1991; He and Beighley, 2008). Some methods may be suited to meet the requirements of a specific task or to solve slope-instability problems in a particular area (Carrara *et al.*, 1991). Many factors strongly influence whether one method or another is implemented successfully, these include data availability, mapping scales and accuracy of expected results (He and Beighley, 2008).

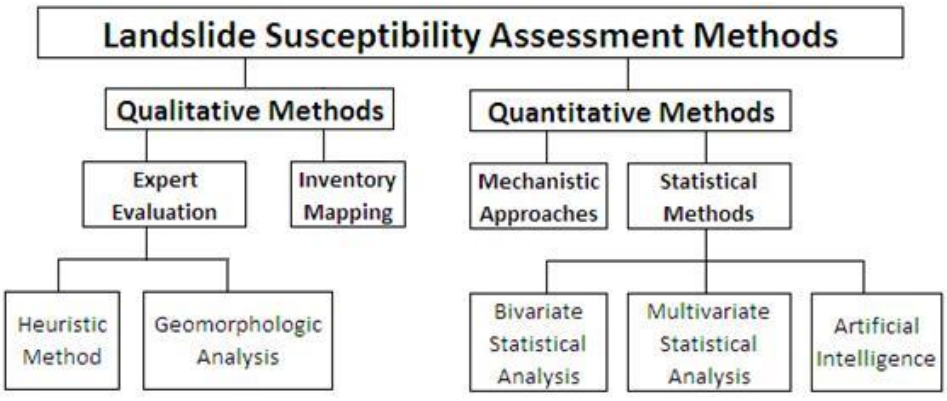


Figure 2.1. An adaptation of the classification of landslide susceptibility assessment methods proposed by He and Beighley (2008).

2.5.1 Qualitative Methods

Qualitative methods rely on expert opinions of the experienced geoscientists involved (Ayalew and Yamagishi, 2005; Dai and Lee, 2002; Ermini *et al.*, 2005) and as such, are limited by subjective decision rules (Dai and Lee, 2002). Several qualitative approaches include ranking and weighting, which may end up evolving to a semi-quantitative nature (Ayalew and Yamagishi, 2005). Of the qualitative methods, the two most common approaches are expert evaluation and landslide inventory mapping (He and Beighley, 2008).

2.5.1.1 Expert Evaluation

Expert evaluation includes heuristic methods and geomorphologic analysis (He and Beighley, 2008; Süzen and Doyuran, 2004). The advantage of expert evaluation is that it can be successfully implemented at any scale, while the disadvantage is the subjectivity involved in the decision making, lengthy field surveys, and the requirement of long term information (He and Beighley, 2008). However, some authors are of the view that when based on an expert's opinion, subjectivity is not necessarily bad (Ercanoglu *et al.*, 2008).

2.5.1.1.1 Heuristic Approach

Heuristic approaches are based on the expert opinions of a geomorphologist to estimate the potential occurrence of a landslide from data on a set of intrinsic variables (relating to the landform etc.), based on the assumption that the relationship between the intrinsic variable and landslide occurrence are known and can be specified by models (Dahal *et al.*, 2008a; Dai *et al.*, 2001; Ermini *et al.*, 2005; He and Beighley, 2008). Instability factors are ranked and weighted (Guzzetti *et al.*, 1999) or rated by the derivation of scores (Ermini *et al.*, 2005) in accordance with the assumed or expected role played by each in controlling the development of mass movements (Ermini *et al.*, 2005; Guzzetti *et al.*, 1999).

Heuristic models require long term information on both landslides and the causal factors for the same site or for a similar geo-environment but, in most cases these are unavailable (Dai *et al.*, 2001). The disadvantage of the heuristic approach is that weighting and ratings are assigned subjectively to the variables and that the

reproducibility of results is limited (Dai *et al.,* 2001). Heuristic methods are also limited in their reliability, which largely depends on the investigator and their knowledge and understanding of the geomorphological processes that may be impacting the terrain (Guzzetti *et al.,* 1999). An experience, knowledge and judgement-driven heuristic approach is generally considered to be too subjective and as such is rarely used now (Ercanoglu *et al.,* 2008).

2.5.1.1.2 Geomorphologic Analysis

Possibly the simplest of the qualitative methods is geomorphologic analysis which was frequently used in the 1970s and 1980s (He and Beighley, 2008). The estimation of actual and potential slope failures in geomorphologic analysis relies on the ability of the investigator (Guzzetti *et al.,* 1999). For qualified scientists, this method can be done in the field rapidly, basing their experience on similar situations (He and Beighley, 2008).

2.5.1.2 Landslide Inventory

Most commonly in qualitative methods, areas susceptible to landsliding are identified based on sites of similar geomorphological and geological properties identified using a landslide inventory (Ayalew and Yamagishi, 2005). A landslide inventory map in its simplest form indicates the location and extent of evident landslides (Cevik and Topal, 2003; Dahal *et al.,* 2008a; He and Beighley, 2008). More elaborate landslide inventories not only portray the spatial distribution of past mass movements, but also the type, activity (Guzzetti *et al.,* 1999; Wang and Sassa, 2006), and age (Guzzetti *et al.,* 1999). A landslide inventory map is achieved through direct landslide mapping of discernable landslides (Cevik and Topal, 2003; Dahal *et al.,* 2008a). Identification of historic landslide events generally involves some combination of field surveys (Lee *et al.,* 2003a, 2006; Lee and Talib, 2005; Yesilnacar and Topal, 2005), interpretation of aerial photographs (He and Beighley, 2008; Lee *et al.,* 2003a; Lee and Talib, 2005), satellite imagery interpretation (He and Beighley, 2008; Lee *et al.,* 2006), and/or previous inventory maps (Yesilnacar and Topal, 2005). Landslide inventory mapping is difficult, subject to uncertainties and prone to error (Guzzetti *et al.,* 1999), so it seems the quality of the landslide inventory depends on time, the

interpreter's experience, and scale of both the inventory and aerial photographs (Guzzetti *et al.*, 1999).

Landslide inventories serve only to indicate locations of past events and are thus limited in their use to effectively assess the location and frequency of future landslides; however these are a useful component in landslide susceptibility assessment (Weirich and Blesius, 2007). For landslide susceptibility or hazard assessment it is important to first obtain and compile a landslide inventory map of the area (Guzzetti *et al.*, 1999; Lee and Talib, 2005; Yesilnacar and Topal, 2005). In landslide susceptibility assessments the landslide inventory is used in the analysis of landslide occurrence in relation to a set of environmental conditions (Wang and Sassa, 2006). In statistical analysis, the landslide inventory is used as the dependent variable, with presence of landslides being assigned the value of 1, and 0 assigned for absence of landslides (Ayalew and Yamagishi, 2005). The landslide inventory is also used in the evaluation of landslide susceptibility assessment outputs to validate (if data is split) or determine a goodness of fit (if using the same dataset) (Lee and Sambath, 2006; Remondo *et al.*, 2003).

2.5.2 Quantitative Methods

Quantitative methods cover both the mechanistic (physically based) approaches involving deterministic or probabilistic models, and a wide range of statistical approaches (Table 2.3) (Ayalew and Yamagishi, 2005; He and Beighley, 2008). Both statistical and mechanistic approaches aim to determine where probable landslides would be expected based on the observed relationships between the presence of landslides and a set of associated physical factors which lead to its occurrence (Ermini *et al.*, 2005; Guzzetti *et al.*, 1999; Neuhäuser and Terhorst, 2007).

Table 2.3 gives the frequency (count) of quantitative approaches applied, discussed but not applied, and those briefly mentioned in the observed literature. The quantitative approaches can be statistical or mechanistic, and within these there are a range of different analyses and these can include a variety of different approaches.

Statistical approaches are more commonly used (Table 2.3) as they are objective, reproducible and easy to update. Mechanistic methods are not as commonly used (Table 2.3), possibly as they are limited to smaller areas and require more intensive data collection which can be costly.

Table 2.3 The count of literature that have been used (C_u), discussed (C_d), and mentioned (C_m) quantitative approaches (compiled from 142 journal articles; Appendix 2.1), and a total count (C_T).

Quantitative Method	Type of Analysis	Approach	C_u	C_d	C_m	C_T
Statistical	In general		75	15	22	112
	Bivariate	In general	31	13	12	56
	Multivariate	In general	25	25	13	63
		Discriminant Analysis	5	7	5	17
		Logistic Regression	29	11	7	47
		Other	3	1	0	4
	Artificial Intelligence	Artificial Neural Network	16	17	6	39
		Fuzzy Logic	6	16	4	26
Mechanistic	In general		1	11	6	18
	Monte Carlo		5	2	2	9
	First-Order-Second-Moment		0	1	1	2
	Deterministic		2	19	12	33
	Geotechnical		1	11	4	16

2.5.2.1 Mechanistic (Physically Based)

Physically based approaches or mechanistic (process-driven) methods aim to evaluate and analyse slope stability in one, two or three dimensional deterministic models in order to obtain insight into the causal and triggering factors of landslides (Guzzetti *et al.*, 1999; He and Beighley, 2008). Mechanistic approaches have several limitations, such as the use of overly simple models to evaluate instability, inability to collect geotechnical data at a reasonable cost (even for small regions), and the geotechnical factors having uncontrolled spatial variability (Guzzetti *et al.*, 1999). Physically based models are limited to small areas (Guzzetti *et al.*, 1999), and thus are commonly used in slope-specific soil engineering studies (He and Beighley, 2008).

The deterministic or geotechnical method is an engineering approach which utilises slope instability analyses to evaluate a factor of safety based on mathematical models of physical mechanisms which control slope failure (Ayalew and Yamagishi, 2005; Ermini *et al.*, 2005; Ohlmacher and Davis, 2003).

Landslide susceptibility by deterministic approaches is mostly applied to translational slides (Soeters and Van Westen, 1996), and have widely been assessed through the use of the infinite slope stability model (Dahal *et al.,* 2008a; Dai *et al.,* 2001; Soeters and Van Westen, 1996), which is a static model that considers the local equilibrium along a potential slip surface (Guzzetti *et al.,* 1999).

Deterministic methods are only applicable if geomorphic and geologic ground conditions are relatively uniform across the study area and landslide types are known and simple (Dahal *et al.,* 2008a; Dai and Lee, 2001, 2002; Soeters and Van Westen, 1996). These methods require correct knowledge of failure mechanisms, detailed geotechnical and hydrological data for individual slopes (Ermini *et al.,* 2005), and often require groundwater models (Soeters and Van Westen, 1996). Slope specific geotechnical and hydrological data is often difficult to acquire over wide areas (Ermini *et al.,* 2005; Ohlmacher and Davis, 2003), and in addition to the other limitations, deterministic approaches can only be effectively applied to small areas (Ayalew and Yamagishi, 2005; Dai *et al.,* 2001; Ermini *et al.,* 2005; Weirich and Blesius, 2007). Despite the problems involved in collecting sufficient and reliable input data, the use of deterministic models in hazard analysis of large areas has increased with GIS techniques that aid in handling more calculations to obtain factors of safety (Soeters and Van Westen, 1996). The main disadvantage of deterministic models is the high degree of over simplification (Dai and Lee, 2002; Dai *et al.,* 2001; Soeters and Van Westen, 1996). The advantage however, is that deterministic approaches permit the quantitative calculation of factors of safety or values of stability (Dai and Lee, 2002; Dai *et al.,* 2001; Soeters and Van Westen, 1996).

2.5.2.2 Statistical Approaches

Statistical approaches involve the analysis of a combination of factors that have led to past landslides (Ayalew and Yamagishi, 2005; Dahal *et al.,* 2008a; Dai and Lee, 2002; Dai *et al.,* 2001; Guzzetti *et al.,* 1999; He and Beighley, 2008; Ohlmacher and Davis, 2003) and the application of quantitative predictions to areas where landslides are currently absent but similar conditions exist (Soeters and Van Westen, 1996). Statistical methods were developed as a less subjective

alternative to the highly subjective expert evaluation techniques (Ermini *et al.*, 2005; He and Beighley, 2008; Thiery *et al.*, 2007). Statistical approaches are now considered the most appropriate techniques for landslide susceptibility assessments (Dai *et al.*, 2001; Ermini *et al.*, 2005) as they are objective, easily updated, and reproducible (He and Beighley, 2008). Variables of a continuous (numerical) or discrete (nominal) nature that lead to the initiation of a landslide can be used in statistical analysis (Ermini *et al.*, 2005). The two overriding drawbacks with statistical methods are that expert or personal opinion is used to select the predictor factors; and data over large spatial and temporal extents are required (He and Beighley, 2008). The data requirement can be problematic, as multisource data whether environmental or otherwise cannot be handled by most statistical approaches which are distribution based (Yesilnacar and Topal, 2005). If applied over large areas, an extensive effort is required to collect and validate data (Weirich and Blesius, 2007).

Bivariate and multivariate statistical methods, and artificial intelligence models, have been widely used throughout the literature for landslide susceptibility mapping (Ayalew and Yamagishi, 2005; Dahal *et al.*, 2008a; Süzen and Doyuran, 2004). A wide range of statistical approaches have been used throughout the literature, and various authors have different opinions on which are most commonly used. Regression based methods such as logistic regression (Guzzetti *et al.*, 1999; Park and Chi, 2008; Weirich and Blesius, 2007) and multiple logistic regression (Neuhäuser and Terhorst, 2007) appear to be the most prevalent (Table 2.3). The other two frequently used statistical approaches are artificial neural networks (Guzzetti *et al.*, 1999; Weirich and Blesius, 2007) and discriminant analysis (Guzzetti *et al.*, 1999; Neuhäuser and Terhorst, 2007). The simplest and most used statistical predictions (such as multiple regression, discriminant analysis, and logistic regression) use a contingency table analysis, which cross-tabulates the one or more predictor variables with the two possible outcomes (landslide or no landslide) (Ohlmacher and Davis, 2003).

2.5.2.2.1 Bivariate Statistical Analysis

Bivariate statistical analysis involves the combination and analysis of each parameter map (for example, geology, slope, land use) with the landslide

inventory map to determine weighting values or rank according to the landslide densities calculated for each parameter class (Ayalew and Yamagishi, 2005; Soeters and Van Westen, 1996; Süzen and Doyuran, 2004). Bivariate approaches are considered to be robust and flexible methods, but they have several limitations (Thiery *et al.*, 2007). Limitations can include a loss of data quality and accuracy with oversimplification of input thematic data, as well as a loss of data sensitivity in forced individual analysis of causative factors (Thiery *et al.*, 2007).

To apply bivariate analysis continuous factor maps have to be converted to categorical (discrete) maps before responsible weights for each class can be computed (Süzen and Doyuran, 2004). The conversion of continuous factor values to discrete tends to rely on expert opinion in setting the class boundaries, such as intervals of 100 m for elevation (Süzen and Doyuran, 2004). A range of different statistical methods can be applied to calculate the weighting values; these include weights of evidence method, landslide susceptibility method, information value method (Soeters and Van Westen, 1996; Süzen and Doyuran, 2004), Bayesian probability models, and certainty factors (Soeters and Van Westen, 1996).

2.5.2.2.2 Multivariate Statistical Analysis

Multivariate statistical analyses determine the weights of landslide causal factors based on the relative contribution of each in the presence or absence of past landslide events within a defined land unit (Donati and Turrini, 2002; Santacana *et al.*, 2003; Süzen and Doyuran, 2004; Yesilnacar and Topal, 2005). The assumption is that in an area of known landslide occurrence, the landslide occurrence (dependent variable) is related to the landslide causal factors (independent variables) generally in a log-linear way (Park and Chi, 2008). The output for each cell should take on a value as a function of the contributing factors present in it (Santacana *et al.*, 2003). The two most commonly implemented multivariate statistical approaches are logistic regression and discriminant analysis (Table 2.3) (Ayalew and Yamagishi, 2005; Can *et al.*, 2005).

Discriminant analysis classes observations into two mutually exclusive possible outcomes, either a landslide has occurred, or has not occurred, and creates coefficients to express the difference between the two outcomes (Ohlmacher and Davis, 2003). Logistic regression requires fewer assumptions than discriminant

analysis and attempts to find the best fitting function that describes the relationship between the set of predictor variables and the dependent variable (Ayalew and Yamagishi, 2005). In logistic regression, the dependent variable can have only two values (occurrence or non-occurrence) and as the predicted values are constrained between 0 and 1 they can be interpreted as probability (Dai and Lee, 2002; Dai *et al.,* 2001). Variables included in logistic regression can be continuous, discrete, or a combination of both and are not necessarily normally distributed, unlike in discriminant analysis, where variables must have a normal distribution (Lee *et al.,* 2006; Lee and Sambath, 2006).

2.5.2.2.3 Artificial Intelligence

Artificial intelligence or data mining techniques such as artificial neural network methods (simulating human thinking) and fuzzy logic have increasingly been applied since the late 1990s (Ercanoglu *et al.,* 2008; Lee, 2007a; Lee and Dan, 2005; Lee and Lee, 2006; Lee and Sambath, 2006 Lee and Talib, 2005). Artificial neural networks are computational mechanisms that can organise and correlate information allowing a different view of complex, poorly understood and/or resource intensive problems that other statistical methods cannot address due to their theoretical limitations (Ermini *et al.,* 2005; He and Beighley, 2008; Lee *et al.,* 2003a, 2006).

2.6 Methods Chosen for this Study

Methods chosen for the modelling landslide susceptibility within the Waikato Region are all of a statistical nature. This is on the basis that they are objective, easily updateable, reproducible (He and Beighley, 2008), and are the most appropriate as discussed earlier (section 2.5.2.2). In this study both continuous and categorical variables have been considered in the analysis. This is problematic in that some statistical approaches cannot handle both types in the same analysis. Logistic regression is an exception as it can handle both types and does not require the data to be normally distributed. For this reason it has been chosen as the multivariate approach in this study. The bivariate approach chosen in this study is the weights of evidence method which determines weights calculated using prior probabilities and is readily implemented using categorical data.

2.7 Conclusion

Landslide susceptibility assessments focus on spatial probability, while landslide hazard assessments in a technical sense produce a probability based on space, time and magnitude. Hazard assessments include both intrinsic (causative) and extrinsic (triggering) variables, whereas susceptibility assessments generally use explanatory causative variables. Time and magnitude are not accounted for in this assessment due to the immense difficulty in determining sensible values for these two factors, landslide susceptibility and not hazard is investigated here.

Landslide susceptibility assessments have increasingly been fully conducted and analysed in GIS or partially processed in GIS with some external statistical package and output displayed using GIS.

A wide range of quantitative and qualitative methods and predictor variables have been investigated and applied to landslide susceptibility assessment. Each method has its advantages and disadvantages, and some are more appropriate in a given situation than others. Statistical approaches were found to be the most commonly used as they are reasonably straightforward to implement and the results can easily be updated (He and Beighley, 2008). In a landslide susceptibility assessment at a regional scale, statistical methods may be the most appropriate, so in this assessment these have been considered the most appropriate for assessing the landslide susceptibility for the Waikato Region.

The most widely utilised methods are the statistical-based ones such as logistic regression, discriminant analysis, weights of evidence, and artificial neural networks. The most commonly used predictor variables are slope, geology and land cover or land use. Some of the statistical approaches such as logistic regression and artificial neural networks can deal with both continuous and discrete data, while bivariate techniques such as weights of evidence require the data to be in discrete (categorical) form. Therefore, in this study weights of evidence and logistic regression have been chosen as the methods of analysis.

CHAPTER THREE

Input Factors

3.1 Introduction

This chapter outlines the collection and manipulation of spatial data in the production of a landslide inventory map, a set of causative factor maps, and the compilation of a database of all input factor data for later statistical analyses. The simplification of categorical data into fewer classes for both statistical approaches is discussed. The grouping of continuous data into classes as a requirement of the weights of evidence approach is discussed with consideration to class types, intervals and the number used in other studies.

Eleven physical parameters were chosen as potentially contributing to landslide susceptibility, these are: slope, elevation, aspect, geology, soil order, mean monthly rainfall, maximum monthly rainfall, land use (land cover), distance from faults, distance from roads, and distance from rivers. A 25 m resolution digital elevation model (DEM) and spatial data of geology, soil, land cover, roads, faults, rivers, land resource inventory, monthly maximum and monthly mean total rainfall, and landslide location data were obtained from various sources (Table 3.1).

A set of scripts (Appendix 3.1) were created to execute automated variable specific manipulations in GIS. For polygon vector data this included the simplification of complex data into fewer classes. Linear vector data were buffered to obtain polygon bands around the linear feature and were assigned a distance value for each polygon. The digital elevation model (DEM) was used to extract slope, aspect and elevation, and a continuous copy and categorical copy of each were made. Similarly, continuous and categorical copies of both raster rainfall datasets were made. A landslide inventory was compiled from the GeoNet landslide catalogue and GNS QMap (Auckland and Waikato) landslide spatial

Table 3.1 Source and format of original data. (Edbrooke, 2001b, 2005b; GeoNet, 2009; IGNS, 2000; Landcare, 2000a; Landcare and DoC, 1998; LINZ, 2000a, b; LINZ and Eagle, 2000; MfE, 2004)

Raster Data:	**File Name**	**Format**	**Resolution**	**Extent**	**Copyright and Ownership Details**
Digital Elevation Model (DEM)	north25	grid	25 m	North Island	© Land Information New Zealand (LINZ)
Mean Monthly Rainfall	mean_rain	grid	1 km	New Zealand	© Landcare Research New Zealand Ltd and Department of Conservation (1998)
Maximum Monthly Rainfall	max_rain	grid	1 km	New Zealand	© Landcare Research New Zealand Ltd and Department of Conservation (1998)
Vector Data:			**Scale**		
Landslides - Auckland QMAP	landsl	cover	1:250,000	Topo 260 Auckland	© Institute of Geological and Nuclear Sciences (2001)
Landslides - Waikato QMAP	landsl	cover	1:250,000	Topo 260 Waikato	© Institute of Geological and Nuclear Sciences (2005)
Geology - Auckland QMAP	geol_units	cover	1:250,000	Topo 260 Auckland	© Institute of Geological and Nuclear Sciences (2001)
Geology - Waikato QMAP	geol_units	cover	1:250,000	Topo 260 Waikato	© Institute of Geological and Nuclear Sciences (2005)
Geology - New Zealand	nzgeology	cover	1:250,000	New Zealand	© Institute of Geological and Nuclear Sciences
Faults - Auckland QMAP	faults	cover	1:250,000	Topo 260 Auckland	© Institute of Geological and Nuclear Sciences (2001)
Faults - Waikato QMAP	faults	cover	1:250,000	Topo 260 Waikato	© Institute of Geological and Nuclear Sciences (2005)
Faults - New Zealand	1millfaults	shapefile	1:250,000	New Zealand	© Institute of Geological and Nuclear Sciences
Roads	ni_roads	shapefile	1:50,000	North Island	© Land Information New Zealand (LINZ)
Land Cover Database 2	ni_nzmg	shapefile	1:50,000	North Island	© Ministry for the Environment (2004)
Soil	nzfsl	cover	1:50,000	New Zealand	© Landcare Research New Zealand Ltd (2000)
Rivers	river_cl	cover	1:50,000	New Zealand	© Land Information New Zealand (LINZ) and Eagle Technology Group Ltd. (Eagle)
Database Data:					
Landslide Catalogue	Landslide_Catalogue_ (SCHICKER_GEONET)	Excel		New Zealand	© GeoNet (2009)

data through a series of processes. All spatial datasets were clipped to the Waikato Region and the vector data converted to raster data. As a result eleven parameter maps and a landslide inventory map were created (large versions can be found in Appendix 3.2).

A layer consisting of a grid of 25 m spaced points covering the whole Waikato region was used to extract the values from each raster layer in order to compile all the data into a database. This database was then used to query all categorical variables for the weights of evidence method, and was also imported into STATISTICA for use in logistic regression.

3.2 Landslide Catalogue

A landslide catalogue (GeoNet, 2009) of 2,786 reported national landslide occurrences organised by year (1990, 1996-2008) was sourced from GeoNet (pers. comm. Grant Dellow, 2009). Each year's dataset was converted to a shapefile with New Zealand Map Grid (NZMG) set as the projection and added to ArcMap. The spatial location and locality data of each point was compared with existing boundary spatial data for regional and district councils nationwide. Two new columns were added to the data set and in these the region and district the point fell in was recorded. These were then compared with the recorded locality to check the points were correctly located as a few points in other regions outside of the Waikato had mistyped coordinates. The points with incorrect coordinates either had an incorrect number or not enough numbers in either the Northing or Easting coordinates. These observed point records were reported to GeoNet with suggestions for modification which have since been implemented. Once the location check was completed, each year's dataset was imported into a single Access database (Appendix 3.3). A query was created to select only data for the Waikato region using the added region column. This singled out 123 landslides with the region being identified as Waikato. The results of the query were exported as a text file then converted to a shapefile using ArcCatalog.

3.3 Waikato Region Boundary

A modified Waikato Regional boundary was created using the coastline from a spatial data layer of territorial authority (District and City Councils') boundaries and the inland boundaries from the national regional council boundaries spatial data (Statistics New Zealand, 2002) available on the university server. The boundary adjustments were necessary as the coastline portion of the regional council spatial data was fairly crude in comparison. The new coastline was based on that of six different district councils (Thames-Coromandel, Hauraki, Franklin, Waikato, Otorohanga, and Waitomo) within the territorial authority spatial layer. As the Regional Council boundaries are based on catchments, the territorial authority boundaries do not align with regional boundaries. As a result, some territorial authorities fall in more than one region. For the Waikato Region there are four such territorial authorities (Franklin, Rotorua, Taupo and Waitomo District Councils) which overlap with the surrounding regions (Auckland, Bay of Plenty, Hawkes Bay and Wanganui-Manawatu). Where the territorial boundaries bisected the regional boundary inland, the regional boundary was retained. In addition to these modifications, all offshore islands were excluded as not all the input data covered them.

3.4 Creation and Use of a Series of Scripts for Automated GIS Processes in ArcGIS

Individual scripts were made to set up the data and process each variable (Appendix 3.1). The set up stage involved first checking that the data is in the folder stated by the script, then converting shapefiles to cover or grid so that further processes can be carried out in ARC (an ESRI™ GIS program). A few supporting scripts (Appendix 3.1) were also created. A master script (Appendix 3.1) was made to run all the individual scripts, which allows the user to determine which scripted processes to run. An existing script was used to define projection (Berkowitz, 2004). The purpose and function of each is explained in each file in the digital appendix.

3.4.1 Digital Elevation Model (DEM) Derived Parameters

3.4.1.1 Basis for Selected Digital Elevation Model Resolution

A digital elevation model (DEM) is essentially a digital representation of a contour map, and with higher resolution (smaller pixel size) DEMs the original contour map is more closely reflected (Jibson *et al.,* 2000). A high resolution DEM is ideal to obtain higher quality data for the parameters derived from it (such as slope, aspect and elevation), however with increased resolution there is an increase in file size. Access to high quality DEMs can be also limited (Dikau *et al.,* 1996). The resolution (pixel size) of the DEM is generally the basis for the mapping/grid unit in landslide susceptibility assessments (Ayalew and Yamagishi, 2005; Ayalew *et al.,* 2005; Duman *et al.,* 2006).

DEM resolutions used in the literature vary quite significantly from 1 m LIDAR derived DEM (Van Den Eeckhaut *et al.,* 2006) to 90 m (Lee and Dan, 2005; Neuhäuser and Terhorst, 2007) or even as coarse as 230 m (Guzzetti *et al.,* 1999), but most commonly a 10 m DEM is used (Ayalew and Yamagishi, 2005; Ayalew *et al.,* 2005; Catani *et al.* 2005; Jiménez-Perálvarez *et al.* 2009; Jibson *et al.,* 2000; Lee, 2005, 2007b; Lee *et al.*,2003b; Lee *et al.,* 2006; Lee and Talib, 2005; Nandi and Shakoor, 2009; Song *et al.,* 2008; Van Beek and Van Asch, 2004; Wang and Sassa, 2006). In this study, a DEM with a resolution of 25 m was used. While this may not be as high in quality as the majority of landslide susceptibility assessments, other studies have proven this to be an acceptable scale (Duman *et al.,* 2006; Guzzetti *et al.,* 2000; Kanungo *et al.,* 2006; Nefeslioglu *et al.,* 2008; Yesilnacar and Topal, 2005). At this resolution the pixel count after being clipped to the Waikato Region is quite high (approximately 39 million). A smaller resolution would be difficult to work with as the mapping resolution was set to that of the DEM, making for a large amount of data.

3.4.1.2 Continuous Topographic Data

Previous international landslide susceptibility assessment studies using logistic regression have used continuous topographic data (Duman *et al.,* 2006; Nandi and Shakoor, 2009; Süzen and Doyuran *et al.,* 2004; Van Den Eeckhaut *et al.,* 2006; Yesilnacar and Topal, 2005). Slope, aspect and elevation were all applied as

continuous data in three of those studies (Duman *et al.*, 2006; Süzen and Doyuran *et al.*, 2004; Yesilnacar and Topal, 2005). Logistic regression is one of the chosen methods for this study, so continuous data for slope, elevation and aspect was used for that approach.

3.4.1.3 Categorical Topographic Data

In many landslide susceptibility and hazard studies topographic data has been classed. In this study categorical topographic data for slope, elevation and aspect is required for the weights of evidence analysis, which is a bivariate approach. In the case of topographic data, this does not appear to be limited to bivariate analyses as several studies that have used logistic regression have also used categorical values for slope, elevation, and/or aspect (Dai *et al.*, 2001; Dominguez-Cuesta *et al.*, 2007; Garcia-Rodriguez *et al.*, 2008; Lee and Sambath, 2006; Ohlmacher and Davis, 2003), although the use of continuous data for logistic regression seems more appropriate.

3.4.1.3.1 Slope Classes

Slope is one of the most widely used factors in landslide susceptibility and hazard assessments. There are many variations of classes used and these can be intervals of equal, mainly equal, or mixed sizes. Most studies seem to use intervals of 10°, starting at 0 – 10° (Can *et al.*, 2005; Cevik and Topal, 2003; Dahal *et al.*, 2008a; Dai *et al.*, 2001; Ercanoglu and Gökceoglu, 2004; Ercanoglu *et al.*, 2008; Jibson *et al.*, 2000; Lan *et al.*, 2004; Nagarajan *et al.*, 1998; Ohlmacher and Davis, 2003; van Westen *et al.*, 2003; Wang and Sassa, 2006; Yalcin, 2008) or ≤ 15° (Arora *et al.*, 2004; Thiery *et al.*, 2007; Sakar and Kanungo, 2004; Tangestani, 2004). However, applying intervals of 10° (starting at 0 – 10 °) would not be appropriate for the Waikato Region, as most slopes (approximately 85%) would be classed in the first two classes, so any effects from slopes in this range as a proportion of the class may be insignificant, but at smaller intervals this may not be the case. Intervals of 5° are also quite popular (Clerici *et al.*, 2006; Dominguez-Cuesta *et al.*, 2007; Lee, 2004, 2005, 2007a, b; Lee and Dan, 2005; Lee and Lee, 2006; Lee and Talib, 2005), but as the slope ranges from 0 – 86° in the Waikato Region, applying intervals of 5° would result a very large number of classes, which was

not desirable. This could have been applied with reduced classes if a cut off point were decided, but this would result in the grouping of a wide range of high slopes together as a single class. Another common approach in other landslide susceptibility and hazard assessment studies is to apply intervals of mixed sizes/ranges which were specific to each study (Abdallah *et al.*. 2005; Ayalew and Yamagishi, 2005; Catani *et al.*, 2005; Ermini *et al.*, 2005; Falaschi *et al.*, 2009; He and Beighley, 2008; Lee and Min, 2001; Lee and Sambath, 2006; Lee *et al.*, 2004b). To combat under or over representing both ends of the skewed slope angle distribution, it was decided that mixed sized classes were probably most appropriate for this study.

The classes applied to the Waikato Region (Table 3.2) are 0 – 4 °, > 4 – 8°, > 8 – 14°, > 14 – 20°, > 20 – 30°, > 30 – 45°, and > 45°. The 0 – 4° class accounts for approximately 50% of the Waikato Region, the next four classes (> 4 – 8°, > 8 – 14, > 14 – 20°, > 20 – 30°) cover between 10 – 13% of the land area each, the 30 – 45° makes up about 4%, while > 45° makes up less than 1% of the total regional area.

3.4.1.3.2 Elevation Classes

There is a wide range of elevation classes used in other landslide susceptibility and hazard studies; these include intervals of 30 m (Ayalew *et al.*, 2005), 50 m (Ayalew and Yamagishi, 2005; Dai and Lee, 2002, 2003; Yilmaz, 2009b), 75 m (Yilmaz, 2009a), 100 m (Dai and Lee, 2001; Dai *et al.*, 2001; Ercanoglu *et al.*, 2008; Wang and Sassa, 2006), 150 m (Cevik and Topal, 2003), 200 m (Clerici *et al.*, 2006; Tangestani, 2004), 300 m (Sakar and Kanungo, 2004), 500 m (He *et al.*, 2003; Lan *et al.*, 2004), and mixed (He and Beighley, 2008). The same problem exists for elevation as for slope, as elevation varies from 0 – 2,747 m in the Waikato Region. The higher elevations mostly represent the Central North Island mountains, while most of the region is at lower elevations. Mixed classes were also applied to elevation, so that the first few classes were in increments of 25 m and 50 m followed by some at 100 m and another at 150 m while the last covers just under 2000 m (Table 3.2). The resulting eight classes each cover between 9 – 17% of the regional land area.

3.4.1.3.3 Aspect Classes

Aspect ranges from -1 to 359° where -1 is flat (Süzen and Doyuran *et al.*, 2004) or non-oriented land (Abdallah *et al.,* 2005). In most studies aspect is divided into nine classes: north, northeast, east, southeast, south, southwest, west, northwest, and flat (Ayalew and Yamagishi, 2005; Can *et al.,* 2005; Cevik and Topal, 2003; Dahal *et al.,* 2008a; Dai and Lee, 2001, 2002; Garcia-Rodriguez *et al.*, 2008; Lee, 2004, 2005, 2007a, b; Lee and Choi, 2004; Lee and Dan 2005; Lee and Lee, 2006; Lee and Sambath, 2006; Lee and Talib, 2005; Wang and Sassa, 2006). The same nine classes have been applied in this study by splitting the eight orientations by their angles and ± 22.5°, and taking -1 as flat (Sakar and Kanungo, 2004; Yilmaz, 2009a, b). As integer values were used in this study any values with a decimal of 0.5 or more attached were rounded up, so the breaks used in this study have been applied to account for this (Table 3.2).

3.4.1.4 Preparing the DEM and Extracting Topographic Variables

The rasterised regional boundary was used as a mask for the DEM to limit the coverage to the area inside the regional boundary. Integer raster layers of elevation, slope and aspect were extracted from the DEM. Classes were assigned to the slope, elevation and aspect raster layers by use of a series of individual conditional statements with defined value ranges and newly assigned values for each class (Table 3.2). For each class within the parameter (slope, aspect and elevation) a new raster layer was created which contained only the data within the defined value ranges specified in the conditional statement, with the assigned code for that class. The raster layers for each class (none of which overlap) within the parameter were then added together to create the reclassified raster layer for that parameter. Classed maps of slope (Map 3.1), elevation (Map 3.2), and aspect (Map 3.3) were then created. A boundary based on the DEM was created as grid (raster), coverage and shapefile (vector) formats in order to clip all other spatial datasets by.

Table 3.2 Classes assigned to slope, elevation and aspect.

Slope		Elevation		Aspect		
Selected Range (°)	**Assigned Code**	**Selected Range (m)**	**Assigned Code**	**Class**	**Selected Range (°)**	**Assigned Code**
0 - ≤ 4	4	0 - ≤ 25	25	N	≥ 0 - < 23, ≥ 338 - < 360	1
> 4 - ≤ 8	8	> 25 - ≤ 50	50	NE	≥ 23 - < 68	2
> 8 - ≤ 14	14	> 50 - ≤ 100	100	E	≥ 68 - < 113	3
> 14 - ≤ 20	20	> 100 - ≤ 200	200	SE	≥ 113 - < 158	4
> 20 - ≤ 30	30	> 200 - ≤ 300	300	S	≥ 158 - < 203	5
> 30 - ≤ 45	45	> 300 - ≤ 400	400	SW	≥ 203 - < 248	6
> 45	90	> 400 - ≤ 550	550	W	≥ 248 - < 293	7
		> 550	2800	NW	≥ 293 - < 338	8
				Flat	< 0	9

3.4.1.5 Spatial Variation of Topographic Variables within the Waikato Region

3.4.1.5.1 Slope

The higher slopes (> 30°) mainly correspond to the various mountains and ranges within the region. The more pronounced mountains with slopes > 30° are Mount Karioi and Mount Pirongia west of Hamilton and Mount Maungatautari south of Hamilton (Map 3.1). Slopes > 30° are extensive throughout the Herangi Range near Mokau, Rangitoto Range northeast of Benneydale and west of Mangakino, Umukarikari Range south of Taupo, Kaimai Range east of Hamilton, and the Coromandel and Moehau Ranges in the northeast (Map 3.1). In some parts of the Hapuakohe Range (west of the Hauraki Plains), Taupiri and Hakarimata Ranges (near Huntly and Ngaruawahia), and Kapamuhunga Range (southwest of Hamilton) slopes of more than 30° can be found (Map 3.1).

Slopes between 14° and 30° occur in areas of hilly topography, most of which occur near areas of higher slopes, and either make up part of the same ranges and mountains mentioned above, or are in the general vicinity of these (Map 3.1). The Hauraki Plains and Hamilton Basin represent a large proportion of flat and very low (0 – 4°) sloping land (Map 3.1). Lake Taupo and the course of the Waikato River also fall in the areas of 0 – 4° slope (Map 3.1), as expected.

3.4.1.5.2 Elevation

Areas of higher elevation in the Coromandel, Moehau, Hapuakohe, and Herangi Ranges take on the form of a series of closely spaced irregular shaped banded features (Map 3.2). Individual mountains such as Pirongia, Karioi, and Maungatautari can be identified on the west coast and inland as localised spots of high elevation. The main effect observed in Map 3.2 however, is the increase in elevation from the Hauraki Plains in the north to the Central Plateau in the south. Incised parts of the course of the Waikato River can be identified by surrounding bands of lower elevations in areas of higher elevation extending from Lake Taupo to Port Waikato in South Auckland (Map 3.2).

3.4.1.5.3 Aspect

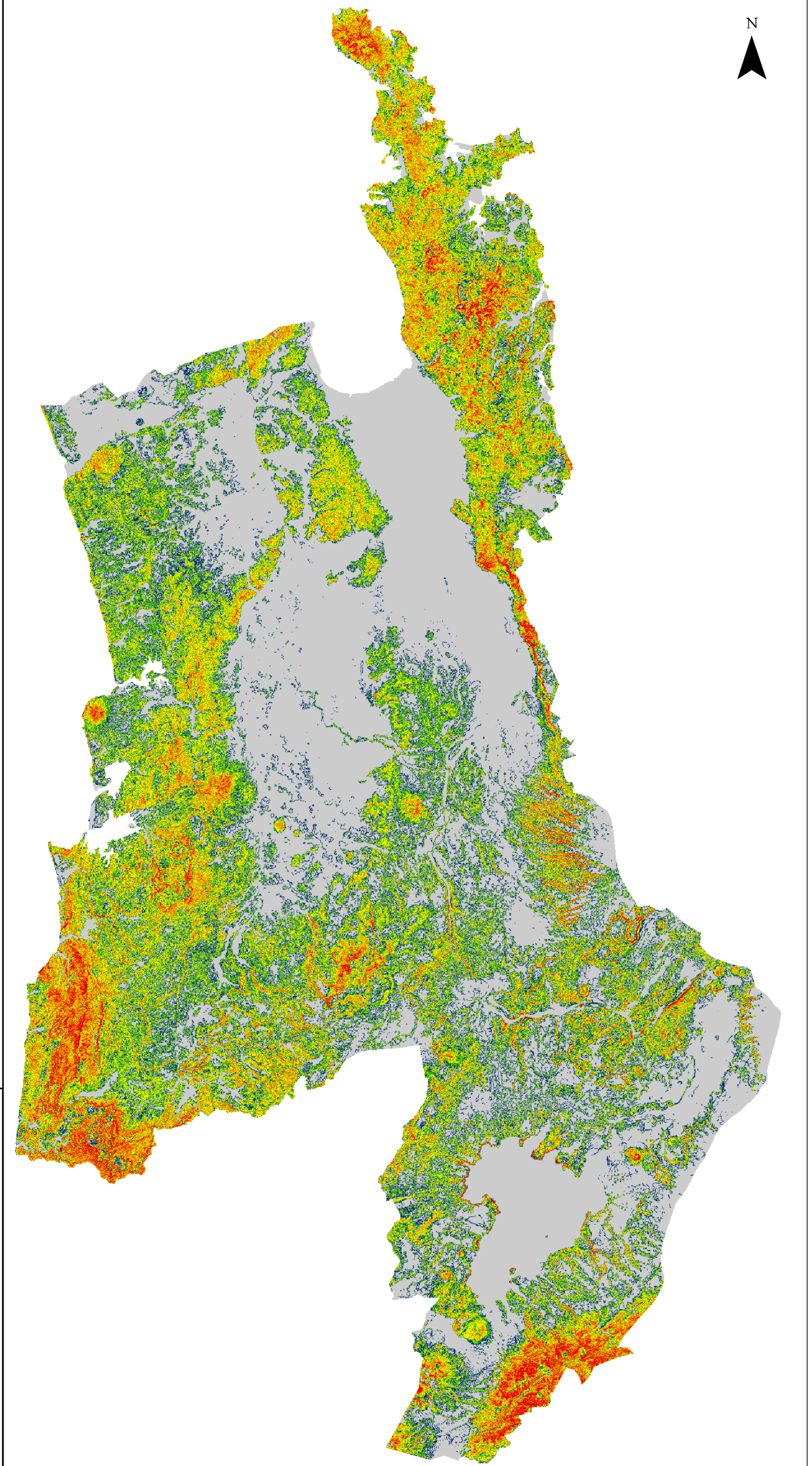
Aspect throughout the region (Map 3.3) exhibits the greatest spatial variation as change in slope orientation varies to different extents for each hill slope within the region. Flat areas are easily identifiable in the Hauraki Plains and Hamilton Basin (Map 3.3) where the slope of the ground surface is flat (0°), aspect is also flat, as no direction of maximum slope exists. Where there is a direction of maximum slope, aspect can be determined (Dahal *et al.,* 2008a), however it can vary over a small area, as hill slopes are irregular shaped three dimensional features with many slope faces and as such have varying geographic aspects.

Map 3.1 Waikato Region - Slope

Slope

Degrees

- 0 - 4
- > 4 - 8
- > 8 - 14
- > 14 - 20
- > 20 - 30
- > 30 - 45
- > 45



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

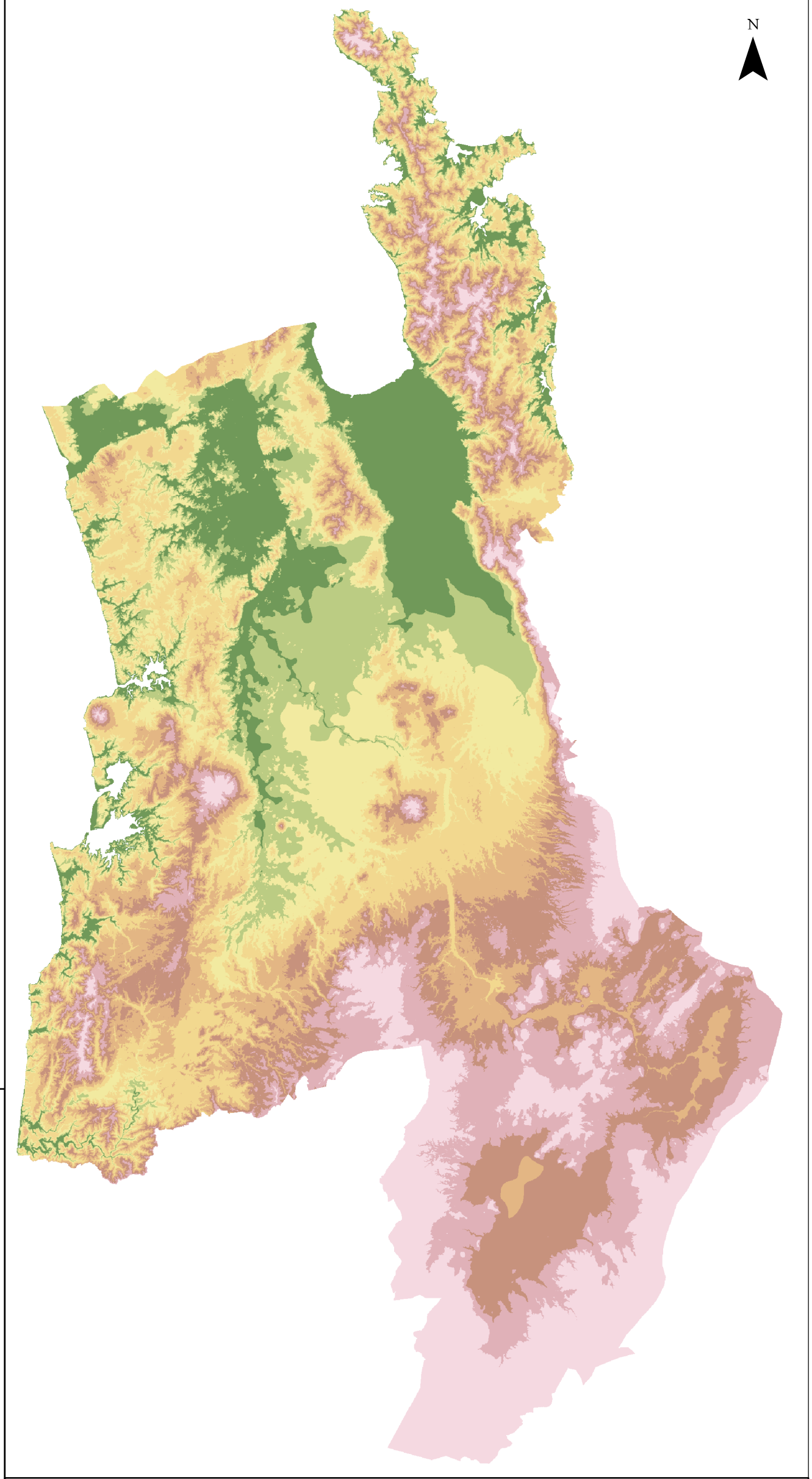
Original Data Source:
North Island Digital Elevation Model (DEM)
Resolution: 25 m

Map 3.2 Waikato Region - Elevation

Elevation

(m)

- 0 - 25
- > 25 - 50
- > 50 - 100
- > 100 - 200
- > 200 - 300
- > 300 - 400
- > 400 - 550
- > 550



Created By: Renée Schicker
Date: 18 February 2010

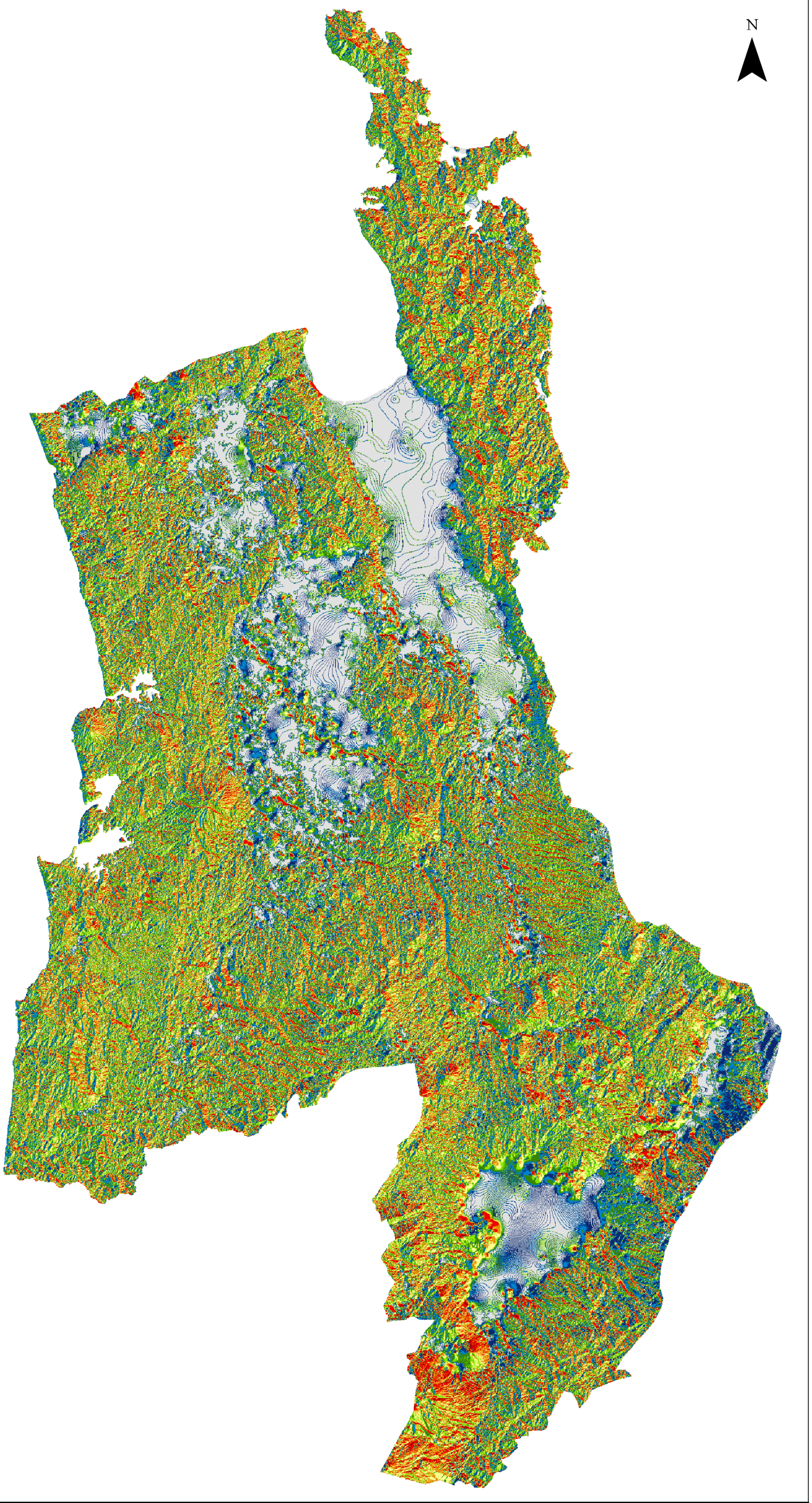
Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
North Island Digital Elevation Model (DEM)
Resolution: 25 m

Map 3.3 Waikato Region - Aspect

Aspect

- North
- Northeast
- East
- Southeast
- South
- Southwest
- West
- Northwest
- Flat



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
North Island Digital Elevation Model (DEM)
Resolution: 25 m

3.4.2 Landslide Inventory

3.4.2.1 Compiling the Landslide Inventory

Two sources of landslide data were used in the construction of the landslide inventory; these were the GeoNet landslide catalogue (GeoNet, 2009) and the Auckland and Waikato QMap landslide data (Edbrooke, 2001b, 2005b). The QMap spatial data is in polygon form and lacks specific information about the landslides. While the GeoNet landslide catalogue has only a single spatial reference (point location) for each entry, it also (for most records), has a distance or radius from each point, within which the landslide is located (pers. comm. Grant Dellow, 2009) and from which a polygon can be created. The point data was transformed to polygon data, essentially by drawing a circle around the point based on the radius record for each point, which was done by using the "buffer" command (Figure 3.1). The identification number (Auto_ID), year of occurrence (Year) and location class (Loc_Class) which gives an indication of the location's accuracy for each record was copied from the point layer to the polygon layer.

Initially a multi-looped process was implemented to transform each individual data point to a polygon in order of ascending landslide record identification number (Auto_ID). The count of the loop cycle corresponded to the landslide record identification number (Auto_ID). The landslide identification number, year and location class (Loc_Class) associated with each reference point were copied to the polygon layer. These edits were replicated in the output of the union process where each individual layer was combined with the others (in ascending order of Auto_ID). In this way each polygon could then be distinguished from the others and more readily identified. If a more simple approach had been used at this stage several of the smaller polygons would have been absorbed by the polygons with larger radii. This was of little surprise as the radius distance for the landslide point data ranges between 0.01 – 25 km, and several points are located near others. A simpler approach without the edits would also mean that clusters of landslides would appear as a singular non-identifiable polygon following a union process. If individual landslide records are not required, the table of the landslide data can be converted into a shapefile, and polygons created for all features based on the radius using a single buffer process.

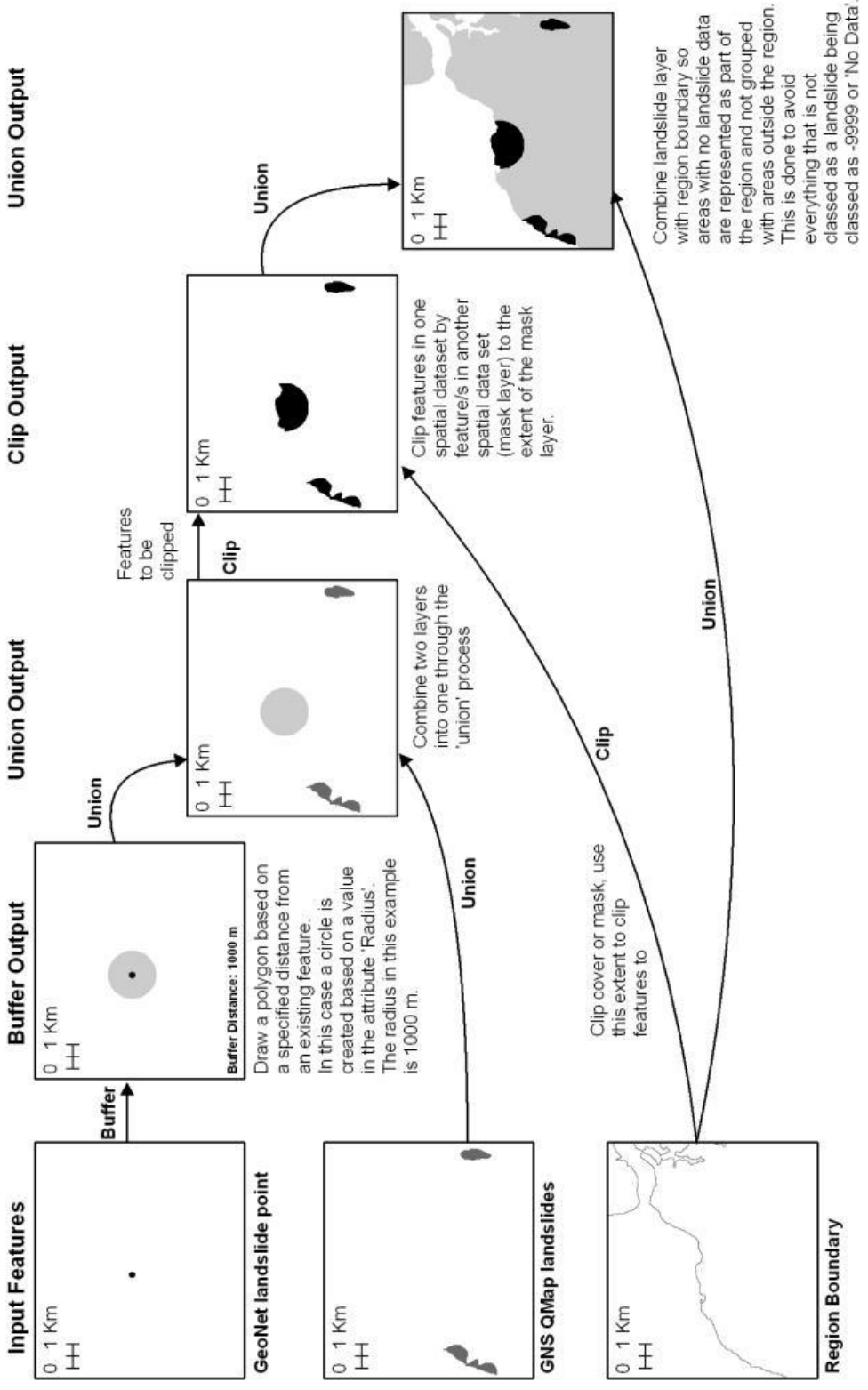


Figure 3.1 Diagram of some of the key GIS processes involved in compiling the landslide inventory.

Of the 123 points 115 polygons were created, the remaining eight points had no radius value to buffer by, so no polygon could be created for any of these. The identifying attributes for the 115 records in the resulting polygon layer were compiled according to descending radius value so the smaller polygons would appear on top of the larger polygons as opposed to being absorbed by the larger polygons if these overlapped. All temporary spatial layers and attribute columns were deleted.

The GeoNet landslide records either have an area greater than 100,000 m^2 (0.1 km^2) or a volume greater than 1,000,000 m^3 (pers. comm. Grant Dellow, 2009). In addition, the radius around the point locations range from 10 m to 25 km, and somewhere within this area there is a landslide. It was thus appropriate to excludethe larger polygons as these had more uncertainty tied to them. A maximum radius limit of 2.5 km was chosen as this would exclude the polygons with a larger radius and would not limit the quantity of data as much as it otherwise would if a smaller radius distance was set as the maximum limit. A radius of 2.5 km results in a circle of 19.6 km^2 around the point. Within this circle, about 0.5% of the area would contain a landslide. If greater radius was accepted there would be more uncertainty as less than 0.5% of the area would have a landslide. By implementing this limit the number of landslides was further reduced from 115 to 73.

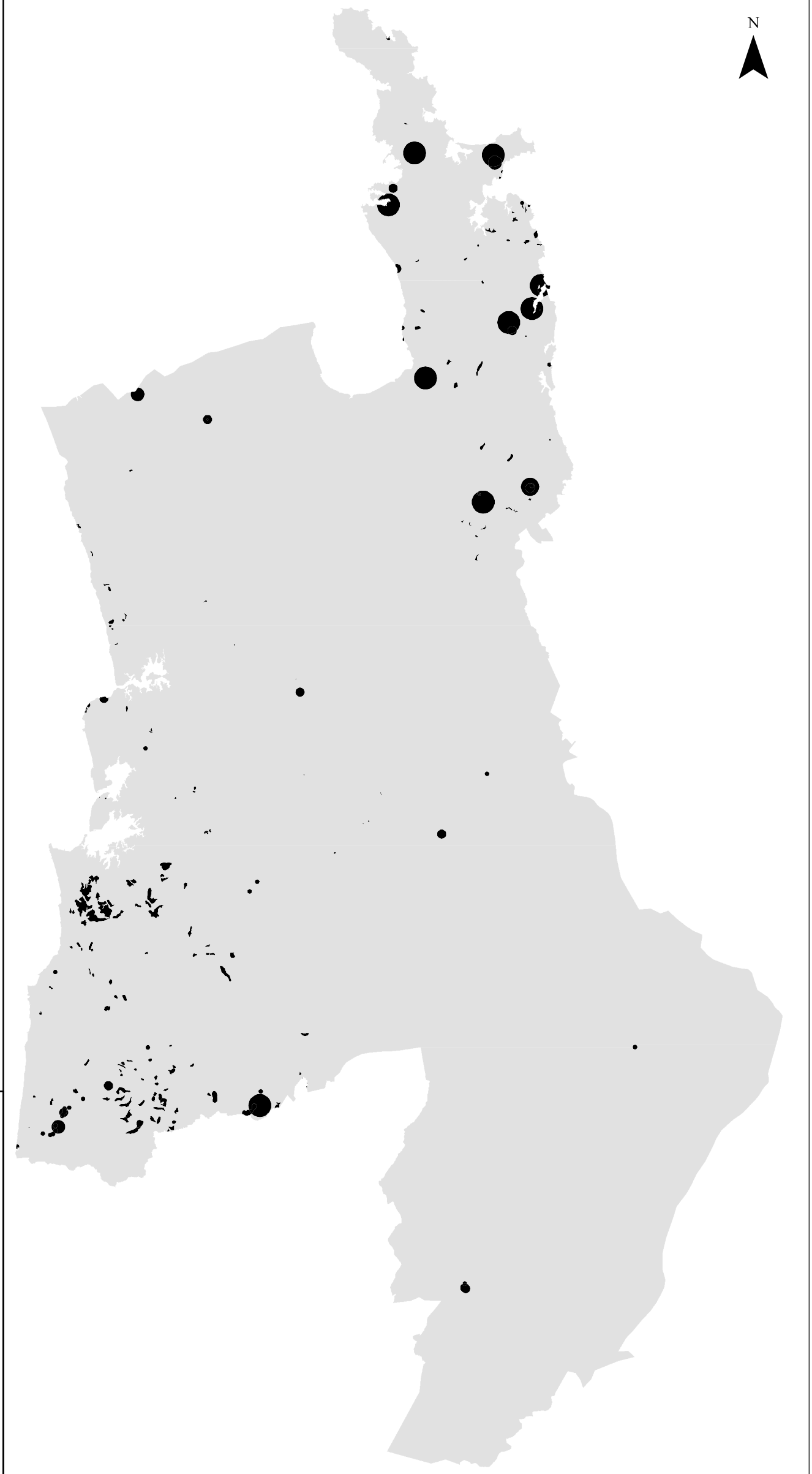
The GNS QMap landslide layers for Auckland and Waikato (Edbrooke, 2001b, 2005b) were combined into a new single layer using a union function. The combined QMap landslide spatial data was then combined (union) with the radius limited landslide catalogue derived spatial coverage and clipped to the DEM boundary (Figure 3.1). The combined product was a landslide inventory map (Map 3.4) based on both the QMap and GeoNet landslide data. For data extraction purposes, the vector data was rasterised (converted to grid format) with a cell size of 25 m to match that of the DEM (Figure 3.1).

3.4.2.2 Waikato Landslide Inventory Observations

Landslides only cover a small proportion of area within the Waikato Region (Map 3.4) which equates to 1.42% or approximately 345 km^2. While these may only account for a small proportion of the total area, their spatial densities vary

considerably, as some areas have had more landslide occurrences than others. The majority of the landslides within the Waikato Region were found in the Thames-Coromandel (northeast) and Waitomo (southwest) districts (Map 3.4). Very few are located in the southeast, but this may not be the case in reality as the Rotorua Topo260 tile area of the QMap project has not yet been released, and as such the data currently available for that area is limited to the reported landslide occurrences recorded in the GeoNet landslide catalogue.

Map 3.4 Waikato Region - Landslide Inventory



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
landsl (Landslides - Auckland and Waikato QMap)
© Institute of Geological and Nuclear Sciences (2001)

Landslide Catalogue © GeoNet (2009)
(Landslides with a radius between 10 m and 2.5 km in the Waikato Region extracted from the database)

3.4.3 Geology

3.4.3.1 Original Geological Spatial Data and Basis for Use

The GNS QMap vector data (Edbrooke, 2001b, 2005b) is currently the most up to date and best source of geological spatial data. The original data has been clipped to cover the extent of the Topo260 tile for each area, and as this project is still in production there are some areas which are yet to be completed. Unfortunately this includes the Rotorua Topo260 tile area which covers approximately one third of the Waikato Region. To address this issue a slightly older GNS geological dataset was incorporated to cover the Rotorua portion of the Waikato Region.

3.4.3.2 GIS Processing of the Geological Spatial Data

To clip the less current New Zealand geology (nzgeology) spatial data to the missing Rotorua portion of the region, a mask was created as a new shapefile consisting of a single polygon which had the vertices snapped to those of the QMap polygons where it would join up. This mask layer was converted to a cover and the nzgeology data successfully clipped, so that all data that fell inside the mask polygon was copied as a new cover. The clipped nzgeology layer and both QMap geology input covers were clipped using the polygon in the DEM boundary coverage.

The two clipped QMap geology layers were unioned together and the rock group (Rock_Group) and main rock (Main_Rock) classes (attributes) for both were compiled into two new columns before combining (union) with the Rotorua geology component. The geological classes were then reclassified to 14 lithological classes (Simple_Lith) classes using the 21 rock group classes for the QMap derived data, and 21 lithological (LITH) classes for the nzgeology derived data. Where the QMap and nzgeology derived datasets were joined there were a few classifications which did not match up uniformly. In addition to this, solely classifying the QMap according to rock group was not ideal in all cases. For instance, the Hauraki Plains which mostly consist of “mud” or “peat” as the main rock type were classed as “mudstone” according to the rock group attribute. To address these issues, the main rocks classed as “mudstone” according to the rock type were singled out and the list of associated geological descriptions

(Description) and stratigraphic units (Strat_Unit) for each class were investigated. Most of the "mud" classes within the "mudstone" rock group were reclassified as "Alluvium" and others as "Engineering Soils" depending on the geological description. All polygons identified as "mudstone" and belonging to the "Newcastle Group" were reclassified as "Argillite". The script was also modified to make these changes, and to retain "peat" as the lithology instead of the "mudstone" rock group it otherwise would have been classed as.

The new simplified lithology (Simple_Lith) classes were dissolved so neighbouring identically classed polygons were amalgamated into single polygons. The output of this was a simplified regional geology map (Map 3.5) for the Waikato Region. In order to convert to a raster (grid) layer the data must be numerical, so each class was assigned a code number (Table 3.3) before being converted to a 25 m cell sized raster layer.

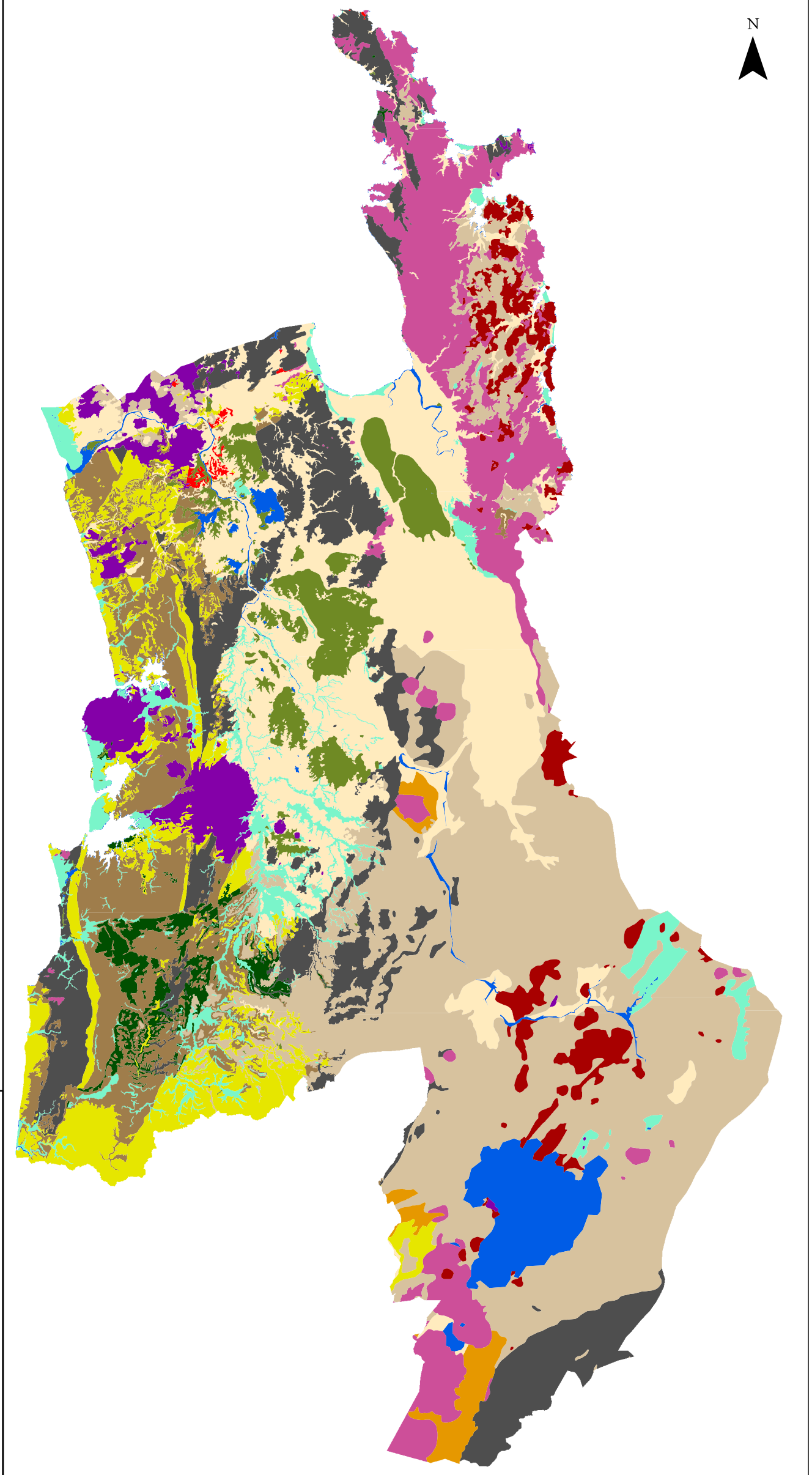
Table 3.3 Geological classes and the assigned code numbers.

Geology	
Class (Simple_Lith)	**Assigned Code**
Water	0
Alluvium	1
Alternating sandstone/siltstone	2
Andesite, dacite and diorite	3
Basalt	4
Engineering Soils	5
Greywacke, argillite and chert	6
Ignimbrite and tuff	7
Laharic colluvium	8
Limestone	9
Mudstone	10
Peat	11
Rhyolite	12
Sandstone	13

Map 3.5 Waikato Region - Geology

Lithology

- Alluvium
- Alternating sandstone/siltstone
- Andesite, dacite and diorite
- Basalt
- Engineering Soils
- Greywacke, argillite and chert
- Ignimbrite and tuff
- Laharic colluvium
- Limestone
- Mudstone
- Peat
- Rhyolite
- Sandstone
- Water



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
QMap Geological Data (Auckland and Waikato)

New Zealand Geology (nzgeology) Data

3.4.3.3 Waikato Regional Geology

3.4.3.3.1 Alluvium

Alluvium can be found quite widespread in the lowlands of the Hauraki Plains, Hamilton Basin, parts of South Auckland, and in localised areas in the Taupo Volcanic Zone (Map 3.5). Alluvium in the Tauranga Group consists of a range of pumiceous deposits, alluvium and colluvium, silts, sands, gravels, silty, sandy, and pumiceous clays, and quite often include interbedded peat (Edbrooke, 2001b, 2005b). Subsidence is common in poorly compacted alluvial deposits when unsupported or overloaded. Sensitive materials are present in the Tauranga Group alluvium in the form of fine-grained pumice beds; these have a tendency to flow if saturated and unconfined, but especially more so when vibrated (Edbrooke, 2001a, 2005a).

3.4.3.3.2 Alternating Sandstone/Siltstone

Alternating sandstone/siltstone lithologies relate to turbidites in the Warkworth Subgroup and date to the Early Miocene (Edbrooke, 2001a). These rocks consist of alternating sandstone and mudstone with volcaniclastic or andestic grits (Edbrooke, 2001b). In the Waikato Region, these rocks belong to the Amokura and East Coast Bays Formations which can be found in a few areas between Te Kauwhata and Bombay, and Colville Formation near Cape Colville in the northern Coromandel Peninsula (Edbrooke, 2001a, b) (Map 3.5).

3.4.3.3.3 Andesite, Dacite and Diorite

In the Waikato Region, andesite, dacite and diorite lithologies are mostly found in the Coromandel Peninsula, but are also found in mountains surrounding Cambridge, Tahuna, and Taupo (Map 3.5). These lithologies commonly form steep slopes and bluffs and in their unaltered and unweathered states have high to very high strengths (Edbrooke, 2001a, 2005a).

The Coromandel Group and Kiwitahi Volcanic Group andesites and dacites date to the Miocene age (Edbrooke, 2001b). The Coromandel Group comprise of several subgroups of andesite and dacite intrusives or lava flows and domes, with some combination of volcaniclastic sediments, tuffs, and breccias (Edbrooke,

2001b). Diorite is also found in the Coromadel Group in the form of quartz diorite to granodiorite rocks (Paritu Plutonics) (Edbrooke, 2001b).

Andesite and dacite of the Orangiwhao Group are Pliocene in Age, and occur in the odd locations around Kawhia Harbour and along the West Coast between Marakopa and Waikawau (Edbrooke, 2005a, b) (Map 3.5). The Orangiwhao Group consist of coarse–grained deeply eroded and weathered remnants of andesitic and dacitic intrusives such as sills, plugs, dikes, necks and plutons (Edbrooke, 2005a, b).

3.4.3.3.4 Basalt

Basalts in the Waikato Region belong to the Alexandra and Kerikeri Volcanic Groups (Edbrooke, 2001b, 2005b). The basalts of the Alexandra Group are Pliocene in age and consist of basaltic andesite and/or basalt lava with tuff, scoria and/or volcanic breccia (Edbrooke, 2005b). The Alexandra Group basalts can be found around Mounts Karioi and Pirongia (Map 3.5).

Plesitocene to Holocene aged basalts of the Kerikeri Volcanic Group are found near Waiuku, Pukakohe, Bombay, Tuakau, and Pukekawa in the South Auckland Volcanic Field (Map 3.5). Basalts in the South Auckland Volcanic Field include young unweathered basaltic rocks, older extensively weathered basalts, and scoria (Edbrooke, 2001a). When loaded, oversteepened, loose deposits of reasonably hard basaltic scoria compress, are likely to slump (Edbrooke, 2001a).

3.4.3.3.5 Engineering Soils

Engineering soils include unconsolidated sediments such as sands, mud, clay, gravel and tephra. These Quaternary sediments are predominantly found in lowland areas, typically in coastal and river environments, where deposition has occurred (Map 3.5). Landslides are commonly experienced in weakly cemented and unconsolidated dune sands as a result of water saturation and oversteepening caused by stream erosion and coastal cliff retreat (Edbrooke, 2005a).

3.4.3.3.6 Greywacke, Argillite and Chert

In the unweathered state, greywacke is generally hard to very hard, and high in strength, but increased weathering can reduce both strength and hardness (Edbrooke, 2001a). Wedge and slab failures can result in cut or unweathered to moderately weathered rock slopes where bulk strength is lowered by shear zones and closed-spaced jointing (Edbrooke, 2001a, 2005a).

Greywacke, argillite and chert are Late Triassic to Early Cretaceous sedimentary rocks which predominantly belong to the Waipapa and Manaia Hill Groups (Edbrooke, 2001a). These lithologies can be observed in parts of the Coromandel Peninsula, Hapuakohe Range, and Rangitoto Range (Edbrooke, 2001a, 2005a) (Map 3.5).

3.4.3.3.7 Ignimbrite and Tuff

Ignimbrites of varying composition and welding are widespread throughout the Waikato Region (Map 3.5). Some ignimbrites are associated with fall deposits and reworked materials of different origins, so are often referred to as formations (Edbrooke, 2005a). In the Waikato Region there are a wide range of ignimbrite formations and these belong to the Pakaumanu Group and Whakamaru Group Ignimbrites of the Taupo Volcanic Centre (Edbrooke, 2005a). Several of the ignimbrites are pumice-rich, and were deposited by pyroclastic flows from the Taupo Volcanic Zone (Edbrooke, 2005a).

The strength of ignimbrites varies between high and low depending on the degree of welding and lithification (Edbrooke, 2005a). Hydrothermal alteration and weathering can reduce the strength of the ignimbrite (Edbrooke, 2001a, 2005a). This is problematic on steeper slopes as slope failure such as debris flows, slides and slumps of varying sizes can result (Edbrooke, 2005a).

3.4.3.3.8 Laharic Colluvium

Laharic colluvium and laharic andesitic colluvium can be observed in the Taupo Volcanic Zone (Map 3.5). These are located in the downslope areas of the Central North Island Mountains, and also Mount Maungatautari (Map 3.5), and are most likely weathered and/or eroded material from these features.

3.4.3.3.9 Limestone

Limestone is more predominantly found in sequences of the Te Kuiti Group in outlying areas of Te Kuiti in South Waikato (Map 3.5). Small indistinguishable areas of Papakura Limestone can be found south of Auckland, near Onewhero and Limestone Downs. The Papakura Limestone is a bioclastic, locally flaggy limestone which often comprises of small greywacke pebbles, shell fragments and sandstone lenses (Edbrooke, 2001a). Limestones of the Te Kuiti Group generally thick, form steep cliffs and bluffs, and are either Orahiri Limestones, Otorohanga Limestones, or part of the Upper Te Kuiti Subgroup (Edbrooke, 2005a, b). The Orahiri Limestone can range in lithology from moderately sandy to glaucontic and pebbly and is characterised by thick oyster beds (Edbrooke, 2005a, b). A well-developed karst topography is characteristic of the flaggy, pure bioclastic Otorohanga Limestone (Edbrooke, 2005a). Limestone of the Upper Te Kuiti Subgroup is sandy and skeletal, and can include conglomerate and calcareous sandstone (Edbrooke, 2005b). In some areas of limestone, there is a potential for subsidence (Edbrooke, 2005b).

3.4.3.3.10 Mudstone

Mudstones in the Waikato Region are widespread along the West Coast from Port Waikato to Mokau (Map 3.5). These are soft to moderately soft and are often unstable and susceptible to failure (Edbrooke, 2005a). Unsupported cuttings and repeated wetting and drying can initiate failure (Edbrooke, 2005a). Hard calcareous rocks which overlie soft mudstones can be undermined by active earth flows in the mudstone and lead to block falls (Edbrooke, 2005a).

The Mangakotuku Formation consists of montmorillinite-rich mudstone and siltstone, and is less stable than the high plasticity kaolinite mudstone of Waikato Coal Measures which it overlies (Edbrooke, 2005a). In natural outcrops and low angle cuttings, the Mangakotuku Formation mudstone has a propensity to slump or collapse (Edbrooke, 2005a).

3.4.3.3.11 Peat

Peat is characterised by its dark brown to black organic muddy content (Edbrooke, 2001b, 2005b). Peat is generally found in the lowland areas such as

the Hauraki Plains, and include features such as the Kopuatai Peat Dome and neighbouring areas to the Rotowaro and Huntly coal mines (Map 3.5).

3.4.3.3.12 Rhyolite

Within the Waikato Region, rhyolite can be found in the Coromandel Volcanic Zone and in some parts of the Taupo Volcanic Zone (Map 3.5). The Coromandel rhyolite (Map 3.5) belongs to the Minden Rhyolite Subgroup of the Whitianga Group (Edbrooke, 2001a, b). These include rhyolite flow and dome complexes, and related breccias and tuffs from the Late Miocene to Early Pliocene era (Edbrooke, 2001a, b). The dome complexes are thought to be a result of rhyolitic caldera eruptions within the Coromandel Volcanic Zone (Edbrooke, 2001a).

3.4.3.3.13 Sandstone

Sandstone is widespread in areas surrounding mudstone along the West Coast from Port Waikato to Mokau (Map 3.5). Some sandstones are interbedded with thin beds of mudstone, siltstone, or coal seams; others can be calcareous or non-calcareous and can include minor siltstone and tuff (Edbrooke, 2001a, b, 2005a, b). Sandstones in the Waikato Region range from moderately soft to hard, and are usually stable (Edbrooke, 2005a).

3.4.4 Soil

3.4.4.1 Simplification of Soil Data Using GIS

The soil coverage used (nzfsl) is an extended multi-factored coverage of the New Zealand Land Resource Inventory (NZLRI) soil description (Landcare, 2000a, b). The New Zealand Soil Classification (NZSC) attribute soil codes were deciphered by referring to the NZLRI metadata (Newsome *et al.,* 2000) and the New Zealand Soil Classification (Hewitt, 1992) before simplifying the 73 different NZSC classes to the soil order. Towns, quarries, lakes, rivers, and ice were reclassified as non-soil features. A parameter map of the soil order (Map 3.6) was then created. The data were then prepared for data extraction and compilation. Each soil order class was assigned a code number (Table 3.4) and the vector data converted to raster.

Table 3.4 Soil order classes and the assigned code numbers.

Soil Class (Soil Order)	Assigned Code
Non-soil features	0
Allophanic Soils	1
Brown Soils	2
Gley Soils	3
Granular Soils	4
Organic Soils	5
Oxidic Soils	6
Pallic Soils	7
Podzols	8
Pumice Soils	9
Raw Soils	10
Recent Soils	11
Ultic Soils	12

3.4.4.2 Waikato Soils

3.4.4.2.1 Non-soil Features

Areas of lakes, rivers and ice are classed as non-soil features as they currently exist as bodies of water and are lacking surface soil. Quarries and mines are non-soil features as a result of the soil being removed from the site. In urban areas, the soil is hidden or lost beneath the mass of concrete and tarseal. Non-soil features account for about 3.7% of the Waikato Region (Figure 3.2) and most of this figure is explained by Lake Taupo (Map 3.6).

3.4.4.2.2 Allophanic Soils

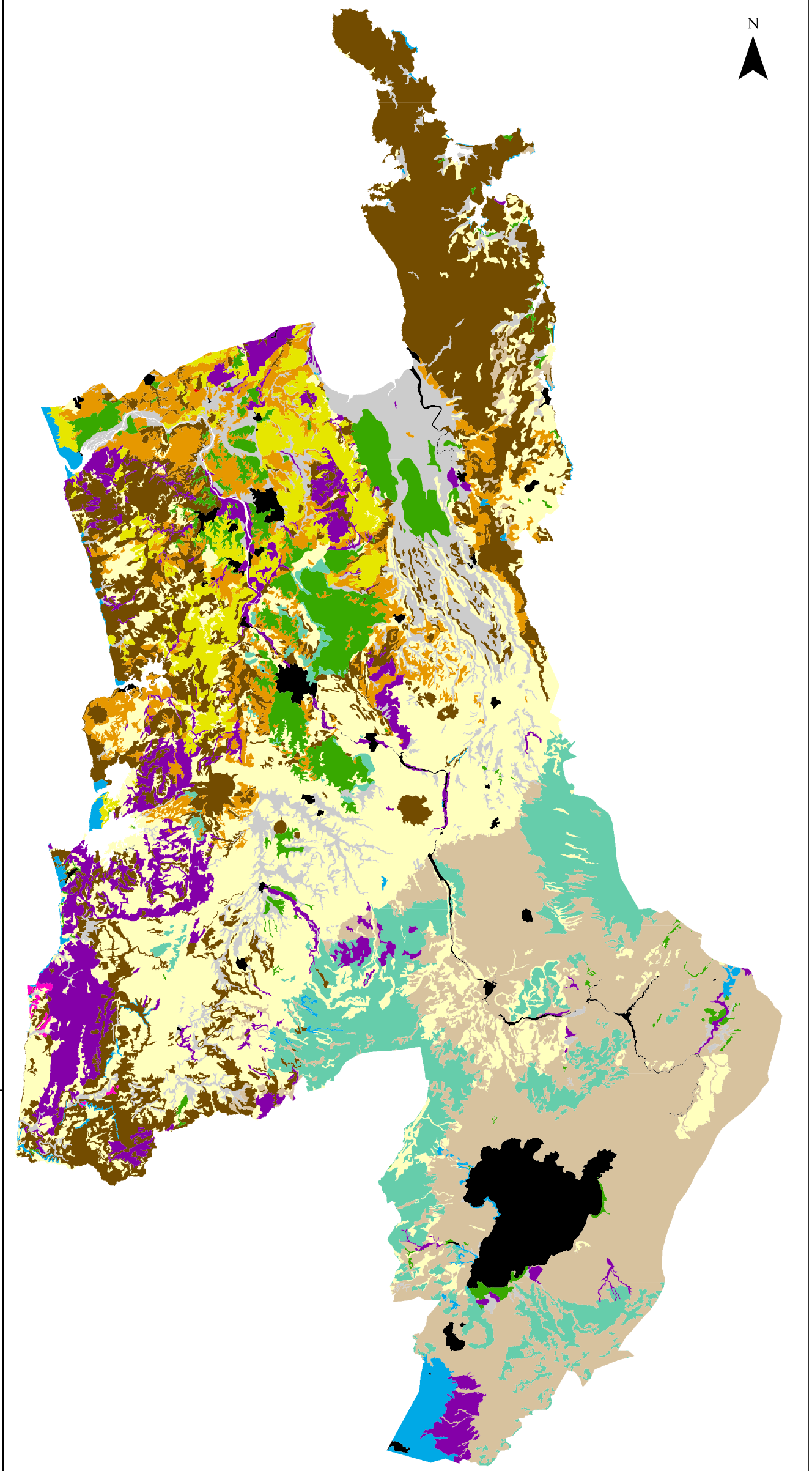
Allophane, imogolite and ferihydrite minerals that have a short-range order strongly influence the properties of allophonic soils (Hewitt, 1992). Allophanic soils are typically weak in strength, have very low bulk densities, are sensitive, and predominantly occur in volcanic parent materials (Hewitt, 1992). Parent materials are mainly basaltic scoria and ash, but in some instances, can be quartzo-feldspathic and tuffaceous (greywacke) sandstone (Hewitt, 1992). Allophanic soils are the most prevalent soil within the Waikato Region and are mostly found in the large space between the Herangi Range in the southwest and Kaimai Range in the east (Map 3.6). Less extensive areas of allophonic soils exist

Map 3.6 Waikato Region - Soil Order

Soil

NZSC Soil Order

- Allophanic soils
- Brown soils
- Gley soils
- Granular soils
- Non-soil features
- Organic soils
- Oxidic soils
- Pallic soils
- Podzols
- Pumice soils
- Raw soils
- Recent soils
- Ultic soils



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
New Zealand Land Resource Inventory (Extended Fundamental Soil Attributes) Data

Classified according to New Zealand Soil Classification (NZSC), Hewitt (1992).

beyond the above mentioned extent, but are smaller in area by comparison (Map 3.6).

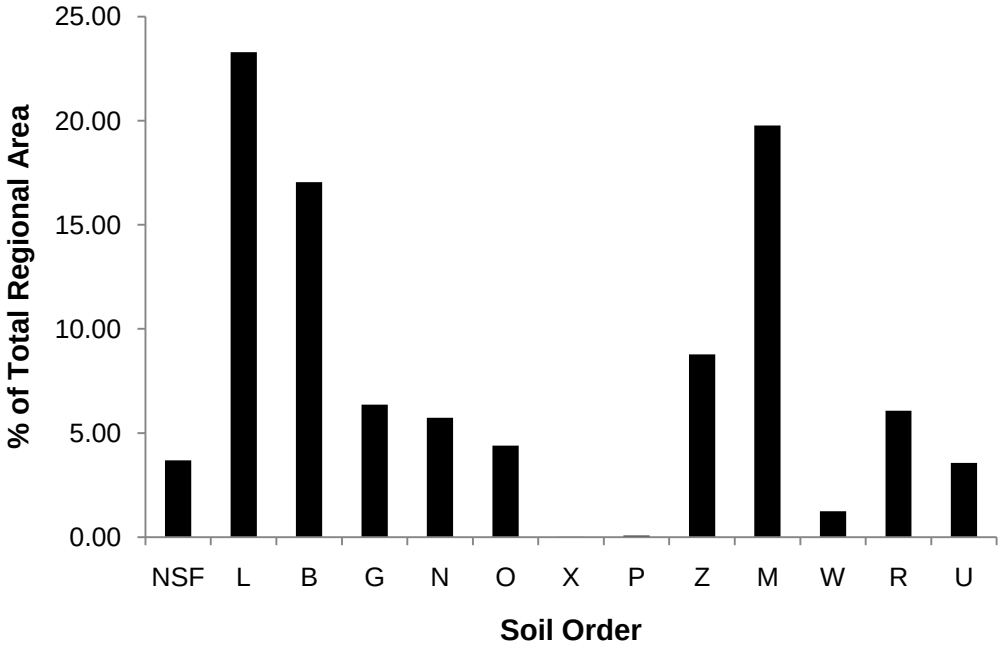


Figure 3.2 Soil orders: non-soil features (NSF), allophanic soils (L), brown soils (B), gley soils (G), granular soils (N), organic soils (O), oxidic soils (X), pallic soils (P), podzols (Z), pumice soils (M), raw soils (W), recent soils (R), and ultic soils (U), in the Waikato Region and the proportion of total land area for each.

3.4.4.2.3 Pumice Soils

Pumice soils occur in pumiceous or sandy tephra which have an age in the range of 700 to 3500 years (Hewitt, 1992). These soils are the second most common in the Waikato Region and are largely found within the Taupo Volcanic Zone, but a small proportion can also be found in the Coromandel (Map 3.6). A pumiceous and glassy skeleton with a low clay content normally containing allophane dominate the soil properties in the pumice soil order (Hewitt, 1992).

3.4.4.2.4 Brown Soils

Brown soils can be found extensively throughout the Waikato Region, but the largest area of brown soils within the study area appears to be in the Coromandel (Map 3.6). Brown soils generally contain 2:1 clay minerals and normally have moderate to very low base saturations (Hewitt, 1992).

3.4.4.2.5 Podzols

Podzols are most dominant in and around the Taupo Volcanic Zone (Map 3.6). These are acid soils of low base saturation which have a horizon of accumulated aluminium in complex forms with short-range-order minerals (usually with silicon as allophane or imogolite) and/or organic matter (Hewitt, 1992). An overlying E horizon which indicates translocation is often associated with the above mentioned horizon but may be masked by organic matter or missing as a result of ploughing, erosion or bioturbation (Hewitt, 1992).

3.4.4.2.6 Gley Soils

Gley soils are poorly to very poorly drained soils that exhibit a greyish colour up to depths of 90 cm or more in the soil solum as a result of reducing conditions from prolonged periods of limited oxygen in saturated conditions (Hewitt, 1992). On Map 3.6, gley soils are found in several dendritic shaped areas which appear to be located in parts of the river networks as well as a large proportion of the Hauraki Plains.

3.4.4.2.7 Recent Soils

The main concept with recent soils is not so much related to the length of time of soil formation, but to the weak soil development (Hewitt, 1992). Only early signs of soil forming processes are shown in these soils, generally as a result of truncation of an older solum, youthfulness, and in exceptional instances, where soil material is resistant to alteration (Hewitt, 1992). In these soils, soil formation has been sufficient to form a distinct topsoil (Hewitt, 1992). Recent soils in the Waikato Region can be found on the Hunua, Hapuakohe, Taupiri, Pakaroa, and Herangi Ranges, base of the Central North Island Volcanoes (Mounts Ruapehu, Ngaruhoe, and Tongariro), small areas around Mounts Karioi and Pirongia, and in some parts along the course of the Waikato River (Map 3.6). These areas have been contributing to soil formation through erosion and transportation of sediments so is not surprising to find recent soils in these vicinities.

3.4.4.2.8 Granular Soils

Granular soils are widespread across the northern parts of the Waikato, in the area bounded by Mount Pirongia in the southwest, Kaimai Range in the west, and the southwestern extent of the Coromandel Range in the northeast (Map 3.6). In these clayey soils, the kaolin-group minerals are dominant, and vermiculite and hydrous-vermiculite are usually associated with them (Hewitt, 1992). Polyhedral peds make up the soil fabric, and the strength of these habitually alter rapidly with water content (Hewitt, 1992).

3.4.4.2.9 Organic Soils

Organic soils on Map 3.6 tend to correspond to areas of peat or swamp where vegetation remains such as wetland plants (peat) or forest litter, that are partly decomposed (Hewitt, 1992). These soils occur in sites in which the decomposition of organic matter is either balanced or exceeded by the production and accumulation rates of plant biomass (Hewitt, 1992). In the Waikato, these correspond to the Rukuhia and Moanatuatua Swamps, the Kapuatai Peat Dome in the Hauraki Plains, and some flat areas surrounding Hamilton, Ngaruawahia, Gordonton, Whangamarino, and Huntly (Map 3.6). As its name indicates, organic soils are dominated by organic soil material, but mineral soil material is also commonly present (Hewitt, 1992).

3.4.4.2.10 Ultic Soils

Ultic soils are mostly located south of Auckland, generally in and around the Hapuakohe, Taupiri, Kaketu, Hakarimata, and Kapamahunga Ranges and Hangawera Hills, with the southernmost extent in the area surrounding Kawhia (Map 3.6). These are acid soils which mostly develop in the clayey weathered products of acid igneous rocks or siliceous sediments, but a few are formed in weathered products of greensands and limestone (Hewitt, 1992). Kaolinite, halloysite, aluminium-interlayered vermiculite and smectite are among the clay mineral mixtures usually contained in ultic soils (Hewitt, 1992). Subsoil horizons consist of clayey and/or organic illuvial features and argillic horizons are often present (Hewitt, 1992).

3.4.4.2.11 Raw Soils

Raw soils are either fluid at depth or lacking distinct topsoil development (Hewitt, 1992). They are found in environments such as lagoons, tidal estuaries, beach sands, alpine rock areas and active screes which are areas of active erosion, deposition, or rockiness that prevent the development of topsoil (Hewitt, 1992). In the Waikato raw soils are found around the alpine areas of Mt Ruapehu, intermittently in sandy coastal areas around the Coromandel and West Coast, some parts of the shores surrounding Lake Taupo, and in a few hilly areas in the south and east (Map 3.6).

3.4.4.2.12 Pallic Soils

Pallic soils only make up 0.08% of the Waikato Region and are predominantly found around Waikawau on the southwest coast but a small area also exists in the Hapuakohe Range near Mt Maungakawa (Map 3.6). These soils are low in secondary iron oxide contents, and have moderate to high base status (Hewitt, 1992). In the subsurface horizons, pallic soils are pale in colour, have high slaking potential and high density (Hewitt, 1992). In winter and spring pallic soils have soil water surpluses, and water deficits in summer (Hewitt, 1992).

3.4.4.2.13 Oxidic Soils

Oxidic Soils are uncommon in the Waikato Region, but a small area exists near Pokeno (Map 3.6). Secondary oxides and low-activity phyllosilicate clays contained in these soils bring about variable charge properties (Hewitt, 1992). These soils develop in weathered clayey products of basic rocks, with clayey surface horizons which increase in clay content with depth (Hewitt, 1992). Oxidic soils have a low plasticity in relation to clay content, in addition to a polyhedral soil fabric which ranges between fine to very fine with friable breakdown to stable microaggregates of 2 mm or less (Hewitt, 1992).

3.4.4 Land Cover (Land Use)

3.4.5.1 Simplifying Land Cover Classes

The 42 land cover database 2 (LCDB2) classes which occurred throughout the Waikato region (out of 43 possible classes) were simplified to the eight first order classes provided in the supporting documentation (MfE, 2004). This resulted in a simplified land cover map (Map 3.7) for the Waikato Region. Code numbers were assigned to the simplified land cover classes (Table 3.5) and a raster copy created.

Table 3.5 First order land cover classes and assigned code numbers.

Land Cover Class	Assigned Code
Water Bodies	0
Artificial Surfaces	1
Bare or Lightly Vegetated	2
Cropland	3
Grassland	4
Sedgeland and Saltmarsh	5
Forest	6
Shrub and Shrubland	7

3.4.5.2 Land Cover /Land Use within the Waikato Region

3.4.5.2.1 Grassland

Grassland covers both high producing exotic and low producing pastoral grasslands, and both tall and depleted tussock grasslands (MfE, 2004). Approximately 53% of the Waikato Region is grassland (Figure 3.3) the majority of which supports dairying and stock grazing land use activities (Environment Waikato, 1998). Being the largest land cover class, grassland is quite widespread throughout the region, and generally appears to be in lower sloped or flat areas (Map 3.7).

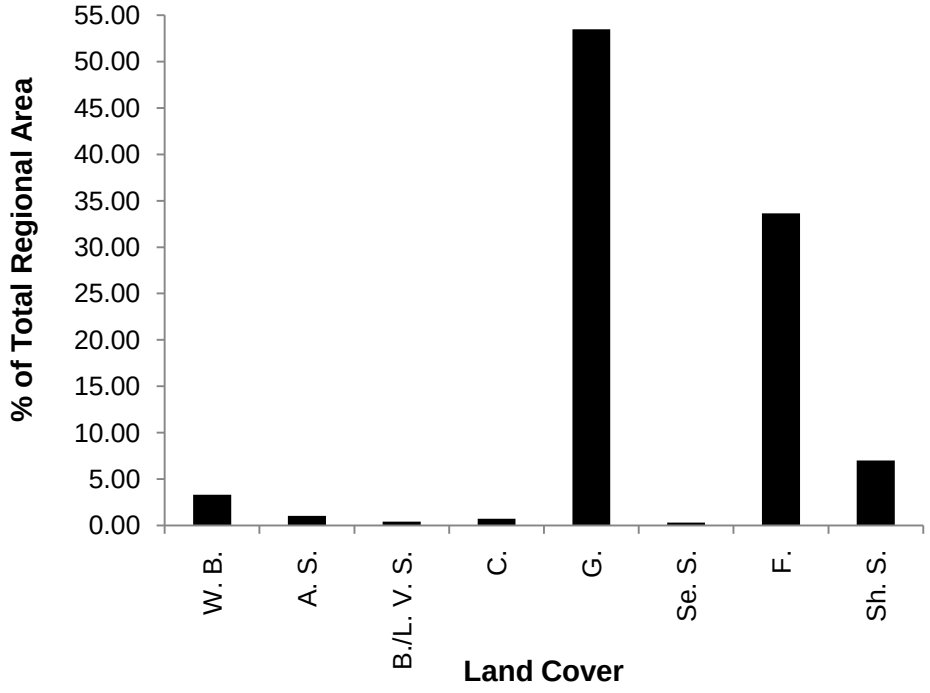


Figure 3.3 Land cover classes: water bodies (W. B.), artificial surfaces (A. S.), bare/lightly vegetated surfaces (B. /L. V. S.), cropland (C.), grassland (G.), sedgeland and saltmarsh (Se. S.), forest (F.), and shrub and shrubland (Sh. S), within the Waikato Region and the proportion of total land area they occupy

3.4.5.2.2 Forest

Forests make up the second largest class with approximately 34% of the total land area (Figure 3.3), and this encompasses indigenous and plantation forests (in various states of growth) as well as the less commonly occurring shelterbelts, deciduous hardwoods, and mangroves (MfE, 2004). Forests are relatively common on sloped areas in the Coromandel Range in the northeast and in many areas of the southwest (Map 3.7). The largest area of forestry appears to be in the southeast in the Taupo Volcanic Zone (Map 3.7) and is a mix of both indigenous and plantation forestry (Environment Waikato, 1998).

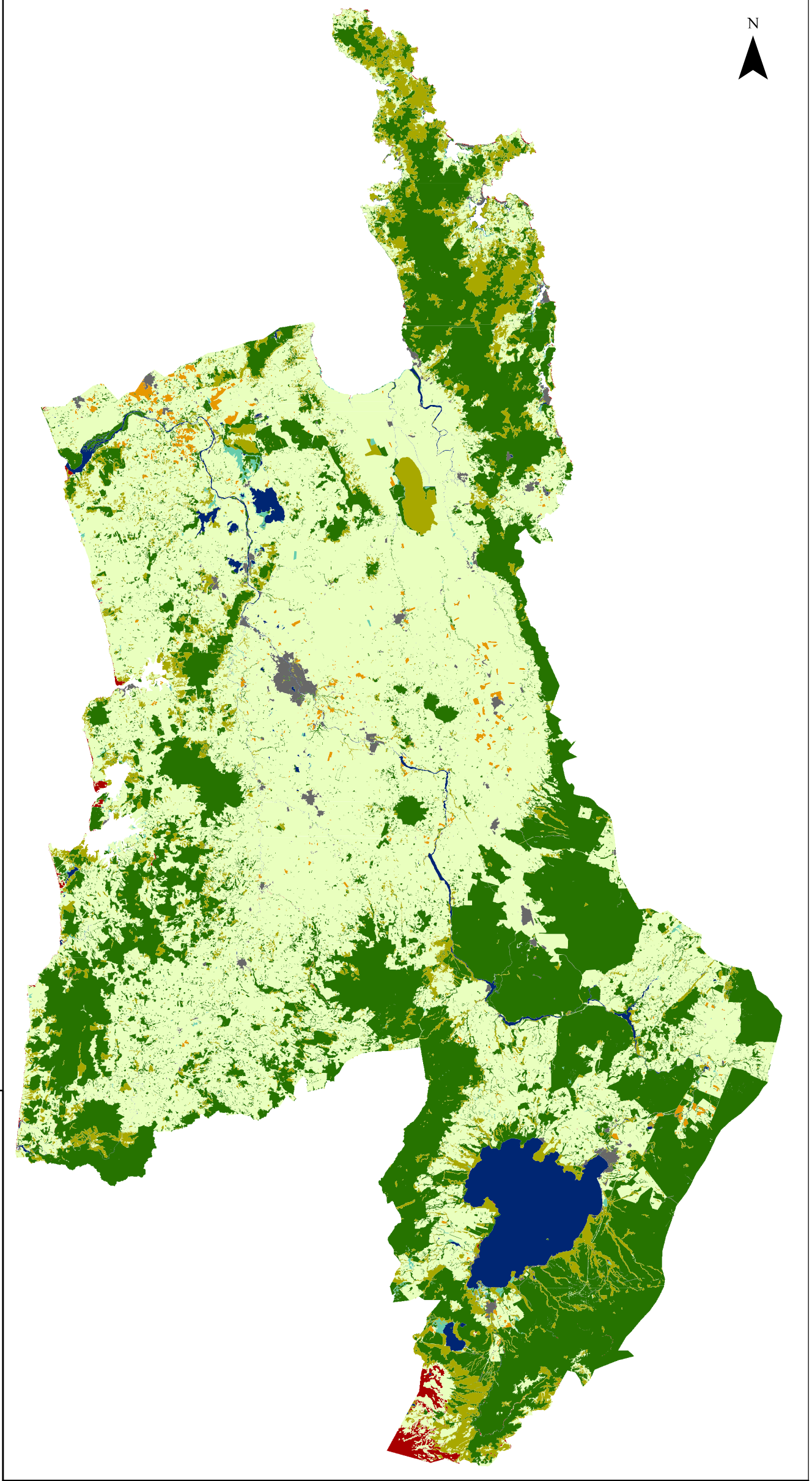
3.4.5.2.3 Shrub and Shrubland

Shrub and shrubland includes a wide range of scrub, such as fernland, gorse and broom, manuka and or kanuka, matagouri, grey scrub, mixed exotic shrubland, sub alpine shrubland and broadleaved indigenous hardwoods (MfE, 2004). This land cover class is mostly dispersed across the region in small sized areas, with the exception of the Kopuatai Peat Dome in the Hauraki Plains consisting of manuka and/or kanuka (Map 3.7).

Map 3.7 Waikato Region - Land Cover

Land Cover

- Artificial Surfaces
- Bare or Lightly Vegetated
- Cropland
- Forest
- Grassland
- Sedgeland and Saltmarsh
- Shrub and Shrubland
- Water Bodies



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
Land Cover Database 2 (LCDB2)

3.4.5.2.4 Water Bodies

Lakes, ponds, rivers and estuarine open water make up the water bodies class (MfE, 2004). The largest of these features is Lake Taupo (Map 3.7) which has an area of approximately 618 km^2 and makes up about 74% of the water bodies class. The second largest water body in the Waikato region is the Waikato River (Map 3.7) which extends 425 km from Lake Taupo to Port Waikato (MfE, 1997), and has a mean discharge of 340 m^3s^{-1} (Duncan and Woods, 2004). Other large (> 1 km^2) lakes, which make up another 8% of this class, include Lake Rotoaira south of Lake Taupo and Lakes Waikare, Whangape, Waahi, and Rotongaro on either side of the Waikato River near Huntly and Te Kauwhata (Map 3.7).

3.4.5.2.5 Artificial Surfaces

Artificial surfaces cover man made features such as built-up areas, urban parklands or open spaces, surface mines, dumps, and transport infrastructure (MfE, 2004). The majority of the artificial surfaces appear to be urban or built up areas. Hamilton City is the largest urban centre in the Waikato Region, accounting for 22% of this class (Map 3.7). Surface mines include several coal mines around Huntly and Rotowaro, gold mines in the Coromandel, and quarries throughout the region (Map 3.7). Other features included in this class such as airstrips, sawmills, golf courses and race courses are quite widespread throughout the region (Map 3.7).

3.4.5.2.6 Cropland

Areas of cropland within the Waikato Region (Map 3.7) include short-rotation cropland, orchards and other perennial crops, and vineyards (MfE, 2004). Cropland makes up less than 1% of the total regional area but is most common in the area south of Auckland (Map 3.7), mostly the Bombay area which has several large market gardening operations. Several occurrences also exist around Hamilton (Map 3.7), some of these are vineyards and berry growing operations among other things. A few areas of cropland can also be found around Taupo (Map 3.7).

3.4.5.2.7 Bare/Lightly Vegetated Surfaces

Areas of coastal sand and gravel, river and lakeshore gravel and rock, alpine gravel and rock, permanent snow and ice, and alpine grass or herbfield are considered as bare or lightly vegetated surfaces (MfE, 2004). The majority of the bare or lightly vegetated surfaces are around the central North Island volcanoes, most of this is alpine gravel and rock but some is permanent snow and ice and alpine grass or herbfield (Map 3.7). Areas of river and lakeshore gravel and rock are found on some lakeshores of Lake Taupo and the Waipa River (Map 3.7). Coastal sands and gravels are found in various parts of the west coast and Coromandel, a significant amount can be found around Aotea Harbour (Map 3.7).

3.4.5.2.8 Sedgeland and Saltmarsh

Sedgeland and saltmarsh cover flaxland and herbaceous vegetation in either freshwater inland wetlands, or saline coastal wetlands (MfE, 2004). Most of the freshwater inland wetlands are found around Whangamarino; Lakes Waikare, Kopuera, and Whangape near Te Kauwhata, Rangiriri, and Ohinewai; Lakes Hakaroa, Kimihia and Waahi near Huntly; and Lake Rotoaira and Lake Taupo near Turangi (Map 3.7). In various locations along the west coast and Coromandel Peninsula small areas of saline coastal wetlands and flaxland can be found (Map 3.7).

3.4.4 Distance from Linear Features (Faults, Roads, and Rivers)

3.4.6.1 The Basis for Chosen Buffer Distances

3.4.6.1.1 Buffer Distances for Distance from Roads

In the studies that have included distance from roads as a predictor, there appears to be no agreement on the size or number of buffer distances used. Some studies used equally sized increments like 25 m (Yalcin, 2008), 50 m (Sharma and Kumar, 2008), 150 m (Yilmaz, 2009a), 250 m (Yilmaz, 2009b), or 1 km (He and Beighley, 2008). Others have used various distances for the buffers such as: 0 – 10 m, 10 – 20 m, 20 – 50 m, 50 – 100 m, 100 – 200 m, and > 200 m (Dahal *et al.*, 2008a); or 0 – 50 m, 50 – 100 m, 100 – 200 m, 200 – 300 m, 300 – 400 m, and

400 – 500 m (Tangestani, 2004). The last class of a chosen set of buffer distances ranges from 25 – 50 m (van Westen *et al.,* 2003) to 2250 – 5203 (Yilmaz, 2009b) and > 4 km (He and Beighley, 2008). As there was no agreement between the studies in the literature, a mix of classes loosely based on the above examples were chosen for this study, 0 – 50 m, 50 – 100 m, 100 – 250 m, 250 – 500 m, 500 – 1000 m, and > 1000 m.

3.4.6.1.2 Buffer Distances for Distance from Faults

Distance from major faults allows the effect of seismicity to be considered indirectly (Gupta and Joshi, 1990). A series of layers of varying (buffer) distances from a linear feature such as fault lines can be created, overlaid and combined to create a single spatial dataset where each distal zone is represented by a band (Figure 3.4). However, in the literature there is no agreement on the size of the buffer distances for use in distance from fault lines, and as a result various different distances have been used in other studies (Ercanoglu and Gökceoglu, 2004). Among the many examples in the literature, Cevik and Topal (2003) and Dhakal *et al.* (1999) both applied classes of: 0 – 100 m, 100 – 250 m, 250 – 500 m, 500 – 1000 m, and > 1000 m; Saha *et al.* (2002) used: < 500 m, 500 – 1000 m, and > 1000 m; Pachauri and Pant (1992) used increments of 2 km from 0 to 10 km; Yilmaz (2009a) applied increments of 150 m up to 1350 m as well as a class of 1350 – 2499 m; and Yilmaz (2009b) used increments of 250 up to 2250 m, with an additional class of 2250 – 4806 m. The classes chosen in this study are loosely based on those of Cevik and Topal (2003) and Dhakal *et al.* (1999) as stated above, so the first four classes match, but the additional three classes were created out of consideration of studies such as Pachauri and Pant (1992) and Yilmaz (2009a, b) which had considered a wider distance; and for this sized study area, a limit of 1000 m may be too low.

3.4.6.1.3 Buffer Distances for Distance from Rivers

In other studies which have taken a categorical approach to distance from streams, rivers, or drainage lines, the most commonly used buffer distances are increments of 50 m (Abdallah *et al.*, 2005; Dai and Lee, 2001; Dai *et al.*, 2001; Sharma and Kumar, 2009; Yalcin, 2008) or 100 m (Cevik and Topal, 2003; Lee, 2007a; Lee

and Lee, 2006; Lee and Talib, 2005; Perotto-Baldiviezo *et al.*, 2004). The size of the buffer distances is one issue; the other issue is deciding how far from the feature these should extend (or how many classes should be created). Despite many of those studies which have used the same increments of distance, few tend to use the same number of classes or upper limit. The most common combination was found to be the use of 5 classes in increments of 100 m (upper limit > 400 m) (Cevik and Topal, 2003; Lee, 2007a; Lee and Lee, 2006; Lee and Talib, 2005) so these classes have been applied to distance from rivers in this study.

3.4.6.2 Creating the Polygon Buffer Zones around Linear Features

The scripted processes for each of the linear features are basically the same; just the chosen distances and/or number of classes vary. The distance values were set as variables at the start, so if any distance values were altered, the variables could be changed without having to change the procedure based content of the script. Five variables were set for roads, six for faults, and four for rivers. The variable distances set were used to buffer the linear features by and were set in a specific order which allows the union process to union in descending order of buffer distances (Figure 3.4).

Editing prior to the union process was carried out so the distance used to buffer the feature by was recorded in the buffered outputs attribute table in a uniquely identifiably column name based on the set variable number. When combined, all data from the two input covers were transferred to the output union covers attribute table (Figure 3.4). The edits made prior to combining (union) all buffer layers together were used to consolidate the buffer distances into a single attribute column, which was also done in descending order of distance values. All temporary and unneeded attribute columns were deleted from the attribute table. The output polygon cover was clipped by the DEM boundary (Figure 3.4). The clipped output cover and the DEM boundary were combined (union) to get complete polygon coverage for the whole region.

The buffer zone distances for roads, faults and rivers, were assigned code numbers in a new column in each spatial dataset based on the upper limit of each buffer zone (Table 3.6). All areas outside the extent of the last buffer zone which make

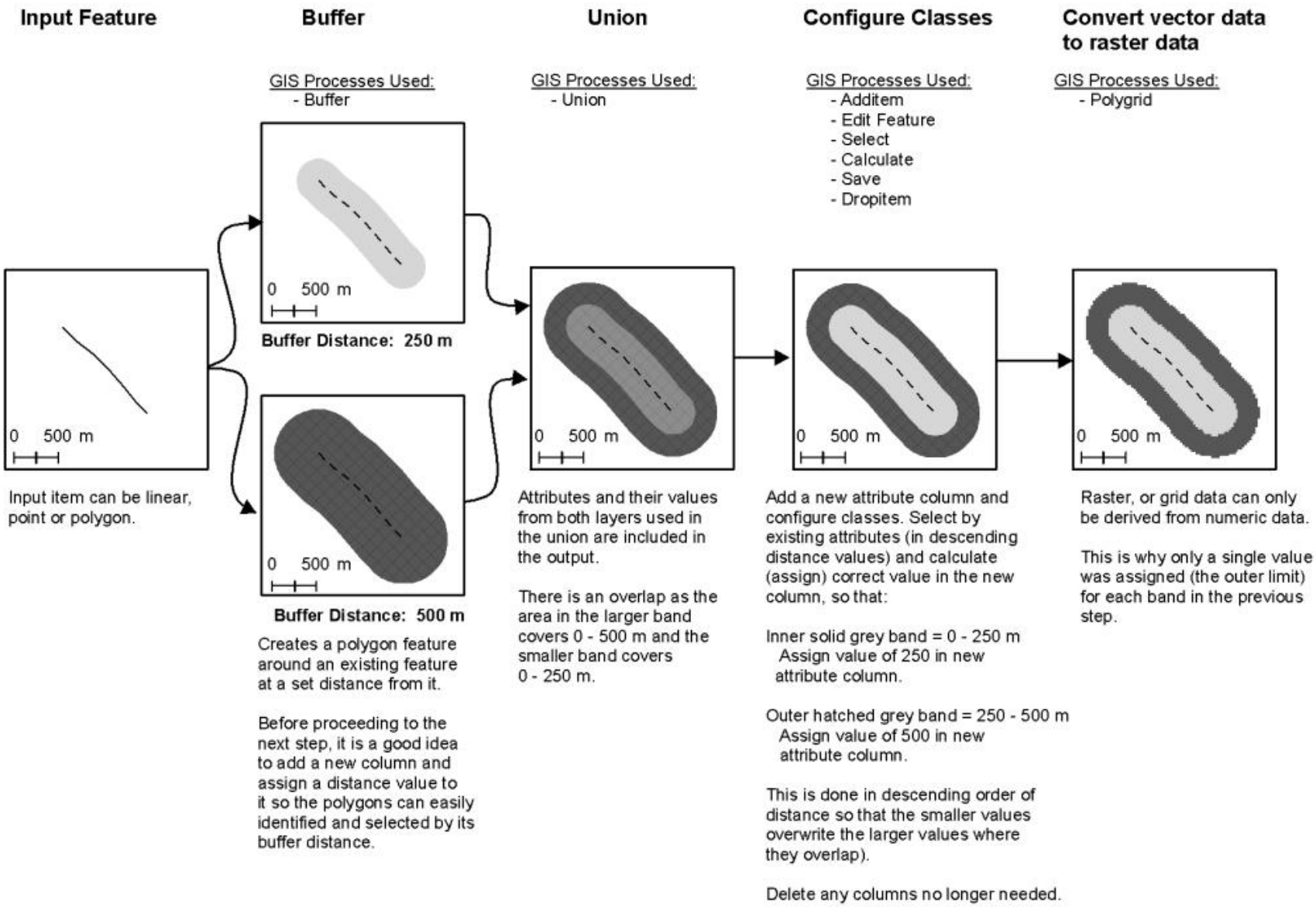


Figure 3.4 Diagram of the key GIS processes used to create the buffered linear feature layers.

up the remainder of the Waikato region were assigned zero as a code number in the same column (Table 3.6). As this data is categorical, zero was chosen as the maximum distance was not determined, and the default value of -9999 for "no data" in GIS would otherwise be assigned. For each buffered linear spatial feature (roads, faults, and rivers), a raster copy with a pixel size of 25 m was made to align with the pixels in the DEM. Fewer classes and smaller distances were used for rivers as anything greater generally took longer to process and in map production ended up being quite large in file size, and the printer sometimes refused to print. The outcome of this process was the creation of a set of buffered linear feature maps representing distance from roads (Map 3.8), distance from faults (Map 3.9), and distance from rivers (Map 3.10).

Table 3.6 Distance from linear features (roads, faults, and rivers) and the buffer distances used (Assigned Code).

Distance from Roads		Distance from Faults		Distance from Rivers	
Selected Range (m)	**Assigned Code**	**Selected Range (m)**	**Assigned Code**	**Selected Range (m)**	**Assigned Code**
0-50	50	0-100	100	0-100	100
50-100	100	100-250	250	100-200	200
100-250	250	250-500	500	200-300	300
250-500	500	500-1000	1000	300-400	400
500-1000	1000	1000-2000	2000	>400	0
> 1000	0	2000-5000	5000		
		>5000	0		

3.4.6.3 Discussion of Buffered Linear Features

3.4.6.3.1 Roads

Most of the roads are concentrated around the urban and built up areas and these tend to take on a grid shaped appearance which works best for areas of greater population (Map 3.8). A lot of the rural roads are more dendritic in appearance and are more widely spaced (Map 3.8). Noticeably there are long road lengths skirting the flanks of the Herangi Range in the southwest and Coromandel Range in the northeast and few roads crossing them (Map 3.8).

Map 3.8 Waikato Region - Distance from Roads

Distance from Roads

(m)

- 0 - 50
- > 50 - 100
- > 100 - 250
- > 250 - 500
- > 500 - 1000
- > 1000

Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

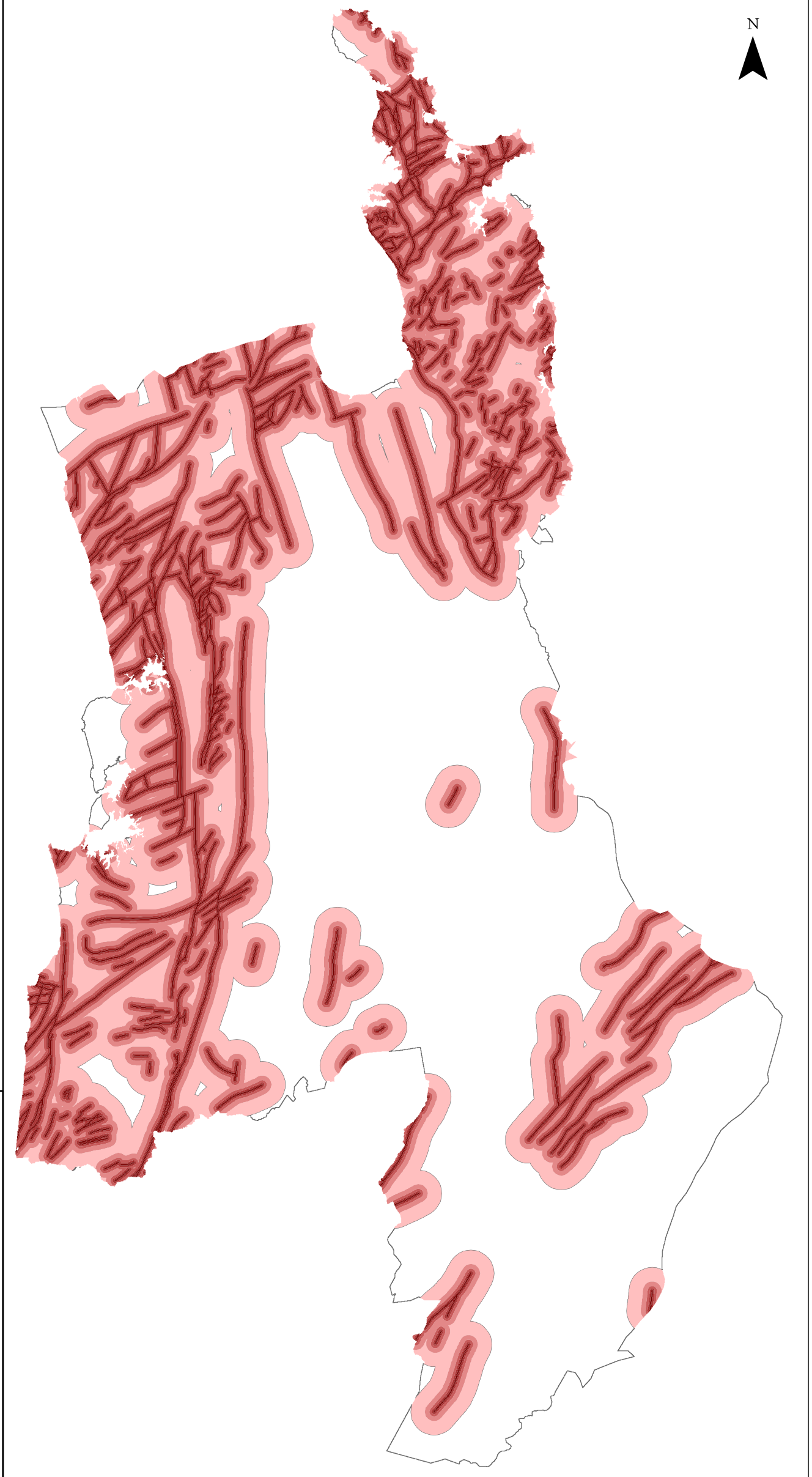
Original Data Source:
New Zealand 1:50,000 Topographic Vector Data

Map 3.9 Waikato Region - Distance from Faults

Distance from Faults

(m)

- 0 - 100
- > 100 - 250
- > 250 - 500
- > 500 - 1000
- > 1000 - 2000
- > 2000 - 5000
- > 5000



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
QMap Fault Data (Auckland and Waikato)

Earlier New Zealand Fault (1millfaults) Data

Map 3.10 Waikato Region - Distance from Rivers

Distance from Roads

(m)

- 0 - 100
- > 100 - 200
- > 200 - 300
- > 300 - 400
- > 400

Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
New Zealand 1:50,000 Topographic Vector Data

0 20 40 60 80 100 Km

3.4.6.3.2 Faults

Slope stability is usually low in slopes close to major faults, and the distance from major faults is considered to be an important factor (Gökceoglu, 2001). In the vicinity of a fault, a higher incidence of discontinuities tends to result in the rock mass in faulted terrain which weakens the material and makes it more susceptible to landsliding (Abdallah *et al.*, 2005; Ercanoglu and Gökceoglu, 2004).

Many of the faults within the Waikato Region can be found along the West Coast, around south Auckland and the Coromandel (Map 3.9). The Kerepehi Fault in the Hauraki Plains and the Wairoa Fault in the Hunua Ranges are the only known active faults in the northern part of the Waikato Region (Edbrooke, 2001a). In the Hauraki Rift, the Kerepehi fault borders half-grabens and is one of the main west-dipping normal faults (Edbrooke, 2001a). While many faults appear in the southwest onshore area, none of these have been identified as being active (Edbrooke, 2005a). The Taupo Volcanic Zone is somewhat under-represented as the data for that area is not as detailed as that derived from the QMap spatial data, but it does give an indication of where some faults are present (Map 3.9).

Most studies which have used faults have found that higher frequencies of landslides occur within close vicinity of faults. Furthermore, as distance from a fault increases there is usually a decrease in landslide occurrences (Abdallah *et al*., 2005; Gökceoglu, 2001; Gökceoglu and Aksoy, 1996; Pachauri and Pant, 1992; Saha *et al.*, 2002; Yilmaz, 2009a), although this is not always the case, and an inverse relationship can occur in which landslide occurrences are more prevalent at greater distances from a fault (Cevik and Topal, 2003; Uromeihy and Mahdavifar; 2000). The frequency of landslides within each buffer distance can easily be determined by obtaining the sum of pixels for each buffer distance, and the count of landslide pixels within each buffer distance. The number of pixels in each class increases moving away from the fault, and the landslide pixels increase to a point, then decrease (Figure 3.5).

Previous studies tend to indicate that most landslides occur within 250 – 1000 m of a fault (Cevik and Topal, 2003). Gupta and Joshi (1990) found 33% of the landslide occurrences in Ramganga Region of the Himalayas, India, were within 1 km of fault lines (Gökceoglu and Aksoy, 1996; Gupta and Joshi, 1990).

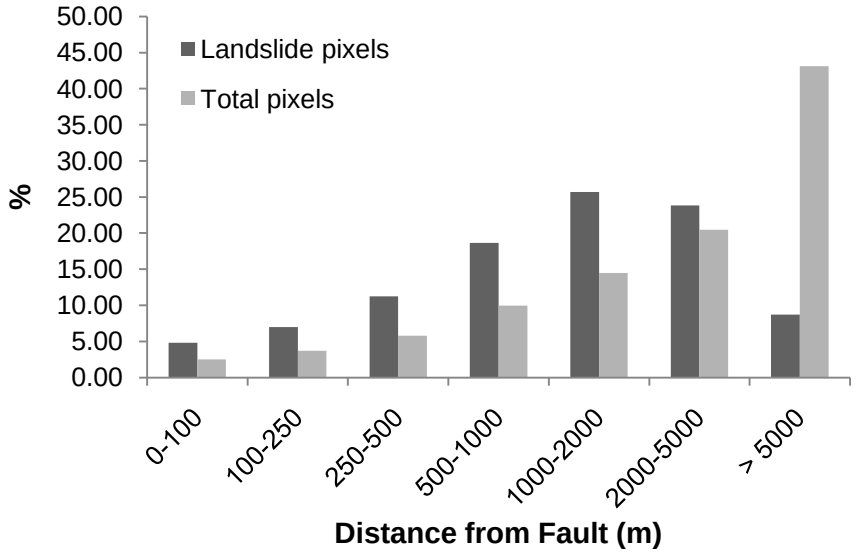


Figure 3.5 Percent of total pixels and landslide pixels with distance from faults (m)

Gökceoglu and Aksoy (1996) found 88% of landslides in the Mengen Region, Turkey, were within 250 m of major faults. If the same two thresholds are compared with values from this study, the following was found:

- approximately 42% of the landslide pixels occur within 1 km,
- ~ 12% of the total landslides occur within 250 m,
- 30% of the landslide pixels occur between 250 – 1000m.

The comparison with Gupta and Joshi (1990) is more appropriate as they used increments of 1 km up to 8 km in a study area of about 3,135 km^2, whereas Gökceoglu and Aksoy (1996) used increments of 50 m in their study area of about 120 km^2.

3.4.6.3.3 Rivers

The proximity of rivers, streams or drainage patterns to slopes is an important consideration in some landslide susceptibility and hazard assessment studies as the stability of slopes can be adversely affected by streams through slope saturation, erosion at the toe of the slope, or both (Gökceoglu, 2000; Gökceoglu and Aksoy, 1996).

The Waikato catchment is made up of a large network of rivers and streams, the largest of which is the Waikato River (Map 3.10). Several of the large areas with an absence of rivers are in them are in fact lakes, such as Lake Taupo, and many

of the lakes near Huntly, Ohinewai, Rangiriri and Te Kauwhata in the north such as Lakes Waikare and Whangape (Map 3.10).

3.4.4 Rainfall

3.4.7.1 Basis for the Choice of Rainfall Data Used

Rainfall is an important factor to consider as it is one of the main factors which trigger landslides (Abdallah *et al.,* 2005). In a region, the areas which receive the highest rainfall would generally have a greater possibility of landslide occurrence (He and Beighley, 2008). In this study, it was considered sensible to include rainfall as an input explanatory factor. While rainfall is an extrinsic variable, this study remains a susceptibility assessment as no time and/or magnitude of events are considered.

In this study, the mean and maximum monthly total rainfall for 1998 was used. This data was readily available, and although it is not as current as would be desired, it should be satisfactory in a susceptibility assessment where location of higher and lower rainfall is possibly more important than the exact rainfall.

Some authors that have included rainfall as an input factor looked at annual rainfalls over a long time frame of many years in their studies (Can *et al.,* 2005; Clerici *et al.,* 2002; Dai and Lee, 2003; Garcia-Rodriguez *et al.,* 2008; Guzzetti *et al.,* 1999; Wang and Sassa, 2006). Dai and Lee (2003) in their study, used a mean annual rainfall between the period of 1961-1991, and has a similar time period to that used in the study by Garcia-Rodriguez *et al.* (2008) who used mean annual rainfall between the years 1961-1990.

3.4.7.2 Processing the Rainfall Data

Mean and maximum monthly total rainfalls (Landcare and DoC, 1998) were clipped by a polygon mask which extends beyond the Waikato Region in order to catch all pixels which may otherwise be excluded. The raster output from this was converted to vector data where edits were made by selecting each defined 50 mm value range and writing the upper limit to a new column in the attribute table (Table 3.7). The vector covers were converted back to raster grid format, with a

cell size of 25 m set to align with the DEM, although the resolution for both sets of rainfall data remain at 1 km. After being clipped to the extent of the DEM, parameter maps of the maximum monthly rainfall (Map 3.11) and mean monthly rainfall (Map 3.12) were created.

Table 3.7 Monthly maximum and monthly mean rainfall classes and assigned code.

Maximum Rain		Mean Rain	
Selected Range (mm)	**Assigned Code**	**Selected Range (mm)**	**Assigned Code**
≤ 150	150	≤ 100	100
> 150 - 200	200	> 100 - 150	150
> 200 - 250	250	> 150 - 200	200
> 250 - 300	300	> 200 - 250	250
> 300 - 350	350	> 250 - 300	300
> 350	400	> 300	350

3.4.7.3 Rainfall in the Waikato Region

The higher mean and maximum monthly rainfalls occur in the Herangi, Kaimai, and Coromandel Ranges, and around Mount Ruapehu (Maps 3.11 and 3.12). The Hamilton Basin and Hauraki Plains receive less rainfall (Maps 3.11 and 3.12). When both maximum (Map 3.11) and mean rainfall (Map 3.12) maps are compared there is only a slight difference in the spatial pattern. This tends to be the case for most of New Zealand, so it is possible to use mean rainfall as a surrogate for maximum rainfall (Dymond *et al.*, 2010). The frequency of pixels for each rainfall value was plotted for both monthly mean (Figure 3.6 A) and monthly maximum (Figure 3.6 B) rainfalls. Overall the distributions for both sets of rainfall data are quite similar, only the monthly maximum rainfall tends to have a more defined second peak (Figure 3.6 B).

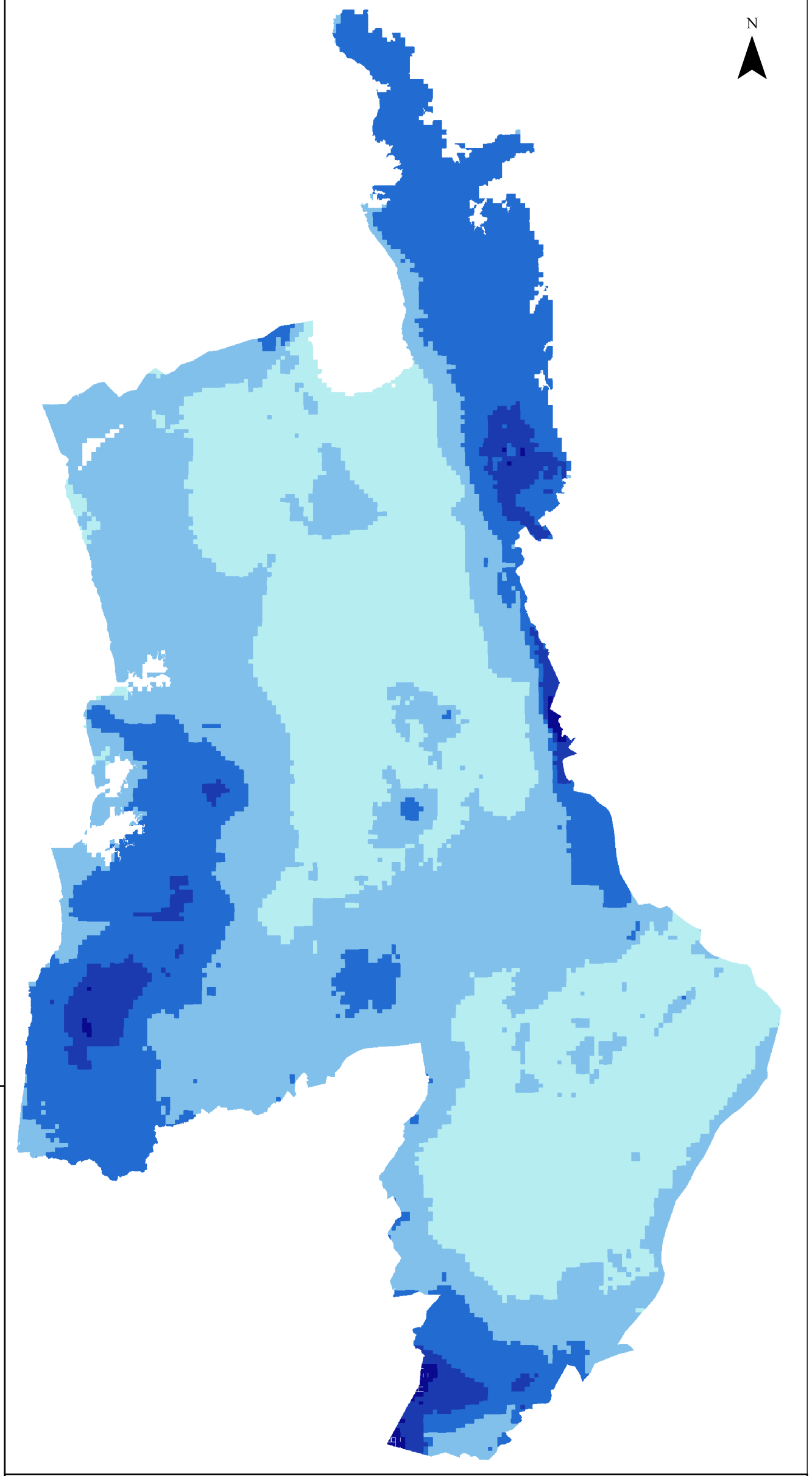
As there is a slight difference between both the mean monthly rainfall and maximum monthly rainfall both cannot be used in the same analysis. Both however, can be trialled separately as one rainfall dataset may have more of an influence than the other.

Map 3.11 Waikato Region - Maximum Monthly Rainfall

1998 Maximum Monthly Rainfall

(mm)

- > 300 - 350
- > 250 - 300
- > 200 - 250
- > 150 - 200
- > 100 - 150



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

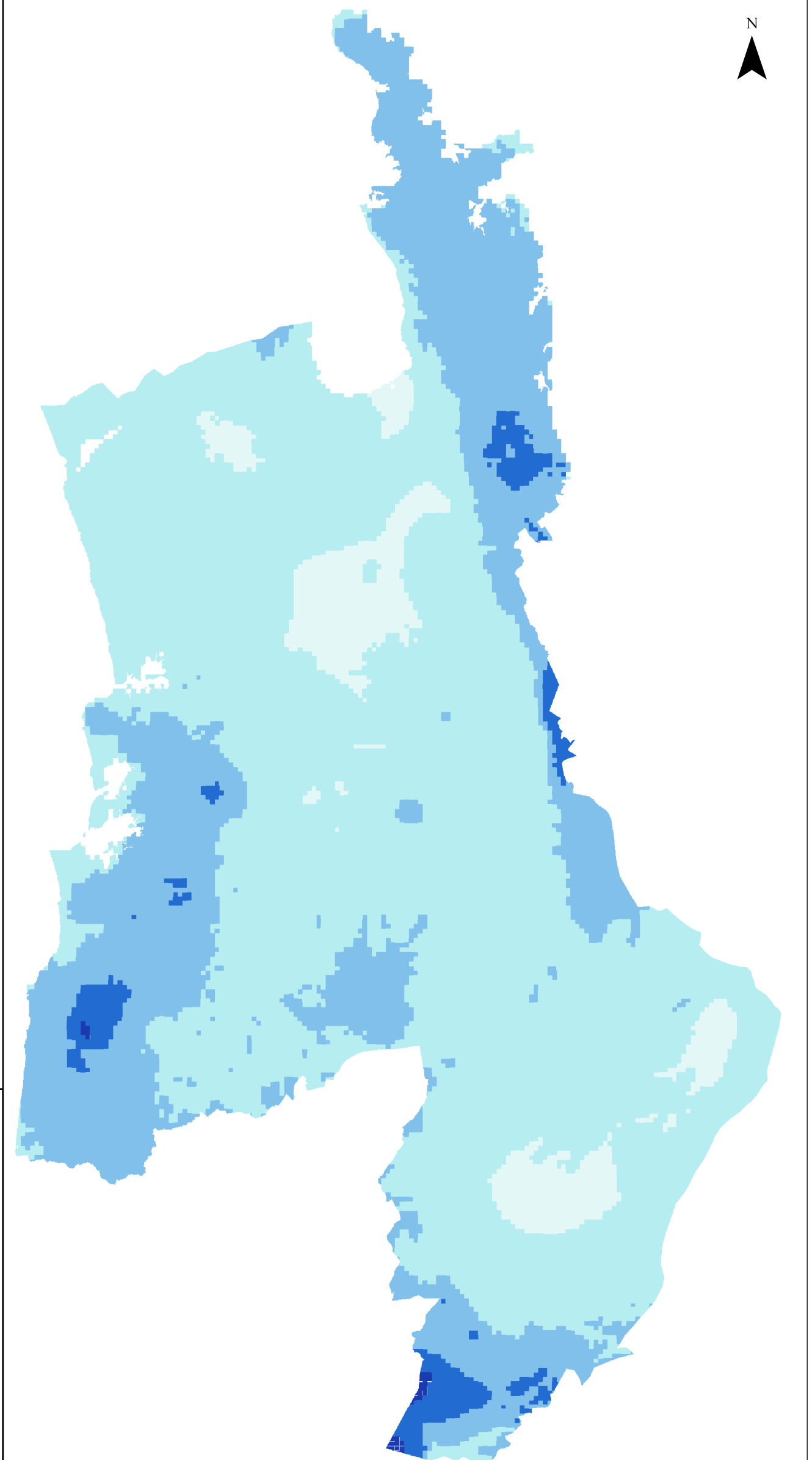
Original Data Source:
New Zealand Climate Estimates data

Map 3.12 Waikato Region - Mean Monthly Rainfall

1998 Mean Monthly Rainfall

(mm)

- > 250 - 300
- > 200 - 250
- > 150 - 200
- > 100 - 150
- > 50 - 100



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Original Data Source:
New Zealand Climate Estimates data

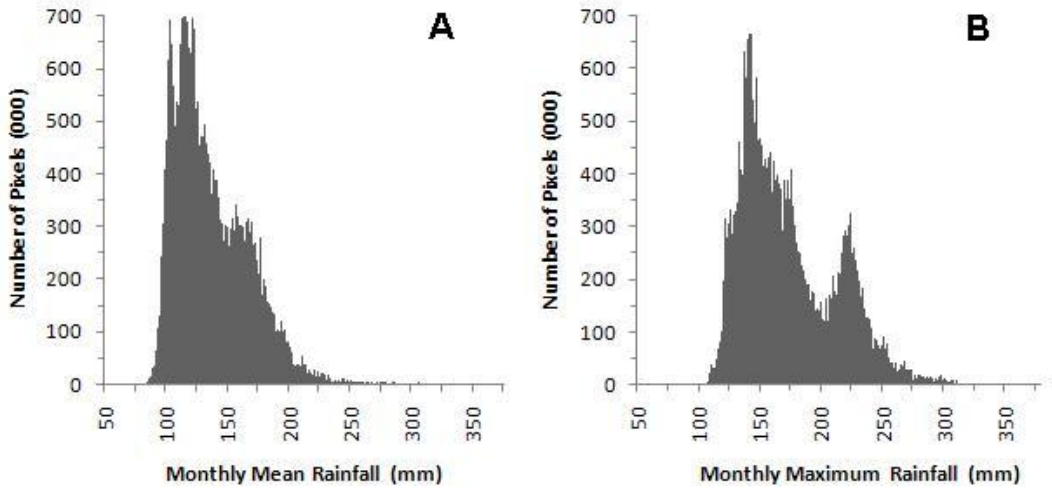


Figure 3.6 Plot of rainfall pixel frequencies within the Waikato Region for both monthly mean rainfall (A) and monthly maximum rainfall (B).

3.4 Data Extraction for External Analysis

A grid of equidistant pixel centroid points spaced 25 m apart in both X and Y directions was created (Figure 3.7). This was done by first using the ungenerate function to convert a raster layer such as aspect to a text file then generate the output from the ungenerate function to create a point layer. The point layer comprises of a single centroid point for each cluster of pixels of the same value (Figure 3.7). This was applied to the aspect raster data as there was more chance of finding singule pixel points. Using the attribute table the X and Y maximum and minimum values can easily be obtained by sorting in ascending or descending order. Highlighting each of these entries makes them easy to locate and possible to zoom to location, which allows a check to see whether the point represents the centre of a single pixel or a group of pixels. The maximum and minimum X and Y coordinates were noted.

An Excel file was set up with ten interlinked worksheets containing five columns of 12,612 points in such a way that for each of the 12,612 rows the Y coordinate value would increase by 25 m and after the last row a new column would begin and the X coordinate value would increase by 25 m. This set up was duplicated to cover the whole extent of the region; the only value manually changed was the

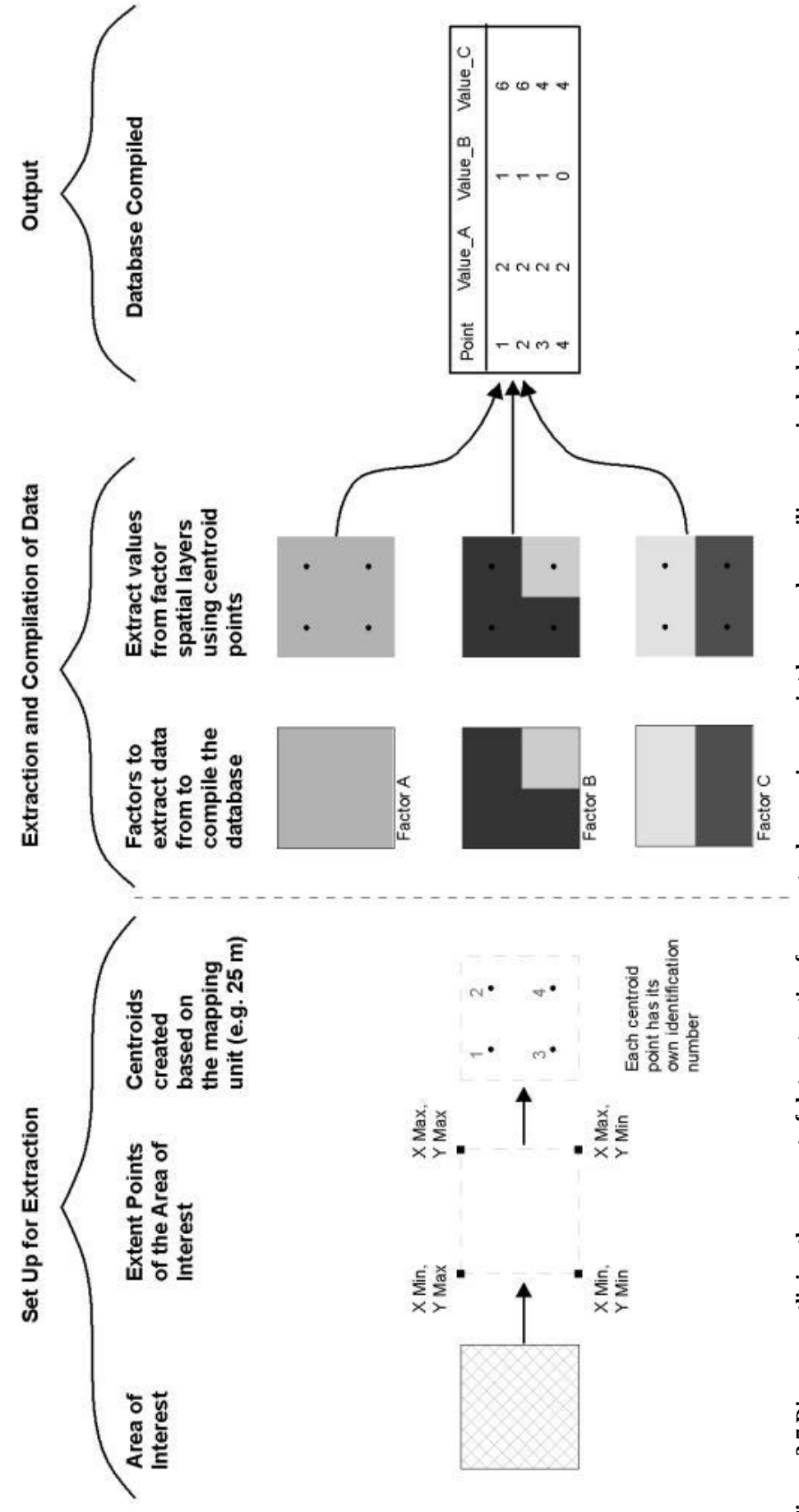


Figure 3.7 Diagram outlining the concept of data extraction from raster layer using a point layer and compiling as a single database.

first X coordinate in the first worksheet. The last X coordinate on the last worksheet of each Excel file was recorded and 25 m added to this value gave the first value of the next Excel file.

All Excel worksheets were imported into Access databases in groups of 1,000 columns of point coordinates (20 Excel files to each initial Access Database). A record was kept as each worksheet was imported, and when each Excel file had been fully imported into Access it was moved to a different folder to avoid any mistakes. Seven Access databases were created, and from this the data was exported as dBASE Tables and converted to shapefiles in ArcCatalog. These point shapefiles were added to ArcMap with the shapefile version of the DEM boundary, which was used to clip the seven sets of points to the region area. The clipped shapefile attribute tables were exported as text files then imported into a single Access database.

The database was divided into three partitions, Auckland, Waikato and Rotorua closely following the QMap divisions as there was too much data for one or two shapefiles. These were added to ArcMap along with each of the output raster layers created through the scripted GIS process (Figure 3.7). The "Extract Values to Points" function in ArcToolbox was utilised so the input point layer and raster layer were specified and from these a new point layer was created with the raster values in a new field under the default title of "RASTERVAL". The title of this field could not be changed in ArcMap but could when the attribute table was imported into Access. The need to do this arises as the extraction process will not work if there is an existing "RASTERVAL" field. The fields were renamed according to the variable the data was taken from. Categorical variables had "Code" appended to the field name (e.g. SlopeCode, SoilCode) and for continuous variables "Val" was appended (e.g. SlopeVal, AvRainVal). As this data juggling process progressed the size of the files increased, and at some point each of the three initial grid point sets were divided in half, this continued until there were eleven separate shapefiles covering different portions of the region.

Eighteen extraction procedures were required to compile the landslide and predictor variable data into the same files. The size of the information was too large to fit into a single Access database, so has been compiled into two separate

databases, Auckland and Waikato in one, and Rotorua in the other (Appendix 3.4). A query was made to exclude any no data values which occur as a result of differing coastlines between the datasets. Both databases were then exported as text files for use in STATISTICA sampling and statistical analysis.

3.6 Conclusion

Spatial data of both the landslide locations and selection of predictor variables were obtained, manipulated, processed and mapped using GIS to obtain eleven parameter maps (Maps 3.1 – 3.3, 3.5 – 3.12; Appendix 3.2) and a landslide inventory map (Map 3.4; Appendix 3.2). The basis for the classification of each variable was discussed, and an outline of the scripted GIS processes (Appendix 3.1) given. Several interesting observations were made regarding the ranges within the Waikato Region, such as the Herangi, Coromandel, and Rangitoto Ranges. The ranges are generally areas of high slope, elevation, and rainfall and fewer roads. The landslide inventory indicated these areas were more frequent to landslide events, especially around the Herangi Range in the southwest and Coromandel Range in the northeast.

The soil and geology vary throughout the region, most the brown soils and andesite, dacite and diorite are found in the Coromandel; while The Taupo Volcanic Zone is predominantly pumice soil, and ignimbrite in its geology. The area south of Auckland and the West Coast are quite a complex mix for both soil order and geology.

Topographic parameters, slope, elevation and aspect, were derived from a digital elevation model and a continuous and categorical data set was obtained for each. A set of both continuous and categorical data was also derived from the monthly mean and monthly maximum rainfall data. Categorical polygon vector data for geology, land cover, and soils was simplified to fewer classes and a raster copy made with a pixel size of 25 m to match that of the DEM. Linear features underwent various GIS processes to create a series of classed polygon bands around them and a raster copy was also made for each.

A point layer was created and used to extract data from the raster versions of the parameters, this included the eleven categorical parameters, five continuous parameters (DEM derived, and rainfall), and the landslide inventory. The extracted data was compiled in a database (Appendix 3.4) to be used in the statistical analyses.

CHAPTER FOUR

A Bivariate Approach Using Weights of Evidence

4.1 Introduction

A bivariate approach is used to determine and map landslide susceptibility based on an individual assessment of each predictor factor using the weights of evidence technique. An outline of the weights of evidence method and how the probabilities and weights are determined is given. In addition a test of conditional independence using chi squares for each factor is investigated as part of the process of finding a significant model. The impact of systematically removing a single factor from the model is also considered. Classification techniques and model goodness of fit are also discussed.

4.2 The Weights of Evidence Model

4.2.1 Background

Using a quantitative data-driven method such as the weights of evidence model it is possible to combine datasets and estimate by statistical means the relative importance of each of the predictor factors in relation to landslide events (Thiery *et al.,* 2007). The weights of evidence model is a (log – linear) Bayesian probability model which was originally developed for mineral potential assessment and has since been applied to landslide susceptibility assessment (Lee and Choi, 2004; Thiery *et al.,* 2007; Dahal *et al.,* 2008a).

The weights of evidence is easily implemented and less time consuming than some other methods (Dahal *et al.,* 2008a; Soeters and van Westen, 1996; Süzen and Doyuran, 2004) which can be applied readily using GIS (Dahal *et al.,* 2008a).

The weights of evidence model calculates the weight for each predictive factor (*B*) based on the probability of its presence or absence with there being a landslide (*L*) or not (*L*-) within the area (Dahal *et al.,* 2008a). The output of this is a landslide susceptibility map which gives an indication of which areas are more likely to experience landslides based on past occurrences and the causative factors (Neuhäuser and Terhorst, 2007).

4.2.2 Assumptions

The fundamental assumption of this approach is that factors and conditions which have resulted in past landslides will have a similar or equal effect in landslide occurrence in the future (Neuhäuser and Terhorst, 2007). Historical landslide data is thus necessary to apply the weights of evidence method, as past landslide occurrences are used to determine the weights for the contributing and/or causative factors (Neuhäuser and Terhorst, 2007). In order to describe future landslide hazard, it is essential that complete and suitably representative input parameter spatial datasets are used and that the investigator has good knowledge of these factors (Neuhäuser and Terhorst, 2007).

Results from the weights of evidence model are greatly dependent on the quality of the landslide inventory map and the number of events and their estimation probabilities included in the model (Thiery *et al.,* 2007). The estimated weights can be stable and realistic if the study area has a reasonable coverage of landslide events, but cautious interpretation of the result is required if the study area is characterised by rare events where probabilities are very low (Thiery *et al.,* 2007).

A further assumption is that for mapped landslides, the causative factors are considered to be reasonably constant with time (Neuhäuser and Terhorst, 2007). However this can only be assumed for a single landslide type as causes for each type vary, and ideally the method should be applied separately to each landslide type (Neuhäuser and Terhorst, 2007).

In weights of evidence, the most important assumption in the application of the Bayes probability theory model is that the factors are conditionally independent of each other with regard to landslide (L) occurrence (Neuhäuser and Terhorst, 2007; Thiery *et al.,* 2007). Meaning that for locations where landslides have been

identified, neither condition B_1 (e.g. Slope) or condition B_2 (e.g. Geology) should be dependent on each other (Neuhäuser and Terhorst, 2007). Neuhäuser and Terhorst (2007) describe this assumption for factors B_1 and B_2 as:

$$P\{B_1 \cap B_2|L\} = P\{B_1|L\} \times P\{B_2|L\} \quad \textbf{1}$$

which is basically a simplification of the relationships in nature, but allows an individual assessment of the factors. In order to assume conditional independence, the causative factors require a check of independence, in which dependent factors are rejected from subsequent analyses (Neuhäuser and Terhorst, 2007). Statistical tests of conditional independence such as χ^2-test, omnibus test and new omnibus test can be implemented to meet this need (Thiery *et al.*, 2007).

4.2.3 Weights of Evidence Statistical Approach in Theory

The two most important concepts of the Bayesian approach are the prior and posterior probabilities (Neuhäuser and Terhorst, 2007; Thiery *et al.*, 2007). The probability that a terrain unit contains a landslide is the prior probability which is determined before predictive variables are investigated and is estimated based on the total landslide area within the study area (Thiery *et al.*, 2007). The prior probability (P_{pr}) of occurrence is expressed by:

$$P_{pr}\{L\} = \frac{N\{L\}}{N\{T\}} \quad \textbf{2}$$

where $N\{L\}$ is the number of cells containing a landslide occurrence, L, and $N\{T\}$ is the total number of cells (or pixels) in the study area (Lee and Choi, 2004; Sharma and Kumar, 2008; Song *et al.*, 2008).

Posterior (or conditional) probability expresses the probability that a landslide event (L) will occur with the presence of a parameter class (B) (Neuhäuser and Terhorst, 2007). The four possible conditional probabilities (Figure 4.1 and Table 4.1) are estimated based on the density (pixel count) of landslides for each class within each predictor parameter (Thiery *et al.*, 2007).

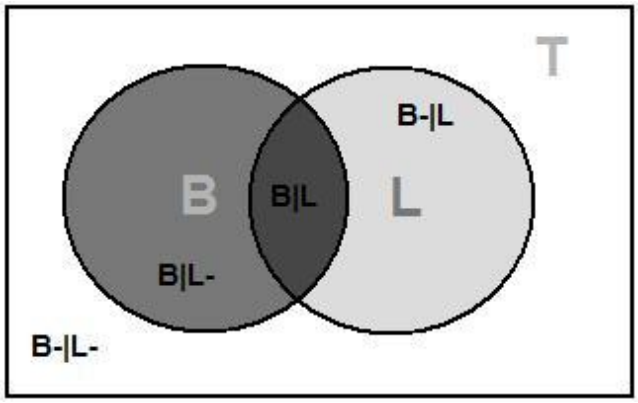


Figure 4.1. A Venn diagram which illustrates the combinational probabilities for the presence (B) or absences (B-) of a class within a parameter map in relation to landslides occurrence (L) or non-occurrence (L-) within the total study area (T). The darker circle (B) represents the presence of a class within a parameter map. In the area outside this circle the class is considered absent (B-), and is explained by the other classes within the parameter map. The same theory applies to the lighter circle (L), which explains landslide occurrence (L), and the area outside this is landslide non-occurrence (L-). Where the two circles overlap (B|L) is the probability the class in the parameter and landslides are both present. The remaining parts of the circles are the probability of the presence of the class with landslide non-occurrence (B|L-); and the probability of a landslide occurrence when the same parameter class is absent (B-|L). The area in white (B-|L-) is the probability there is an absence of both landslide (L) and the parameter class (B).

Table 4.1 The four possible combinations based on the presence or absence of the predictor variable and the presence or absence of a landslide.

Landslide	Class within a given variable	
	Present (B)	Absent ($\bar{B}$)
Present (L)	$B\|L$	$\bar{B}\|L$
Absent ($\bar{L}$)	$B\|\bar{L}$	$\bar{B}\|\bar{L}$

The conditional probabilities (P_c) of being inside and outside the predictor pattern B (the presence or absence of a given parameter class) given the presence of landslide occurrence (L) (Table 4.1) are $P_c\{B|L\}$ and $P_c\{\bar{B}|L\}$, respectively (Lee and Choi, 2004). Both prior and posterior probabilities are integrated into Bayes theorem (Equations 3 and 4) to determine the posterior probability $P_c\{L|B\}$ and $P_c\{L|\bar{B}\}$ of a landslide occurrence given the presence and absence of a given class within a parameter map (Lee and Choi, 2004; Neuhäuser and Terhorst, 2007):

$$P_c\{L|B\} = \frac{P\{L \cap B\}}{P\{B\}} = P\{L\}\frac{P\{L|B\}}{P\{B\}} \quad \textbf{3}$$

4

$$P_c\{L|\bar{B}\} = \frac{P\{L \cap \bar{B}\}}{P\{\bar{B}\}} = P\{L\}\frac{P\{L|\bar{B}\}}{P\{\bar{B}\}}$$

In the application of the weights of evidence method, the presence of the predictor variable at the landslide locations is indicated by a positive weight (W^+), while a negative weight (W^-) indicates the absence of the predictor variable (Dahal *et al.*, 2008a). W^+ gives an indication of the level of positive correlation between the presence of the predictor factor and landslides, and similarly, W^- the magnitude of negative correlation between the absence of the predictor factor and landslides (Neuhäuser and Terhorst, 2007; Dahal *et al.*, 2008a).

Three similar expressions appear in the literature; however each appears to give the same value for the weighted contrast (W_f). The first is based on the binary predictor pattern (factor) being present or absent (Lee and Choi, 2004; Thiery *et al.*, 2007; Dahal *et al.*, 2008a, b; Sharma and Kumar, 2008; Zahiri *et al.*, 2006), and is given by:

5

$$W^+ = ln\frac{P_c\{B|L\}}{P_c\{B|\bar{L}\}} = ln\frac{P_c(factor\ and\ landslide\ both\ present)}{P_c(factor\ present\ and\ landslide\ absent)}$$

6

$$W^- = ln\frac{P_c\{\bar{B}|L\}}{P_c\{\bar{B}|\bar{L}\}} = ln\frac{P_c(factor\ absent\ and\ landslide\ present)}{P_c(factor\ and\ landslide\ both\ absent)}$$

where P_c is the conditional probability, B the presence of the predictive factor, and $\bar{B}$ the absence of the predictive factor, L is the presence of a landslide event, while $\bar{L}$ is the absence of a landslide event (Dahal *et al.*, 2008a).

The second expression (Neuhäuser and Terhorst, 2007) appears to be based on the landslide event being present or absent, using the values obtained from Equations 3 and 4 and variations thereof, and is expressed as:

7

$$W_j^+ = ln\frac{P_c\{L|B_i\}}{P_c\{L|\bar{B}_i\}} = ln\frac{P_c(landslide\ and\ factor\ class\ both\ present)}{P_c(landslide\ present\ and\ factor\ class\ absent)}$$

8

$$W_j^- = ln\frac{P_c\{\bar{L}|B_i\}}{P_c\{\bar{L}|\bar{B}_i\}} = ln\frac{P(landslide\ absent\ and\ factor\ class\ present)}{P(landslide\ and\ factor\ class\ both\ absent)}$$

where W_j^+ is the likelihood ratio in which landslide presence is evident in the presence or absence of the predictor *B* from a number of *i* evidences, and W_j^- expresses the same relationship in the absence of landslide evidence (Neuhäuser and Terhorst, 2007).

However, on closer inspection both the W_j^+ and W_j^- equations are divided by the absence of *B*. The above two equations thus appear to mean that provided a landslide occurs either the predictive variable does or does not occur.

The third expression (Dahal *et al.*, 2008b; Song *et al.*, 2008) is based on the odds but simplifies to the first expression:

9

$$W_i^+ = ln\frac{O(L|B)}{O(L)} = ln\frac{\left\{\frac{P_{pr}(L)}{[1-P_{pr}(L)]}\right\}\times\left[\frac{P_c(B|L)}{P_c(B|\bar{L})}\right]}{P_{pr}(L)/[1-P_{pr}(L)]} = \ln\frac{P_c(B|L)}{P_c(B|\bar{L})}$$

10

$$W_i^- = ln\frac{O(L|\bar{B})}{O(L)} = ln\frac{\left\{\frac{P_{pr}(L)}{[1-P_{pr}(L)]}\right\}\times\left[\frac{P_c(\bar{B}|L)}{P_c(\bar{B}|\bar{L})}\right]}{\left\{\frac{P_{pr}(L)}{[1-P_{pr}(L)]}\right\}} = ln\frac{P_c(\bar{B}|L)}{P_c(\bar{B}|\bar{L})}$$

where $O(L|\bar{B})$ is the odds of landslide to no landslide for a factor class, and *O*(L) is the odds of landslide presence to landslide absence for the total area (Song *et al.*, 2008).

The weight contrast (W_f) is the difference between the two weights (Equation 11), and the overall spatial association between landslides and the predictor variables is represented by the magnitude of this value (Dahal *et al.*, 2008a; Zahiri *et al.*, 2006).

11

$$W_f = W^+ - W^-$$

The W_f weights for each class of each causative factor are assigned to their respective thematic layer to produce weighted thematic maps which when combined, are numerically added (Equation 12) to produce a landslide susceptibility index map (Dahal *et al*., 2008a).

$$LSI = W_f Slope + W_f Aspect + \ldots + W_f Land\ Use \quad (12)$$

4.2.4 Normalised Contrast

If W_f is positive, there is a positive spatial association between the predictor and landslide occurrence, but if W_f is negative the spatial association will be negative (Neuhäuser and Terhorst, 2007). A positive spatial association would indicate that there has been past evidence of landslide occurrence in the presence of the same predictor. A negative spatial association would indicate few or no known past occurrences in the presence of the predictor. A significant positive contrast would suggest the pattern is a useful predictor, based on its normalised contrast (Raines, 1999). The normalised contrast is a ratio of the contrast to its standard deviation which can provide a measure of confidence (Ghosh *et al.*, 2009; Neuhäuser and Terhorst, 2007; Raines, 1999):

$$Normalised\ C = \frac{W_f}{sC} \quad (13)$$

where sC is the standard deviation of the contrast (W_f):

$$sC = \sqrt{s^2(W^+) + s^2(W^-)} \quad (14)$$

where $s^2(W^+)$ and $s^2(W^-)$ are the variances of W^+ and W^- (Ghosh *et al.*, 2009; Lee and Choi, 2004):

$$s^2(W^+) = \frac{1}{N(B \cap L)} + \frac{1}{N(B \cap \bar{L})} = \frac{1}{N(B \cap L)} + \frac{1}{[N(B) - N(B \cap L)]} \quad (15)$$

16

$$s^2(W^-) = \frac{1}{N(\bar{B} \cap L)} + \frac{1}{N(\bar{B} \cap \bar{L})}$$

$$= \frac{1}{[N(L) - N(B \cap L]} + \frac{1}{[N(T) - N(L) - N(B) + N(B \cap L)]}$$

For an overall positive significance at a confidence of approximately 97.5%, the positive normalised contrast should be > 1.96 (Ghosh *et al.,* 2009; Raines, 1999). A negative normalised contrast < -1.96 indicates a negative overall significance (Ghosh *et al.,* 2009).

4.3 Conditional Independence

In many probability and statistically applied methods there may be errors as the assumption is that the population has a normal distribution, whether the distribution is known or not it, it is set as a normal distribution (Lee and Choi, 2004). To address these errors non-parametric statistics such as a test for dependence using χ^2 can be employed (Lee and Choi, 2004). To assess this, a contingency table is used where classes within a factor are compared with classes in another factor in a row by column basis using only the data for which landslides are present under the assumption that both factors are independent of each other (Lee and Choi, 2004; Neuhäuser and Terhorst, 2007). The contingency table (Table 4.2) looks at the relationship between each class of one causative factor (F_1) and with those of another (F_2) in four scenarios with landslide occurrence (L). For instance, the presence of both class 1 of F_1 (F_1C_1) and class 1 of F_2 (F_2C_1); class 1 of F_1 with all other classes within F_2 except class 1; class 1 of F_2 with all other classes within F_1 except class 1; and all other classes (except class 1 for both) within both F_1 and F_2.

Table 4.2 Example adapted from Lee and Choi (2004) of a contingency table in which the number of occurrences where a class in one factor intersects (∩) or overlaps with a class in another factor, within landslide only areas are tabulated. Where N(L) is the total number of landslide pixels, the row totals give the total pixels with landslides in each class of factor 2, and the column totals give the total pixels with landslides in each class of factor 1. The four combinations look at the count of landslide pixels in each class of each factor in relation to those of the other factor where both factor may be present in the same location.

		Factor 1 (F_1)			
		Class 1 (C_1)	~	Class n (C_n)	Totals
Factor 2 (F_2)	Class 1 (C_1)	$N(F_1C_1 \cap F_2C_1 \cap L)$	~	$N(F_1C_n \cap F_2C_1 \cap L)$	$N(F_2C_1 \cap L)$
	~	~	~	~	~
	Class n (C_n)	$N(F_1C_1 \cap F_2C_n \cap L)$	~	$N(F_1C_n \cap F_2C_n \cap L)$	$N(F_2C_n \cap L)$
	Totals	$N(F_1C_1 \cap L)$	~	$N(F_1C_n \cap L)$	N(L)

The observed values, or count of landslides, for each scenario are used to calculate the expected values (Moore and McCabe, 2003):

$$Expected\ Count = \frac{(Row\ Total - Column\ Total)}{Total\ Observations} \quad \textbf{17}$$

The observed and expected values are then used to determine the chi square (Lee and Choi, 2004; Moore and McCabe, 2003; Neuhäuser and Terhorst, 2007):

$$\chi^2 = \sum \frac{(Observed\ Count - Expected\ Count)^2}{Expected\ Count} \quad \textbf{18}$$

The degrees of freedom can also be determined (Lee and Choi, 2004; Moore and McCabe, 2003):

$$df = (r-1)(c-1) \quad \textbf{19}$$

where r is the number of rows, and c the number of columns (Lee and Choi, 2004; Moore and McCabe, 2003). In a direct comparison of landslide frequencies between a class in one factor and a class in another factor, there is only 1 degree of freedom. The measured χ^2 is then compared with the theoretical χ^2 at a given level of significance and based on the same degrees of freedom (Neuhäuser and Terhorst, 2007).

4.4 Applying the Weights of Evidence Method

4.4.1 Determining Probabilities and Weights

The database was queried and the count of pixels with and without landslide occurrence was obtained (Appendix 4.1) and Equation 2 applied so that:

20

$$P\{L\} = \frac{N\{L\}}{N\{T\}} = \frac{551{,}924}{38{,}952{,}689} = 0.01417\ (4\ sf)$$

A series of queries was then run to obtain the count of pixels with and without landslides for each class of each variable. The probabilities of each combination (Table 4.1) were calculated (Appendix 4.1) using the following equation:

21

$$P\{C\} = \frac{NPix(C)}{NPix(Total)}$$

Where *NPix*(Total) is the total number of pixels in the study area, *NPix(C)* is the number of pixels for any combination (C) in Table 4.1.

The probability of occurrence for each class was calculated by substituting L for B in Equation 2. Following this Equations 3 and 4 were then implemented to determine the conditional probabilities (Appendix 4.1). The positive and negative weights were calculated using the three expressions (Equations 5 – 10) and while the resulting values for the positive and negative weights differed, the resulting weight contrast (Equation 11) values were identical (Appendix 4.1).

4.4.2 Interpreting Positive Weightings and Contrasts

The magnitude of the positive weight (W^+ or W_j^+) is said to give an indication of the positive correlation between a predictor variable's class and landslide occurrence (Neuhäuser and Terhorst, 2007; Dahal *et al.,* 2008a). In addition, W_f is said to be an indication of positive and negative spatial association, and the normalised contrast the significance of the relationship (Neuhäuser and Terhorst,

2007). To investigate this, the three variations of the technique were trialled (Appendix 4.1). It was found that the W^+ and W^- values varied for each factor (Figure 4.2 A and B), but the difference between them (W_f) did not (Figure 4.2 C).

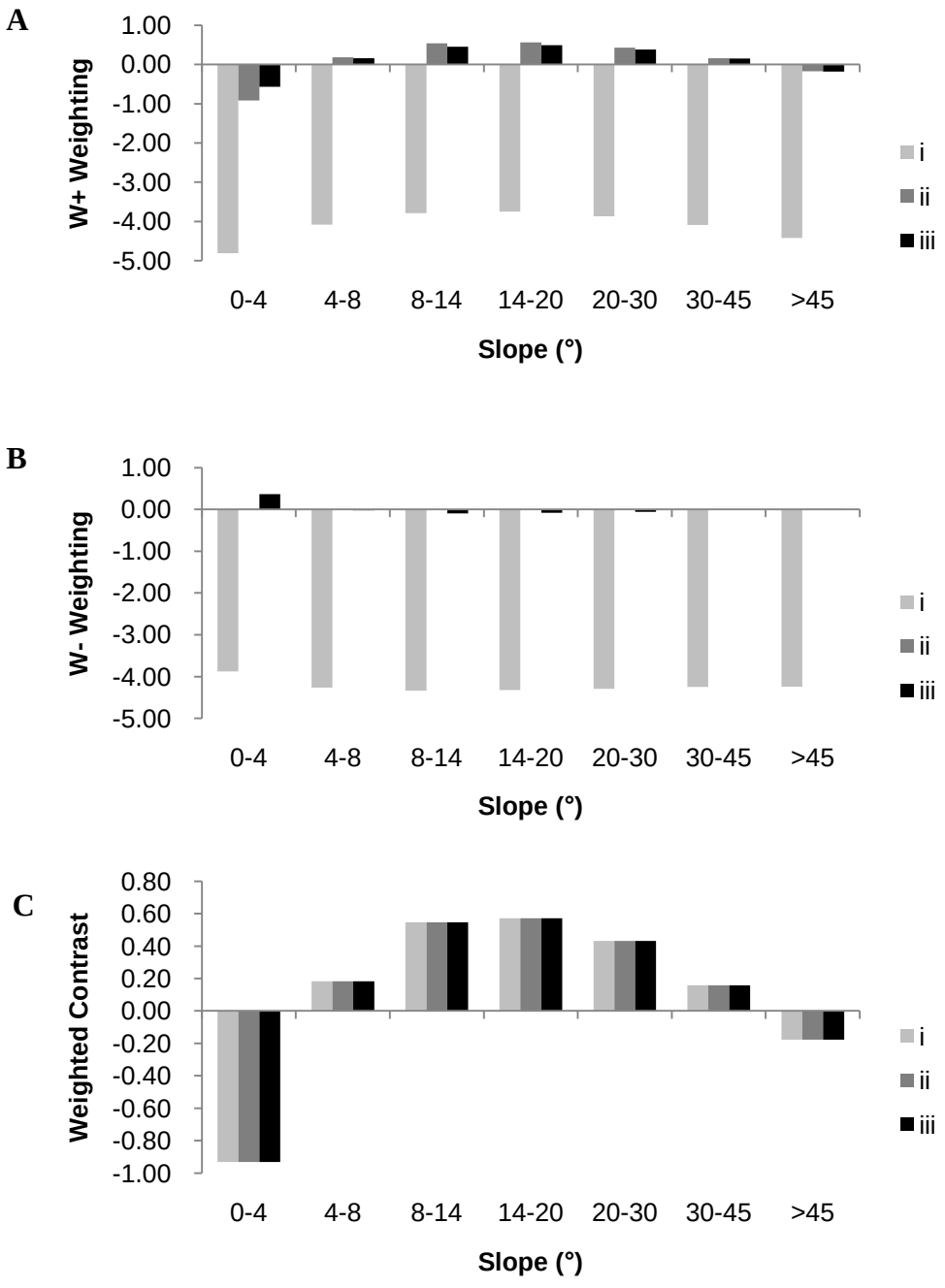


Figure 4.2 An example of varying positive weightings (A); varying negative weightings (B); and identical weighted contrasts (C), when slope classes are considered, following the expressions given in Lee and Choi (2004) (i); Sharma and Kumar (2008) and Neuhäuser and Terhorst (2007) (ii); and Song *et al.*(2008) and Ghosh *et al.* (2009) (iii).

The first approach (i) resulted in negative values for both positive and negative weightings (Figure 4.2 A and B) not only for all slope classes, but for all classes in each of the other predictor factors also. The other two approaches (ii and iii) resulted in a mix of positive and negative values for both positive and negative weightings (Figure 4.2 A and B), and give more of an indication of which classes in a given factor are more frequently present or absent in an area of landslide occurrence. As the W^+ varies between each of the three variations in techniques followed, it was decided that the weighted contrast (W_f) would give a better measure of influence as this value remained constant for each (Figure 4.2. C). Based on this, the resulting trends in weighted contrast for each factor will be explored, with the inclusion of the normalised contrast as an indication of the "significance".

Slope angle has a positive effect in the range between 4 – 45° with most influence in the 8 – 20° range based on the positive weighted contrasts (Figure 4.3, A). Slopes of very low gradient (0 – 4°) are expected to have a low frequency of landslides (Lee and Sambath, 2006; Lee and Talib, 2005), and this is reflected by the negative contrast (Figure 4.3, A). Steeper slopes (> 45°) such as natural outcropping bedrock, may not necessarily be susceptible to landslides (Lee and Sambath, 2006; Lee and Talib, 2005) so the negative contrast obtained is quite feasible (Figure 4.3, A).

In terms of aspect, flat or non-orientated areas exhibit a significant negative spatial association (normalised contrast < -1.96) with landslide occurrence (Figure 4.3, B). In other landslide susceptibility assessments (Abdallah *et al.*, 2005; Lee and Dan, 2005; Lee and Lee, 2006; Lee and Talib, 2005) that have investigated aspect, south facing slopes were found most susceptible to landslides. In this study, south-facing slopes have the greatest contrast value which implies these slopes are most susceptible, but as indicated by the normalised contrast it is not that significant when compared to the other aspect classes (Figure 4.3 B).

Elevations between 0 – 25 m and 50 – 300 m display some positive association with landslides, but are not considered to be significant as the normalised contrast is < 1.96 (Figure 4.3 C). However, a significant negative spatial association exists for elevations > 550 m (Figure 4.3 C). There are several mountain tops in

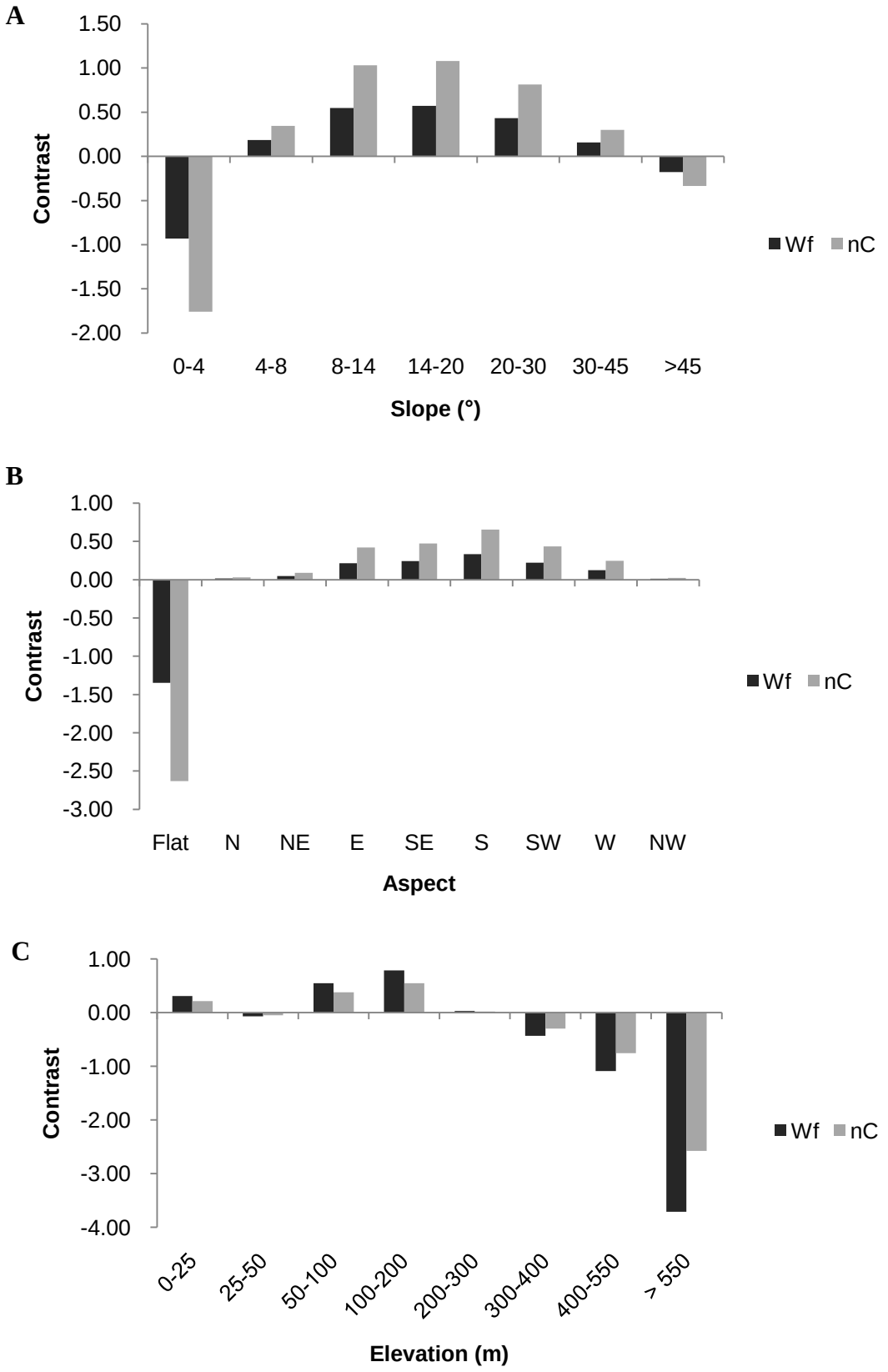


Figure 4.3 An indication of the relationships between landslides and slope classes (A), different aspects (B), and elevation ranges (C) by investigating the weighted (Wf) and normalised (nC) contrasts.

elevations > 550 m which may be less susceptible to landsliding, but most of this class corresponds to the Taupo Volcanic Zone, where landslide data is limited.

The highest W_f value determined in the geology dataset (Figure 4.4) was for mudstone, while peat had the lowest. Most of the other geology classes appear to show a negative W_f value, with the exception of andesite, dacite and diorite, rhyolite, and sandstone. Laharic colluvium was only present in the older geology (nzgeology) data set and displayed no presence of landslides (Appendix 4.1). Areas classed as water, such as Lake Taupo and the Waikato River, have a negative W_f value as expected.

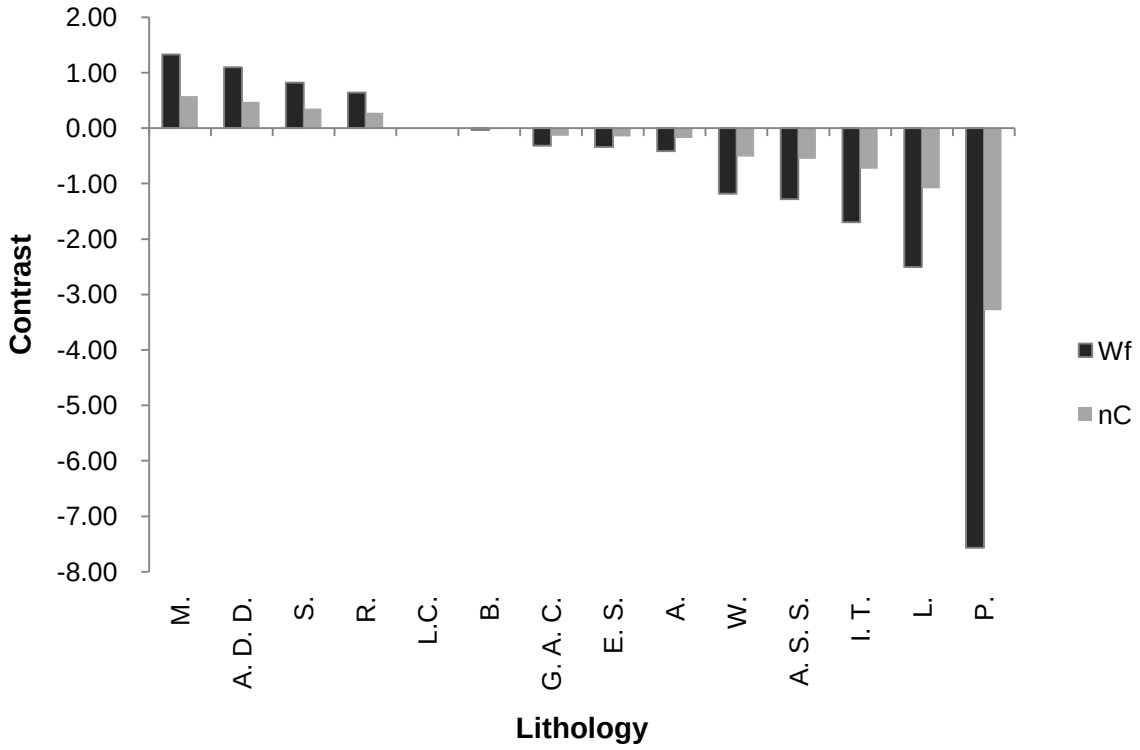


Figure 4.4 Weighted and normalised contrasts for the lithology classes: mudstone (M.); andesite, dacite and diorite (A. D. D.); sandstone (S.); rhyolite (R.); laharic colluvium (L. C.); basalt (B.); greywacke, argillite and chert (G. A. C.); engineering soils (E. S.); alluvium (A.), water (W.); alternating sandstone and siltstone (A. S. S.); ignimbrite and tuff (I. T.); limestone (L.); and peat (P.).

Of the soil orders, brown soils have the greatest W_f value followed by recent soils (Figure 4.5), which indicates a good positive association. Interestingly, organic soils, pumice soils, ultic soils and podzols have comparatively large negative weighted contrast values, but of these only ultic soils are considered to be

significantly negatively associated (nC < -1.96) with landslide occurrence (Figure 4.5).

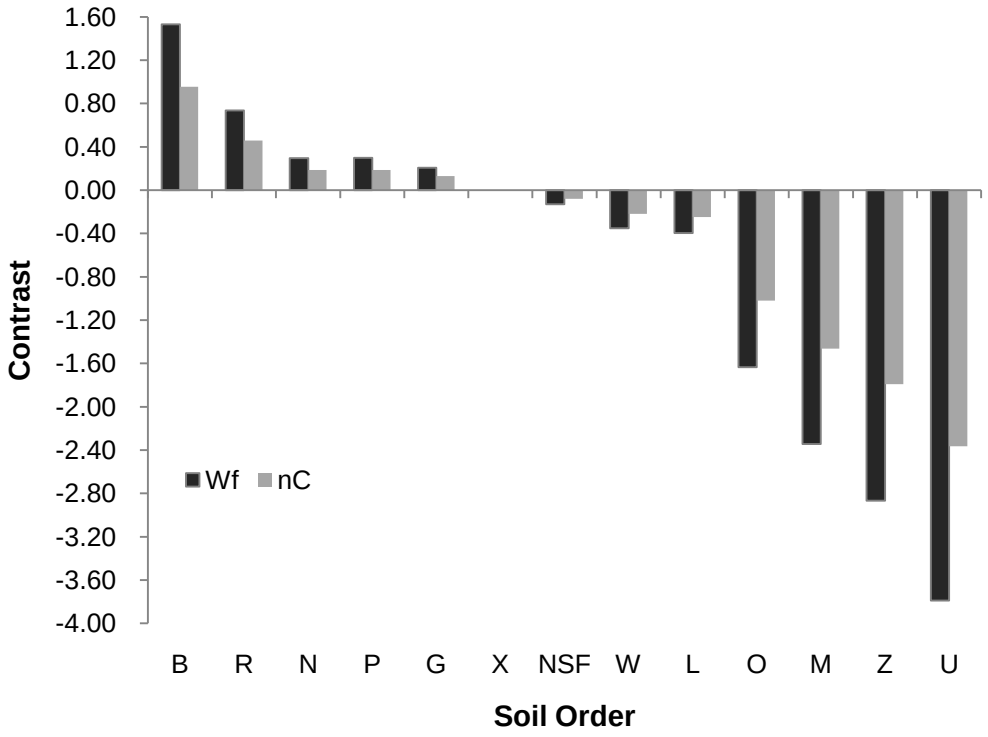


Figure 4.5 Weighted contrast and normalised contrasts for each of the soil orders: brown soils (B); recent soils (R); granular soils (N); pallic soils (P); gley soils (G); oxidic soils (X); non-soil features (NSF); raw soils (W); allophonic soils (L); organic soils (O); pumice soils (M); podzols (z); and ultic soils (U).

In terms of land cover, half the land cover classes resulted in negative weighted contrasts. These were grassland, forest, water bodies, and bare or lightly vegetated surfaces (Figure 4.6). Grassland is the largest land cover class in the Waikato region (13,019 km^2), with the majority present on low slopes where landslides are less common. Bare or lightly vegetated surfaces predominantly occur in the mountainous region in the south. Trees are considered, for the most part, to help with slope stability so it would be expected that forest would have some negative association with landslides; but not one of significance that would prove it is a great indicator of landslide absence as landslides can still occur under forestry.

Sedgeland and saltmarsh, shrub and shrubland, cropland, and artificial surfaces all show a positive spatial association with landslides and are represented by the positive weighted contrast values (Figure 4.6).

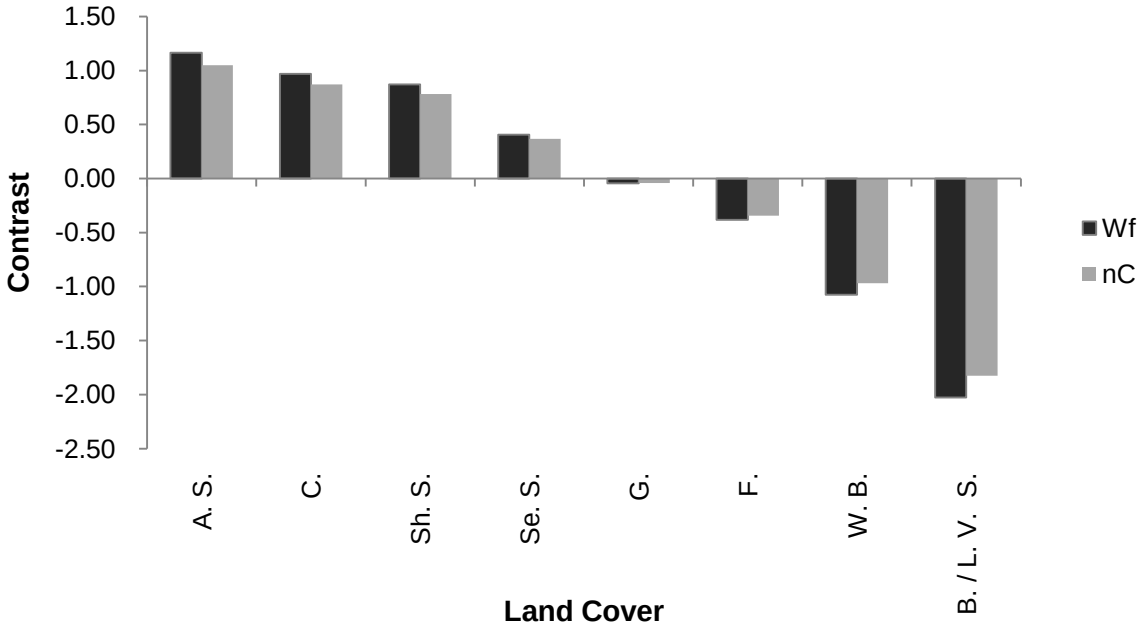


Figure 4.6 Weighted contrasts and normalised contrasts (W_f and nC) for the land cover classes: artificial surfaces (A. S.); cropland (C.); Shrub and shrubland (Sh. S.); sedgeland and saltmarsh (Se. S.); grassland (G.); forest (F.); water bodies (W. B.); and bare or lightly vegetated surfaces (B. / L. V. S.).

When distance from rivers is considered, it can be observed that the weighted contrast and hence spatial association declines with increasing distance from rivers (Figure 4.7). Also, the normalised contrast appears to be at least twice the value of the weighted contrast.

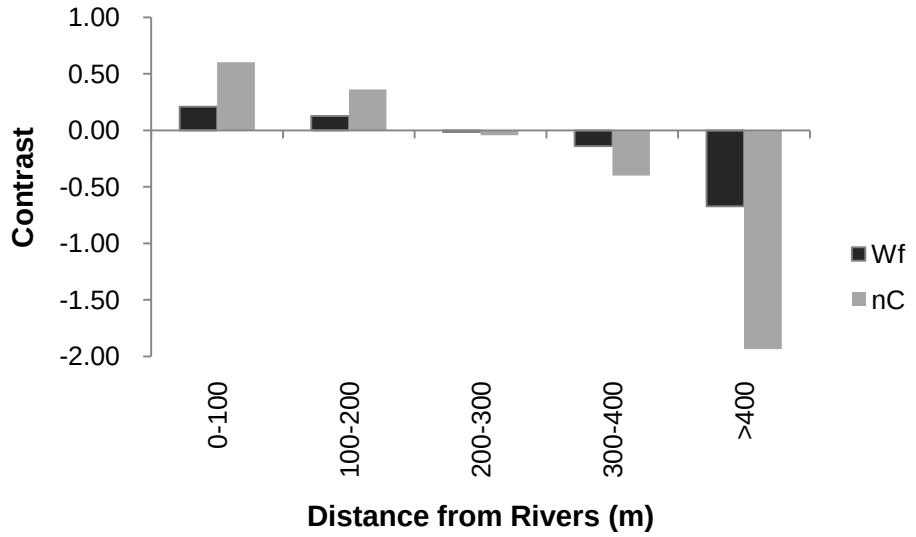


Figure 4.7 Trends in weighted contrast and normalised contrast for distance from rivers.

Both 1998 monthly mean and maximum rainfall display a similar trend when the weighted contrasts are graphed (Figure 4.8). Both have negative values in the first two classes, a large positive spike in the third category (150 – 200 mm and 200 – 250 mm respectively), and nothing in the last two classes (Figure 4.8). These two higher classes in both data sets generally occur in the area where there is limited landslide data. As both rainfalls exhibit a very similar trend with landslide occurrence, both should not be included in the same model. Based on the weighted contrasts and normalised contrasts, it may be best to exclude monthly mean rainfall, but the test of conditional independence is still required in order to be certain.

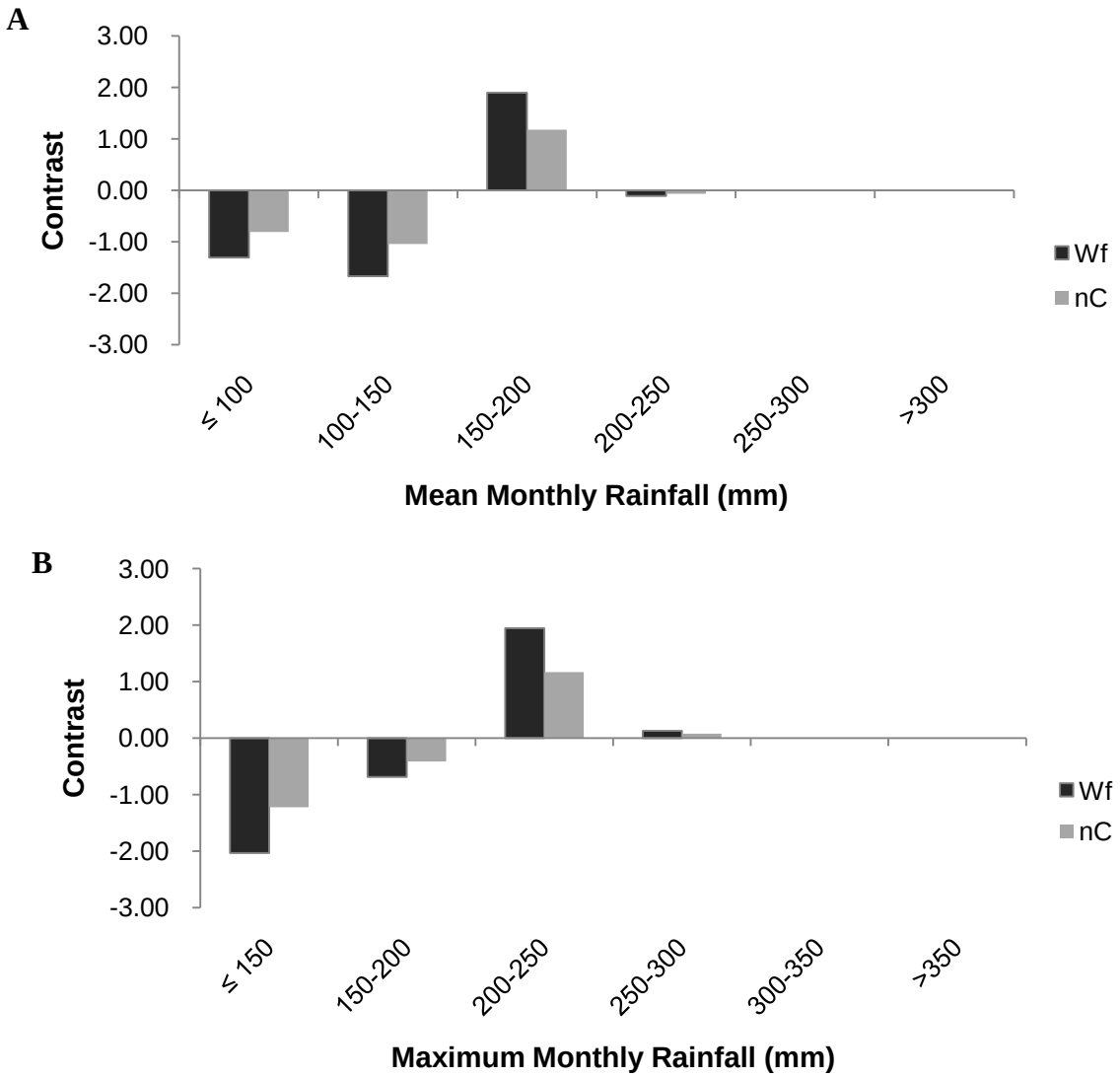


Figure 4.8 Weighted (Wf) and normalised (nC) contrasts for both mean monthly rainfall (A) and max monthly rainfall (B).

4.4.3 Conditional Independence and Chi Square Tests

All pixels with the presence of landslides were subjected to conditional independence and chi square tests so that a comparison of each factor with each of the other factors could be made. A count of landslide pixels in each paired comparison was readily obtained through summary statistics in GIS, and compiled in Excel as a base for a set of linked tables (Appendix 4.2). The observed values were used to determine expected values, and both observed and expected values used to obtain the resulting chi square. In each table, each comparison between each pair of factors and their classes was established. 3,230 paired comparisons with four possible outcomes for each were made based on the eleven factors and a total of 89 classes.

The derived chi squares for each paired comparison were compared with a theoretical χ^2 at $p = 0.01$. If the derived χ^2 is less than the theoretical χ^2 of a given significance the binary predictors are conditionally independent and can be used in the same analysis to map landslide susceptibility (Lee and Choi, 2004). High chi square values indicate the joint probability of the two factors in relation to landslide presence or absence is not significant. High χ^2 values are indicative of conditional dependence, which would imply the binary predictors cannot be used together to derive a landslide susceptibility map (Lee and Choi, 2004; Neuhäuser and Terhorst, 2007).

Approximately 40% (5,138) of the pairwise comparisons between different classes in different parameter sets were found to be significant at the 99% significance level. Some parameters such as, slope, geology, and aspect had a greater number of significant chi squares, whereas other parameters such as roads had fewer. Both rainfall parameters and their classes were compared and only two outcomes of 64 showed significance when compared to the theoretical χ^2. This proves that the two rainfall parameters are conditionally dependent and both cannot be used in the same analysis. As only one rainfall parameter should be used, maximum monthly rainfall was considered more appropriate, so mean monthly rainfall was excluded.

4.4.4 Applying in GIS

The weight contrast values were assigned to each respective class within each of the predictive factor thematic layers by script (Appendix 4.3) in ArcGIS. The resulting weighted raster layers were added together to obtain a single raster layer of the landslide susceptibility index:

$$LSI = W_f Slope + W_f Aspect + W_f Elevation + W_f MaxRain + W_f Geology + W_f LandCover + W_f Soil + W_f Faults + W_f Roads + W_f Rivers \quad \textbf{22}$$

The LSI raster layer was imported to ArcMap where landslide susceptibility classes were set and compared with the landslide inventory.

4.4.4.1 Classification

Four different types of classifications (equal intervals, geometric intervals, natural breaks, and quantile) were implemented to determine five classes from very low to very high. The equal intervals classification splits the data into classes of equal value ranges based on the number of classes specified (ESRI, 2008). Natural breaks are best for data which has jumps in the data values, as it can identify class breaks between these natural groupings and pick the best group of similar values while maximising the difference between classes (ESRI, 2008). The geometric intervals classification was designed to give a visually appealing representation of continuous data, but it can also work quite well on non-normally distributed data (ESRI, 2008). Using geometric intervals the class breaks are determined geometrically in such a way that the variance within classes is minimised so that each class has approximately the same number of values, and the intervals between them are relatively consistent (ESRI, 2008). The quantile classification divides the data into equal proportions, so that each class has the same number of features (ESRI, 2008). Classifying by quantile may not be appropriate with few classes as it can lead to greater distortion and result in a misleading map (ESRI, 2008).

4.4.4.1.1 Visual Comparison of the Goodness of Fit for each Classification

When applied to the same resulting data, each classification produced a slightly different map (Figure 4.9) as a result of splitting the data distribution differently. For each classification, the landslide inventory was overlain to obtain a visual estimation of the goodness of fit (Figure 4.9). Most of the landslide events appear in the high to very high susceptibility classes for all four classifications. The map classed by equal intervals has the least area in the very high and very low classes, and most in the moderate class (Figure 4.9 A). The geometric classification (Figure 4.9 B) appears to have overestimated the high and very high susceptibility classes, and underestimated areas of moderate susceptibility. The map classed by quantile (Figure 4.9 D) looks to give a good result but is likely to be greatly distorted. Natural breaks appears to be the most appropriate classification, as it results in a more realistic image (Figure 4.9 C) and is based on natural, non-uniform groupings.

4.4.4.2 Percent Area and Relative Landslide Density

A good fitting model should in theory show that the higher susceptibility classes explain a large proportion of the landslides in a relatively small proportion of the total area. Based on this assumption, a plot of the percent of landslide only area compared with percent of the total study area within each class should give a basic indication of how well the model fits.

Copies of the weights of evidence model output were classed using each classification technique, and the output layers combined (union) with the landslide inventory. Summary statistics of the total area within each of the susceptibility classes that are classed as landslide only areas and non-landslide only areas were obtained (Appendix 4.4). The percentages of landslide area and total area for each class was determined and plotted (Figure 4.10).

When compared to the other classification techniques, geometric intervals class the majority of the landslide area in the very high class (Figure 4.10 A) but this is within a large proportion of the total area (Figure 4.10 B) which is not preferable. The quantile classification would appear to be the next best (Figure 4.10 A), but the class sizes are near identical (Figure 4.10 B), which may not be appropriate for

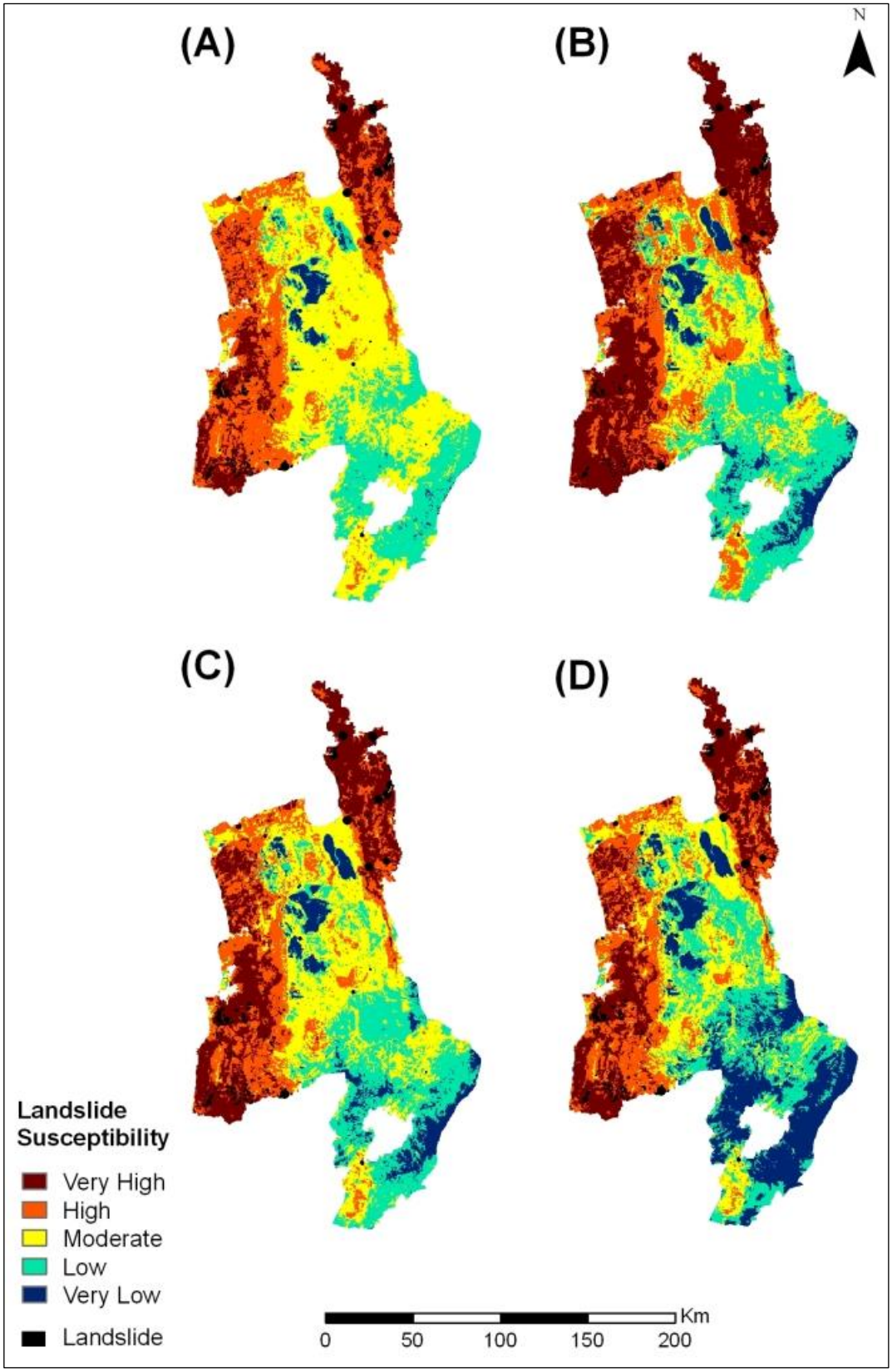


Figure 4.9 Landslide susceptibility for the Waikato Region determined by weights of evidence using all parameters except mean monthly rainfall, classified by (A) equal intervals, (B) geometric intervals, (C) natural breaks, and (D) quantile, with the landslide inventory (QMap landslide data © GNS 2001, and landslide catalogue © GeoNet 2009) overlain in black.

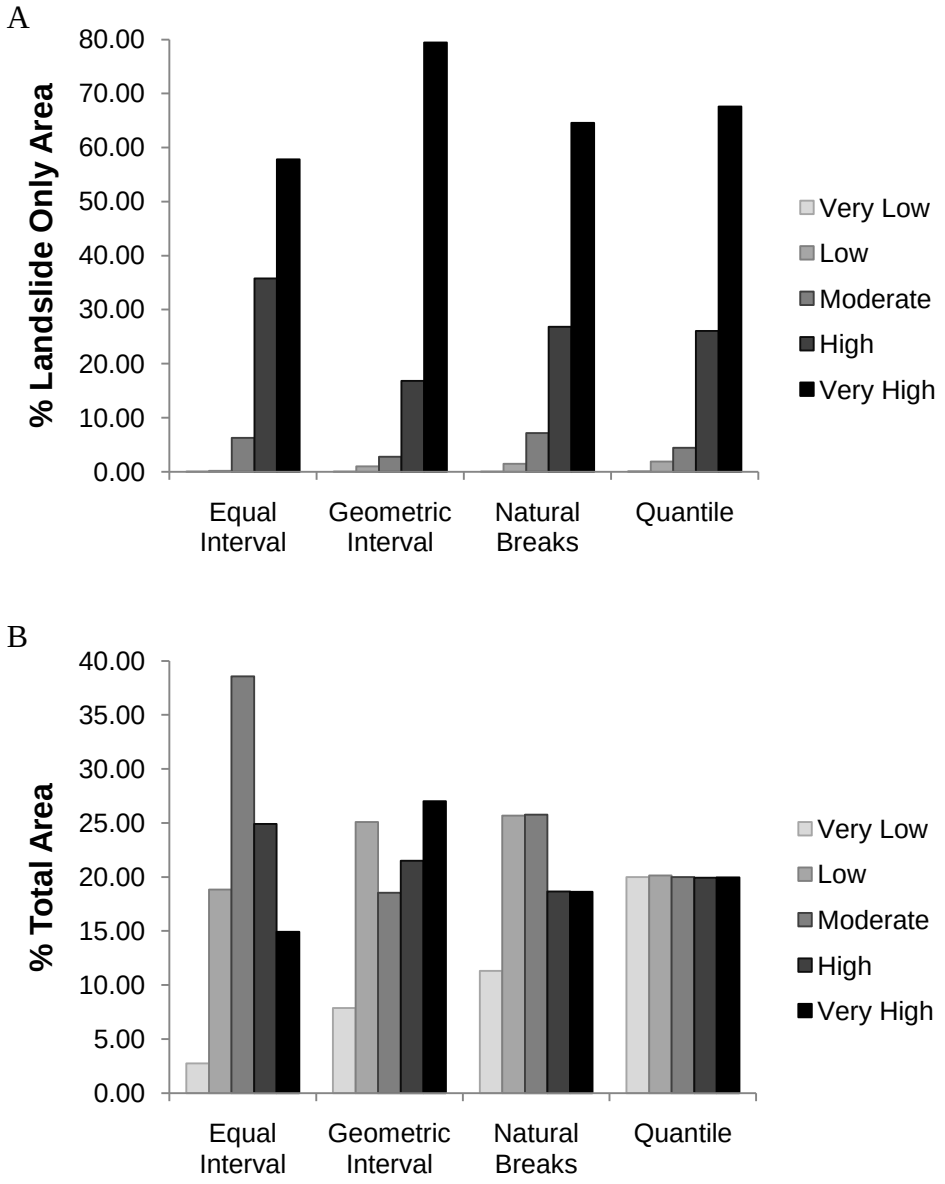


Figure 4.10 Percentage of landslide only area (A) and total area (B) for each of the five susceptibility classes determined by the four classification techniques (equal interval, geometric interval, natural breaks, and quantile), when applied to the combination of all factors except mean monthly rainfall.

presenting this data. The quantile classification may be misleading as few classes have been used, and distortion is not uncommon when fewer classes are used (ESRI, 2008). Natural breaks classify a larger proportion of the landslide area (65%) than equal intervals (58%) (Figure 4.10 A). However, equal intervals have classed a lower percentage of the total area (15%) in the very high class than natural breaks (19%) (Figure 4.10 B). If the combined total areas for the high and very high susceptibilities are compared, natural breaks have a smaller proportion than equal intervals (Figure 4.10 B). When the lower susceptibility (very low and

low) classes are considered, natural breaks class a greater percentage of the total area (37%) than equal intervals (22%), but include a slightly greater percentage of the landslide area. Natural breaks look to be slightly better than equal intervals. Natural breaks have 7% more of the total landslide area in 4% more of the total area. Natural breaks also class 15% more of the total area and 1% more of the landslide area in the lower susceptibility classes. Based on these findings the natural breaks classification appears to be most appropriate. However, the percentage of landslide area is based on landslide only areas and so does not give any indication of landslides as a proportion of the total area or class area. Instead, the two sets of percentages graphed (Figure 4.10) can be used to determine the relative landslide frequency (*R)*, which gives an indication of the goodness of fit (Arora *et al.,* 2004):

$$R = \frac{\%\ Landslide\ Only\ Area\ in\ Class\ i}{\%\ Total\ Area\ in\ Class\ i} \quad \textbf{23}$$

R is also referred to as the relative landslide density, and returns the same values as determined for relative landslide frequency. Relative landslide density is defined as:

$$R = \frac{(n_i/N_i)}{\sum(n_i/N_i)} \quad \textbf{24}$$

where n_i is the sum of land area classed as landslide in a susceptibility level or class "*i"*, N_i is the total area occupied by the susceptibility class "*i"* (Santacana *et al.,* 2003).

It is normally expected that areas classed as higher susceptibilities (moderate to very high) should have a greater proportion of the landslides (He and Beighley, 2008; Santacana *et al.,* 2003). The proportion of landslide area should increase with increasing susceptibility, so that the majority of the landslide area is associated with the very high landslide susceptibility class as these areas are generally more prone to landslides (Arora *et al.,* 2004).

The relative landslide densities for each class within the derived susceptibility map were determined (Table 4.3; Appendix 4.3). The *R* values obtained display

an increase in landslide density with increasing susceptibility classes (Table 4.3). When plotted, the trend of increasing relative frequency or density with increasing susceptibility can be more readily observed (Figure 4.11).

Table 4.3 Landslide area, total area and relative landslide density for each landslide susceptibility class for each classification applied to the susceptiblity map derived by weights of evidence using all parameters except mean monthly rainfall.

Classification	Susceptibility Class	Landslide Area (km^2)	Total Area (km^2)	*R*
Equal Interval				
	Very Low	0.00	650.49	0.00
	Low	0.44	4464.27	0.01
	Moderate	21.37	9140.81	0.16
	High	121.86	5903.72	1.44
	Very High	197.13	3534.30	3.88
Geometric Interval				
	Very Low	0.01	1863.90	0.00
	Low	3.40	5945.80	0.04
	Moderate	9.36	4392.39	0.15
	High	57.25	5092.83	0.78
	Very High	270.78	6398.69	2.94
Natural Breaks				
	Very Low	0.01	2675.15	0.00
	Low	5.04	6082.52	0.06
	Moderate	24.42	6106.30	0.28
	High	91.34	4419.43	1.44
	Very High	220.00	4410.19	3.47
Quantile				
	Very Low	0.22	4736.45	0.00
	Low	6.37	4771.51	0.09
	Moderate	15.13	4735.56	0.22
	High	88.71	4722.38	1.31
	Very High	230.38	4727.71	3.39

The model classed by geometric interval resulted in a poor fit to the landslide data from which it was derived, when compared to the other classifications. The equal interval, natural breaks and quantile classifications give a similar trend. The model classified by equal intervals looks to give the best fit as it has the highest relative landslide density for the very high susceptibility class, and some of the lowest relative landslide densities at lower susceptibilities.

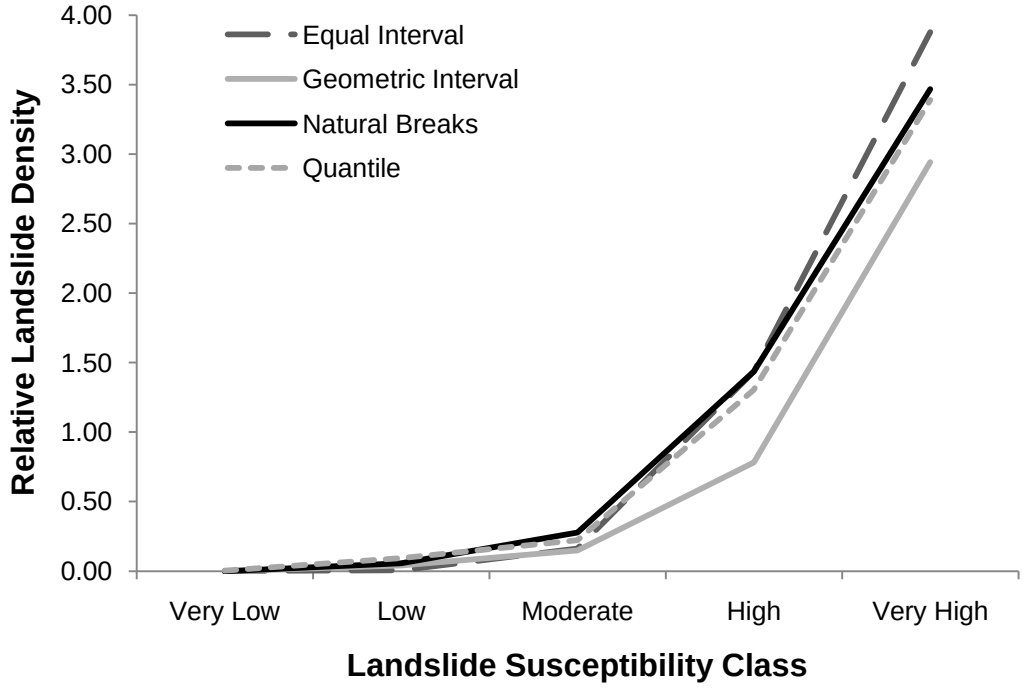


Figure 4.11 A comparison of relative landslide density with increasing landslide susceptibility for each classification applied to the weights of evidence model with no mean monthly rainfall.

Equal intervals and natural breaks both look to be quite suitable classifications. The natural breaks classification is possibly more appropriate, as classes reflect breaks in the data distribution. Although the user specifies the number of classes, natural breaks does not introduce bias from expert opinion as the class range cannot be predetermined by the user like equal intervals can.

4.4.5 Assessing the Impact from the Removal of Factors

Variations of equation 22 were applied in GIS with the exclusion of one or more variable in an attempt to find a better model (Appendix 4.3). Roads and rivers were found to make little difference in the end result so were excluded. This left eight parameters to consider: slope, elevation, aspect, geology, land cover, maximum monthly rainfall, soil order, and distance from faults. Slope, geology and land cover were included in most the combinations of variables, as these resulted in many cases of significance in the paired conditional independence tests. These variables were removed from some combinations but often resulted in a poorer fitting model. Of the 25 or so combinations trialled, three looked

potentially better than the initial model that was derived from all parameters except mean monthly rainfall. These combinations were:

Model A: slope, maximum monthly rainfall, land cover, and geology;

Model B: the same as Model A, but with distance from faults included;

Model C: the same as Model A, but with aspect included.

Each combination was classified using natural breaks, and combined with the landslide inventory (Figure 4.12). The relative landslide densities were determined for the susceptibility classes in each model (Appendix 4.5) and compared to those derived from the first model (Figure 4.13).

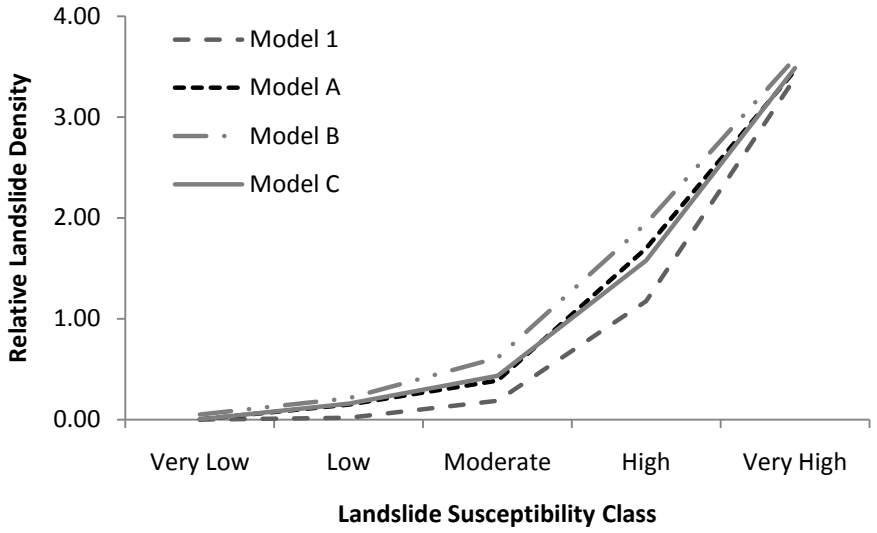


Figure 4.13 Comparison of the initial model with no mean monthly rainfall (Model 1) and each of the three new models (Models A, B, and C) with respect to the relative landslide densities determined for each susceptibility class.

Based on the relative landslide densities, all of the models (A, B, and C) gave a better fit than the initial model (Model 1) (Figure 4.13). When compared to Models A, B, and C, Model 1 classed less of the total landslide area in the very high susceptibility class, and also resulted in the lowest *R* values for all susceptibility classes. Model B resulted in the highest *R* values and most landslide area in the higher susceptibility classes (Figure 4.13). Based on these findings, Model B was chosen as the final landslide susceptibility map determined by weights of evidence (Map 4.1; Appendix 4.6).

4.5. Discussion of the Weights of Evidence Map

4.5.1 Identified Artefacts

On closer inspection of the landslide susceptibility map (Map 4.1), a few artefacts from different input factors can be observed. The large area of very low landslide susceptibility in Hauraki Plains is likely to be a result of peat in the geology map (Map 3.5), while the area of low susceptibility surrounding the Kopuatai Peat Dome in the Hauraki Plains, resembles some of the features of the distance from faults map (Map 3.9). Likewise, north of Taupo the low susceptibility is also likely to be associated with the distance from faults map.

4.5.2 Susceptibilities and the Dominant Factors Observed

Areas of very high susceptibility were observed in the Coromandel, West Coast, and a small portion of south Auckland (Map 4.1). The geology is mainly andesite, dacite and diorite; mudstone, or sandstone on 14 - 30° slopes with brown or recent soils and a monthly max rainfall greater than 200 mm. While there are several fault lines in these zones, they do not appear to be the main driver as they occur in all of the landslide susceptibility classes.

The areas of high susceptibility (Map 4.1) appear to be dominantly influenced by the geology. In the Coromandel these areas are predominantly classed as areas of rhyolite or andesite, dacite and diorite, whereas along the west coast, mudstone and basalt dominate.

The majority of the low susceptibility areas correspond to flat areas or low slope angles such as the Hamilton Basin and Hauraki Plains, and parts of the Taupo Volcanic Zone (Map 4.1). These areas all have a comparatively low monthly maximum rainfall. Areas of ignimbrite lithology and/or pumice soils with forest and/or grassland in the Taupo Volcanic Zone relate to areas of low landslide susceptibility. The Hamilton Basin and Hauraki Plains are quite different, as the

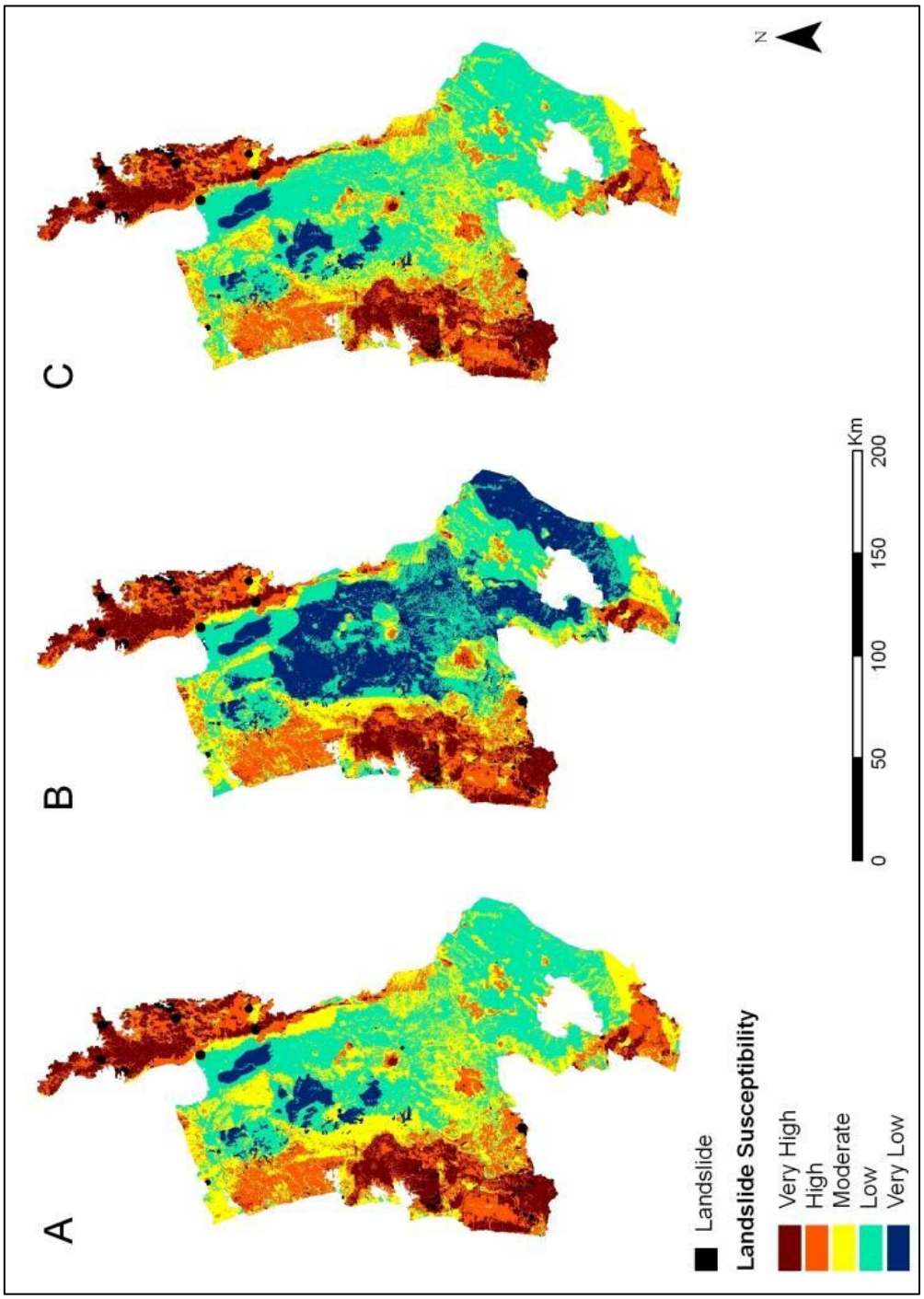
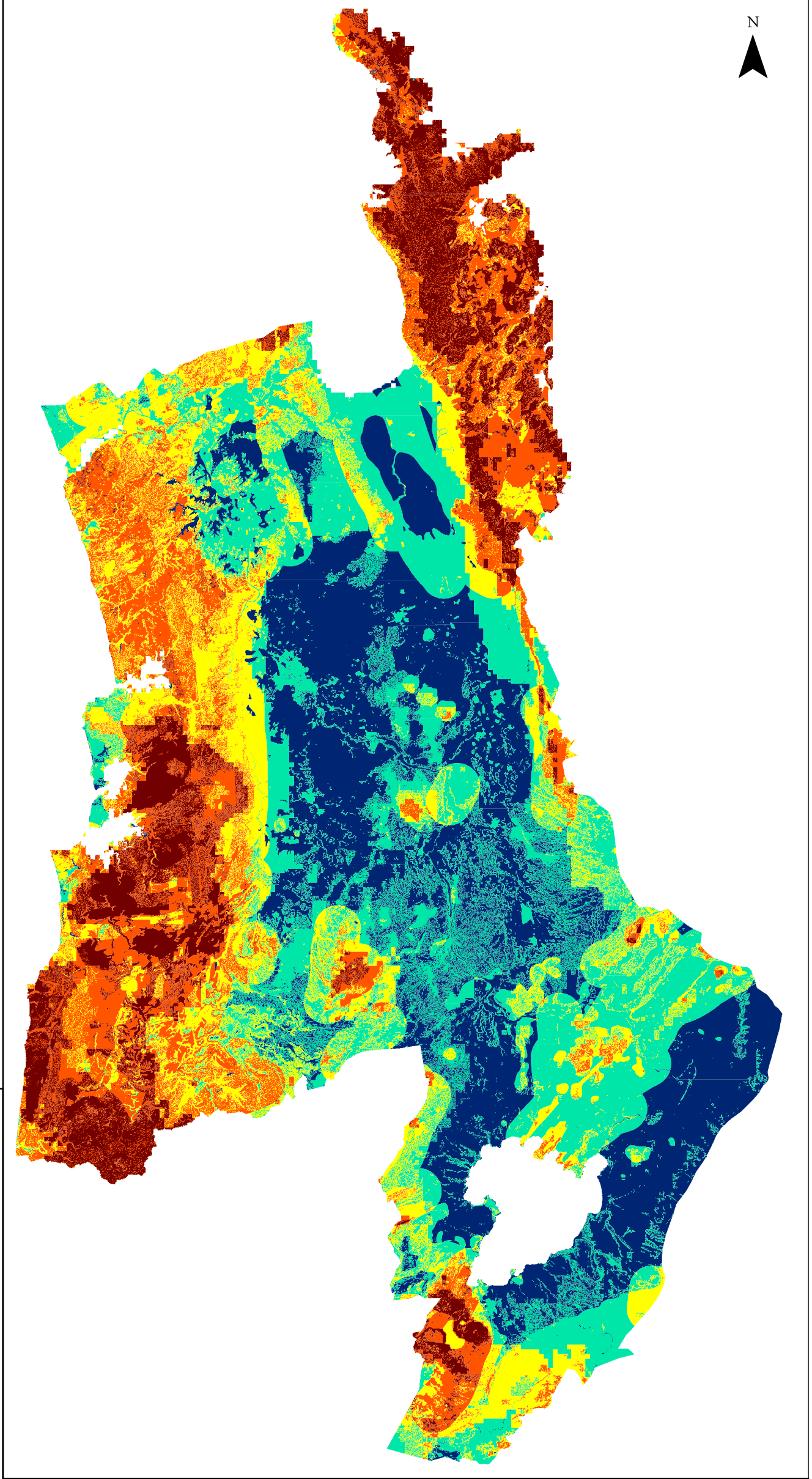


Figure 4.12 Landslide susceptibility maps derived from Model A (A), Model B (B), and Model C (C), with the landslide inventory (QMAP landslide data © GNS 2001, and landslide catalogue © GeoNet 2009) overlain in black.

Map 4.1 Landslide Susceptibility for the Waikato Region Determined by Weights of Evidence

Landslide Susceptibility

- Very Low
- Low
- Moderate
- High
- Very High



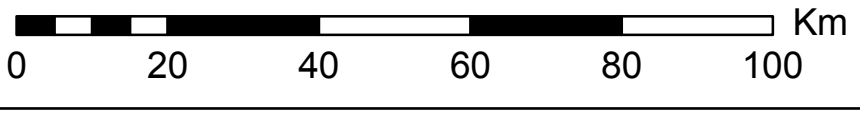
Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Model Used: Weights of Evidence, Model B

Predictor Parameters:
Slope;
Maximum Monthly Rainfall;
Land Cover (Land Use);
Geology;
Distance from Faults

Classification: Natural Breaks



areas classed as low landslide susceptibility consist of a wide range of soil orders, and the land cover is mostly grassland.

The areas of very low susceptibility in the vicinity of Taupo (Map 4.1) look to be mostly influenced by geology. These could be an influence of different data sources, and a lack of data, as there is an absence of QMap landslide and geology data for this area as it is still in production. As a result, some geological classes report no occurrences of landslides, when there may or may not be more in these localities.

4.6 Conclusion

Bivariate analyses such as the weights of evidence can be used to model where landslides will likely occur in the future based on the conditioning factors present at locations of past landslide occurrences. The weights of evidence were determined by the probability of the presence and absence of each of the classes within each of the eleven predictor factors in relation to the presence or absence of a past landslide occurrence at the same location. Parameters used in the analysis should be conditionally independent, yet mean monthly rainfall and maximum monthly rainfall showed dependence. Both mean and maximum monthly rainfall could not be used in the same analysis, and as maximum monthly rainfall was considered more appropriate, mean monthly rainfall was excluded.

The weighted spatial layers of the remaining ten parameters were added and an initial landslide susceptibility map created. Four classification techniques were trialled, and the most appropriate was found to be the natural breaks classification. An assessment of the goodness of fit showed the classed mapped output fit was satisfactory. Although this initial model showed a good fit, it was not the best model. The exclusion of various parameters led to three potentially better mapped outputs, all of which gave a better goodness of fit than the initial model. Model B, which consisted of slope, maximum monthly rainfall, land cover, geology and distance from faults, proved to be the better model and was chosen as the final susceptibility map derived from the weights of evidence approach.

CHAPTER FIVE

A Multivariate Approach Using Logistic Regression

5.1 Introduction

Logistic regression is a generalised linear model that can be used to develop a predictive model of landslide susceptibility based on a binary dependent variable (landslide or non-landslide, 1 or 0) and a set of independent variables for the study area under investigation. Different combinations of predictor variables (independent variables) are analysed to find a combination which produces a significant model (χ^2 significant at $p \leq 0.05$) and correctly predicts a large proportion of the observed landslides. The model coefficients (beta values) determined by logistic regression for each variable and the model intercept are used in an equation applied in GIS to determine the linear combination (z value or logit Y) and the associated probability. A map of continuous values is created which is then classified after having converted these values from floating point to integers (the equivalent percentages if the values had been rounded to two decimal places). The four different classification techniques that were explored in the previous chapter are again considered here.

5.2 Logistic Regression Method in the Literature

5.2.1 Variable Requirements for Logistic Regression

The intended use of logistic regression modelling in landslide susceptibility assessments is to describe the likelihood of landslide occurrence at a regional scale (Dai and Lee, 2002; Dai *et al.,* 2001). The advantage of logistic regression over other multivariate techniques is that it can be used to predict a result measured by a binary variable such as absence or presence of landslides (0 or 1)

based on a set of one or more independent variables (physical parameters) (Ayalew and Yamagishi, 2005; Can *et al.*, 2005; Dai and Lee, 2002; Dai *et al.*, 2001; Lee, 2004; Lee and Sambath, 2006; Menard, 1995; Nandi and Shakoor, 2009; Süzen and Doyuran, 2004; Yesilnacar and Topal, 2005). The arbitrary assignment of codes for the presence (1) and absence (0) of landslides is a matter of convenience and carries no intrinsic meaning (Menard, 1995; Ohlmacher and Davis, 2003), and it is recommended that equal proportions of both landslide (1) and non-landslide (0) pixels are used (Ayalew and Yamagishi, 2005; Nandi and Shakoor, 2009). The independent variables used to predict the dependent variable can be nonlinear, non normally distributed, continuous, categorical, or a combination of both continuous and categorical (Lee, 2004; Nandi and Shakoor, 2009; Süzen and Doyuran, 2004; Yesilnacar and Topal, 2005).

5.2.2 The Logistic Regression Model

The aim of logistic regression is to identify the best fitting, yet reasonable model which describes the relationship between the dependent variable and the independent variables (Ayalew and Yamagishi, 2005). In its most simple form, the relationship between landslide occurrence and its dependency on several variables can quantitatively be expressed as:

25

$$P = \frac{1}{1 + e^{-z}}$$

where *p* is the probability of a landslide occurring, and z is the linear combination (Dai and Lee, 2002; Dai *et al.*, 2001; Lee, 2004; Lee and Sambath, 2006; Süzen and Doyuran, 2004):

26

$$z = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

Where α is the intercept of the model, n is the number of variables, X_i (i = 1, 2, ..., n) are the independent variables and β_i (i = 1, 2, ..., n) are the beta values associated with each of the independent variables, and subsequently the partial slope coefficients of the model (Dai and Lee, 2002; Dai *et al.*, 2001; Lee, 2004; Lee *et al.*, 2006; Lee and Sambath, 2006; Menard, 1995; Süzen and Doyuran,

2004). The probability (*P*) varies between 0 and 1 on an *S*-shaped curve as the linear logistic model, *z*, varies from -∞ to +∞ (Dai and Lee, 2002; Dai *et al.*, 2001; Garcia-Rodriguez *et al.*, 2008; Lee, 2004; Lee *et al.*, 2006; Lee and Sambath, 2006; Süzen and Doyuran, 2004).

It is important to note that susceptibility is not directly defined by logistic regression, but that an inference can be made based on the probability (Ayalew and Yamagishi, 2005).

5.2.3 Odds, Logit, and Probability

If the probability of landslide occurrence [P(*Y* = 1)] is known, the probability of landslide non-occurrence [P(*Y* = 0) = 1 – P(*Y* = 1)] is also known (Menard, 1995; Ohlmacher and Davis, 2003). The probability of occurrence could be modelled by regression as:

27

$$\mathrm{P}(Y = 1) = \alpha + \beta \mathrm{X}$$

However, a problem arises, as the predicted values may be less than 0 or greater than 1, and the probability of there being a landslide occurrence [P(*Y* = 1)] must lie between 0 and 1 (Menard, 1995; Ohlmacher and Davis, 2003). To combat the issue of predicted values falling outside the constraints, the probability that *Y* = 1 can be replaced with the odds that *Y* = 1 (Dai and Lee, 2003; Menard, 1995; Ohlmacher and Davis, 2003). The odds (or likelihood) ratio gives the probability of landslide occurrence to the probability of landslide non-occurrence:

28

$$Odds(Y = 1) = \frac{p(Y = 1)}{1 - p(Y = 1)}$$

and like probability, the odds ratio has a minimum value of 0 but differs in that it has no fixed maximum (Garcia-Rodriguez *et al.*, 2008; Menard, 1995; Ohlmacher and Davis, 2003).

The logit form, in which the logit is a transformation of the probability that there is a landslide occurrence [P(Y=1)] is another way the logistic regression model

can be written (Garcia-Rodriguez *et al.*, 2008). The logit of *Y* is determined as the natural logarithm of the odds:

29

$$Logit(Y) = \ln\left[\frac{p(Y=1)}{1-p(Y=1)}\right] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

This gives the variable z (from Equation 26) which varies from negative infinity to positive infinity (Menard, 1995; Ohlmacher and Davis, 2003).

By using the logit value as the dependent variable the problem of the estimated probability exceeding the maximum and minimum probability is avoided (Menard, 1995; Ohlmacher and Davis, 2003). The logit can be converted back to odds through exponentiation (Equation 30) which results in Equation 31 (Menard, 1995).

30

$$Odds(Y=1) = e^{Logit\ (Y)}$$

31

$$Odds(Y=1) = e^{\ln\left[\frac{P(Y=1)}{1-P(Y=1)}\right]} = e^{\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n}$$

Odds can then be converted back to probability (Menard, 1995) by:

32

$$P(Y=1) = \frac{Odds(Y=1)}{1+Odds(Y=1)} = \frac{e^{\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n}}{1+\ e^{\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n}}$$

Converting from logit to probability via the odds results in the expression:

33

$$P(Y=1) = \frac{1}{1+\ e^{-(\alpha+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_n X_n)}}$$

(Dai and Lee, 2003) which is an expanded version of the simplified equation in Equation 25.

It is important to note that the *probability*, the *odds*, and the *logit* are three different ways of expressing the exact same thing, and can easily be converted from one to the other (Menard, 1995; Ohlmacher and Davis, 2003). This would explain why different variations of the same thing can be found in other landslide susceptibility or hazard assessments where logistic regression has been used, as some give the odds equation (Ohlmacher and Davis, 2003), some the logit (Equation 29) (Ayalew and Yamagishi, 2005; Can *et al.,* 2005; Duman *et al.*, 2006; Garcia-Rodriguez *et al.,* 2008) some the probability based on odds (Equation 32) (Falaschi *et al.,* 2009; Lee *et al.,* 2006; Ohlmacher and Davis, 2003), and some the probability derived from logit in the form of Equations 25 and 33 (Dai and Lee, 2002, 2003; Dai *et al.,* 2001; Greco *et al.,* 2007; Lee, 2004, 2005; Lee and Sambath, 2006; Ohlmacher and Davis, 2003; Süzen and Doyuran, 2004).

5.2.4 Obtaining a Model

In order to obtain a model, the binary dependent variable is transformed into a logit variable and the maximum likelihood estimation (the natural log of the event of landslides occurring or not) is applied (Süzen and Doyuran, 2004). The maximum likelihood method estimates coefficients for the contributing parameters in the logistic regression model and selects those coefficients which make the observed results most likely (Dai *et al.,* 2001; Lee and Sambath, 2006; Ohlmacher and Davis, 2003). The statistical significance of the coefficient for each independent variable included in the model is assessed using the Wald test (Nandi and Shakoor, 2009). To determine whether variables should be added to, or removed from the model the likelihood-ratio test is used (Dai *et al.,* 2001). In this test, the model is estimated with the observed change in the logarithm of likelihood as each variable is removed from the model (Dai *et al.,* 2001). To make the logistic regression model more complete, it is best to begin with a tentative solution which includes a larger number of independent variables that play a major role in the determination of the dependent variable (Ayalew and Yamagishi, 2005; Menard, 1995). Finding the selection of predictors which give the best solution can be difficult. It is possible, however, by repeating with slightly revised

combinations and assessing the improvement until the change in the likelihood becomes negligible (Ayalew and Yamagishi, 2005; Menard, 1995).

5.2.5 Goodness of Fit and Significance of the Model

In logistic regression the maximum likelihood estimation is used to test the goodness of fit (Nandi and Shakoor, 2009) and the chi square test is used to determine the overall significance (Lee and Sambath, 2006). An important concept in understanding the tests in logistic regression is the log likelihood which acts as the criterion for selecting parameters in the model while indicating how likely the independent variables match the observed values of the dependent variable (Lee and Sambath, 2006; Menard, 1995). The maximum likelihood estimation sets out to maximise the value of the log likelihood function (Menard, 1995). The log likelihood multiplied by -2 gives the -2LL statistic which approximately has a chi square (χ^2) distribution, and as the log likelihood value is negative the -2LL value will be positive (Menard, 1995). For the full model, the -2LL statistic is referred to as the deviance or deviation of χ^2 (D_M). The deviance is an indicator of how poorly the independent variables are fit in the equation by the model. In terms of the maximum likelihood estimation, a smaller -2LL value indicates a better prediction of the dependent variable and a model which better fits the data (Menard, 1995; Nandi and Shakoor, 2009). The initial (intercept-only) -2LL is the -2 log likelihood statistic with none of the independent variables and is referred to as D_0 (Menard, 1995). If the difference between D_0 and D_M is taken, the model chi square (G_M) can be determined:

$$G_M = D_0 - D_M \quad \textbf{34}$$

By treating G_M as a chi square statistic, the null hypothesis that $B_1 = B_2 = ... = B_n = 0$ for the logistic regression model can be tested (Menard, 1995). The null hypothesis is rejected if the model chi square is significant ($p \leq 0.05$) and it can be concluded that the independent variables included in the model allow for a better prediction of the dependent variable than what could be made without the independent variables (Menard, 1995).

5.2.6 Model Accuracy

In some cases the accuracy of the model (whether predictions are correct or incorrect) may be of greater concern than the model fit (Menard, 1995). Classification tables for the case applied in logistic regression can be generated which indicate the predicted and observed values of the dependent variable (Menard, 1995).

5.3 Applying Logistic Regression to the Waikato Region

5.3.1 Overview

Using STATISTICA, several approaches were explored in the sampling of data and defining the model to use for the Waikato Region. In terms of sampling, different sample set sizes, extents, and methods of sampling were trialled. Defining the model involved various analytical options in the building of the model. These included: the model type, type of analysis, distribution, and link functions such as logit, probit, loglog, and complementary log-log. Once the sample set and the model had been configured the analysis was carried out. This involved the trialling of different combinations of predictor variables and assessing the significance of the model and the model fit. After finding the most significant and best fitting model, the model coefficients (beta values) were applied to the parameter's spatial data in GIS and the logistic regression equation analysed. A second sample set created in the same fashion was used to validate the results from the previous sample set. In addition to validation, various classification techniques were investigated in order to determine which would be the most appropriate for the resulting landslide susceptibility map based on five susceptibility classes.

5.3.1.1 Sampling

Sampling in STATISTICA involved both systematic random sampling and stratified random sampling. Systematic random sampling generates a random subset using the k value specified to select a starting point somewhere in the first

k number of cases then it systematically takes each k^{th} value from then on. So if the k value is set as 10 the first case is randomly chosen from the first ten cases and from there every tenth case is selected (Statsoft, 2009). Stratified random sampling was then used on the resulting subset of data to obtain a sample set with an even proportion of landslide and non-landslide data by specifying a percentage for each to give the desired number of samples.

5.3.1.2 Generalised Linear/Nonlinear Models

There are a wide range of options to choose from in the building of the model. Some of these were obvious choices, and others required some experimentation. In the setup menu, the distribution was set as binomial for the dependent variable as it has two possible outcomes: landslide or non-landslide. Multiple regression was selected as the type of analysis in order to find the best combination of predictor parameters that explain the dependent variable. In a situation where the dependent variable is assumed to be nonlinearly related to the predictor parameters (as in this study) the link function is used to model the responses (Statsoft, 2009). The selection of link functions is dependent on the assumed distribution of the dependent variable. As the distribution was defined as binomial in the setup menu this led to a choice of four link functions, these were logit, probit, log-log, and complimentary log. Each link function uses a different equation (Table 5.1) which results in different responses.

Table 5.1 Link functions available when using a dependent variable with a binomial distribution and their equations adapted from (Statsoft, 2009).

Link Function	Equation	
Logit link	f(z) = log(z/1-z)	
Probit link	f(z) = invnorm(z)	where *invnorm* is the inverse of the standard normal cumulative distribution function
Complementary log-log link	f(z) = log(-log(1-z))	
Loglog link	f(z) = -log(-log(z))	

5.3.2 The Initial Trial

5.3.2.1 Sampling across the Whole Region

The Auckland, Waikato, and Rotorua datasets, after being imported to STATISTICA, underwent systematic random sampling using a k value of 2 which would equate to 50 m spacing based on the input data being derived from 25 m pixels. The output of this was exported to an Access database, and a single, slightly simplified, dataset covering the entire Waikato Region was compiled. This was imported to STATISTICA where it underwent an initial attempt at stratified random sampling to obtain a subset where half the data had landslides and half had none.

5.3.2.2 Defining the Model Type

The first attempt at applying a generalised linear/nonlinear model to carry out multiple regression analysis over a binomial distribution was done to get an indication of which link functions (i.e. logit, probit, log-log, and complimentary log) to use and whether the sampling method was appropriate or not. Of the four link functions the logit and log-log link functions returned the better results, based on the percentage of landslides correctly identified. Using the model output from the logit analysis, a script was created and applied in GIS. The spatial output showed much of the Rotorua portion of the region to have lower susceptibility than the Hauraki Plains. This appears to be as a result of the very limited landslide data in the Rotorua area. To address this problem a different approach to sampling was taken.

5.3.3 The Consideration of Limited Landslide Data in a Second Trial

It was decided that as the Rotorua dataset is limited in landslide data, it may be more appropriate to derive a sample set from the Auckland and Waikato area where there is QMap landslide data. The Auckland and Waikato database consisting of 25 m pixels underwent systematic random sampling using a k value of 40, which equates to a spacing of 1 km between each pixel selected. This reduced the data set from 22,839,716 pixels to 570,993 pixels. This approach is

similar to that of Nandi and Shakoor (2009) who used limited landslide data from a portion of the study area in logistic regression and following the successful test on this training set, the results were extended to their entire study area. Nandi and Shakoor (2009) also recommend systematic sampling from a smaller, representative area to produce realistic results with less time and effort.

5.3.3.1 The Implementation of Stratified Random Sampling

Over sampling and under sampling of some parts of the area can be an issue with simple random sampling, but if stratified random sampling is implemented this problem can be overcome (Dhakal *et al*., 2000). To make the analysis easier to follow, the column containing the landslide code (LSlideCode) in the 570,993 dataset was selected and using "Find and Replace", the code assigned for the presence of a landslide (1) was replaced with "Landslide", and for the absence of landslide (0) this was replaced with "Non-Landslide". The frequency of landslides and no landslides was found to be 13,226 and 557,767 respectively. The frequency of landslides is considerably lower than that of non-landslides, and only accounts for 2.37% of this sample dataset. In accordance with the recommendations of using equal proportions of landslides and non-landslides (Ayalew and Yamagishi, 2005; Nandi and Shakoor, 2009), a subset consisting of 50% each "Landslide" and "Non-Landslide" was obtained by the implementation of stratified random sampling.

In STATISTICA, stratified sampling is used to obtain a range of different sized sample sets based on a percentage of each feature within a group. If the dataset being sampled was evenly split between landslides and non landslides then 50% could be used to sample for each feature. But as there were more records of non-landslides than landslides, to obtain a sample set which consists of 50% of each, the maximum number was limited to all landslide records from the data set being sampled. This also meant that in order to obtain a sample set consisting of 50% landslide data and 50% non-landslide data the percentage of each group sampled was different. The proportion of landslide records which make up half the sample is determined by:

35

$$S_L = \frac{\left(\frac{S_s}{2}\right)}{L_T}$$

where S_L is proportion of total landslide records, S_s is the selected sample size, and L_T is the total number of landslide records. Similarly, proportion of non-landslides which make up the other half of the sample set can be determined by:

36

$$S_{NL} = \frac{\left(\frac{S_s}{2}\right)}{NL_T}$$

where S_{NL} is the proportion of total non-landslide records, and NL_T is the total number of landslide records.

As there are fewer landslides, the half value as a percent of total landslide records is higher than the other half as a percent of total non-landslides when extracting the same number of records for each class in a new sample set. For each of the different sized sample sets, the number of landslides and non-landslides were determined by a percentage (to six decimal places) of their frequency in the sample set being sampled. Six decimal places are used in STATISTICA as rounding to anything less generally led to a disproportionate dataset of an unintended size. Nine sample sets of various sizes made up of half landslides and half non-landslides were created using the stratified random sampling (Table 5.2).

Table 5.2 An example, using the two largest sample sets of the nine different sized sample sets which were obtained from a sample set of 570,993 pixels through stratified random sampling, showing the percentages of both non-landslide and landslide data used to achieved the correct number in each class.

Sample set size	Number in each outcome	% of the 557,767 Non-Landslide pixels	% of the 13,226 Landslide pixels
25,000	12,500	2.241079	94.510812
26,452	13,226	2.371241	100.000000

5.3.3.2 A More Rigorous Application of the Logit and Log-Log Link Functions

5.3.3.2.1 Model Input and Specifications

Using the generalised linear/nonlinear model tool in STATISTICA backwards stepwise multiple regression analyses using both the logit and log-log link functions were carried out on each of the sample sets. The landslide code (LSlideCode) was set as the dependent variable and the response codes for this defined. Backwards stepwise regression with the p to enter and p to remove both set to 0.005 (default was 0.05) was used to determine which of the selected predictor variables should be removed or included. A set of nine predictor variables were used to begin with, these included slope, aspect, elevation, land cover, geology, soil order, distance from fault lines, distance from rivers, and either mean monthly rainfall or maximum monthly rainfall. Both rainfall variables were not used together in the analyses as these were quite highly correlated, the use of one or the other however is acceptable. Distance from roads was excluded from the analyses as the landslide catalogue data is likely to be correlated to roads as a result of the reporting of landslide occurrence rather than actual susceptibility.

5.3.3.2.2 Assessment of the Model Fit and Significance

After each model run, the results of the selected predictors in their efficiency to predict the dependent variable were viewed in two parts. Firstly, the 'goodness of fit' summary was examined, which gave the log likelihood, Deviance (-2LL), degrees of freedom, and Pearson chi square. Based on the degrees of freedom in the model, the Pearson chi square of the model can be compared to the chi square value that would be obtained at the $p = 0.05$ significance level. Secondly, the 'odds ratio' summary was considered, which gives the odds and log ratios as well as an indication of the model accuracy based on the percentage of both the observed landslide and non-landslide outcomes correctly identified by the predicted outcomes. To help improve the model, the 'model building' summary gives an indication of which predictors to keep and which should be removed based on the Wald p. The Wald p statistic is based on the maximum likelihood and acts as a test of significance of the regression coefficient (Statsoft, 2009). This helped as a guide in the right direction. The best model ultimately, was found

through trialling by removing an individual predictor and assessing the impact on the model as a result of its exclusion.

For all sample sets trialled, the combination of nine predictor variables which includes mean monthly rainfall but excludes maximum monthly rainfall returned a better Pearson chi square than the inclusion of maximum monthly rainfall and exclusion of mean monthly rainfall. Excluding both rainfall predictors led to a much worse model than what was obtained by the exclusion of either one alone. Based on this it was decided that mean monthly rainfall would remain in the model and maximum monthly rainfall would be excluded. Distance from rivers and distance from faults were commonly suggested to be removed from the model based on the high Wald p values exceeding the p to remove and p to enter constraints. In addition to these findings, it was found that the sample sets of 25,000 and 26,452 consistently returned more significant results than the smaller sample sets. The larger of the two (26,452) was submitted to more rigorous trialling using the remaining seven predictors in various combinations where one or more predictors were removed.

5.3.3.2.3 Comparison of the Results of the Logit and Log-Log Link Functions

The best outcome found using the log-log link function was for the 26,452 sample set using the combination of slope, elevation, mean monthly rainfall, land use, and geology which resulted in a chi square of 27,266.70. The chi square value for p = 0.05 based on 26,446 degrees of freedom equated to 26,825.42. As the model chi square is larger than the chi square at p = 0.05 significance level the model is considered to be significant. Using an online chi square calculator (Walker, 2009) with the model chi square (27,266.70) and degrees of freedom (26,446) as input gave a p value of 0.0002. The odds ratio obtained was 5.21 and the proportion of observed ‘landslide’ and ‘non-landslide’ pixels correctly predicted were 69.02% and 70.03% respectively.

The log-log equation (Chen and Shao, 2001) was applied (following Equations 37 and 38; Appendix 5.1) in GIS which gave a result where the resulting probabilities fell across a very small range between 0.996445 and 1.0. There is very little distinction between high and low as the range between them is 0.003555, this would indicate that many of the values have been grossly overestimated as the

probability should theoretically cover a range of values between 0 and 1 and not constrained to the top 0.0005% within that range.

37

$$z = 3.4376 + (-0.0055 \times Slope) + (0.0032 \times Elevation) + (-0.0245 \times Mean\ Monthly\ Rainfall) + (0.0683 \times Land\ Use) + (-0.0166 \times Geology)$$

38

$$p = F(z) = \exp[-\exp(-z)]$$

In comparison, the best outcome when the logit link function was applied to the 26,452 sample set was the combination of slope, mean monthly rainfall, land use, and geology. This gave a chi square of 27,016.60 for the model which was found to be significant when compared to the p = 0.05 chi square of 26,826.43 (based on 26,447 degrees of freedom). A p value of 0.0069 was given when the chi square of the model and degrees of freedom were input into the online chi square calculator (Walker, 2009). An odds ratio of 4.14 was obtained and of the observed 'landslide' and 'non-landslide' groups 71.18% and 62.62% respectively were predicted correctly. While log-log link function may be more significant based on the chi square values, the logit link function returns a better odds ratio and correctly predicts a greater proportion of landslides. The model using the logit link function was applied in GIS (Equation 39 and Equation 25; Appendix 5.1) and the resulting susceptibility map had a probability range of 0.101556 to 0.994192 which appeared to be more visually correct than the one derived from the log-log link function.

39

$$z = -3.9092 + (-0.0024 \times Slope) + (0.0305 \times Mean\ Monthly\ Rainfall) + (-0.1292 \times Land\ Use) + (0.0009 \times Geology)$$

Using the logit link function is in keeping with the bulk of the literature, and also gave a better mapped outcome than the log-log link function; from here on only the logit link function will be used.

5.3.4 The Final Sampling Strategy and Logistic Regression Model

5.3.4.1 The Creation of Separate Testing and Validation Sample Sets

An improved method of sampling which would include a separate sample set for model validation was implemented. While the previous trial resulted in a successful looking model, the method of sampling used to obtain it lacked some consideration of spatial location. A crucial flaw was found in the implementation of systematic random sampling as a result of it systematically working through in the order the dataset is sorted by. To better explain this, the flaw was that while the data were organised in ascending Easting coordinates, no consideration was given to the Northing coordinates. This meant the data were only spaced by the k value in the horizontal direction and not also in the vertical. Following this systematic random sampling based on the Easting coordinate, the resulting sample set should have been sorted by Northing coordinates and a second systematic random sample carried out before continuing on to the stratified random sampling.

The double systematic random sampling was applied to the Auckland and Waikato data set using a k value of 8 (spacing of 200 m) in both horizontal (sorted by Easting coordinates) and vertical (sorted by Northing coordinates) directions. This reduced the sample size from 22,839,716 to 356,871. The frequency of landslides and non-landslides for this data set was found to be 8,282 and 348,589 respectively. As the best results from the previous trial were obtained using all the landslides from the systematic random sample set in the stratified random sampling, this was repeated. Stratified random sampling based on all 8,282 landslides and the equivalent number of non-landslides gave a sample set of 16,564 samples which will now be referred to as Set 1 (Appendix 5.2). Following basically the same process, a second sample for validation which will now be referred to as Set 2 was created (Appendix 5.2). As the same k value of 8 is being used there was a one in eight chance (as the first record can be any within the first k value range) the results of the first systematic random sampling (by Easting coordinates) in the creation of Set 2 would be the same as that obtained for the first step of Set 1. If this had eventuated the systematic random sampling would have been rerun, but a different sample set was obtained for Set 2. If the first of

the duel systematic random sampling in the creation of Set 2 was found to be different to the same step of that to derive Set 1 the second systematic random sampling (by Northing coordinates) should have no chance of resulting in any of the same records. Following the duel systematic random sampling the Set 2 sample set was reduced from 22,839,716 to 356,870. Using GIS, this sample set was compared with the sample set from which Set 1 was derived and none of the samples shared identical locations. The frequency of landslides and non-landslides in this second set was 8,333 and 348,537 respectively. Following stratified random sampling based on the 8,333 landslides a sample set of 16,666 samples, which will now be referred to as Set 2, was created.

5.3.4.2. Applying Logistic Regression to Sample Set 1

Using Set 1 the generalised linear/nonlinear model was used to carry out multiple regression of a binomial distribution using the logit link function. Backwards stepwise regression using both a p to enter and a p to remove of 0.05 was set to determine at which significance level predictors should be added to or be removed from the model. Several combinations of predictors were trialled, starting with the best combination from the previous trial using the logit link function. This combination of slope, mean monthly rainfall, land cover (land use) and geology was found to give a significant model. Several variations with one or more variables excluded or added were trialled but none were found to be more significant than the combination of slope, mean monthly rainfall, land cover and geology. The model resulted in a chi-square of 16,952.2 and 16,559 degrees of freedom (Table 5.3). When compared to the chi square at p = 0.05 of 16,859. 5 (Walker, 2009), the model was found to be significant with a p value of 0.0158 (Table 5.3). The model correctly predicted 71.36% of the observed landslides and 63.15% of the observed non-landslides, and returned an odds ratio of 4.27 (Table 5.3).

Table 5.3 Summary of statistics for Set 1 following logistic regression of slope, mean monthly rainfall, land cover (land use) and geology as predictors of landslide occurrence or non-occurrence.

Statistic	Set 1
Model chi square	16952.2
p = 0.05 chi square	16859.5
Log likelihood (LL)	-10291.1
Deviance (-2LL)	20582.2
Degrees of freedom	16559
p value of model	0.0158
Odds ratio	4.27
Landslides correctly predicted (%)	71.36
Non-landslides correctly predicted (%)	63.15

The linear combination for the model was determined:

40

$$z_{set1} = -4.0652 + (-0.0068 \times Slope) + (0.0318 \times Mean\ Monthly\ Rainfall) + (-0.1208 \times Land\ Use) + (-0.0029 \times Geology)$$

and the logistic regression model (Equation 25) was applied in GIS (Appendix 5.3).

5.4 Model Validation

The model obtained using the data from Set 1 was validated by applying it to the Set 2 data and comparing the percentage of observed landslides and non-landslides correctly predicted by the model (Table 5.4).

Table 5.4 Percent landslide and non-landslide correctly predicted by both Set 1 and Set 2 sample sets using the model derived from Set 1.

	Landslide	Non-Landslide
Set 1 Sample Set	71.36	63.15
Set 2 Sample Set	71.32	62.56

The Set 1 model only differs in percentage values by 0.04% in correctly predicting the observed landslides when the results of Set 1 (71.36%) and Set 2 (71.32%) predictive capabilities are compared. 62.56% of the observed non-landslides in Set 2 were correctly predicted using the Set 1 model and when

compared to Set 1 (63.15%) this only differed by 0.59. As there is very little difference between the predictive capabilities when the Set 1 model is applied to both Set 1 and Set 2 this indicates the model fits well and is not a result of chance. While the model fits the data well and can now be mapped, further validation in the form of ground-truthing and validation curves will be carried out in chapter 6 to assess how realistic the model is.

5.5 Mapping the Set 1 Model

After the Set 1 model was applied in GIS, the resulting landslide susceptibility spatial layer was compared to the factor maps of those parameters used in the model prior to classification. Aspects of the map, such as artefacts, were identified and the possible causes were considered. Following this the map was classified into five classes by a selection of techniques to determine which was most appropriate for the Set 1 logistic regression model landslide susceptibility map.

5.5.1 Visual Observations of the Model Prior to Classification

An unusual straight edged feature of high susceptibility is immediately apparent in the model output when mapped. In order to get a better understanding of what is causing this, the parameters used in the creation of the map were mapped alongside the susceptibility map as insets of this area of interest (Figure 5.1). All insets are at the same scale and cover the exact same area, and each represents something different.

When comparing the landslide susceptibility output within the area selected (Figure 5.1 A) it is apparent that the shape is very similar to that of the mean monthly rainfall (Figure 5.1 B). The shape appears to have been introduced by the mean monthly rainfall which also happens to be the only variable in the model to be multiplied by a positive beta value (Equation 40). The rainfall data is of a lower resolution (1 km) than the rest of the spatial data, so is less detailed and slightly more generalised. The pixel size (resolution) may explain the straight shape, but not the stark contrast between higher and lower rainfall. This is more likely a result of the rainfall being estimated based on records from 2,202 rain gauge stations spread around the whole of New Zealand (Leathwick *et al.*, 1998).

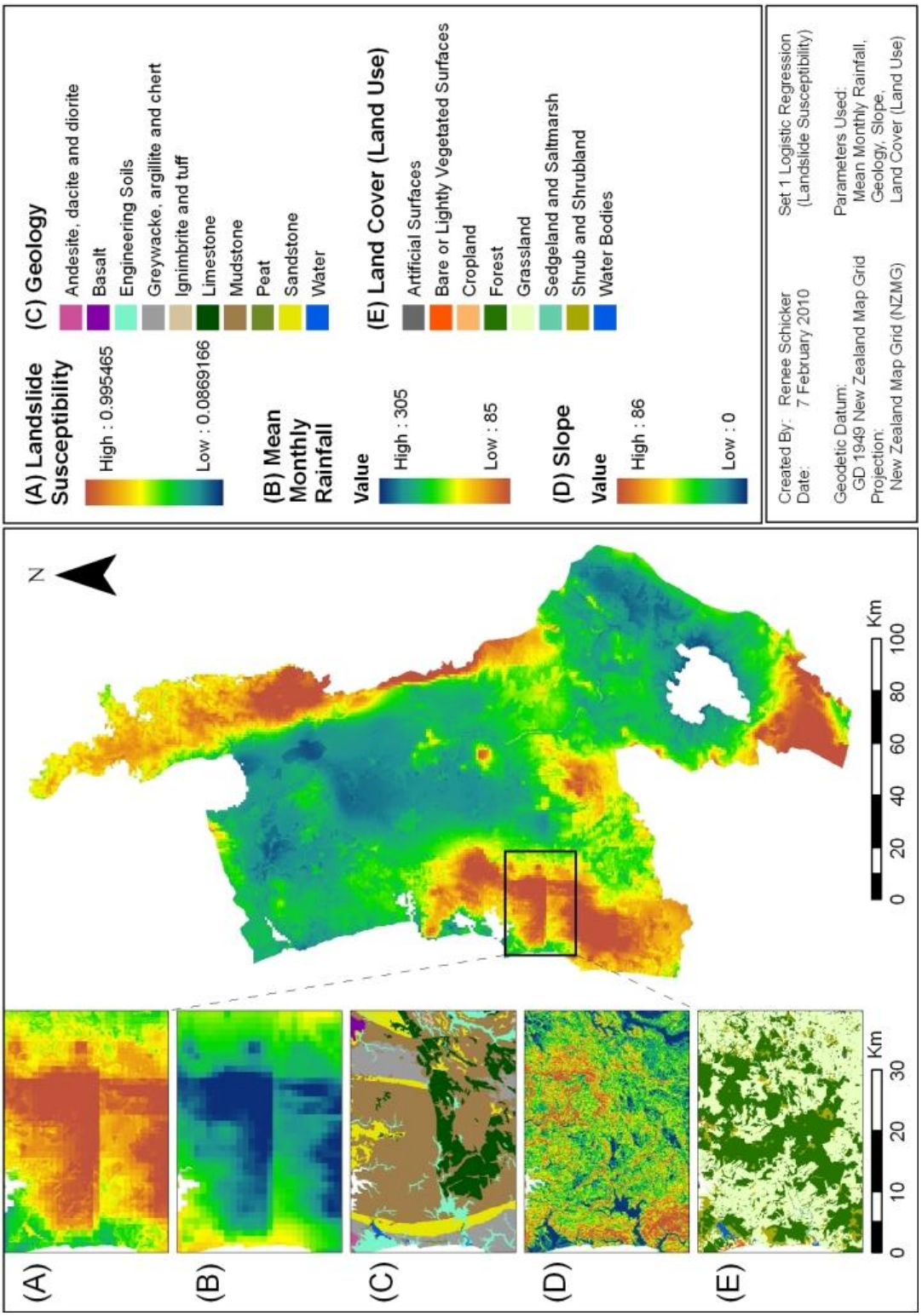


Figure 5.1 Landslide susceptibility based on the Set 1 logistic regression model and a series of insets (A-E) displaying the composition of parameters within the area indicated on the regional map of landslide susceptibility where an unusual straight edged feature has resulted (A). The parameters are (in order of inset), mean monthly rainfall (B), geology (C), slope (D), and land use (E) and the classes for each are given in the legend which has also been labelled according to Inset.

The number of rain gauges for the size of the area is quite small. The rainfall estimates are derived through correlation of nearby rain stations with some regard to the topology (Leathwick *et al.*, 1998). There is potentially an issue in areas with few rain gauges to determine rainfall estimates from.

The area of focus in the inset (Figure 5.1) looks to be the only case where the mean monthly rainfall has brought in a strange effect (Map 3.12). The remaining parameters (Figure 5.1 C-E) involved in the model appear to have a more subtle effect or have been overshadowed by the dominant effect of the mean monthly rainfall as these are harder to observe.

5.5.2 Classifying the Susceptibility Map

5.5.2.1 Defining Susceptibility using Different Classification Techniques

In order to classify the results of logistic regression in GIS the data had to first be simplified as ArcMap could not classify the many values with six decimal places. The simplest approach was to convert the resulting raster layer to integer after multiplying all values by 100 (to avoid a binary outcome), thereby creating a percentage form of the values if they had been rounded to two decimal places. The mean, standard deviation and distribution of the data (Figure 5.2) could then be observed in ArcMap and classifications applied.

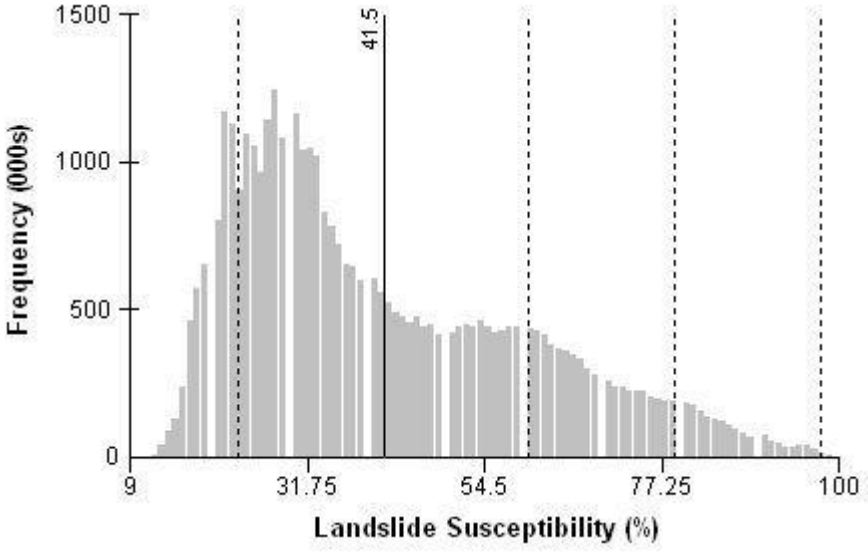
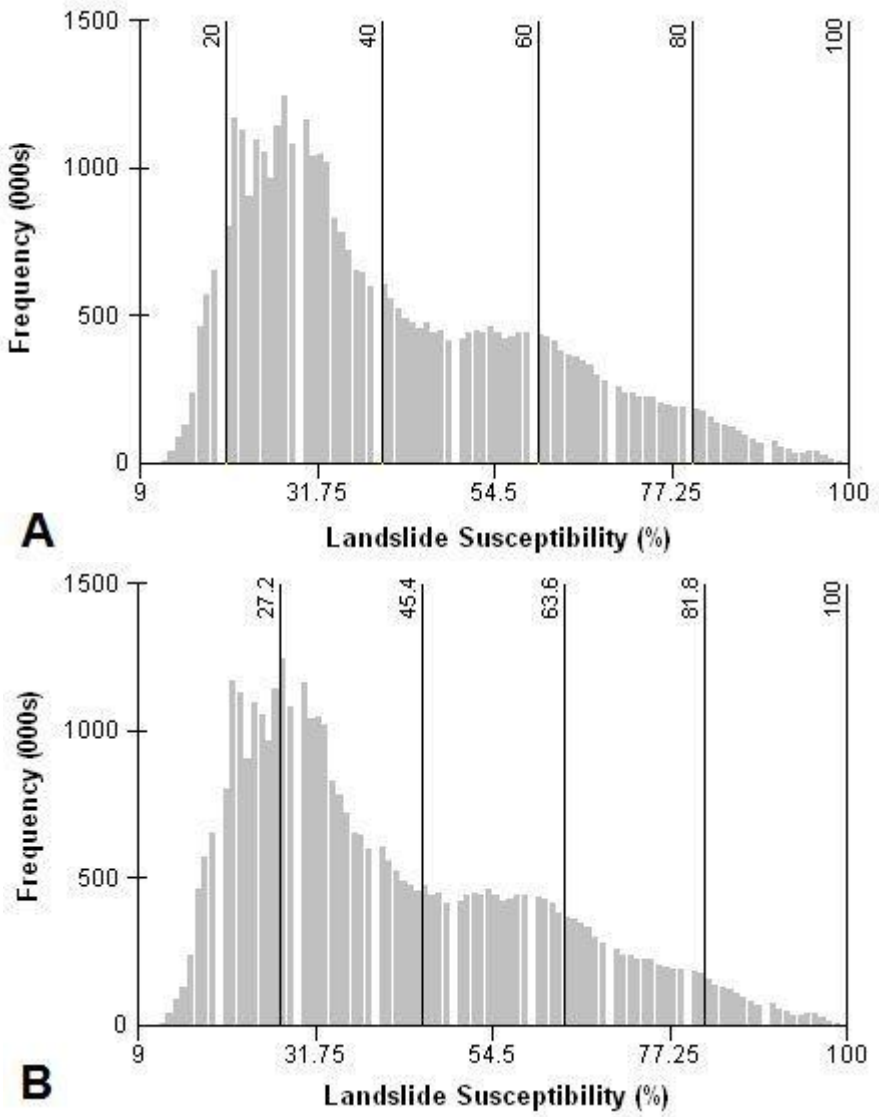


Figure 5.2 Probability distribution (percent form) for the Set 1 logistic regression results. The distribution has a mean (black solid line) of 41.5 (1 d.p.) and a standard deviation (black dashed lines) of 18.7 (1 d.p.).

Five different classification techniques, all of which each split the distribution of the data into five classes (very low, low, moderate, high and very high) differently, were investigated. These were: defined intervals of 20% (Figure 5.3 A); equal intervals (Figure 5.3 B); geometric intervals (Figure 5.3 C); natural breaks (Figure 5.3 D); and quantile (Figure 5.3 E).



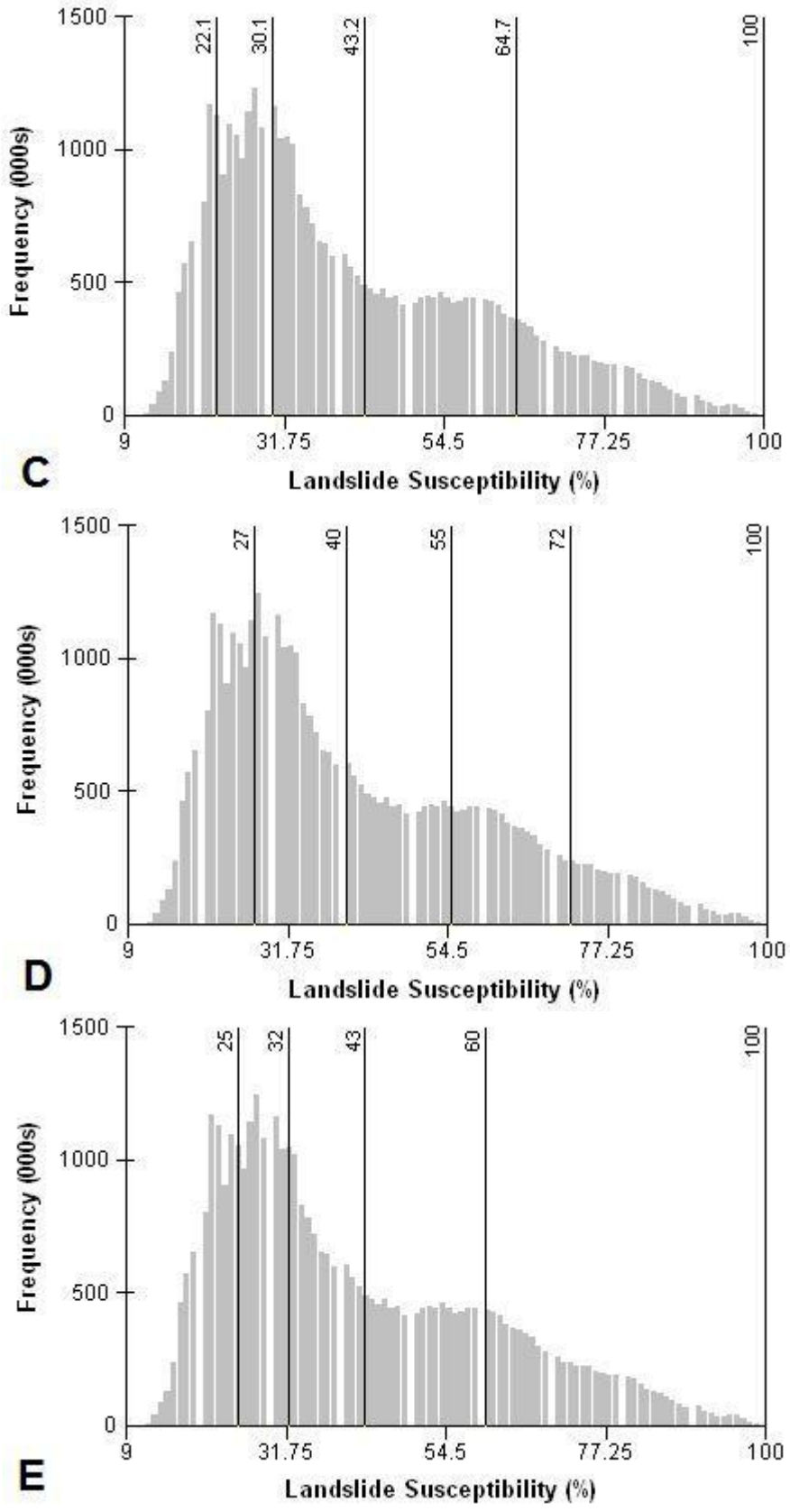


Figure 5.3 Distribution of the logistic regression results based on the class breaks for defined interval (A), equal interval (B), geometric interval (C), natural breaks (D), and quantile (E) classification techniques.

The resulting maps based on each of the five classification methods (Figure 5.4) are similar in some of the susceptibility classes as expected, but the extent of coverage for the classes varies.

All five classifications appear to represent the flat areas of the Hauraki Plains, Hamilton Basin, and the Central Plateau well with very low to low susceptibility. Most of the high to very high susceptibility areas appear where expected in the Coromandel, South Waikato, and in the vicinity of the Central North Island volcanoes. The defined interval classification using increments of 20% has classed the majority of the region as low susceptibility as a result of a large proportion of values falling in the 20 – 40% range due to the skewed distribution (Figure 5.3 A). The geometric interval (Figure 5.4 C) and quantile (Figure 5.4 E) classifications look to classify a greater area in the moderate to very high susceptibility classes which may not reflect reality. The geometric interval classification also has the least area in the very low susceptibility class which is an effect of the classification (Figure 5.3 C).

For each of the five classified maps the landslide inventory was overlain (Figure 5.4) to get a visual assessment of the goodness of fit. For all five classifications it appears most the landslides fall in the high to very high susceptibility classes. However, fewer landslides correspond with the high to very high susceptibility classed areas in the defined interval and equal interval classifications (Figures 5.4 A and B). For both defined interval and equal interval classifications most of the region has been classed as the lower susceptibility classes and less as higher susceptibility, which is a result of the type of classifications being applied to the skewed distribution (Figures 5.4 A and B). These two classifications do not appear appropriate as the data is not normally distributed. Natural breaks therefore make the most sense based on visual observations of the splitting of the distribution and the resulting maps.

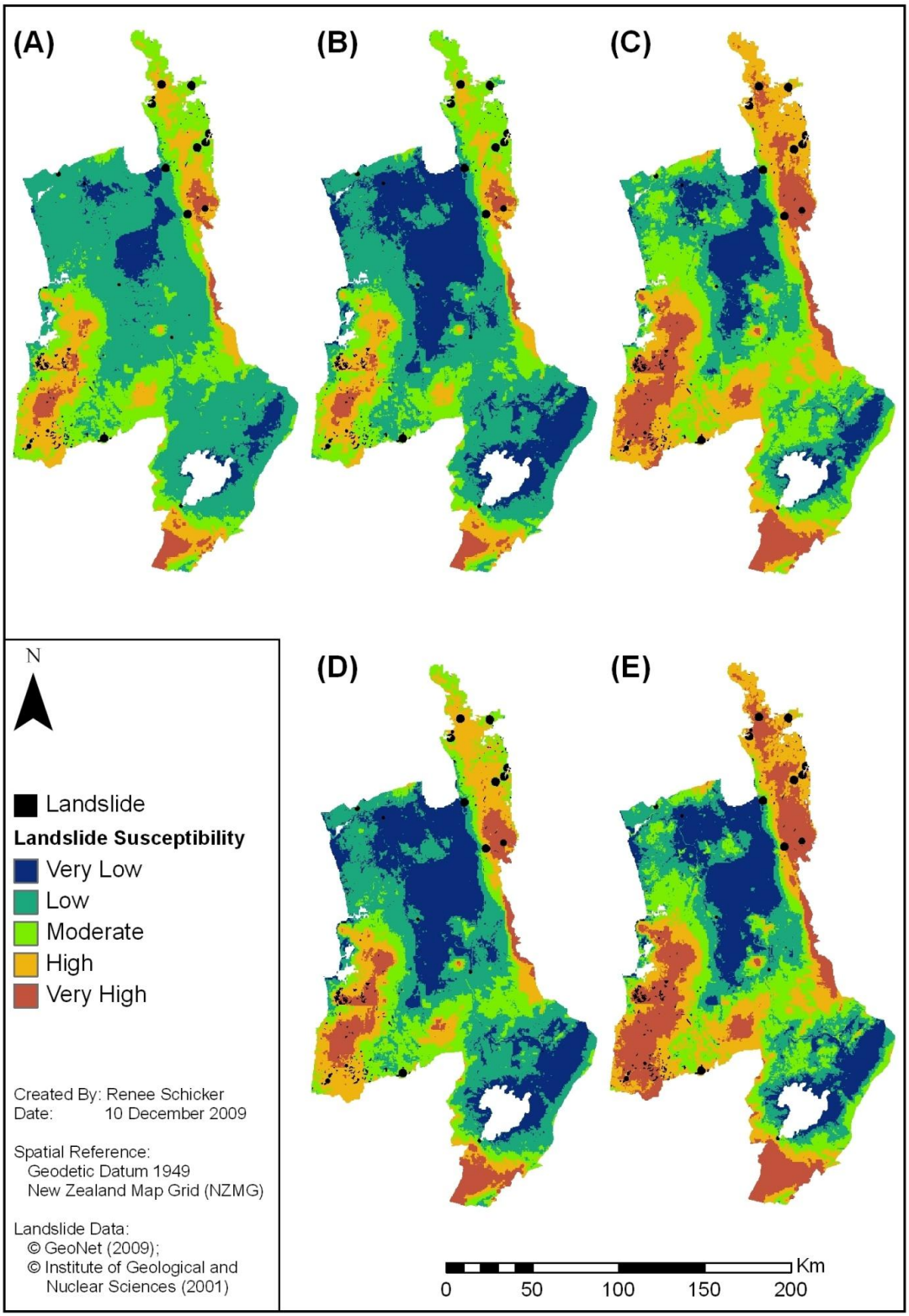


Figure 5.4 Landslide susceptibility for the Waikato Region determined by logistic regression (using Set 1 model), classified by (A) defined intervals, (B) equal intervals, (C) geometric intervals, (D) natural breaks, and (E) quantile with the landslide inventory (QMap landslide data © GNS 2001, and landslide catalogue © GeoNet 2009) overlain in black.

5.5.2.2 Statistical Evaluation of the Classification Techniques using GIS

In GIS, the landslide susceptibility raster data was classified according to each classification technique to create five different raster layers. These raster layers were then converted to vector format and individually unioned with the landslide inventory. Using these union vector coverages, summary statistics were carried out in ArcMap to calculate the area occupied by landslides and non-landslides for each susceptibility class. The percent of total area (Figure 5.5), non-landslide area (Figure 5.6), and landslide area (Figure 5.7) within each class for each classification was determined. Each classification split the data differently and this is best observed when the total area in each class is observed (Figure 5.5).

Despite the distirbution being skewed (Figure 5.2), the percent of total area in each class using geometric interval classification appears normally distributed (Figure 5.5). Using the quantile classification, there appears to be an even proportion of the total area in each class (Figure 5.5). Some studies have found that the disadvantage of using the quantile based classification is that it groups widely different values into the same class (Akgün and Bulut, 2007; Ayalew and Yamagishi, 2005). Defined intervals, equal intervals and natural breaks display the skewedness of the data, but to different extents (Figure 5.5) as a result of different class break values.

The percent of non-landslides (Figure 5.6) follows the same trend seen with the total area (Figure 5.5) as the majority (98.56%) of the Waikato region has no record of landslides according to the landslide inventory. While the total landslide area is only a small part of the total area, the percent of this area in each susceptibility class of each classification technique (Figure 5.7) is quite important. The quantile and geometric interval classifications which have the greater percentage of landslide area classed in the high and very high classes would appear to be the best. However, based on the choice of class breaks, the resulting area in each susceptiblity class, visual observations, and findings from the literature it was determined that neither of the quantile or geometric interval classifications would be used in the final map. Similarly, defined intervals and equal intervals will not be used as these have a tendency to under-represent the

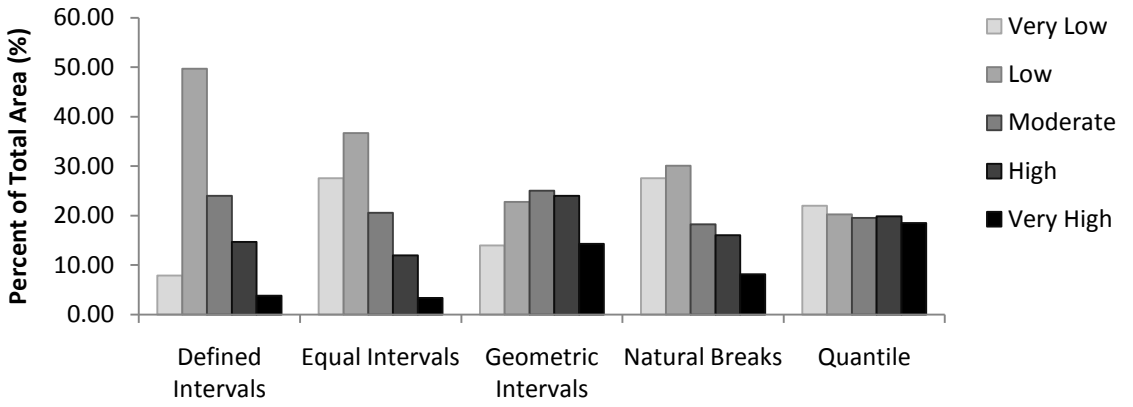


Figure 5.5 Percentage of the total area in each susceptibility class (very low to very high) for each of the five classification techniques.

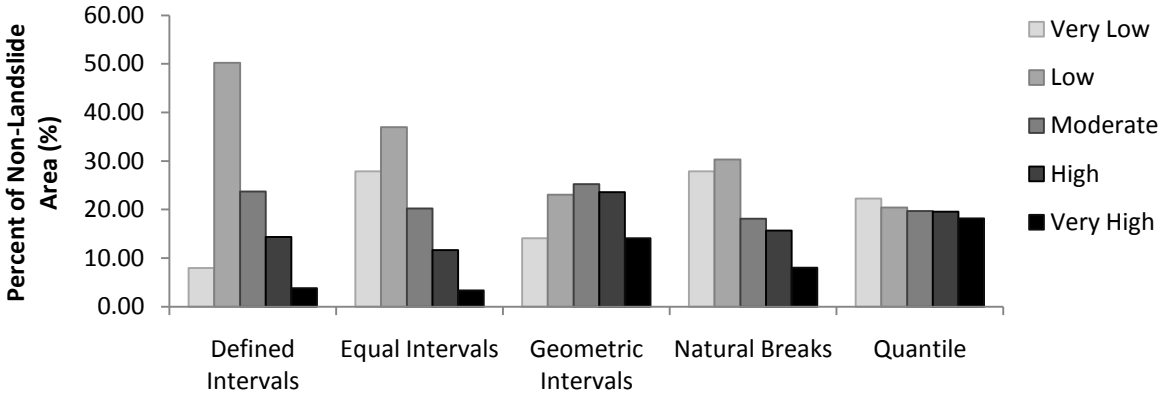


Figure 5.6 Percentage of the total non-landslide area in each susceptibility class (very low to very high) for each of the five classification techniques.

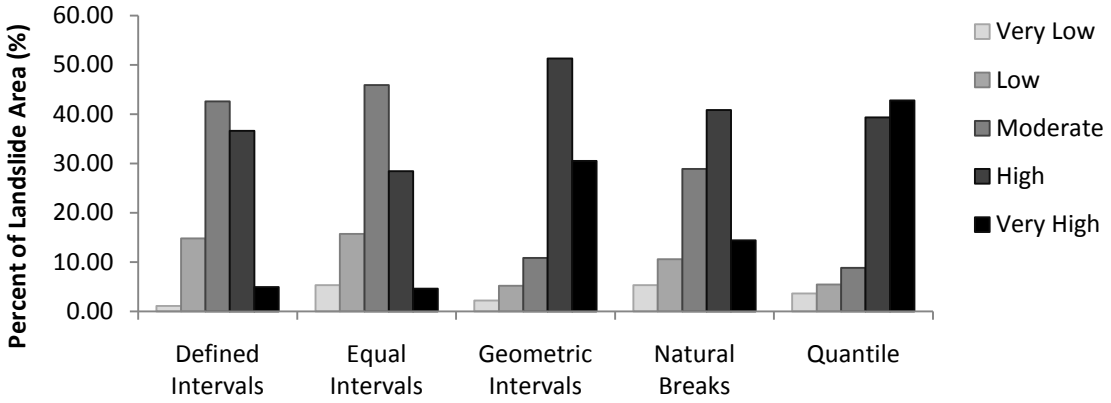


Figure 5.7 Percentage of the total landslide area in each susceptibility class (very low to very high) for each of the five classification techniques.

landslide area. This is a result of both classifications using equally spaced class breaks. The defined interval, being based on the probability as a percentage falling between 0 and 100, uses increments of 20% for the five classes. Equal intervals have been found unhelpful in some studies as it emphasises that each class is relative to the others (Akgün and Bulut, 2007; Ayalew and Yamagishi, 2005; Nandi and Shakoor, 2009) despite this, it has been used when comparing two statistical models (Nandi and Shakoor, 2009; Yesilnacar and Topal, 2005). When applied, equal intervals splits the range of the data into five classes, and as the data ranges from 9 – 100% these do not result in the same class breaks as the defined interval. The data are skewed in favour of the lower classes, so if equal or defined intervals are applied, a large proportion of the data is classed by the lower susceptibility classes, and as the spacing is consistent, a smaller proportion is classed by the higher suscetpibility classes.

The natural breaks classification looks the most appropriate as it classifies according to breaks in the data. Natural breaks has managed to classify in a manner in which the percent of total area and non-landslide area decrease from the low to very high susceptibility classes while the percentage landslide area increases with increasing susceptibility (with the exception of the very high class). Natural breaks are more appropriate when obvious jumps in the data values exist (Akgün and Bulut, 2007; Ayalew and Yamagishi, 2005; Nandi and Shakoor, 2009) which is the case in this study. Therefore the resulting logistic regression map (Map 5.1; Appendix 5.4) has been classified using natural breaks to determine the susceptibility classes.

5.6 Conclusion

A thorough approach has been taken at each stage of the application of logistic regression. The type of model, most appropriate link function, sampling technique and sampling size have all been considered leading up to the production of an acceptable model. A generalised linear/non-linear model for multiple regression analysis using the logit link function to predict the presence or absence of landslides was chosen. The logit link function was chosen as it gave the best results and is in keeping with the bulk of the literature.

A significant model with a χ^2 of 16952.2 resulting in a p value of 0.0158 was obtained from the Set 1 sample set using slope, mean monthly rainfall, land use and geology as predictors in the model. The model derived from Set 1 was applied to the Set 2 sample set and was found to have a very similar predictive accuracy of observed landslide and non-landslide samples.

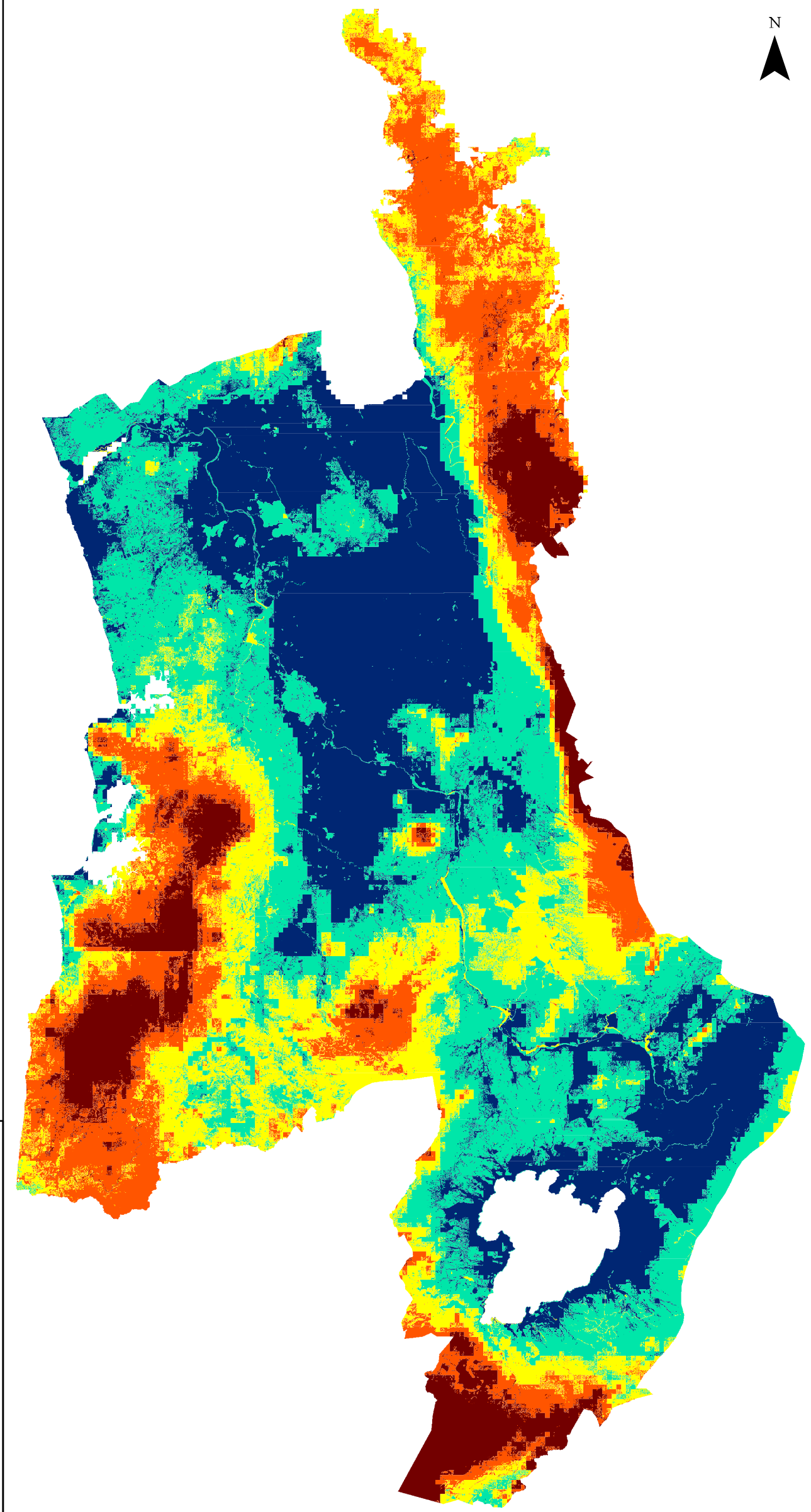
The logistic regression model was applied in GIS and the resulting map was visually assessed before it underwent classification. The visual assessment was carried out to determine the cause of a discontinuity which is apparent in an area of high to very high susceptibility in the South Waikato. This feature was found to be mainly a result of the mean monthly rainfall spatial data and is thought to be a combined product of how the data was originally derived and low resolution (pixel size of 1 km).

Five different classifications were applied and compared both visually and statistically to determine which would be the most appropriate. The quantile and geometric interval classifications while both having the larger proportions of landslide area classed as high or very high landslide susceptibility do not reflect reality. Defined interval and equal interval were not appropriate given they both use equal sized class breaks, and the distribution is skewed with a larger proportion of low values, which results in map that underestimates the observed landslide areas. The natural breaks classification was found to be the most appropriate choice and was applied to produce the final map (Map 5.1).

Map 5.1 Landslide Susceptibility for the Waikato Region Determined by Logistic Regression

Landslide Susceptibility

- Very low
- Low
- Moderate
- High
- Very High



Created By: Renée Schicker
Date: 18 February 2010

Projection: New Zealand Map Grid (NZMG)
Datum: Geodetic Datum 1949

Model Used: Logistic Regression of Set 1 Data

Equation: P = 1/ [1 + EXP(-z)]

Predictor Parameters:
Slope;
Mean Monthly Rainfall;
Land Cover (Land Use);
Geology.

Classification: Natural Breaks

CHAPTER SIX

Evaluation

6.1 Introduction

This chapter outlines the need for validation/evaluation of landslide susceptibility maps and various techniques that exist for this purpose. Validation and evaluation differ in their meanings but are used to explain the same purpose of testing the model output and obtaining an indication of how well the model fits and how good its predictive capability is. This is done by testing the model using a different set of landslide location data to that used to derive the model, and can be done both quantitatively using validation curves and qualitatively using simple map overlays. This chapter will explore some of these techniques in the process of validating the landslide susceptibility maps derived from the weights of evidence and logistic regression approaches to determine which map is best.

6.2 Validation or Evaluation

Landslide susceptibility assessments commonly require some form of assessment to test the model accuracy. Some studies refer to this process as validation, but this implies there is an absolute acceptability as in a case of yes or no, right or wrong, when in fact what is actually being conducted is an evaluation of relative degree of acceptability as a case of being better or worse (Carrara and Pike, 2008). A true validation of landslide susceptibility models is only realistically possible using landslides which occur in the time following the creation of the map, which means adopting a "wait and see" approach (Ermini *et al.*, 2005; Guzzetti *et al.*, 1999; Neuhäuser and Terhorst, 2007). The problem with that approach is that there could be a long wait and even then it may never be validated, and a model which has not been validated is considered to have no scientific significance (Van Den Eeckhaut *et al.*, 2006). The "wait and see" approach is unacceptable as some measure of validity and predictive power

should be offered with the map, especially if it is to be used in land use planning decisions (Neuhäuser and Terhorst, 2007). Instead of waiting for nature to prove the model right or wrong, landslide modellers have turned to numerically testing their models to obtain an indication of better or worse predictive capability (Carrara and Pike, 2008). In order to numerically test the mapped model output, it is advised that a landslide population independent to the one used in deriving the model output be used in some qualitative or quantitative evaluation (Remondo *et al.,* 2003). A good model should be able to differentiate between significantly different landslide density conditions, as well as being statistically reliable, in which case it should have a great dispersion or spread around the mean density value (Clerici *et al.,* 2006).

6.2.1 Assumptions

Two basic assumptions are required when evaluating susceptibility models; the first is that landslides are related to the spatial information for the predictor factors (such as slope, geology, and land cover), the second is that a specific impact factor such as rainfall or earthquake will trigger future landslides (Lee *et al.,* 2003a; Lee 2004, 2005, 2007b; Lee and Lee, 2006). This follows with the principle that "the past is the key to the future" which assumes that future landslides are more likely to occur in areas consisting of the same, or similar, geomorphic, geologic, and hydrologic factors that have led to past and present failures in the study area (Remondo *et al.,* 2003; Varnes, 1984). A landslide susceptibility map is considered satisfactory if there is reasonable agreement between it and the locations of existing landslide movements in the area, but this is more a validation of its 'success rate' than the predictive value (Remondo *et al.,* 2003).

6.3 Evaluation Techniques

Several methods of evaluation exist, and these can be either qualitative which could just be a simple visual overlay; or quantitative which may involve looking at the area of class affected by landslides as a ratio of total class area expressed by some means of indices (Remondo *et al.,* 2003).

6.3.1 Ground-truthing and Simple Overlay

In some landslide susceptibility assessments, validation is done by ground-truthing through field surveys (Abdallah *et al.,* 2005; Yesilnacar and Topal, 2005) and/or photo interpretation (Weirich and Blesisus, 2008). In some studies, field surveys are employed to create a validation sample set (Dominguez-Cuesta *et al.*, 2007; Kanungo *et al.,* 2006). The validation sample set can be applied in a qualitative approach of simple overlay, or in a quantitative approach to construct validation curves. In the simple overlay approach, the validation landslide dataset is separately overlain each susceptibility map and the percentage of landslides in each susceptibility class determined (Akgün and Bulut, 2007; Arora *et al.,* 2004; Ayalew *et al.,* 2005).

6.3.2 Validation Curves

There are two decision rules to consider: (1) the majority of the mapped landslides should be located in pixels of high susceptibility classes; and (2) the high susceptibility classes should cover a relatively small area on the map (Can *et al.*, 2005; Duman *et al.,* 2006). This can be observed through the construction of validation or rate curves which give an indication of how well the output from the chosen model and variables predict landslides (Lee and Dan, 2005).

Validation curves can be constructed by comparing the validation sample of landslide occurrence with the susceptibility classes in the mapped output (Remondo *et al.,* 2003). This can easily be done in GIS by combining the susceptibility layer with the landslide occurrence layer (union) and obtaining the summed area for each susceptibility class with and without landslides (summary statistics). Validation curves depict the landslides in the validation sample as a cumulative percentage (y-axis) with respect to decreasing susceptibility levels (x-axis) which are often expressed as cumulative percentages of the study area (Conoscenti *et al.,* 2008; Remondo *et al.,* 2003).

6.3.2.1 Success and Prediction Rate Curves

6.3.2.1.1 Similarities and Differences

Both success rate and prediction rate curves are applied to the landslide susceptibility model output in the same fashion, just the landslide data being used in the comparison varies (Chung and Fabbri, 1999, 2003). Success rates give an indication of the goodness of fit, or how well the predictive model fits the landslides from which it was derived (Chung and Fabbri, 1999; Conoscenti *et al.,* 2008). Prediction rates measure how well the derived landslide susceptibility model predicts future landslides using an independent landslide dataset (Chung and Fabbri, 1999; Lee *et al.,* 2003a). As the same process is used to apply both success rate and prediction rate curves, it should be clearly stated that an independent landslide location data set is being used; however this is not always the case.

6.3.2.1.2 Causes for Confusion

Some studies discuss high prediction accuracy with reference to cumulative frequency diagrams or rate curves verified using "known" landslide locations (Lee, 2005; Lee and Evangelista, 2006; Lee and Pradhan, 2007) which, on appearance could be either success rate or prediction rate curves but are not implicitly stated as one or the other. The use of "known" landslide locations does not give any differentiation between the landslide locations used to derive the model and those to validate it. It remains unclear if there even was an independent set of landslide location data and whether the curve presented is explaining the accuracy of the model fit (goodness of fit) or an indication of its predictive capability (validation). Without the independent landslide location it is only a goodness of fit and not an indication of how good its predictive capability is (Remondo *et al.,* 2003). The other issue is the use of success rate curves and subsequent reporting of predictive accuracy (Lee, 2005) when in fact the curve is displaying model accuracy.

6.3.2.1.3 Constructing Either of the Curves

Prior to constructing a success curve, the landslide inventory is overlain and the joint frequency of landslide presence with each susceptibility value is calculated

(van Westen *et al.,* 2003). The same process follows for prediction rate curves using a separate set of landslide location data. The pixels in the landslide susceptibility are sorted in descending order and the cumulative percentages are calculated for both landslides and proportion of susceptibility classes in descending order (Dahal *et al.,* 2008a; Lee, 2005; Lee *et al.,* 2003a; van Westen *et al.,* 2003). The cumulative frequency of landslides (y-axis) is plotted against the landslide susceptibility index rank (Lee, 2005) or percentage of susceptibility map (van Westen *et al.,* 2003) (x-axis) (Figure 6.1). The susceptibility values decrease from left to right on the x-axis while the cumulative percent susceptibility values increases. For instance, the 90-100% susceptibility values equate to the top 10% of the susceptibility classes (at the very high end).

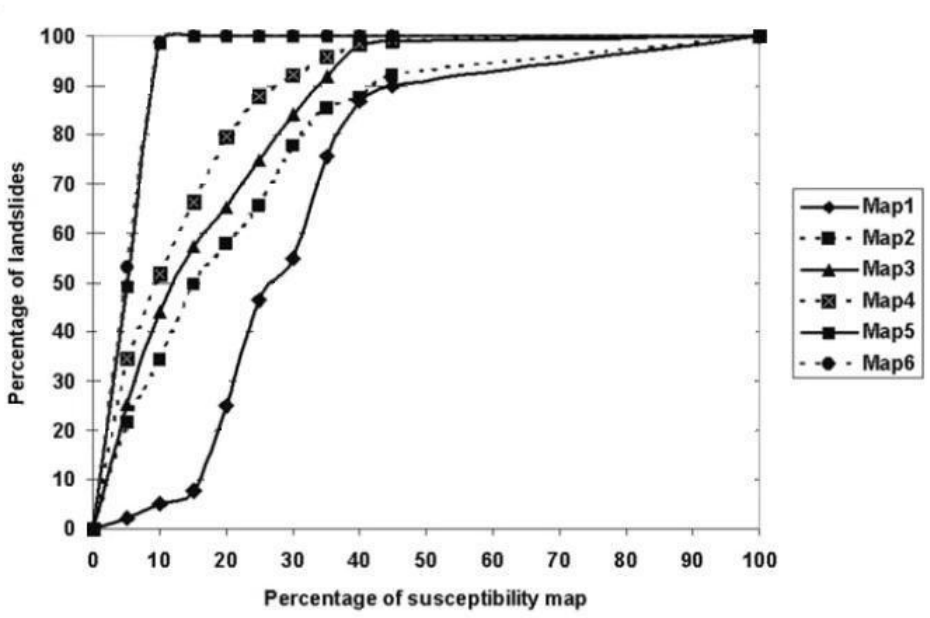


Figure 6.1 Success rate curves adapted from van Westen *et al.* (2003) for six different mapped landslide susceptibility models. The cumulative percentage of all landslides in the study area (Y-axis) is plotted against the proportion or percentage of susceptibility map (X-axis) in descending order (higher susceptibilities on the left and lower susceptibilities on the right).

6.3.2.1.4 Interpreting Prediction and Success Rate Curves

If the prediction rate curve coincided with a diagonal from 0 to 1 (or 0 to 100%) the prediction would be considered to be totally random, whereas a validation curve for a given model, further up and away from the diagonal will result in a better predictive value (Conoscenti *et al.,* 2008; Remondo *et al.,* 2003). In theory, this applies to success rate curves too, but instead of random prediction, the

diagonal would indicate a random model fit, and further up and away from this diagonal would indicate a better fit. The gradient in the first part of the curve is often examined as a greater gradient is indicative of a greater predictive capability (Conoscenti *et al.,* 2008; Remondo *et al.,* 2003). For both success and prediction rate curves, how large the percent of landslides explained by the classes of highest value (often the top 10% , 20% or 40%) of the susceptibility classes is generally of interest (Conoscenti *et al.,* 2008; Dahal *et al.,* 2008a; Lee, 2005; van Westen *et al.,* 2003).

Both success rate and prediction rate curves can be plotted on the same graph and compared (Figures 6.2 and 6.3). A high success rate and prediction rate is most desirable (Figure 6.2), but a high success rate alone is not indicative of a high prediction rate (Figure 6.3).

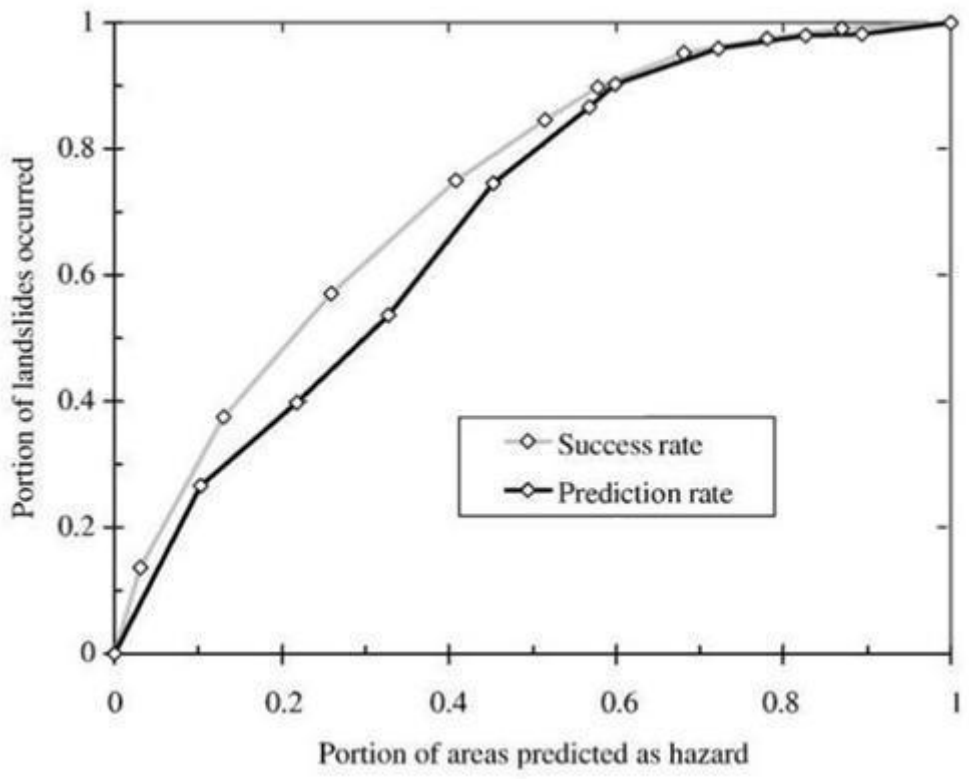


Figure 6.2 Example adapted from Conoscenti et al. (2008) of a prediction rate curve and a success rate curve, where both the success rate and predictive rate are quite high and very similar.

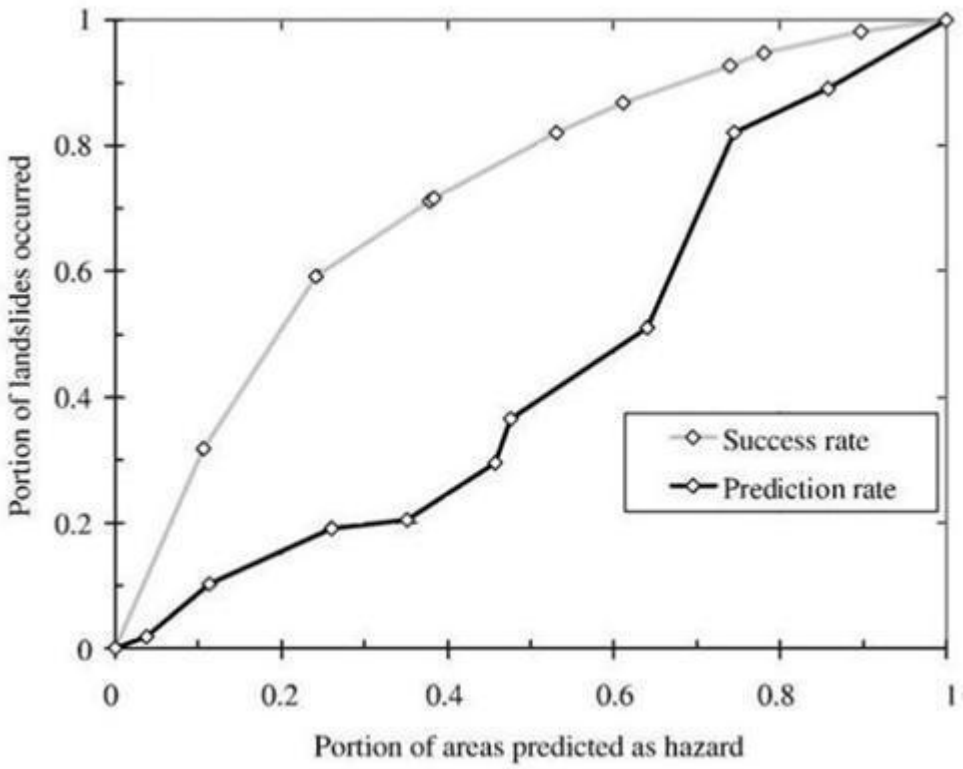


Figure 6.3 Example adapted from Conoscenti *et al.* (2008) of a prediction rate curve and a success rate curve where the success rate is high but the prediction rate is low.

6.3.2.2 Areal Cumulative Curve

Unlike the previous cumulative success and prediction rate curves, the cumulative curve in this section does not involve sorting or pixel ranking. Instead of ranking pixels in descending order, the susceptibility values are plotted in ascending order on the x-axis, and the cumulative percentage (y-axis) of both landslides in the study area and total susceptible area are plotted (Figure 6.4). The first curve (Figure 6.4, curve-a) represents the cumulative percentage of pixels containing only landslides versus landslide susceptibility, and the second (curve-b) represents the cumulative percent of pixels within the susceptibility classes which defines the areal distribution throughout the region (Duman *et al.,* 2006).To satisfy the decision rules (section 6.3.2), the value obtained for the first curve (landslides versus susceptibility) at the cut-off (of 50%) should be as small as possible, while the value for the second curve should be relatively high in comparison (Duman *et al.,* 2006).

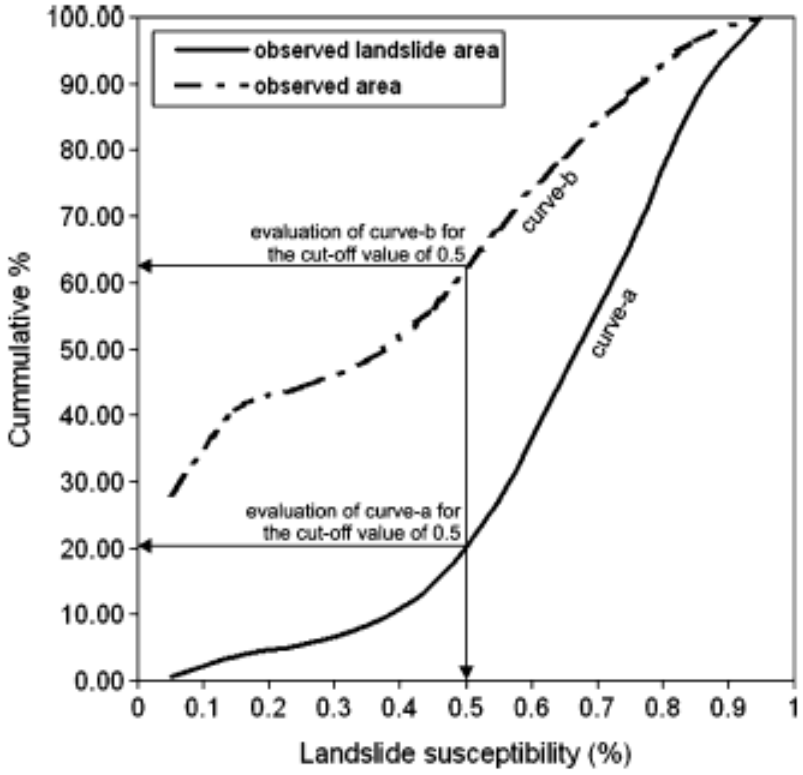


Figure 6.4 Cumulative frequency graph from Duman *et al.* (2006) which depicts the two curves and the cut-off of 0.5, which indicates at 0 – 50% susceptibility that in ~ 63% of the study area ~20% of the landslides are explained by the susceptibility model. This then means that ~ 80% of the landslides are explained by the higher susceptibilities in the remaining 37% of the study area.

6.4 Evaluation Trials

An evaluation of both logistic regression and weights of evidence derived landslide susceptibility maps was required using an independent landslides dataset. The cumulative area curve and prediction rate curve were considered in order to ascertain the predictive capacity of the two models and to determine which would be better. A random sample set of individual 1 km^2 boxes was created and applied in Google Earth. A record of the number, widths, and coordinates of landslides within each sample box was obtained (Appendix 6.1), and used to create the validation spatial dataset.

6.4.1 Google Earth Landslide Density

In order to represent each of the five classes fairly, it was decided that each class be randomly sampled to obtain ten randomly spread locations. A boxed area of

1 km^2 around each of the fifty sampling point locations was chosen. The database compiled in chapter three was converted to a set of spatial layers (shapefiles) of point locations spaced 25 m apart. Each point aligns with the centre of each pixel in each of the raster datasets, and in an extraction the point would take a record of the value of the pixel in the raster layer being extracted from. This means the susceptibility classes from each susceptibility map could be compiled into a database which could then be randomly sampled.

The logistic regression and weights of evidence derived susceptibility classes were extracted using the point layer and added to the database (Appendix 3.4). The database was then divided into the five susceptibility classes and these new databases were added to STATISTICA. Stratified random sampling was used to derive 20 samples from each class, which were then converted to a shapefile in ArcCatalog. A looped script (Appendix 6.2) was created to pull each point out separately, and convert it to a raster with a pixel size of 1 km by 1 km. Based on the idea that each point would act as a centroid, this would be the most effective method of creating a 1 km^2 sampling grid around a point. The raster image was converted to vector and labelled according to sampling number. One sampling box had a large proportion of ocean in it, so was excluded.

Stratified random sampling was again carried out in STATISTICA to obtain ten samples from the samples of twenty for each susceptibility class. While there were two occurrences of sampling grid overlaps when all 100 sampling grids were observed in ArcMap, the final random sample of 50 obtained did not result in any overlaps.

The available version of Google Earth did not allow for GIS data to be uploaded, although this is possible in more advanced versions. Instead the extent of each 1 km^2 sampling box was recorded using ArcMap to obtain coordinates in latitude and longitude (degrees, minutes and seconds), which were then manually entered into Google Earth to replicate the extent. A sampling box was then drawn using the extent points as markers for the vertices, and assigned a label. Landslides within the sampling box were marked and numbered according to the susceptibility class, sampling box and landslide number. For sample boxes with good imagery in which landslide features could clearly be identified, a count of

the landslides was tabulated and the class total and average number of landslides per square kilometre determined (Table 6.1; Appendix 6.1).

Table 6.1 Count of landslides and average number of landslides per square kilometre for each susceptibility class based on the number of sample boxes with good imagery.

Classed Susceptibility	Number with good imagery	Landslide Count	Count Range	Average per Km^2	Standard Deviation	Standard Error
Very High	8	62	2-18	7.75	5.15	1.82
High	10	78	1-16	7.80	5.49	1.74
Moderate	10	28	0-12	2.8	3.88	1.23
Low	7	1	0-1	0.14	0.38	0.14
Very Low	9	0	0	0.00	0.00	0.00

There seems to be little distinction between the very high and high classes as indicated by the similar average number of landslides per square kilometre (Table 6.1), so it may be feasible to classify the map by four classes instead of five.

6.4.2 Recreating the Validation Landslide Dataset in GIS

A record of the coordinates of each landslide marked in Google Earth and the approximate width (measured with the Google Earth scale bar) were made and applied in GIS. This required converting longitude and latitude in degrees, minutes and seconds to degrees (Appendix 6.1) and having to apply a coordinate system (New Zealand Geodetic Datum 1949) before defining the projection (New Zealand Map Grid). Circular areas were drawn around the points using the approximate width as a radius (buffer) before they were combined (union) with the layer of all sample grids and later combined with the landslide susceptibility maps.

Summary statistics were used in ArcMap to calculate the sum of area based on the susceptibility percentage, sample grid presence, and ground-truthed landslide presence.

6.4.3 Cumulative (Area) Curve

The susceptibility output raster layers (prior to classifying into five classes) were converted to vector format and combined (union) with the polygon layer of ground-truthed landslide locations. The area occupied by the presence and absence of these landslides within each percentage of model derived susceptibility (1% intervals) was obtained through summary statistics in ArcMap. The cumulative percent of area for both landslide only area and total sample box area were calculated and plotted (Appendix 6.3) against the derived model susceptibility values for both weights of evidence and logistic regression approaches (Figures 6.5 and 6.6). In the lower susceptibility values (0 – 50%) it is desirable to have a small percentage of the total landslide area and a large percentage of the total area, so that the higher susceptibilities (50 – 100%) have a small percentage of the total area and a large proportion of the total landslide area. The cumulative curves for logistic regression (Figure 6.5) give a good result, as over 55% of the total sample box area is explaining approximately 11% of the landslides, meaning the higher susceptibilities account for about 44% of the sample area and about 89% of the total ground-truthed landslide area.

The cumulative curves for the weights of evidence give a poor result (Figure 6.6) in comparison to the set of curves obtained for the logistic regression model (Figure 6.5). The weights of evidence cumulative curves (Figure 6.6) may show a great result for proportion of ground-truthed landslides (< 1%) in the lower susceptibility classes, but the proportion of the sample area is also quite low (< 29%). This looks to be a poor result as 71.3% of the sample area is explaining approximately 99% of the landslide area. When the cumulative total area is instead considered, landslide susceptibilities > 72% account for approximately 50% of the total area. While the model agrees with the first assumption in section 6.3.2, it violates the second assumption as the higher pixels cover a large proportion of the map.

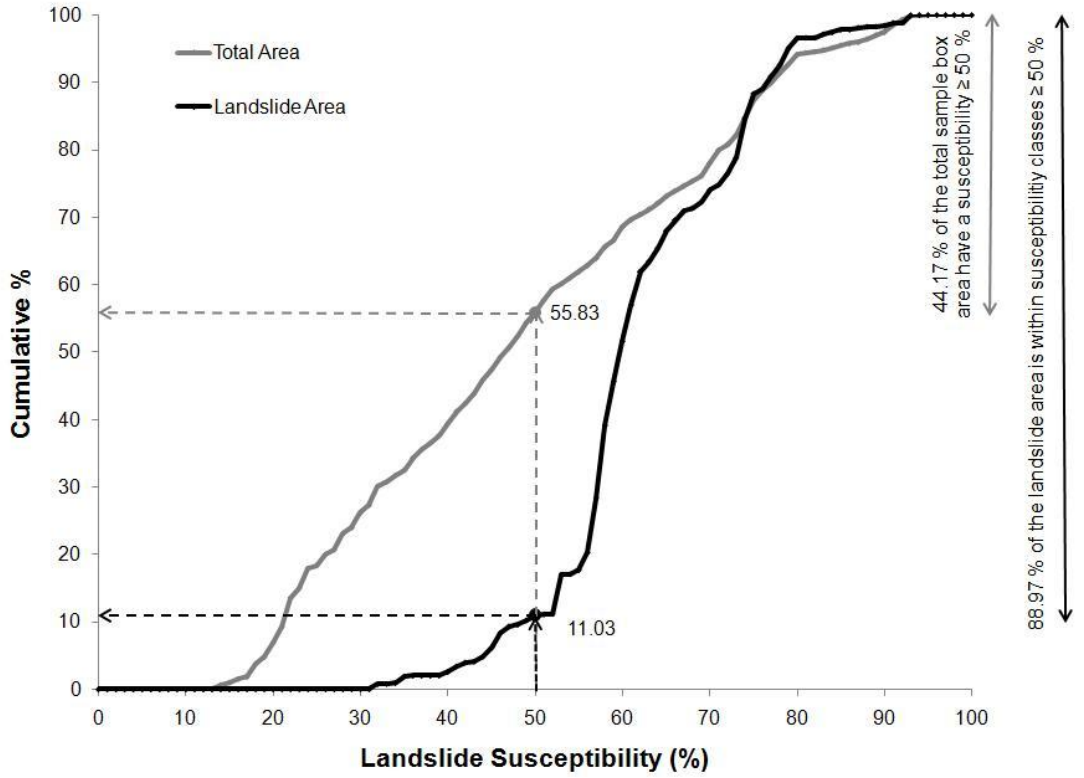


Figure 6.5 Cumulative area curve of ground-truthed landslides and total sample box area with increasing susceptibility values for the logistic regression approach.

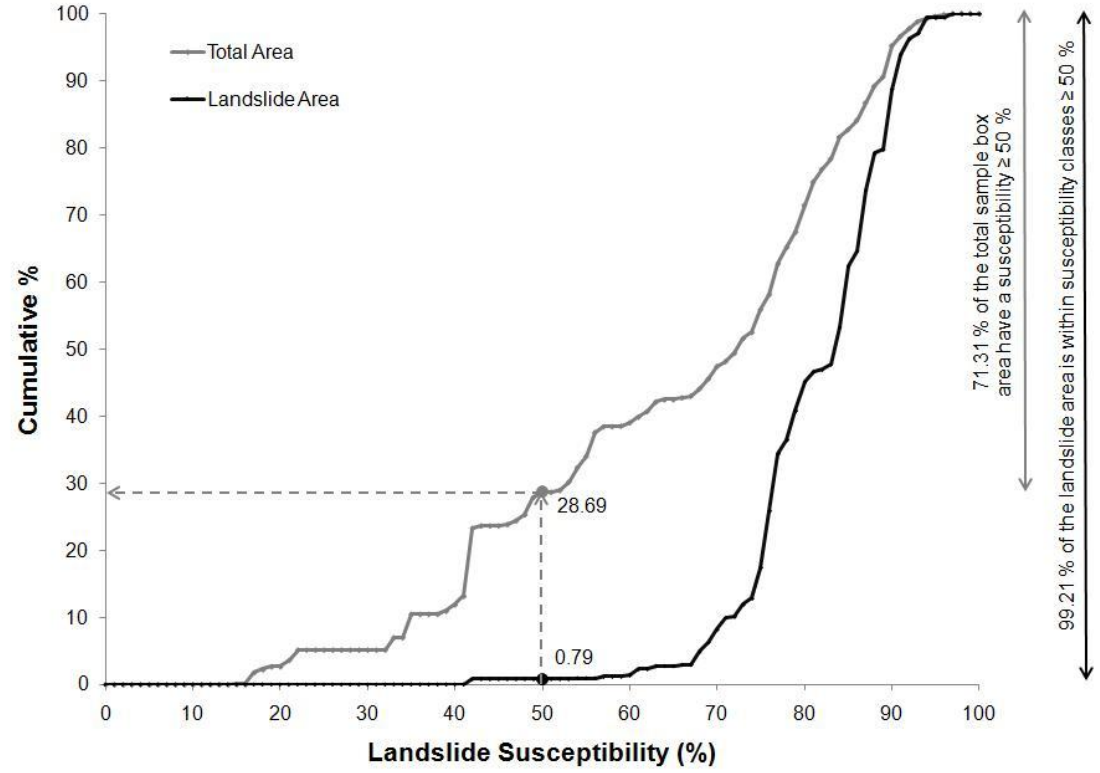


Figure 6.6 Cumulative area curve of the ground-truthed landslides and total sample box area with increasing values of susceptibility for the weights of evidence approach.

6.4.4 Prediction Rate Curve

6.4.4.1 Compiling and Ranking

A prediction rate curve is based on a proportion of susceptible pixels and the cumulative proportion of these with landslide. A simple approach to obtaining the data required is to extract values from the raster layer using a vector layer of points centred to the middle of each pixel. The table of data attached to the point layer (attribute table) can then be exported, sorted and ranked and a prediction rate curve drawn (Appendix 6.4).

The ranking procedure requires that the total pixels investigated in the validation be ordered from highest to lowest and a rank applied based on its predictive value (Chung and Fabbri, 2003). The pixel most susceptible to landslides is assigned 1 as it has the highest predictive value (Chung and Fabbri, 2003). The revised pixel rank value can be calculated based on its ranked placing and dividing by the total number of pixels under consideration (Chung and Fabbri, 2003). In this validation of 50 sampling boxes, 84,050 pixels are considered, so the lowest predictive value (lowest value of susceptibility) was assigned 0.000012. Both the weights of evidence and logistic regression methods yield different spatial results, so the pixels will not be ranked in the same order, so each is method's susceptibility values were ranked separately. The number of pixels in each set is identical, so the values of the ranks will be the same, just a pixel in one model may not have the same predictive value in the other model. Plotting by pixel rank allows different models to be compared on the same scale.

Prediction rate curves with the proportion ground-truthed landslides plotted against the proportion of susceptibility classes in 1% intervals were constructed for both models (Figure 6.7).

6.4.4.2 Model Prediction Rate Interpretations

Both the weights of evidence and logistic regression models indicate a good predictive capability (Figure 6.7). Based on both the highest 10% and 20% of the susceptibility classes, the weights of evidence model identified a greater proportion of the ground-truthed landslides (29% and 51% respectively) than the

logistic regression model (16% and 35% respectively) (Figure 6.7). However, when the highest 40% of the susceptibility classes is considered, the logistic regression model identifies approximately 81% of the landslides, while approximately 73% were identified using the weights of evidence model (Figure 6.7).

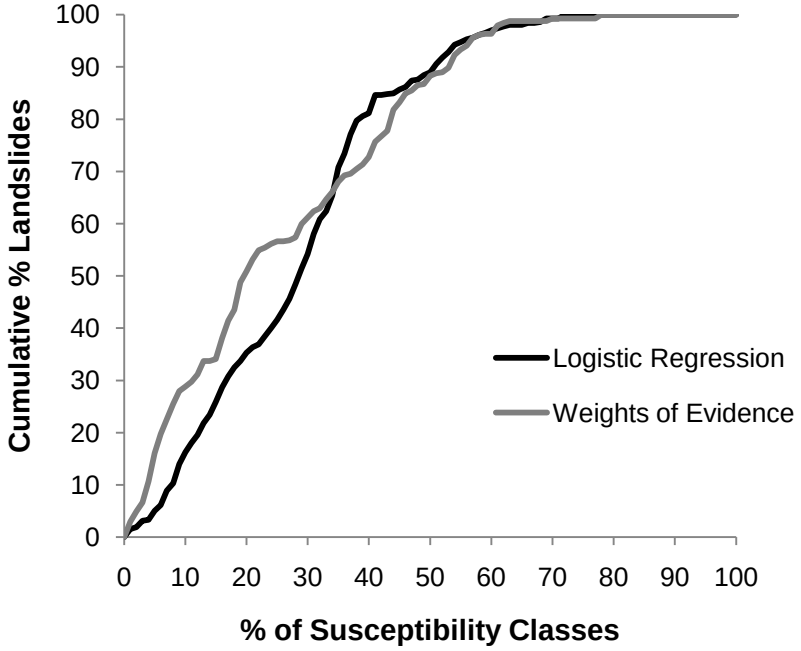


Figure 6.7 Prediction rate curve comparing both logistic regression and weights of evidence methods predictive capability in determining landslide susceptibility for the Waikato Region.

6.5 Discussion

Based on the prediction rate curves, both maps give a good level of predictive capability. The weights of evidence map resulted in a better prediction rate when the highest 10% and 20% of the susceptibility classes were considered. But when the 40% highest susceptibility classes were considered, the logistic regression map gave a better prediction rate. Although a good result was observed in the prediction rate curve for the weights of evidence map, it gave a poor result in the cumulative area curve analysis.

The cumulative area curves for the weights of evidence map, showed that 99.2% of the landslide area, and 71.3% of the total area was identified within landslide susceptibilities $\geq$ 50%. The first assumption (Section 6.3.2) which states that the majority of landslides should occur in the areas of higher landslide susceptibilities

was satisfied. However, it was also observed that landslide susceptibilities > 72 % accounted for approximately 50% of the total area. As the high susceptibility classes cover a relatively large proportion of the area on the weights of evidence map, the second assumption (Section 6.3.2) was violated. Unlike the weights of evidence map, the map derived by logistic regression gave a satisfactory result. Both assumptions were satisfied as the approximately 89% of the landslide area and 44% of the total area were in landslide susceptibilities ≥ 50%.

Based on the observations from both evaluation approaches, the better landslide susceptibility map is the one derived by logistic regression.

6.6 Conclusion

Evaluation is an important part of susceptibility assessments as some indication of the landslide susceptibility map's predictive capability should be provided with the map. The terminology of validation is somewhat misleading as a relative predictive capability is given making it more an evaluation than validation which would imply the predictive capability is absolute which is unlikely to be the case. In addition to misleading terminology, there appears to be some confusion in what an evaluation is. Two types of "evaluation" tend to be referred to in the literature one is a true evaluation using an independent landslide location data set to determine predictive capability, and the other is a test of model fit (goodness of fit) where the landslide data used to derive the model is then used to "verify" it.

Prediction rate curves and cumulative area curves were applied in this study to validate the two landslide susceptibility maps based on the weights of evidence and logistic regression approaches. The cumulative area curve was gave a good result for the logistic regression model, but not for the weights of evidence. Both landslide susceptibility maps resulted in good prediction rate curves, and weights of evidence looked to be a slightly better model based on the highest 30% of the susceptibility classes. However, it was observed that the weights of evidence map violated one of the assumptions in the cumulative area curve, so its prediction capability is questionable at best. Whereas, in comparison, the logistic regression map satisfied both assumptions, and both prediction rate curve and cumulative area curves gave a good result. Based on these findings, the susceptibility map

obtained by the logistic regression method is the better landslide susceptibility map for the Waikato Region.

CHAPTER SEVEN

Summary

7.1 Introduction

Landslide susceptibility assessments allow a prediction of where potential future landslides are likely to occur based on evidence of past landslide events and the likely causative factors identified at these locations. Susceptibility assessments differ from hazard assessments as neither time nor magnitude are accounted for. A landslide susceptibility assessment was conducted for the Waikato Region using past landslide events and a set of likely predictor parameters. The types of approaches and variables employed, and those most commonly used in other landslide susceptibility assessments were explored. Following this, two statistical approaches and eleven causative factors were chosen for this study. A bivariate approach (weights of evidence), and a multivariate approach (logistic regression) were investigated and applied to the Waikato Region and a landslide susceptibility map created for each. Evaluation of the two resulting susceptibility maps was conducted using an independent set of landslide data to identify the map with the best predictive capability.

7.2 Data Collection and Transformation

A geographical information system (GIS) was used to prepare a set of parameter maps and a landslide inventory map using available spatial data. A set of automated procedures (scripts) was created to conduct and replicate the parameter transformations in a time efficient and accurate manner.

A landslide inventory was created using two sources of landslide information, the GNS QMap landslide vector data and the GeoNet landslide catalogue. The Waikato landslide point locations were extracted from the GeoNet database, converted to polygon using the radius of each record, and combined with the QMap landslide data. Topographical parameters of slope, elevation and aspect

were derived from a 25 m resolution DEM. The rainfall data for mean monthly rainfall and maximum monthly rainfall were of a lower resolution (1 km) so the pixel size was converted to 25 m to align with the DEM derived data. For both the topographical and rainfall data, continuous datasets were obtained for logistic regression and a categorical dataset created for weights of evidence. The classes for each of the topographical data sets were chosen based on examples of those used in other studies and identifying which would be most appropriate for this study. Classes were chosen on a similar basis to determine the size and number of buffer zones (bands of distance around a linear feature) for the distance from roads, faults and rivers data used in this study. Existing polygon data for geology, land cover and soil order were simplified into broader classes, based on the information in the supporting documentation for each dataset. Both the GNS QMap geological unit spatial data and an older GNS New Zealand geology dataset for the Rotorua area were combined to obtain the geology spatial layer in this study. The older geology data was used in the Rotorua Topo 260 area as the QMap spatial data for this area was not yet released.

All polygon vector spatial layers were converted to raster data using a 25 m pixel size to match that of the DEM derived data and modified rainfall data. A grid of 25 m spaced points was created and overlain each of the twelve raster layers, and the pixel values for each were recorded and a database compiled for statistical analysis.

7.3 Weights of Evidence

The weights of evidence approach and conditional independence was applied. Three variations of the same method were identified, and although each was observed to give different positive and negative weightings to each class in each factor map, the weighted difference for each was identical in all approaches. Conditional independence was investigated using a contingency table of observed and expected probabilities in paired comparisons between each of the predictor factors and their classes. Chi squares determined in each comparison were compared with a theoretical χ^2 at the 99% significance level ($p = 0.01$) for 1 degree of freedom. When compared, mean monthly rainfall and maximum monthly rainfall displayed conditional dependence (high χ^2 values), so both

predictors could not be used in the same analysis. Maximum monthly rainfall was considered more appropriate so mean monthly rainfall was excluded.

The ten remaining predictors were used to derive an initial susceptibility map, and this looked to give a good fit to the landslide inventory data. Four different classifications which split the data differently were applied to determine which would give the best representation and be most appropriate. The equal intervals and natural breaks classification gave the best fits when the relative landslide densities for each class were compared. Equal intervals may have given a better result but it classed a smaller proportion of the landslide area in the very high class as well as a smaller proportion of the total area in the very low and low susceptibility classes. The equal intervals classification involves an element of expert opinion, in that the data is split according to the number of intervals set, and the size of the classes can be predetermined. The natural breaks classification was considered more appropriate as it is less influenced by expert opinion. Although the user can set the number of classes, the size of the classes cannot be predetermined, as the data is split by natural groupings inherent in the data (ESRI, 2008).

Several variations of the initial model were trialled with the exclusion of one or more variables. Three of the resulting combinations were compared to the initial model and all three proved to be better when the relative landslide densities were compared. The model derived from maximum monthly rainfall, slope, geology, land cover and distance from faults identified the greatest proportion of landslide area in the smallest proportion of total area for the higher susceptibility classes. This model was thus chosen as the final weights of evidence derived landslide susceptibility map.

7.4 Logistic Regression

Logistic regression is a generalised linear model that can be used to obtain a predictive model of landslide susceptibility based on a binary predictor (the presence or absence of past landslides). The model obtained involves a linear combination (z) and beta coefficients (β_n) which can be integrated into the logistic regression equation using GIS and a map of susceptibility created.

Logistic regression requires an even number of both landslide and non-landslide data. Landslides only make up a small proportion (~1.4%) of the Waikato Region so sampling was required in order to obtain a sample set with the same number of landslide and non-landslide records for logistic regression. Several random sampling strategies were trialled in STATISTICA using both systematic random sampling and stratified random sampling. The most effective method involved three stages of random sampling. This involved systematic random sampling by easting coordinates, followed by a second systematic random sample by northing coordinates, and finally stratified random sampling. Two sample sets were obtained in this manner, one to derive a significant model from (Set 1), and another to validate the derived model (Set 2).

For a binomial distribution, four link functions: logit, probit, complementary log-log and loglog, were available in STATISICA. Each link function was trialled using the same combination of predictors, and gave a different response. The logit link function proved to be the most appropriate, and was also in keeping with the bulk of the literature. Backwards stepwise logistic regression was used to determine the best combination of predictors by removing those which had a p value less than the p to enter and p to remove values (both set at 0.05).

The most significant model obtained for the Set 1 sample set was derived from mean monthly rainfall, geology, slope, and land cover. When compared to the theoretical χ^2 at p = 0.05, this combination resulted in a χ^2 with a p value of 0.0158. The beta values for each predictor variable in this model and the resulting linear combination were applied in GIS and a susceptibility map created. To evaluate the model accuracy, the model was applied to Set 2. The percentage of landslides and non-landslides correctly predicted by the model for both Set 1 and Set 2 were compared, and proved to be very similar. 71.36% of the landslides in the Set 1 sample, and 71.32% in the Set 2 sample were correctly identified by the model. For the same model, 63.15% of the non-landslide data in Set 1 sample, and 62.56% in the Set 2 sample were correctly identified. As the difference between both Set 1 and Set 2 landslides and non-landslides was minute (0.04% and 0.59% respectively) the model was found to have a good fit. The model derived from mean monthly rainfall, geology, slope, and land cover based on the Set 1 sample set was thus chosen as the final logistic regression derived susceptibility map.

7.5 Evaluation

Evaluation is an important stage in any landslide susceptibility assessment as it gives an indication of the map's predictive capability. To determine the predictive capability, an independent set of landslide data is used to test the landslide susceptibility map. Some confusion in appears in the literature in regards to evaluation, the landslide data used, and the type of evaluation conducted. Validation implies an absolute yes/no answer whereas evaluation indicates a relative better/worse answer. As a landslide susceptibility map is unlikely to have an absolute predictive capability, the term validation is not appropriate. Similarly, there appears to be some confusion in what an appropriate evaluation is, as some studies present a goodness of fit as the predictive capability which is not the same. An evaluation of predictive capability is determined using a different set of landslide data to that used in the model. Whereas the goodness of fit is determined using the same landslide data the model was derived from, and as such, is not an indication of how good the map represents reality, but how good the model fits the data.

Both statistically derived landslide susceptibility maps underwent an evaluation to ascertain which of the maps resulted in the better predictive capacity. An individual landslide dataset was obtained through ground-truthing. Google Earth was used to identify landslides in 1 km^2 sample boxes, determined by random sampling by susceptibility class, throughout the Waikato region. A table of the ground-truthed landslides identified, their coordinates, and widths was established. The tabulated data was used to create a new landslide spatial layer which would be used to evaluate both maps with the aid of GIS.

Two approaches to evaluation were employed: prediction rate curves and cumulative area curves. Both the weights of evidence and logistic regression methods displayed a good result in the prediction rate curves, and it was observed that the two models resulted in very similar prediction rate curves. When the 10, 20 and 30% most susceptible pixels were considered, weights of evidence displayed a better prediction rate; but when the 40% most susceptible pixels were considered logistic regression displayed a better prediction rate. It was decided

that one model could not justifiably be chosen over the other based on the prediction rate curves alone.

The cumulative area curves for both maps gave very different results. Both weights of evidence and logistic regression models correctly identified the majority of the landslide area in areas of high (≥ 50%) susceptibility (99.21% and 88.97% respectively). The total area in the high susceptibility classes was lowest for logistic regression (44.17%) than weights of evidence (71.31%). The best model should identify a high proportion of the landslide area in a small proportion of the total area in susceptibilities ≥ 50%. Weights of evidence did not meet this criterion, as there was a very large proportion of the total area in the high susceptibility classes.

Based on the observations from both approaches to evaluation, logistic regression was considered the better model, and was thus presented as the final landslide susceptibility map for the Waikato Region.

7.6 Recommendations for Further Research

- Further work could include a landslide hazard assessment in order to estimate the return period and magnitude of a future event, however this could be a challenging task as temporal data can be a limitation. If however, it was achieved, a risk assessment which estimates the cost in relation to the hazard and vulnerability could also be applied.
- The landslide and geology data was limited in the area surrounding Taupo, as the QMap data for this area has not yet been released. If it had been available, a better representative result could have been obtained. Once available it would be possible to refine the analysis.
- The rainfall data was of a lower resolution (1 km) than the DEM used (25 m), and not as current as desired, although it did give an indication of areas which receive more and less rainfall. If however, a landslide hazard assessment were conducted the rainfall data used in this study may not be appropriate.
- The scripts created in this study can readily be updated, adapted and modified to conduct similar or the same GIS processes in other studies.

These could be utilised in another landslide susceptibility assessment, or in other GIS-based studies, and would save the user time.

- Causative factors such as ground water, peak ground acceleration, soil drainage and slope curvature were not considered in this assessment as a result of data availability or time constraints but may be of some importance to consider in future assessments.
- The use of artificial neural networks in future landslide susceptibility assessments could be applied. These were briefly explored in STATISTICA in this assessment but were found to be quite time consuming to run, and somewhat complex to understand and apply, so were not included in the analysis.

References

Abdallah, C., Chorowicz, J., Kheir, R. B., & Khawlie, A. (2005). Detecting major terrain parameters relating to mass movements' occurrence using GIS, remote sensing and statistical correlations, case study Lebanon. *Remote Sensing of Environment, 99*(4), 448-461.

Akgün, A., & Bulut, F. (2007). GIS-based landslide susceptibility for Arsin-Yomra (Trabzon, North Turkey) region. *Environmental Geology, 51*(8), 1377-1387.

Alexander, D. E. (2008). A brief survey of GIS in mass-movement studies, with reflections on theory and methods. *Geomorphology, 94*(3-4), 261-267.

Arora, M. K., Das Gupta, A. S., & Gupta, R. P. (2004). An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. *International Journal of Remote Sensing, 25*(3), 559-572.

Ayalew, L., & Yamagishi, H. (2005). The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan. *Geomorphology, 65*(1-2), 15-31.

Ayalew, L., Yamagishi, H., Marui, H., & Kanno, T. (2005). Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. *Engineering Geology, 81*(4), 432-445.

Berkowitz, P. (2004). Def_proj.aml. from \\liby-travis.liby.waikato.ac.nz\gis_data\GIS_Tools

Can, T., Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H., & Duman, T. Y. (2005). Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. *Geomorphology, 72*(1-4), 250-271.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., & Reichenbach, P. (1991). GIS techniques and statistical-models in evaluating landslide hazard. *Earth Surface Processes and Landforms, 16*(5), 427-445.

Carrara, A., Guzzetti, F., Cardinali, M., & Reichenbach, P. (1999). Use of GIS technology in the prediction and monitoring of landslide hazard. *Natural Hazards, 20*(2-3), 117-135.

Carrara, A., & Pike, R. J. (2008), GIS technology and models for assessing landslide hazard and risk. *Geomorphology, 94*(3-4), 257-260.

Catani, F., Casagli, N., Ermini, L., Righini, G., & Menduni, G. (2005). Landslide hazard and risk mapping at a catchment scale 2005 Arno River basin. *Landslides, 2*(4), 329-342.

Cevik, E., & Topal, T. (2003). GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). *Environmental Geology, 44*(8), 949-962.

Chen, M. H., & Shao, Q. M. (2001). Propriety of posterior distribution for dichotomous quantal response models. *Proceedings of the American Mathematical Society, 129*(1), 293-302.

Chung, C. J. F., & Fabbri, A. G. (1999). Probabilistic prediction models for landslide hazard mapping. *Photogrammetric Engineering and Remote Sensing, 65*(12), 1389-1399.

Chung, C. J. F., & Fabbri, A. G. (2003). Validation of spatial prediction models for landslide hazard mapping. *Natural Hazards, 30*(3), 451-472.

Clerici, A., Perego, S., Tellini, C., & Vescovi, P. (2002). A procedure for landslide zonation by the conditional analysis method. *Geomorphology, 48*(4), 349-364.

Clerici, A., Perego, S., Tellini, C., & Vescovi, P. (2006). A GIS-based automated procedure for landslide susceptibility mapping by the Conditional Analysis method: the Baganza valley case study (Italian Northern Apennines). *Environmental Geology, 50*(7), 941-961.

Conoscenti, C., Di Maggio, C., & Rotighano, E. (2008). GIS analysis to assess landslide susceptibility in a fluvial basin of NW Sicily (Italy). *Geomorphology, 94*(3-4), 325-339.

Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Masuda, T., & Nishino, K. (2008a). GIS-based weights-of-evidence modelling of rainfall-induced landslides in small catchments for landslide susceptibility mapping. *Environmental Geology, 54*(2), 311-324.

Dahal, R. K., Hasegawa, S., Nonomura, A., Yamanaka, M., Dhakal, S., & Paudyal, P. (2008b). Predictive modelling of rainfall-induced landslide hazard in the Lesser Himalaya of Nepal based on weights-of-evidence. *Geomorphology, 102*(3-4), 496-510.

Dai, F. C., & Lee, C. F. (2001). Terrain-based mapping of landslide susceptibility using a geographical information system: a case study. *Canadian Geotechnical Journal, 38*(5), 911-923.

Dai, F. C., & Lee, C. F. (2002). Landslide characteristics and, slope instability modeling using GIS, Lantau Island, Hong Kong. *Geomorphology, 42*(3-4), 213-228.

Dai, F. C., & Lee, C. F. (2003). A spatiotemporal probabilistic modelling of storm-induced shallow landsliding using aerial photographs and logistic regression. *Earth Surface Processes and Landforms, 28*(5), 527-545.

Dai, F. C., Lee, C. F., Li, J., & Xu, Z. W. (2001). Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. *Environmental Geology, 40*(3), 381-391.

Dai, F. C., Lee, C. F., & Ngai, Y. Y. (2002). Landslide risk assessment and management: an overview. *Engineering Geology, 64*(1), 65-87.

Dellow, G. (2009). Re: Landslide Inventory for the Waikato query. In R. Schicker (Ed).

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

Dhakal, A. S., Amada, T., & Aniya, M. (2000). Landslide hazard mapping and its evaluation using GIS: An investigation of sampling schemes for a grid-cell based quantitative method. *Photogrammetric Engineering and Remote Sensing, 66*(8), 981-989.

Dikau, R., Cavallin, A., & Jager, S. (1996). Databases and GIS for landslide research in Europe. *Geomorphology, 15*(3-4), 227-239.

Dominguez-Cuesta, M. J., Jimenez-Sanchez, M., & Berrezueta, E. (2007). Landslides in the Central Coalfield (Cantabrian Mountains, NW Spain): Geomorphological features, conditioning factors and methodological implications in susceptibility assessment. *Geomorphology, 89*(3-4), 358-369.

Donati, L., & Turrini, M. C. (2002). An objective method to rank, the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines, (Valnerina; Perugia, Italy). *Engineering Geology, 63*(3-4), 277-289.

Duman, T. Y., Can, T., Gökceoglu, C., Nefeslioglu, H. A. & Sonmez, H. (2006). Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Instanbul, Turkey. *Environmental Geology, 51*(2), 241-256.

Duncan, M., & Woods, R. (2004). Flow Regimes. In J. Harding, P. Mosley, C. Pearson & B. Sorrell (eds.), *Freshwaters of New Zealand* (pp. 7.1-7.14). Christchurch, New Zealand: New Zealand Hydrological Society Inc. & New Zealand Limnological Society Inc.

Dymond, J. R., Betts, H. D., & Schierlitz, C. S. (2010). An erosion model for evaluating regional land-use scenarios. *Environmental Modelling & Software, 25*(3), 289-298.

Earthquake Commission. (2004). *Annual Report 2003-2004.* Retrieved. from http://www.eqc.govt.nz/downloads/pdfs/eqc-annual-report-03-04.pdf.

Earthquake Commission. (2005). *Annual Report 2004-2005*. Retrieved. from http://www.eqc.govt.nz/downloads/ar-0405/annual-report-2005.pdf.

Earthquake Commission. (2006). *Annual Report 2005-2006.* Retrieved. from http://www.eqc.govt.nz/downloads/pdfs/eqc-annual-report-05-06.pdf.

Earthquake Commission. (2007). *Annual Report 2006-2007*. Retrieved. from http://www.eqc.govt.nz/downloads/pdfs/eqc-annual-report-06-07.pdf.

Earthquake Commission. (2008). *Annual Report 2007-2008.* Retrieved. from http://www.eqc.govt.nz/downloads/pdfs/annual-report-07-08.pdf.

Earthquake Commission. (2009). *Annual Report 2008-2009.*Retrieved from http://www.eqc.govt.nz/downloads/ar-0809/annual-report-2008-09.pdf.

Edbrooke, S. W. (compiler) 2001a: *Geology of the Auckland area.* Institute of Geological & Nuclear Sciences 1:250 000 geological map 3. 1 sheet + 74 p. Lower Hutt, New Zealand. Institute of Geological & Nuclear Sciences Limited.

Edbrooke, S. W. (compiler) 2001b: Geology of the Auckland area. Institute of Geological & Nuclear Sciences 1:250 000 geological map 3. Version 1.

Edbrooke, S. W. (compiler) 2005a: *Geology of the Waikato area.* Institute of Geological & Nuclear Sciences 1:250 000 geological map 4. 1 sheet + 68 p. Lower Hutt, New Zealand. Institute of Geological & Nuclear Sciences Limited.

Edbrooke, S. W. (compiler) 2005b: Geology of the Waikato area. Institute of Geological & Nuclear Sciences 1:250 000 geological map 4. Version 1.

Environment Waikato. (1998). *Waikato state of the environment report.* Hamilton East, [N. Z.]: Environment Waikato.

Ercanoglu, M., & Gökceoglu, C. (2004). Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). *Engineering Geology, 75*(3-4), 229-250.

Ercanoglu, M., Kasmer, O., & Temiz, N. (2008). Adaptation and comparison of expert opinion to analytical hierarchy process for landslide susceptibility mapping. *Bulletin of Engineering Geology and the Environment, 67*(4), 565-578.

Ermini, L., Catani, F., & Casagli, N. (2005). Artificial Neural Networks applied to landslide susceptibility assessment. *Geomorphology, 66*(1-4), 327-343.

ESRI (2008) ArcGIS 9 ArcMap™ Version 9.3.

Falaschi, F., Giacomelli, F., Federici, P. R., Puccinelli, A., D'Amato Avanzi, G., Pochini, A., and Ribolini, A. (2009). Logistic regression versus artificial neural networks: landslide susceptibility evaluation in a sample area of the Serchio River valley, Italy. *Natural Hazards, 50*(3), 551-569.

Fernandez, C. I., Del Castillo, T. F., El Hamdouni, R., & Montero, J. C. (1999). Verification of landslide susceptibility mapping: A case study. *Earth Surface Processes and Landforms, 24*(6), 537-544.

Garcia-Rodriguez, M. J., Malpica, J. A., Benito, B., & Diaz, M. (2008). Susceptibility assessment of earthquake-triggered landslides in El Salvador using logistic regression. *Geomorphology, 95*(3-4), 172-191.

GeoNet. (2009). Landslide Catalogue. From Dellow, G. (2009). Re: Landslide Inventory for the Waikato Region query. In R. Schicker (Ed.)

Ghosh, S., van Westen, C. J., Carranza, E. J. M., Ghoshal, T. B., Sarkar, N. K., & Surendranath, M. (2009). A quantitative approach for improving the BIS (Indian) method of medium-scale landslide susceptibility. *Journal of the Geological Society of India, 74*(5), 625-638.

Gökceoglu, C. (2001). Discussion on "Landslide hazard zonation of the Khorshrostam area, Iran" by A. Uromeihy and M. R. Mahdavifar, Bull Eng Geol Environ 58:207-213. Bulletin of Engineering Geology and the Environment, 60(1), 79-80.

Gökceoglu, C., & Aksoy, H. (1996). Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. *Engineering Geology, 44*(1-4), 147-161.

Gorsevski, P. V., Gessler, P. E., Boll, J., Elliot, W. J., & Foltz, R. B. (2006). Spatially and temporally distributed modeling of landslide susceptibility. *Geomorphology, 80*(3-4), 178-198.

Greco, R., Sorriso-Valvo, M., & Catalano, E. (2007). Logistic regression analysis in the evaluation of mass movements susceptibility: The Aspromonte case study, Calabria, Italy. *Engineering Geology 89*(1-2), 47-66.

Gupta, R. P., & Joshi, B. C. (1990). Landslide hazard zoning using the GIS approach - a case study from the Ramganga catchment, Himalayas. *Engineering Geology, 28*(1-2), 119-131.

Guzzetti, F., Cardinali, M., Reichenbach, P., & Carrara, A. (2000). Comparing landslide maps: A case study in the Upper Tiber River basin, central Italy. *Environmental Management, 25*(3), 247-263.

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology, 31*(1-4), 181-216.

He, Y. P., & Beighley, R. E. (2008). GIS-based regional landslide susceptibility mapping: a case study in southern California. *Earth Surface Processes and Landforms, 33*(3), 380-393.

He, Y. P., Xie, H., Cui, P., Wei, F. Q., Zhong, D. L., & Gardner, J. S. (2003) GIS-based hazard mapping and zonation of debris flows in Xiaojiang Basin, southwestern China. *Environmental Geology, 45*(2), 286-293.

Hewitt, A. E. (1992). *New Zealand Soil Classification.* Lower Hutt, N. Z.: DSIR Land Resources.

Institute of Geological & Nuclear Sciences, (2000). Geological Spatial Layers. from \\liby-travis.liby.waikato.ac.nz\gis_data\Geology.

Jibson, R. W., Harp, E. L., & Michael, J. A. (2000). A method for producing digital probabilistic seismic landslide hazard maps. *Engineering Geology, 58*(3-4), 271-289.

Jiménez-Perálvarez, J. D., Irigaray, C., El Hamdouni, R., & Chacón, J. (2009). Building models for automatic landslide-susceptibility analysis, mapping and validation in ArcGIS. *Natural Hazards, 50*(3), 571-590.

Kanungo, D. P., Arora, M. K., Sarkar, S., & Gupta, R. P. (2006). A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology, 85*(3-4), 347-366.

Lan, H. X., Zhou, C. H., Wang, L. J., Zhang, H. Y., & Li, R. H. (2004). Landslide hazard spatial analysis and prediction using GIS in the Xiaojiang watershed, Yunnan, China. *Engineering Geology, 76*(1-2), 109-128.

Land Information New Zealand (LINZ). (2000a) North Island 25 m DEM. from \\liby.trvais.liby.waikato.ac.nz\gis_data\Elevation\NZMG2000\North\north25

Land Information New Zealand (LINZ). (2000b). NZTopo. from \\liby-travis.liby.waikato.ac.nz\gis_data\Roads.

Land Information New Zealand (LINZ), & Eagle Technologies (Eagle). (2000). NZTopo. from \\liby-travis.liby.waikato.ac.nz\gis_data\Topo_50k\nz.

Landcare Research New Zealand Ltd. (2000a). Land Resource Information System (extended fundamental soil attributes) Spatial Data Layers. from \\liby-travis.liby.waikato.ac.nz\ gis_data\Soils\NZ

Landcare Research New Zealand Ltd. (2000b). _readme.txt. from \\liby-travis.liby.waikato.ac.nz\ gis_data\LRI

Landcare Research New Zealand Ltd. & Department of Conservation. (1998). Climate Surfaces for New Zealand. from \\liby-travis.liby.waikato.ac.nz\ gis_data\Climate\Landcare_Climate_Rasters_1998

Leathwick, J. R., Wilson, G., Rutledge, D., Wardle, P., Morgan, F., Johnston, K., McLeod, M., & Kirkpatrick, R. (2003). *Land Environments of New Zealand = Ngā taiao o Aotearoa.* Auckland, N.Z.: David Bateman, Landcare Research, & Ministry for the Environment.

Leathwick, J. R., Wilson, G., Stephens, R. T. T., Landcare Research New Zealand Ltd, & Department of Conservation. (1998). Climate Surfaces for New Zealand. from \\liby-travis.liby.waikato.ac.nz\gis_data\Climate\Landcare_Climate_Rasters_1998

Lee, S. (2004). Application of likelihood ratio and logistic regression models to landslide susceptibility mapping using GIS. *Environmental Management, 34*(2), 223-232.

Lee, S. (2005). Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data journals. *International Journal of Remote Sensing, 26*(7), 1477-1491.

Lee, S. (2007a). Application and verification of fuzzy algebraic operators to landslide susceptibility mapping. *Environmental Geology, 52*(4), 615-623.

Lee, S. (2007b). Comparison of landslide susceptibility maps generated through multiple logistic regression for three test areas in Korea. *Earth Surface Processes and Landforms, 32*(14), 2133-2148.

Lee, S., & Choi, J. (2004). Landslide susceptibility mapping using GIS and the weight-of-evidence model. *International Journal of Geographical Information Science, 18*(8), 789-814.

Lee, S., Choi, J., & Min, K. (2002). Landslide susceptibility analysis and verification using the Bayesian probability model. *Environmental Geology, 43*(1-2), 120-131.

Lee, S., Choi, J., & Min, K. (2004a). Probabilistic landslide hazard mapping using GIS and remote sensing data at Boun, Korea. *International Journal of Remote Sensing, 25*(11), 2037-2052.

Lee, S., & Dan, N. T. (2005). Probabilistic landslide susceptibility mapping on the Lai Chau province of Vietnam: focus on the relationship between tectonic fractures and landslides. *Environmental Geology, 48*(6), 778-787.

Lee, S., & Evangelista, D. G. (2006). Earthquake-induced landslide susceptibility mapping using artificial neural network. *Natural Hazards and Earth System Sciences, 6*(5), 687-695.

Lee, S., & Lee, M. J. (2006). Detecting landslide location using KOMPSAT1 and its application to landslide-susceptibility mapping at the Gangneung area, Korea. *Advances in Space Research, 38*(10), 2261-2271.

Lee, S., & Min, K. (2001). Statistical analysis of landslide susceptibility at Yongin, Korea. *Environmental Geology, 40*(9), 1095-1113.

Lee, S., & Pradhan, B. (2007). Landslide hazard mapping at Selangor, Malaysia using frequency ratio and logistic regression models. *Landslides, 4*(1), 33-41.

Lee, S., Ryu, J. H., Lee, M. J., & Won, J. S. (2003a). Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. *Environmental Geology, 44*(7), 820-833.

Lee, S., Ryu, J. H., Min, K. D., & Won, J. S. (2003b). Landslide susceptibility analysis using GIS and artificial neural network. *Earth Surface Processes and Landforms, 28*(12), 1361-1376.

Lee, S., Ryu, J. H., Won, J. S., & Park, H. J. (2004b). Determination and application of the weights for landslide susceptibility mapping using an artificial neural network. Engineering *Geology, 71*(3-4), 289-302.

Lee, S., Ryu, J. H., Lee, M. J., & Won, J. S. (2006). The application of artificial neural networks to landslide susceptibility mapping at Janhung, Korea. *Mathematical Geology, 38*(2), 199-220.

Lee, S., & Sambath, T. (2006). Landslide susceptibility mapping in the Damrei Romel area, Cambodia using frequency ratio and logistic regression models. *Environmental Geology, 50*(6), 847-855.

Lee, S., & Talib, J. A. (2005). Probabilistic landslide susceptibility and factor effect analysis. *Environmental Geology, 47*(7), 982-990.

Menard, S. (1995). Applied logistic regression analysis. Sage *University Paper Series on Qunatitative Applications in Social Sciences, 07-106*. Thousand Oaks, CA: Sage.

Ministry for the Environment. (1997). The state of our waters. In Ministry for the Environment (Ed.), *The state of New Zealand's Environment 1997* (pp.7.1-7.100). Wellington, New Zealand, Ministry for the Environment and GP Publications.

Ministry for the Environment. (2004). New Zealand Land Cover Database (LCDB2). from \\liby-travis.liby.waikato.ac.nz\gis_data\Landcover\Land cover2

Ministry for the Environment. (2004). New Zealand Land Cover Database 2 User Guide. from \\liby-travis.liby.waikato.ac.nz\gis_data\Landcover\ Landcover2\metadata\LCDB2_User_Guide.pdf

Moore, D. S., & McCabe, G. P. (2003). *Introduction to the Practice of Statistics* (4th ed.). New York: W. H. Freeman and Company.

Nagarajan, R., Mukherjee, A., Roy, A., & Khire, M. V. (1998). Temporal remote sensing data and GIS application in landslide hazard zonation of part of Western ghat, India. *International Journal of Remote Sensing, 19*(4), 573-585.

Nandi, A., & Shakoor, A. (2009). A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. *Engineering Geology, 110*(1-2), 11-20.

Nefeslioglu, H. A., Gökceoglu, C., & Sonmez, H. (2008). An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology, 97*(3-4), 171-191.

Neuhäuser, B., & Terhorst, B. (2007). Landslide susceptibility assessment using "weights-of-evidence" applied to a study area at the Jurassic escarpment (SW-Germany). *Geomorphology, 86*(1-2), 12-24.

Newsome, P. F. J., Wilde, R.H., & Willoughby, E. J. (2000). Land Resource Information System Spatial Data Layers. from \\liby-travis.liby.waikato. ac.nz\ gis_data\LRI\lri new.doc

Ohlmacher, G. C. (2007). Plan curvature and landslide probability in regions dominated by earth flows and earth slides. *Engineering Geology, 91*(2-4), 117-134.

Ohlmacher, G. C., & Davis, J. C. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology, 69*(3-4), 331-343.

Pachauri, A. K., & Pant, M. (1992). Landslide hazard mapping based on geological attributes. *Engineering Geology, 32*(1-2), 81-100.

Park, N. W., & Chi, K. H. (2008). Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model. *International Journal of Remote Sensing, 29*(1), 247-264.

Perotto-Baldiviezo, H. L., Thurow, T. L., Smith, C. T., Fisher, R. F., & Wu, X. B. (2004). GIS-based spatial analysis and modeling for landslide hazard assessment in steeplands, southern Honduras. *Agriculture Ecosystems & Environment, 103*(1), 165-176.

Raines, G. L. (1999). Evaluation of Weights of Evidence to Predict Epithermal-Gold Deposits in the Great Basin of the Western United States. *Natural Resources Research, 8*(4), 257-276.

Remondo, J., Bonachea, J., Cendrero, A., (2008). Quantitative landslide risk assessment and mapping on the basis of recent occurrences. *Geomorphology, 94*(3-4), 496-507.

Remondo, J., Gonzalez, A., De Teran, J. R. D., Cendrero, A., Fabbri, A., & Chung, C. J. F. (2003). Validation of landslide susceptibility maps; Examples and applications from a case study in northern Spain. *Natural Hazards, 30*(3), 437-449.

Saha, A. K., Gupta, R. P., & Arora, M. K. (2002). GIS-based Landslide Hazard Zonation in the Bhagirathi (Ganga) Valley, Himalayas. *International Journal of Remote Sensing, 23*(2), 357-369.

Sarkar, S., & Kanungo, D. P. (2004). An integrated approach for landslide susceptibility mapping using remote sensing and GIS. *Photogrammatic Engineering and Remote Sensing, 70*(5), 617-625.

Santacana, N., Baeza, B., Corominas, J., De Paz, A., & Marturia, J. (2003). A GIS-based multivariate statistical analysis for shallow landslide susceptibility mapping in La Pobla de Lillet area (Eastern Pyrenees, Spain). *Natural Hazards, 30*(3), 281-295.

Selby, M. J. (1993). *Hillslope materials and processes* (2nd ed.). Oxford, England: Oxford University Press.

Sharma, M., & Kumar, R. (2008). GIS-based landslide hazard zonation: a case study from the Parwanoo area, Lesser and Outer Himalaya, H.P., India. *Bulletin of Engineering Geology and the Environment 67*(1), 129-137.

Smith, N. J. (1999). *Landslide susceptibility mapping and risk assessment for the Waikato Region, New Zealand* University of Waikato, Hamilton, New Zealand.

Soeters, R., & Van Westen, C. J. (1996). Slope instability recognition, analysis, and zonation. In A. K. Turner & R. L. Schuster (Eds.). *Landslide Investigation and Mitigation* (pp. 129-177). Washington D.C.: Transport Research Board, National Research Council.

Song, R. H., Hiromu, D., Kazutoki, A., Usio, K., & Sumio, M. (2008). Modeling the potential distribution of shallow-seated landslides using the weights-of-evidence method and a logistic regression model: a case study of the Sabae Area, Japan. *International Journal of Sediment Research, 23*(2), 106-118.

Statistics New Zealand. (2002). New Zealand Digital Boundaries spatial dataset metadata. From http:\\www.stats.govt.nz

Statsoft. (2009). STATISTICA Electronic Manual.

Süzen, M. L., & Doyuran, V. (2004). A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. *Environmental Geology, 45*(5), 665-679.

Tangestani, M. H. (2004). Landslide susceptibility mapping using the fuzzy gamma approach in a GIS, Kakan catchment area, southwest Iran. *Australian Journal of Earth Sciences, 51*(3), 439-450.

Thiery, Y., Malet, J. P., Sterlacchini, S., Puissant, A., & Maquaire, O. (2007). Landslide susceptibility assessment by bivariate methods at large scales: Application to a complex mountainous environment. *Geomorphology, 92*(1-2), 38-59.

United Nations. (2009). ISDR (2009) Global Assessment Report on Disaster Risk Reduction. Geneva, Switzerland: Oriental Press, Manama, Kingdom of Bahrain.

Uromeihy, A., & Mahdavifar, M. R. (2000). Landslide hazard zonation of the Khorshrostam area, Iran. *Bulletin of Engineering Geology and the Environment, 58*(3), 207-213.

Van Beek, L. P. H., & Van Asch, T. W. J. (2004). Regional assessment of the effects of land-use change on landslide hazard by means of physically based modelling. *Natural Hazards, 31*(1), 289-304.

Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., & Vandekerckhove, L. (2006). Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). *Geomorphology, 76*(3-4), 392-410.

van Westen, C. J., Rengers, N., & Soeters, R. (2003). Use of geomorphological information in indirect landslide susceptibility assessment. *Natural Hazards, 30*(3), 399-419.

Varnes, D. J., & International Association of Engineering Geology Commission on Landslides and Other Mass Movements on Slopes. (1984). Landslide hazard zonation: a review of principles and practice. *Natural Hazards*, 3.

Walker, J. (2009, 28 August 2009). Chi-Square Calculator. Retrieved 30 November, 2009, from http:\\www.fourmilab.ch/rpkp/experiments/analysis/chiCalc.html

Wang, H. B., & Sassa, K. (2006). Rainfall-induced landslide hazard assessment using artificial neural networks. *Earth Surface Processes and Landforms, 31*(2), 235-247.

Weirich, F., & Blesius, L. (2007). Comparison of satellite and air photo based landslide susceptibility maps. *Geomorphology, 87*(4), 352-364.

Wieczorek, G. F. (1996). Landslide Triggering Mechanisms. In A. K. Turner & R. L. Schuster (Eds.). *Landslide Investigation and Mitigation* (pp. 76-90). Washington D.C.: Transport Research Board, National Research Council.

Wilson, R. C. (2004). The Rise and Fall of a Debris-flow Warning System for the San Francisco Bay Region, California. In T. Glade, M. Anderson & M. J. Crozier (Eds.), *Landslide Hazard and Risk* (pp. 496). West Sussex, England: John Wiley & Sons, Ltd.

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

Yalcin, A. (2008). GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): Comparisons of results and confirmations. *Catena, 72*(1), 1-12.

Yesilnacar, E., & Topal, T. (2005). Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). *Engineering Geology, 79*(3-4), 251-266.

Yilmaz, I., (2009a). Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat- Turkey). *Computers & Geosciences 35*(6), 1125-1138.

Yilmaz, I. (2009b). A case study from Koyulhisar (Sivas-Turkey) for landslide susceptibility mapping by artificial neural networks. *Bulletin of Engineering Geology and the Environment 68*(3), 297-306.

Zahiri, H., Palamara, D. R., Flentje, P., Brassington, G. M., & Baafi, E. (2006). A GIS-based Weights-of-Evidence model for mapping cliff instabilities associated with mine subsidence. *Environmental Geology*, *51*(3), 377-386.