LUMA-GIS Thesis nr 6

Shallow Landslide Susceptibility Modelling and Validation

Edgar Pimiento

2010
Dept. of Earth and Ecosystem Sciences
Division of Physical Geography and Ecosystem Analysis
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Shallow Landslide Susceptibility Modelling and Validation

Edgar Pimiento

Department of Physical Geography and Ecosystem Analysis
Centre for Geographical Information Systems

Lund University

A thesis submitted in partial fulfilment for the degree of

Master in Geographical Information Science

Supervisors:

Ulrik Mårtensson, Lund University

Prof Dr.Sc. Hiromitsu Yamagishi, Ehime University (Japan)

Acknowledgements

I thank Prof. Hiromitsu Yamagishi, my supervisor in Japan, for his support during this research.

Abstract

Rainfall frequently triggers shallow landslides in mountainous areas worldwide. Landslide susceptibility maps express the probability of occurrence of landslides based on terrain conditions; they are useful for disaster prevention and land use planning. This report is about validating a qualitative approach to map global landslide susceptibility, based on the weighted linear combination (WLC) of slope gradient, soil type, soil texture, elevation, land cover and drainage density. The parameters are derived from digital global databases. The accuracy assessment was based on a detailed landslide inventory of a 160-km^2 area in Japan, using the receiver-operating characteristic (ROC) plot area under the curve (AUC). The AUC permitted to compare analysis approaches and different parameter combinations. The AUC for the WLC model was 0.47, below a random classification. Two approaches improved the model accuracy, using the weights of evidence (WOE) approach raised the accuracy to 0.64, and using a higher resolution DEM raised the accuracy to 0.66. On the other hand, a quantitative approach based on logistic regression (LR) and using the software package Spatial Data Modeller (SDM) produced models with AUC between 0.67 and 0.71. The highest accuracy for a model including lithology, slope gradient, profile curvature, plan curvature and elevation. The reason for the higher accuracy of the LR models is that the occurrence of landslides depends on local conditions, expressed by the quantitative relations, while the qualitative weights of the WLC model were developed for a global model using different criteria.

Keywords: Landslide susceptibility, weights of evidence, logistic regression, ROC plot AUC, validation

Contents

List of Figures

List of Tables

Part I

Background and Theory

Chapter 1

Introduction

1.1 Background

Landslide is defined as "movement of a mass of rock, debris or earth down a slope" (Cruden, 1991). Landslides are natural phenomena related to mass wasting processes that model the earth surface. Conversely to other movements, like soil creep, some landslides occur suddenly and move fast, and sometimes cause great damage. The identification of areas where landslides are likely to occur is important for the reduction of potential damage. This report is about mapping shallow landslide prone areas, using qualitative and quantitative models, and validating the results.

Occurrence of shallow landslides depends on local terrain conditions, such as slope-forming materials, topography, groundwater, and land cover; in addition to triggering events, like intense rainfall or earthquakes, that modify those characteristics and produce changes that cause slope instability (Soeters and van Westen, 1996).

Assessment of landslide hazard is difficult due to the lack of historical data of triggering events; instead, landslide susceptibility assessments are common (van Westen et al., 2003). Landslide susceptibility maps express the likelihood of occurrence of landslides (spatial probability), estimated from local terrain conditions. The temporal probability of occurrence of landslides, which depends on the recurrence and intensity of triggering factors, is not considered (Soeters and van Westen, 1996).

Medium-scale (1:25,000-1:50,000) landslide susceptibility maps are useful in preliminary impact assessment and delimitation of areas for detailed surveys, and for regional physical planning, and development (transportation corridors, urbanization, large engineering projects). The main users are planners, decision makers, engineers, development enterprises, insurance companies, community, and scientists.

Qualitative (also known as knowledge-driven) methods and quantitative (or data-driven), methods permit the assessment of landslide susceptibility. Qualitative methods are based on opinions of experts, are subjective and it is difficult to apply them in separate areas. Quantitative methods are based on the relationships between the occurrence of landslides and terrain conditions, derived from the analysis of landslide inventories, and they are straightforward. However, landslide inventories are usually not available or difficult to obtain (Soeters and van Westen, 1996).

Assessment of the accuracy of landslide susceptibility models, the capacity to differentiate landslide-free areas from landslide-prone areas, is fundamental for decision-makers during hazard management. Accuracy depends mainly on accuracy of the models, accuracy of the input data, experience of the earth scientist (modeller) and size of the study area (Soeters and van Westen, 1996).

The purpose of this research is to validate a qualitative method based on the weighted linear combination of terrain parameters (slope, soil type, soil texture, elevation, land cover, and drainage density) derived from digital global databases (Hong et al., 2007); and compare it to a quantitative method based on logistic regression of parameters related to the occurrence of shallow landslides.

The principal advantages of the qualitative approach proposed by Hong et al. (2007) are that it can be applied in areas where there are no landslide observations, and that the model data are available as digital global databases. These allow the construction of landslide susceptibility maps for any area. However, according to Hong et al. (2007), it is necessary to validate the model with local landslide inventories.

Furthermore, the occurrence of landslides depends on local conditions (van Westen et al., 2003), and a model to map global landslide susceptibility needs to be adjusted to local conditions.

The accuracy assessment is based on the comparison of the model with a detailed distribution of shallow landslides from a 160-km^2-target area in Japan. It is assumed that the landslide distribution map is the ground truth.

The results will provide information on the accuracy of the qualitative modelling approach and how it can be improved with techniques used in quantitative landslide susceptibility mapping, and using higher resolution data.

1.2 Objectives

Validate a landslide susceptibility mapping approach that is based on expert knowledge and uses satellite remote sensing data. Compare the qualitative landslide susceptibility mapping method with a quantitative method and evaluate the results. The research should answer the following questions:

- How accurate is the global landslide susceptibility mapping (qualitative) method proposed by Hong et al. (2007)?
- How can the qualitative model accuracy be improved?

1.3 Methods

Produce a landslide susceptibility map using the qualitative approach and validate it with a detailed landslide distribution map.

Modify the qualitative method by selecting parameters and their weights based on the weights of evidence approach, a statistical analysis of landslide distribution and related parameters. Validate the resulting landslide susceptibility map.

Make a landslide susceptibility map based on logistic regression, a quantitative approach, using landslide observations; and validate the resulting model.

Finally, compare the qualitative and quantitative models.

1.4 Thesis structure

The thesis consists of two parts. Part I describes background and theory; the chapters are Introduction, Landslides, Landslide susceptibility mapping, GIS landslide susceptibility modelling, Validation of susceptibility models and Target area.

Part II describes landslide susceptibility modelling and validation in the target area. The chapters are Weighted linear combination, Weights of evidence, Logistic regression, Discussion, Conclusions, Logistic regression coefficients as an appendix, and References.

Chapter 2

Landslides

2.1 Occurrence

Landslide is a general term for a wide variety of movements of slope materials due to gravity. Landslides not only occur in mountainous regions, but also occur in gentle slope terrain. Landslide type and occurrence depend on local geomorphology, hydrology, geology, vegetation, land use, and the characteristics of triggering events (Soeters and van Westen, 1996)

The trigger is an external stimulus that modifies the slope stability conditions, increasing the material stress or reducing its strength, and causes the landslide. Intense rainfall, earthquakes, volcanic eruptions, storm waves and rapid erosion are natural triggers; activities, such as excavation or irrigation can be human triggers (Wieczorek, 1996).

According to Cruden and Varnes (1996) landslide classification is based on the types of movement and material. A landslide can be described with two nouns, first the material (rock, debris, earth) and then the movement (fall, topple, slide, spread and flow). Fig. 2.1 shows a schematic representation of general landslides.

2.2 Shallow landslides

Shallow landslide, also known as slope failure, is a movement that involves earth or debris from superficial deposits (mainly soil and colluvium) and does not affect

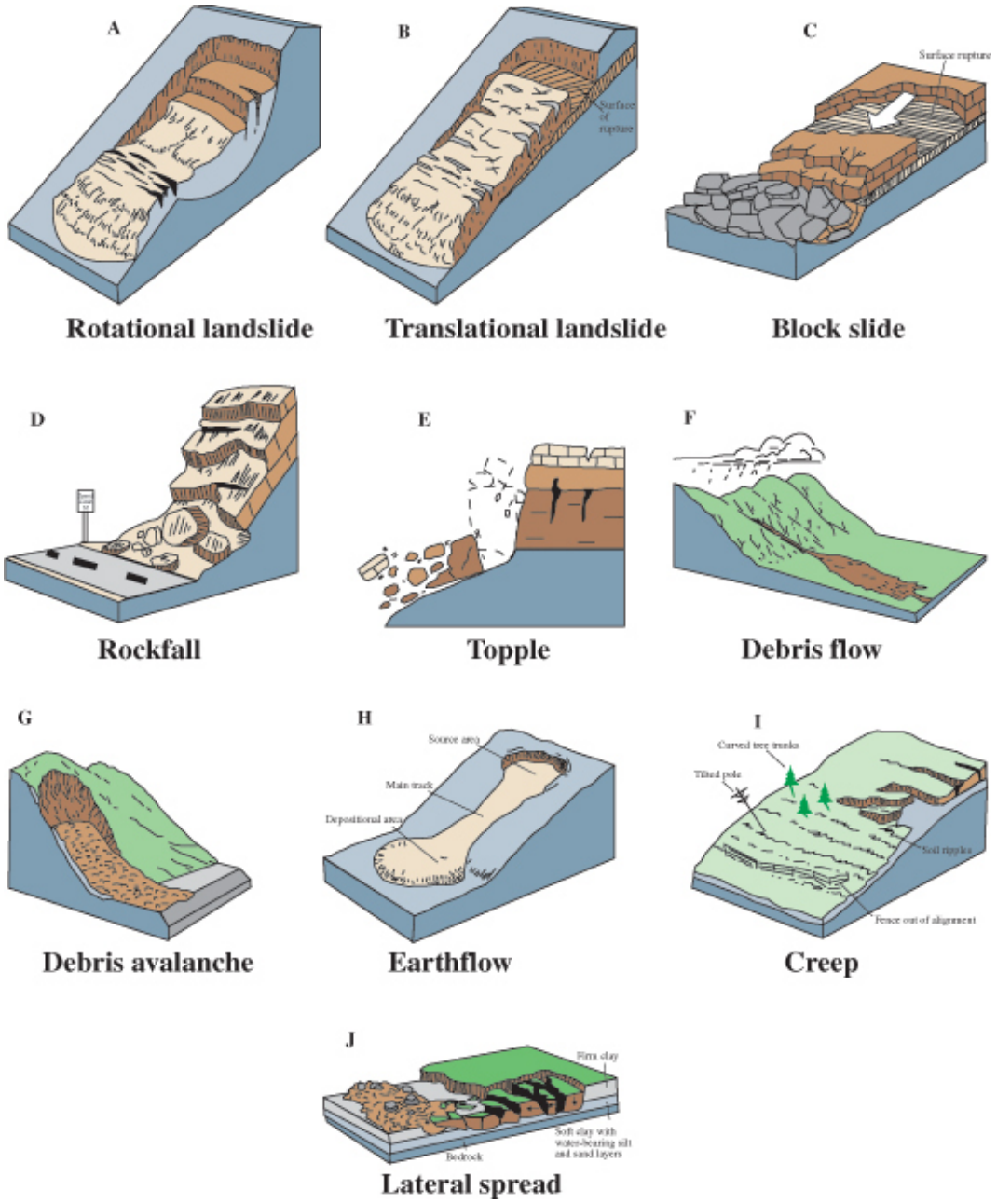

Figure 2.1: Schematic representation of landslides. Source U S Geological Survey (2004)

the bedrock. Shallow refers to the depth of the displaced mass.

Shallow landslides occur frequently in mountainous terrain worldwide triggered by earthquakes, or intense rainfall. Their occurrence greatly depends on slope topography and the presence of weathered rock mass or superficial deposits. They occur suddenly and usually move fast; and can cause great damage. Fig. 2.2 shows shallow landslides in southwester Colombia that occurred in June 1994 during the rainy season, triggered by an earthquake.

Figure 2.2: Shallow landslides in southwester Colombia triggered by an earthquake during a rainy season in 1994. Source Shuster and Highland (2001)

2.3 Impact of landslides

Impact of landslides on society can be huge, because of great economic and social loss. Effects can revert development and cause many casualties. Crozier and Glade (2005) listed a selection of landslide events that caused more than 1,000 casualties; 22 events worldwide, between 1919 and 1999, caused over 513,000 casualties.

Terms related to the effects of landslides on human activity and the environment are used here according to the definitions by Committee on the Review of the National Landslide Hazards Mitigation Strategy (2004).

- **Landslide hazard** is the potential for occurrence of a damaging landslide within a given area. Damage refers to loss of life or injury, damage of property, social and economic disruption, or environmental degradation.
- **Landslide vulnerability** is the extent of potential loss of a given element within the area affected by landslide hazard, expressed on a scale from 0 (no loss) to 1 (total loss). It depends on physical, social, economical and environmental conditions.
- **Landslide risk** is the probability of damaging consequences within a landslide-prone area. It is the product of hazard and vulnerability.
- **Landslide risk evaluation** is the application of analysis and assessments to determine risk management alternatives, which may include the decision that the risk is acceptable or tolerable.
- **Landslide hazard zonation** is the division of the terrain in homogeneous areas and ranking them according to their degree of actual or potential hazard or susceptibility to landslides.

2.4 Impact assessment

The increasing need of territory for development demands the assessment of potential landslide impacts. Furthermore, global warming is likely to increase the occurrence of rainfall related triggers; therefore it is important to assess the potential impacts from landslides.

Maps are appropriate tools for communicating the potential impact of landslides. There are three types of landslide maps useful to planners and general public, 1) landslide inventories, 2) landslide susceptibility maps, and 3) landslide hazard maps (Highland and Bobrowsky, 2008).

- **Landslide inventory maps** depict areas where landslides have occurred. The maps can simply denote areas of past landslides or include detailed information such as components of individual landslides (scarp and accumulation zones), type of movement, activity, geological age, rate of movement, and other characteristics. Inventory maps help identifying areas for detailed studies, and are fundamental for producing other potential impact maps (Highland and Bobrowsky, 2008).

- **Landslide susceptibility maps** denote areas ranked according to the tendency to the occurrence of landslides; based on local conditions (geology, topography, groundwater, vegetation). The temporal probability of occurrence of landslides, which depends on triggering events (rainfall, earthquakes), is not considered (Soeters and van Westen, 1996). Susceptibility maps only expresses the spatial probability of occurrence of landslides, however, they provide information on areas where landslides have not occurred yet.

- **Landslide hazard maps** delineate areas of past, and recent landslides and the probability of occurrence of potential landslides. For a given area, hazard maps contain detailed information on type of landslides, extent of failure, and maximum extent of ground movement (Highland and Bobrowsky, 2008).

Landslide hazard implies the assessment of spatial and temporal probabilities and the definition of type, magnitude, size and velocity of landslides (Wu et al., 1996). It requires the analysis of probability of triggering factors. Furthermore, landslide hazard is difficult to assess because the occurrence of landslides is complex and terrain conditions vary with space and time. Therefore, landslide susceptibility maps are commonly used to express relative stability of slopes. Fig. 2.3 shows an example of a regional landslide susceptibility map.

Landslide susceptibility maps provide information for delimitation of landslide-prone areas and the definition of development restriction areas. They are also important in physical planning, for reduction of costs of construction and maintenance of engineering structures (Soeters and van Westen, 1996).

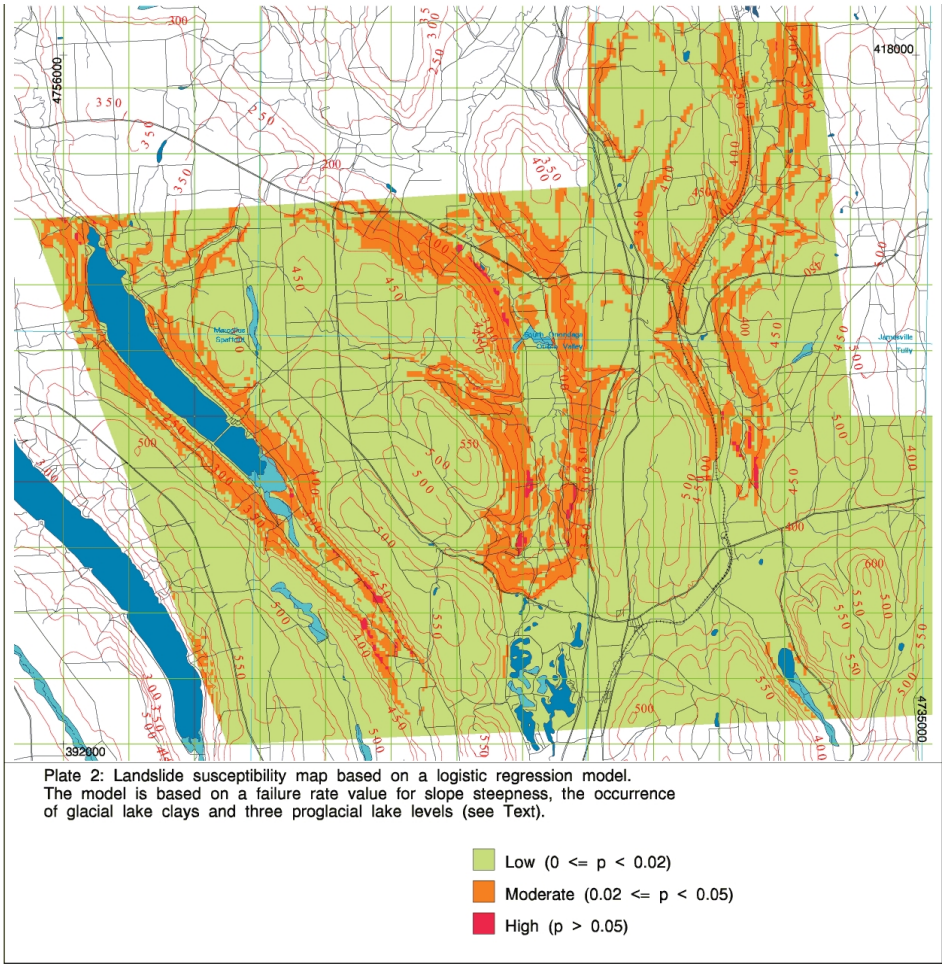

Figure 2.3: Regional landslide susceptibility map of a 415 km^2 area in United States. Susceptibility was estimated as a probability function of the presence of glacial clays, slope gradient and glacial lake levels. Source Jager and Wieczorek (2001)

In the literature, the terms hazard and vulnerability are often used incorrectly as synonymous terms (Committee on the Review of the National Landslide Hazards Mitigation Strategy, 2004). In this thesis, when making reference of other reports, hazard may refer to hazard or susceptibility.

Chapter 3

Landslide Susceptibility Mapping

This chapter consists of fundamentals of susceptibility mapping, a general classification and the description of three approaches that are applied in the target area.

3.1 Principles

Three fundamental principles are applied in landslide susceptibility modelling (Varnes, 1984).

1. The past and present are the keys to the future. This means that future landslides will occur in similar geologic, geomorphic, and hydrologic conditions that present and past landslides occurred. This assumption permits to estimate future occurrences based on historical data.

2. The main conditions that caused landslides can be identified; such as surficial material conditions, topography, effect of groundwater, and triggering mechanisms. This principle permits to make predictions in larger areas based on site observations.

3. Degrees of hazard can be estimated. Hazard (or susceptibility) can be derived from the relative contribution of the conditions that cause landslides, and it can be expressed qualitatively or quantitatively as a map.

Primary aspects in landslide susceptibility modelling are the separation of landslide scars (or scarps) from the deposits, for mapping and analysis. Because both processes occur under different conditions (slope gradient, soil characteristics, elevation). It is also necessary to separate the landslide types during the analysis, because different types of landslides occur under different conditions and mechanisms (shallow landslides, rock falls, deep seated landslides) (Chung and Fabbri, 2005).

Landslide susceptibility assessment approach depends on the scale of the target. The spatial scale of the target influences on the data availability and the model complexity. In site investigations (detailed analysis of individual landslides) many data are available and it is possible to develop complex models. While for large areas, it is difficult and expensive to obtain detailed data. Therefore, more generalized models are applied (Glade and Crozier, 2005).

In landslide hazard analysis, the scales are national (<1:1,000,000); regional scale (1:100,000-1:500,000) for thousands of square kilometres; medium scale (1:25,000-1:50,000) for few hundreds of square kilometres; and large scale (1:5,000-1:10,000) for tens of square kilometres (Soeters and van Westen, 1996). This thesis focuses on medium scale susceptibility analysis.

3.2 Susceptibility analysis approaches

Glade and Crozier (2005) reviewed the techniques to produce landslide susceptibility and hazard maps, and recommended approaches based on analysis scale (Fig.3.1). Techniques are classified in qualitative and quantitative methods.

3.2.1 Qualitative methods

Qualitative methods are based on expert knowledge and experience. They are subjective and difficult to apply to different areas. However, they can be accurate if the person or group who make the analysis know well the processes and the area (Glade and Crozier, 2005).

Landslide inventories and heuristic methods are qualitative methods. Landslide inventories are spatial databases, often used for modelling and for validation.

Scale	Qualitative methods		Quantitative methods		
	Inventory	Heuristic analysis	Statistical analysis	Probabilistic prediction analysis	Deterministic analysis
>1:10,000	Yes	Yes	Yes	Yes	Yes
1:25,000-1:50,000	Yes	Yes	Yes	Yes	Probable
1:100,000-1:500,000	Yes	Yes	Probable	Probable	No
<1:1,000,000	Yes	Yes	No	No	No

Figure 3.1: Recommended approaches for landslide susceptibility analysis. Modified from Glade and Crozier (2005)

Heuristic approaches can be geomorphic analysis or qualitative map combination (Soeters and van Westen, 1996).

Geomorphic analysis consists in mapping hazard in the field. Qualitative map combination consists in selecting parameters related to the occurrence of landslides, assigning weights to parameters classes, combining maps and classifying results to express qualitative degrees of hazard. Topographic, geological, hydrological, geomorphic or geotechnical parameters are often used to estimate susceptibility or hazard (Soeters and van Westen, 1996).

3.2.2 Quantitative methods

Quantitative approaches are based on objective criteria, producing the same results for similar data sets, and it is possible to reproduce them in other areas. Statistical, probabilistic, and deterministic approaches are quantitative methods (Glade and Crozier, 2005).

Statistical approaches are based on the analysis of landslide distribution and maps of factors related to their occurrence, such as lithology, slope gradient, land cover, profile and plan curvature, elevation, slope aspect, etc. Bivariate analyses

are common to estimate landslide parameter weights based on landslide density (Soeters and van Westen, 1996).

In multivariate analysis, parameter maps are combined and compared to the presence and absence of landslides. Then the resulting matrix is analysed using multiple regression or discriminant analysis, to classify landslide free areas as hazardous or safe according to their scores (Soeters and van Westen, 1996).

Probabilistic methods for hazard assessments are based on Bayesian probability and fuzzy logic. Results are probabilistic prediction models (Glade and Crozier, 2005).

Deterministic methods use topographic factors and hydrological conditions with generalized geotechnical information on soil properties to carry out stability analysis. The infinite slope approach is a common model to estimate factor of safety. However, in large areas data are usually difficult to collect (Glade and Crozier, 2005).

3.3 Weighted linear combination

Weighted linear combination (WLC) is based on the qualitative map combination approach (heuristic analysis). However, two types of parameters weights are used, primary-level weights for parameter classes, that can based on practical data like landslide density; and secondary-level weights for model parameters based on expert opinion. Parameters weights are combined to estimate landslide susceptibility and classify areas in relative susceptibility categories (Ayalew et al., 2004; Hong et al., 2007).

Susceptibility $S(i,j)$ in each pixel (i,j) can be expressed as the combination of the product of primary and secondary level weights:

$$S(i,j) = \sum_{k=1}^{n} w_k y_k(i,j) \tag{3.1}$$

and

$$\sum_{k=1}^{n} y_k = 1 \tag{3.2}$$

where w is the primary-level weight of parameter k, and y is the secondary-level weight of parameter k (Hong et al., 2007).

3.4 Weights of evidence

Weights of Evidence (WOE) is a quantitative statistical approach based on the Bayesian probability model, where conditional probability is based on evidence. WOE is an objective approach for the definition and selection of parameter weights in prediction modelling, and has been extensively used in mineral potential modelling (Bonham-Carter, 1994). It is also appropriate for landslide susceptibility modelling (Lee and Choi, 2004; Robinson and Larkins, 2007; Soeters and van Westen, 1996; van Westen et al., 2003).

Positive and negative weights, denoted as W^+ and W^- respectively, express the importance of parameter classes and are used to estimate susceptibility. A positive W^+ indicates that the parameter class is favourable for the occurrence of landslides. A negative W^- means that the absence of the class reduces the landslide susceptibility. Similarly, a negative W^+ means that the class is not favourable for the occurrence of landslides, and a positive W^- means that the absence of the parameter class increases landslide susceptibility. Weights around zero indicate that there is no relation between the occurrence of landslides and the factor class.

The calculation of weights is based on an area cross-tabulation of landslide area and the parameter class, as shown in Table 3.1 (Bonham-Carter, 1994) as follows:

Table 3.1: Area cross-tabulation

	Parameter class		
Landslide area	Present	Absent	Total
Present	T_{11}	T_{12}	$T_{1.}$
Absent	T_{21}	T_{22}	$T_{2.}$
Total	$T_{.1}$	$T_{.2}$	$T_{..}$

$$W^+ = \ln[\frac{\frac{T_{11}}{T_{21}}}{\frac{T_{1.}}{T_{2.}}}] = \ln[\frac{T_{11}T_{2.}}{T_{21}T_{1.}}] \tag{3.3}$$

$$W^- = \ln[\frac{\frac{T_{12}}{T_{22}}}{\frac{T_{1.}}{T_{2.}}}] = \ln[\frac{T_{12}T_{2.}}{T_{22}T_{1.}}] \tag{3.4}$$

The contrast is the addition of the W^+ and W^- weights of the class. The contrast is helpful for reclassification of parameters in more significant classes (Bonham-Carter, 1994). A positive contrast means that the class is favourable for the occurrence of landslides, while a class with a negative contrast is not favourable for the occurrence of landslides.

Using the weights, landslide susceptibility is calculated as the posterior probability (Bonham-Carter, 1994) as follows:

Estimation of prior probability $p(l)$, based on landslide density

$$p(l) = \frac{landslide - area}{total - area} \tag{3.5}$$

Calculation of natural logarithm of odds, or logits, $logit(l)$. Odds is the ratio of the probability that a landslide will occur to the probability that landslides will not occur.

$$logit(l) = \ln[\frac{p(l)}{1 - p(l)}] \tag{3.6}$$

Estimation of the posterior $logit(l|f)$, that is the $logit(l)$ plus the factor weights, W_f^+

$$logit(l|f) = logit(l) + W_f^+ \tag{3.7}$$

Calculation of posterior odds $O(l|f)$, odds given the presence of factor f

$$O(l|f) = \exp(logit(l|f)) \tag{3.8}$$

Estimation of posterior probability $p(l|f)$, the probability given the presence of factor f

$$p(l|f) = \frac{O(l|f)}{(1 + O(l|f))} \tag{3.9}$$

The calculated posterior probability represents the susceptibility expressed by the model parameters.

3.4.1 Overall conditional independence test

WOE assumes that probabilities from the parameter maps are conditional independent. This means that the probability of each parameter is not affected by the presence of the other parameters. However, there is always some degree of dependence between the parameter maps (Bonham-Carter, 1994). For example, a land cover type might present only within a certain altitude range. Lack of conditional independence (CI) affects the posterior probability results, producing higher posterior probabilities.

The overall evaluation of CI consists in comparing the number of predicted events and number of observed events (observations used to estimate the posterior probability model). If the number of predicted observations is 10-15% larger than the observed events, $E(l)_{calc}$, results are biased because of conditional dependence. The calculated number of events is obtained by adding the product of the area in units cells, $n(a)$, times the posterior probability, $p(l)$, for all pixels on the map

$$E(l)_{calc} = \sum_{k=1}^{m} p_k n(a)_k \tag{3.10}$$

where p_k is the posterior probability for pixel k, and k is the number of pixels of pixels in the map (Bonham-Carter, 1994).

3.5 Logistic regression

Logistic regression (LR) is a multivariate statistical technique appropriate for estimating the probability of a dichotomous dependent variable, such as the occurrence or absence of landslides, from its relations with independent variables, like slope gradient, lithology, land cover, etc. The result is probability values between 0 and 1. The probability values can be assimilated as landslide susceptibility.

LR has been applied to mineral potential prediction (Agterberg et al., 1993) and landslide susceptibility assessment (Ayalew et al., 2005; Brenning, 2005; Dai and Lee, 2002; Robinson and Larkins, 2007).

Advantages of LR are that it can be applied when the variables show conditional dependence, contrary to WOE method; and it can be used when variables have many classes or are continuous. LR posterior probabilities are lower than posterior probabilities estimated using the WOE method (Agterberg et al., 1993).

LR is based on the logistic function, $f(z)$

$$f(z) = \frac{1}{1 + e^{-z}} \tag{3.11}$$

where

$$z = \alpha_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k \tag{3.12}$$

LR consists in the calculation of constant or intercept, α_0, and the coefficients β_1, β_2, $\ldots \beta_k$. Variables X_1, X_2, X_k, correspond to the independent variables, or landslide related factors here.

Chapter 4

GIS Landslide Susceptibility Modelling

This chapter is about the application of a geographical information system (GIS) for landslide susceptibility mapping, and a software package for prediction modelling that can be used for susceptibility mapping.

4.1 Geographic information systems and modelling

A GIS is "a powerful set of tools for collecting, storing, retrieving at will, transforming and displaying spatial data from the real world for a particular set of purposes" (Burrough and MacDonnel, 1998).

GIS modelling is the process of combining databases of different spatial variables using a function to obtain an output map. Depending on the function type, GIS models are theoretical, empirical or hybrid. Theoretical models have relationships based on theoretical understanding or physical or chemical principles; empirical models are based on statistical or heuristic relationships; and hybrid models employ semi-empirical relationships, theoretical relationships with empirical functions (Bonham-Carter, 1994).

4.2 Susceptibility modelling

Before the application of GIS, most landslide hazard (or susceptibility) assessments were based on geomorphic and geologic analysis from interpretation of aerial photographs and field observations; that is using qualitative approaches. Landslide hazard (or susceptibility) was expressed as relative qualitative classes.

The use of GIS in landslide hazard began in the late 1970s with simple qualitative models and it had evolved to quantitative models and expert systems (Soeters and van Westen, 1996). Now almost all landslide susceptibility modelling employ GIS. The spatial nature of the landslide process and the amount of data require the use of GIS for its analysis. However, principles for GIS-based susceptibility modelling are the same as those employed in general susceptibility models described in Section 3.1.

GIS-based hazard analysis can be applied from national scale to large-scale analysis. All methods for medium-scale analysis described in Section 3.2 can be implemented in GIS. However, in landslides inventories, GIS is basically used to store and display data. There are many examples of qualitative and quantitative applications. Qualitative approaches are presented by van Westen et al. (2003) and Ayalew et al. (2004). In other qualitative analysis, GIS is used to analyse the difference of landslides triggered by rainfall and earthquakes (Yamagishi and Iwahashi, 2007).

Quantitative approaches include bivariate statistical analysis (Fernandez et al., 2003; Suzen and Doyuran, 2004); multivariate statistical analysis (Carrara et al., 1995; Clerici et al., 2002; Dai and Lee, 2002; Guzzetti et al., 1999; Mark and Ellen, 1995); probabilistic methods (Brenning, 2005; Chung and Fabbri, 1999; Remondo et al., 2003) and relative new approaches, such as neural networks (Lu and Rosenbaum, 2003), and decision tree modelling (Saito et al., 2009).

On the other hand, large-scale quantitative analyses are less common (Iida, 1999; Laouafa and Darve, 2002).

4.3 Visualization of susceptibility

Results from susceptibility approaches are generally continuous numerical values. However, for visualization they are usually divided in classes to express relative degrees of susceptibility. Categorization also permits comparing different susceptibility maps.

Arbitrary classifications are still common; however, the main classification approaches are ranking, natural breaks, equal interval classes, equal area classes, and mean value and standard deviation intervals (Bonham-Carter, 1994; Chung and Fabbri, 1999, 2003; Fabbri and Chung, 2008).

Classification determines the spatial distribution of susceptibility; classifying a susceptibility map using two methods produces two different susceptibility maps. Recommended approaches are ranking based on equal area classes (Chung and Fabbri, 2003) and classification based on mean value and standard deviation intervals (Ayalew et al., 2004).

4.4 Advantages and disadvantages

Main advantages of using GIS in susceptibility analysis are that GIS permits systematic prediction modelling, applying different techniques or factor combinations, and evaluation of models. GIS facilitates data sharing. Scripts facilitate performing calculations And batch command options and analysis data logs also permit to make repeated calculations easily (Chung and Fabbri, 1999, 2005; Remondo et al., 2003; Soeters and van Westen, 1996; van Westen et al., 2003).

Furthermore, availability of global and national digital databases (elevation, soils, geology, remote sensed imagery, etc.) facilitates the application of GIS-based approaches. In some cases, the production of landslide databases is based on detection of landslides on remote sensed data using GIS (Ayalew et al., 2004; Saito et al., 2009).

Disadvantages of GIS-based approaches are that producing digital databases is time-consuming, and some parameters are not possible to express as maps (like proximity to old landslides) and only the parameters that are easy to map are included in the analysis; and statistical analysis usually requires using external

packages (Guzzetti et al., 1999; Soeters and van Westen, 1996; van Westen et al., 2003).

4.5 Spatial Data Modeller

Spatial Data Modeller (SDM) is a set of software tools for prediction modelling using categorical and numerical (interval, ordinal, or ratio) maps. Implemented modelling approaches are weights of evidence, logistic regression, fuzzy logic and neural networks (Sawatzky et al., 2009).

SDM works as an extension for ArcGIS 9.3 and requires Spatial Analyst extension. SDM is public domain and it is available in the Internet at `http://www.ige.unicamp.br/sdm/default_e.htm`.

Applications of SDM include probabilistic spatial prediction modelling for mineral potential mapping (Nykanen and Ojala, 2007; Raines et al., 2007) for groundwater contamination vulnerability assessment (Arthur et al., 2007; Masetti et al., 2007) for aggregate quarry sitting (Robinson and Larkins, 2007); and for landslide analysis (Nelson et al., 2007; Poli and Sterlacchini, 2007).

SDM is used here to make susceptibility maps based on logistic regression.

Chapter 5

Validation of Susceptibility Models

5.1 Definition and importance

Model validation is comparing the results with real world data to assess the model accuracy. Validation of landslide susceptibility models gives information about the confidence of the model to the user. Validation also permits to compare different models or model parameter variables (Begueria, 2006).

In landslide susceptibility assessment, accuracy is the capacity of the map to differentiate landslide-free from landslide-prone areas. Accuracy and objectivity depend on model accuracy, input data, experience of earth scientist and size of the study area (Soeters and van Westen, 1996).

On the other hand, model evaluation is the assessment of its adequacy to the needs of the final user. In landslide susceptibility modelling, it is mainly used to define hazard classes for practical purposes, such as to prioritize areas with the highest susceptibility for further investigations (Begueria, 2006).

5.2 Cross-area tabulation derived statistics

In landslide susceptibility assessment there are two types of prediction errors; 1) landslides may occur in areas that are predicted to be stable, and 2) landslides

may actually not occur in areas that are predicted to be unstable (Soeters and van Westen, 1996). The first type of errors is a false positive (type I error) and the second type is a false negative (type II error).

Validation of landslide susceptibility maps is commonly based on statistics from cross-area tabulation, also known as the confusion matrix or contingency table (Bonham-Carter, 1994). Based on a threshold, continuous susceptibility values are categorized in a binary map (susceptible and not susceptible classes) and then compared with a binary landslide distribution map (presence or absence of landslides).

Cross-tabulation consists in the calculation of overlap areas between the two binary maps. The possible combinations are as follows: landslide areas are classified as susceptible areas (true positive observations); landslide-free areas are classified as no susceptible areas (true negative observations); landslide areas are classified as no susceptible areas (false negative observations); and landslide-free areas are classified as susceptible areas (false positive observations) as shown in Table 5.1.

Table 5.1: Confusion matrix. a, true positive observations; d, true negative observations; b, false negative observations (error type II); and c, false positive observations (error type I) Source Begueria (2006)

	Observed	
Predicted	X_1	X_0
X_1'	a	b
X_0'	c	d

Success and prediction and rates are the most common approaches (Brenning, 2005; Chung and Fabbri, 1999, 2003; Robinson and Larkins, 2007; van Westen et al., 2003). The success rate or model efficiency is the proportion of correctly classified observations. It is calculated comparing the model to the modelling dataset. The prediction rate is calculated comparing the model to a dataset different from the modelling dataset, a validation dataset.

$$Efficiency = \frac{(a+d)}{(a+b+c+d)} \tag{5.1}$$

However, Begueria (2006) noted that the model efficiency greatly depends on prevalence, the relation between false positives (type I errors) and false negatives (II errors). Begueria recommends using statistics not affected by prevalence, proportion of positive and negative cases. They are sensitivity, the proportion of positives observations correctly identified; specificity, the proportion of negatives observations correctly identified; false negative rate, false positive rate, and likelihood ratio. Sensitivity and specificity are estimated as follow:

$$Sensitivity = \frac{a}{(a+c)} \tag{5.2}$$

$$Specificity = \frac{d}{(b+d)} \tag{5.3}$$

5.3 The receiver-operating characteristic (ROC) plot

ROC plot is a graph of sensitivity versus specificity, statistics not affected by prevalence. It is calculated by estimating the parameters for many thresholds.

The area under the ROC plot (AUC) is a statistic accuracy of the model and it is independent of the prediction threshold. AUC is 0.5 when there is no variation with threshold definition and 1 when the model makes a perfect prediction. AUC below 0.5 indicates that performance is lower than classification by chance. The higher the AUC, the higher the model accuracy (Fawcett, 2006).

Fig. 5.1 shows the ROC of two models, the AUC of model A is 0.5 and the AUC of model B is 0.71.

AUC is calculated by adding the areas of the polygons between the thresholds (Begueria, 2006).

$$AUC = \sum_{i=1}^{n+1} \frac{1}{2} \sqrt{(x_i - x_{i+1})^2}\ (y_i + y_{i+1}) \tag{5.4}$$

where x_i is specificity and y_i is sensitivity at threshold i and $x_{n+1} = 0$, $y_{n+1}= 1$.

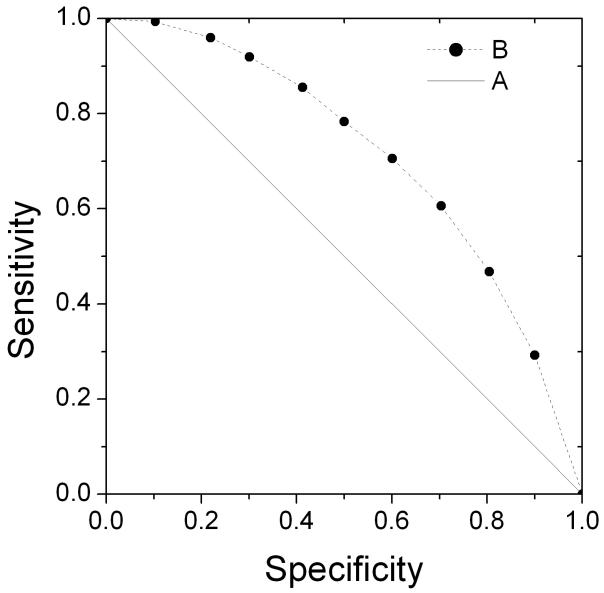

Figure 5.1: ROC plot. Line A is line of no discrimination (AUC is 0.5). Line B represents the accuracy of model B with the values of specificity and sensitivity calculated for different thresholds, and the AUC is 0.71

5.4 Database partition

In quantitative landslide susceptibility prediction modelling, it is not possible to compare the model with future landslides. Therefore, it is necessary to divide the landslide database in a modelling dataset and a validation dataset (Chung and Fabbri, 2003).

In landslide hazard assessment, database partitions are based on time, space and random techniques. Time partition is using databases from different time periods; the older one for modelling and the later for validation. Space partition is dividing the study area in two sub areas, and using one for prediction and the other for validation. Random partition consists in dividing randomly the landslides in two groups, one for prediction and the other for validation. It is pretended that the landslides from the validation dataset have not occurred yet (Chung and Fabbri, 2003).

Partition approach depends on the data available. When databases from different time periods are available, time partition is the best approach (Fabbri and

Chung, 2008). Otherwise it is better to use random partition. Space partition is not appropriate, because sub areas usually present different conditions regarding geology, geomorphology, and hydrology.

Random partition is the most common approach (Brenning, 2005; Chung and Fabbri, 2003; Fabbri and Chung, 2008; Remondo et al., 2003). On the other hand, the partition dataset size affects the results; larger prediction datasets produce better results, and some authors recommend a half and half partition (Brenning, 2005; Fabbri and Chung, 2008).

Chapter 6

Target Area

Landslide susceptibility modelling and validation was carried out in a target area in southwester Japan, defined here as Hamada area.

6.1 Location

The 157-km^2-target area is located in Hamada city, Shimane prefecture, south west Japan (Fig. 6.1). The area corresponds to two 1:25,000-scale topographic quadrangles, 523117 (Misumi) and 523210 (Kitsuka).

6.2 Geological and geomorphological setting

The western part is a coastal area of plains and hills, with elevations generally lower than 150 m. However, there are isolated hills (monadnocks) with elevations between 260 and 400 m. The eastern part is a higher mountainous terrain with elevations of up to 700 m. In general, elevation increases gradually from sea level in the west to about 700 m in the central to eastern part of the target area.

According to Wada et al. (1984), in the region there are low-relief topographic surfaces (peneplains) produced by erosion processes, with different weathering characteristics that may control the occurrence of landslides. They are Takasu plain (between 70 and 100 m), Iwamikogen plain (200-300 m) and Takauchi plain (350-420 m).

Figure 6.1: Location of target area. World map source Blue Marble (2002); Japan map source International Steering Committee for Global Mapping (2009)

Lithologies in the target area are mainly Paleozoic to Mesozoic pelitic and psammitic schists; Paleogene diorites and granitic rocks; and Paleogene rhyolitic to dacitic pyroclastic and volcanic rocks. Quaternary deposits form terraces and alluvial plains within valleys (Fig. 6.2).

The schists form the low and hilly terrain in the western and southeaster parts of the target area. Dioritic and granitic intrusions form high peaks and monadnocks in the central area. Volcanic and pyroclastic rocks constitute a higher mountainous terrain with steep slopes in the northeaster area.

6.3 Rainfall triggering event

Intense rainfall between July 20^{th} and 23^{rd}, 1983, triggered many shallow landslides in western Shimane prefecture. The highest recorded total rainfall was 742 mm (Misumi area) and maximum hourly intensity was 90 mm/h. Shallow landslides, debris flows, and flooding led to 91 deaths, and the overall economic

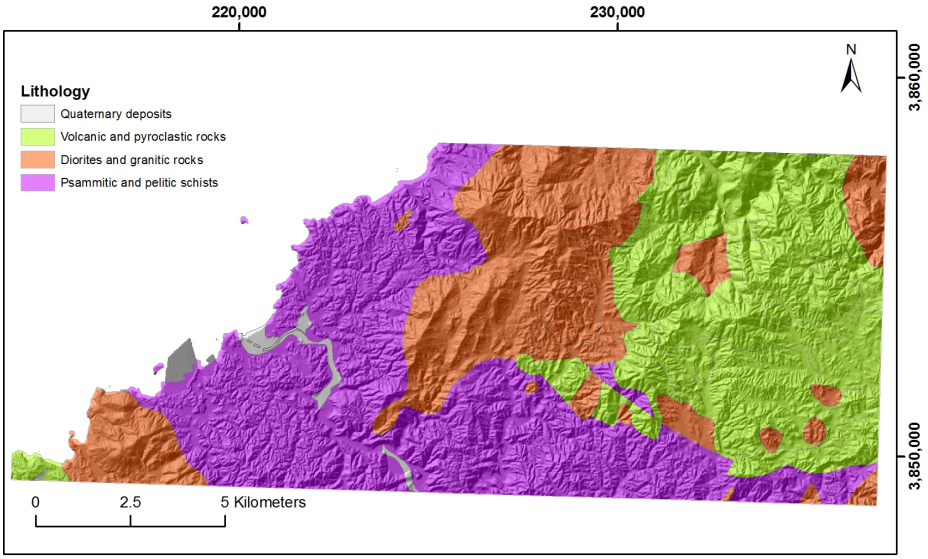

Figure 6.2: Generalized lithology based on 1:200,000-scale Digital Geologic Map of Japan downloaded from Geological Survey . Source Geological Survey Japan (2009)

loss was 360,000 million yen (Research Group of San-in Heavy Rainfall Disaster, 1984), about 2,700 million Euro.

The maximum hourly precipitation (90 mm/h) corresponds to a 150-year recurrence, and the maximum daily precipitation (372 mm/day) corresponds to over a 200-year recurrence. Shallow landslides occurred from 100 mm of total rainfall and from 40 mm/h rainfall intensity (Wada et al., 1984). The event is known as the 1983 San'in heavy rainfall disaster.

6.4 Landslide occurrences

According to Research Group of San-in Heavy Rainfall Disaster (1984), most landslides were shallow, the failure materials were colluvium and residual soil, and the slope failures were related to topographic and geologic conditions and rainfall intensity. The highest frequency of slope failures occurred on 30° to 40° slopes, and the largest landslides occurred in granitic rock regions. With the

increase of water content, many landslides developed into debris flows.

According to Okuda and Okimura (1984), the highest frequency of shallow landslides occurred on slopes with gradients between 15° to 25° slopes, and debris flows on slopes with gradients between 10° and 20°. Wada et al. (1984) pointed out that landslides were also related to geomorphology and weathering zones of topographic levels.

6.5 Landslide inventory

Pimiento and Yokota (2006) produced a 10 m grid inventory map from stereoscopic interpretation of 1:8,000-scale black-and-white aerial photographs. The inventory map consists of the landslide source areas. 2,411 landslides were inventoried in the target area.

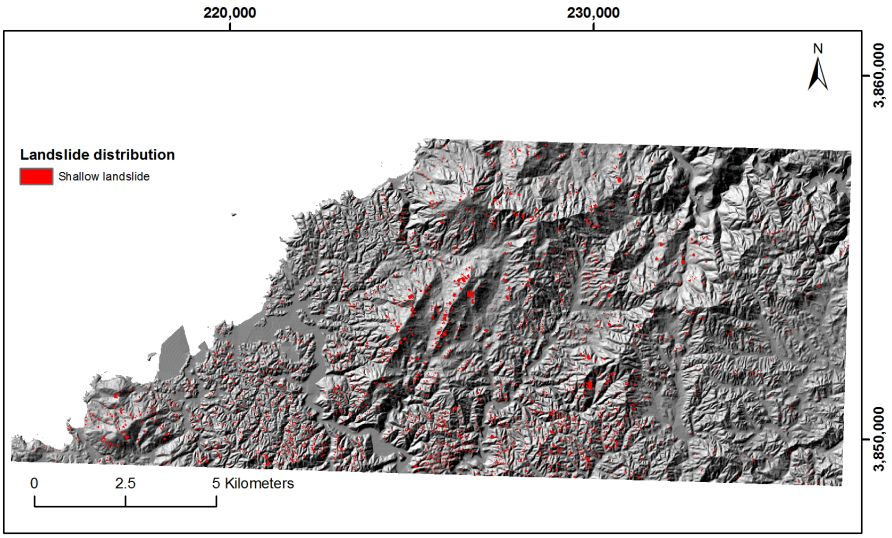

Figure 6.3: Landslide inventory map over a hillshade model. Based on the (10-m grid) landslide distribution map, the predominant size of the source areas was 1,400 m^2 (Pimiento and Yokota, 2006). Elevation model derived from 10 m DEM (Geographical Survey Institute Japan, 2009)

Fig. 6.4 shows a detail of the landslide map over 1 km mesh orthophotos from the Ministry of Land, Infrastructure, Transport and Tourism of Japan.

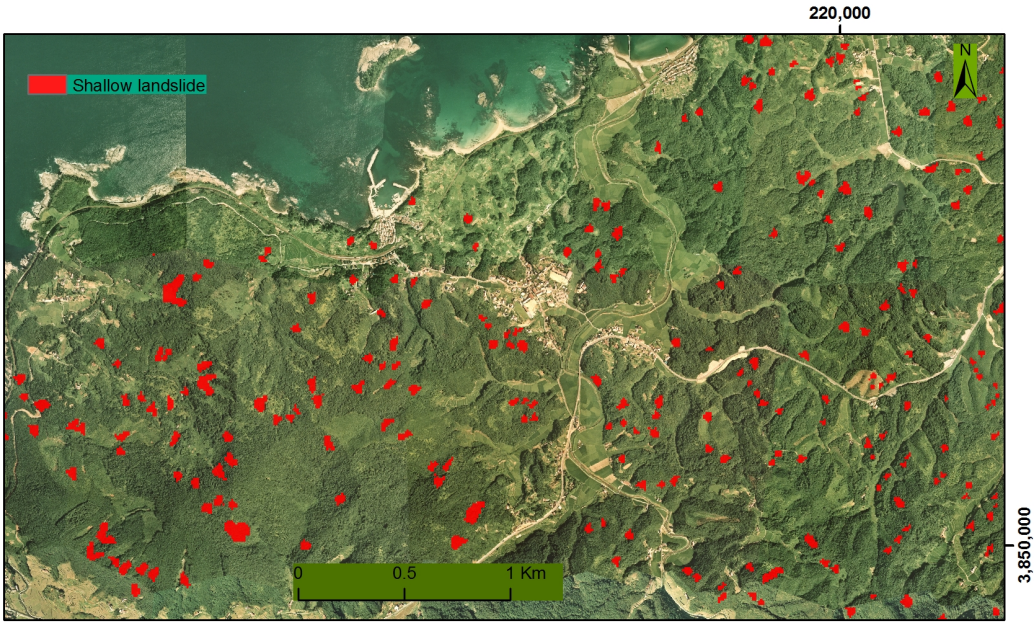

Figure 6.4: Detail of landslide map overlaid on 1 km mesh orthopothos from Ministry of Land, Infrastructure, Transport and Tourism (2009)

6.6 Shallow landslide related parameters

For landslide susceptibility modelling, the landslide mechanism in the target area is analysed and then the factors related to their occurrence are selected.

It is clear from previous reports (Research Group of San-in Heavy Rainfall Disaster, 1984; Wada et al., 1984) that the shallow landslides were triggered by intense rainfall. A common cause for the occurrence of shallow landslides triggered by intense rainfall is the increase of pore-water pressure in soil (reducing its strength), produced by rainfall infiltration and soil saturation. Triggering thresholds depend on rainfall intensity and local geological and geomorphological conditions (Wieczorek, 1996).

According to Turner (1996), parameters influencing the occurrence of shallow landslides in colluvium and residuum are soil properties, geomorphology, vegetative cover and triggering events. However, in regional scale analysis detailed soil properties data are not available, and it is necessary to use substitute variables.

Since soil characteristics depend on bedrock, discontinuities, topography and

climate; lithology can be a surrogate variable of soil properties. Geomorphology may be expressed by elevation, slope gradient, slope aspect, and slope curvature, among others; and land use is generally a function of slope gradient and elevation.

The presence of colluvial deposits greatly influences the occurrence of shallow landslides. Colluvium is abundant in hollows, head of first order drainages where contour lines are concave facing away from the ridge crest (Turner, 1996).

Also, run-off and groundwater flow in soil depend on plan and profile curvatures. In depressions, pore-water pressure increases with rainfall infiltration; therefore, plan and profile curvature should influence the occurrence of rainfall triggered shallow landslides (Ayalew et al., 2004).

Considering the characteristics of the slope failures from the 1983 San'in heavy rainfall disaster in the target area, and the parameters related to the general occurrence of shallow landslides, significant parameters for susceptibility modelling of rainfall triggered shallow landslides in the target area could be lithology, slope gradient, profile curvature, plan curvature and elevation.

Part II

Modelling and Validation

Chapter 7

Weighted Linear Combination

This chapter describes the landslide susceptibility model based on parameter weights defined by expert knowledge, the databases used to produce the model, processing of the databases, the susceptibility map, and finally, the model validation using practical data.

7.1 Definition

Hong et al. (2007) produced a global landslide susceptibility map with 30 m spatial resolution using the WLC approach, and physical parameters related to the occurrence of shallow landslides. The parameters are obtained from global remote sensed databases and ancillary data.

The general steps to produce the landslide susceptibility map are: selection of parameters; classification of parameters and assignment of weights according to contribution to shallow landslide susceptibility; and estimating susceptibility using a weighted linear combination.

The selection of parameters is based on literature and empirical assumptions. Parameters considered are slope gradient, soil type, soil texture, elevation, land cover, and drainage density. The definition of parameter weights is based on other works and the comparison of results from different parameter combinations and the landslide susceptibility map of the United States (Hong et al., 2007).

The landslide susceptibility criteria are higher slope gradient, higher susceptibility; coarser and looser soil, higher susceptibility; higher relative elevation,

higher susceptibility and decreasing susceptibility for higher drainage density. Definition of parameter (primary) weights is based on these criteria.

The model databases have resolutions between 30 meters and 0.25 degrees and consist of elevation from Shuttle Radar Topography Mission (SRTM) and GTOPO30; land cover from Moderate Resolution Imaging Spectroradiometer (MODIS), soil type from Digital Soil Map of the World (DSMW); and soil texture data from International Satellite Land Surface Climatology Project (ISLSCP) Initiative II Data Collection (Hong et al., 2007). All the data are available in the Internet.

Landslide susceptibility calculated for the land areas is normalized from zero to one and classified in six categories, no susceptible, very low, low, moderate, high, and very high susceptibility.

7.2 Modelling database

This section describes the database to derive the parameters for the WLC approach in the target area.

7.2.1 Elevation data

Shuttle Radar Topography Mission version 2 is a global elevation database produced by NASA. The SRTM3 data has spatial resolution of 3 arc-seconds, equivalent to a horizontal resolution of 90-meter in equator areas. However, the original SRTM data contains (data) voids caused by shadowing, and data or radar problems. Therefore I used here SRTM 90m version 4.1 from the Consortium for Spatial Information (CGIAR-CSI).

SRTM 90m version 4.1 is a void-filled database using auxiliary elevation data (DEMs) and interpolating results (Jarvis et al., 2008). The database is available for download in Internet at `http://srtm.csi.cgiar.org`.

7.2.2 MODIS land cover data

Land Cover Type 1 (MODIS 12) is a global land cover database of 17 land cover classes according to International Geosphere-Biosphere Programme (IGBP)

global vegetation database. I used a MCD12Q1 V005 dataset, in raster tiles with 500 m spatial resolution and 1-year temporal granularity (Land Processes Distributed Active Archive Center, 2009).

7.2.3 Digital Soil Map of the World

The soil type data is from the Digital Soil Map of the World (DSMW) produced by the Food and Agriculture Organization of the United Nations (FAO), Version 3.6, and completed in January 2003 (FAO, 2009). The map consists of FAO soil mapping units with texture and slope information. The vector data has shapefile format and is available in the Internet for download at: `http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116&currTab=distribution`

7.2.4 Soil texture data

The soil texture database is one of the 18 raster data soil characteristics from the International Satellite Land Surface Climatology Project (ISLSCP) Initiative II Data Collection (Global Soil Data Task, 2000). The data represents the characteristics of soil at depth between 0 and 150 cm.

The soil texture consists of 12 classes based on the United States Department of Agriculture (USDA), from sand (1) to clay (12); water class is 0, and permanent ice is 13. The dataset with horizontal spatial resolution of 1 degree has ASCII grid file format, and is available for downloaded from: `http://islscp2.sesda.com/ISLSCP2_1/html_pages/groups/hyd/islscp2_soils_1deg.html`.

However, according to Hong et al. (2007), data are available with resolution of 0.25 degrees.

7.3 Processing of model database

General processing consisted in data import, definition of spatial reference, selection of area corresponding to the target area, projection to Universal Transverse Mercator (UTM) 53N zone with Tokyo datum; reclassification to landslide susceptibility (Hong et al., 2007); and resampling to 10 m grid, the cell size of the landslide database.

7.3.1 Elevation data SRTM3

The original Geotiff format image is in decimal degrees and datum WGS84; it is a tile of 5 degrees of longitude by 5 degrees of latitude. It was projected to Tokyo UTM 53N Zone; then resampled to 10 m grid and clipped to the target area.

Elevation values were between -2 (near the coast) and 690; maximum elevation agrees with a printed topographic map. Negative values were replaced with 1, and elevation was normalized between 0 and 1 to express susceptibility. Higher elevation representing higher susceptibility. Fig. 7.1 shows the elevation map in the target area.

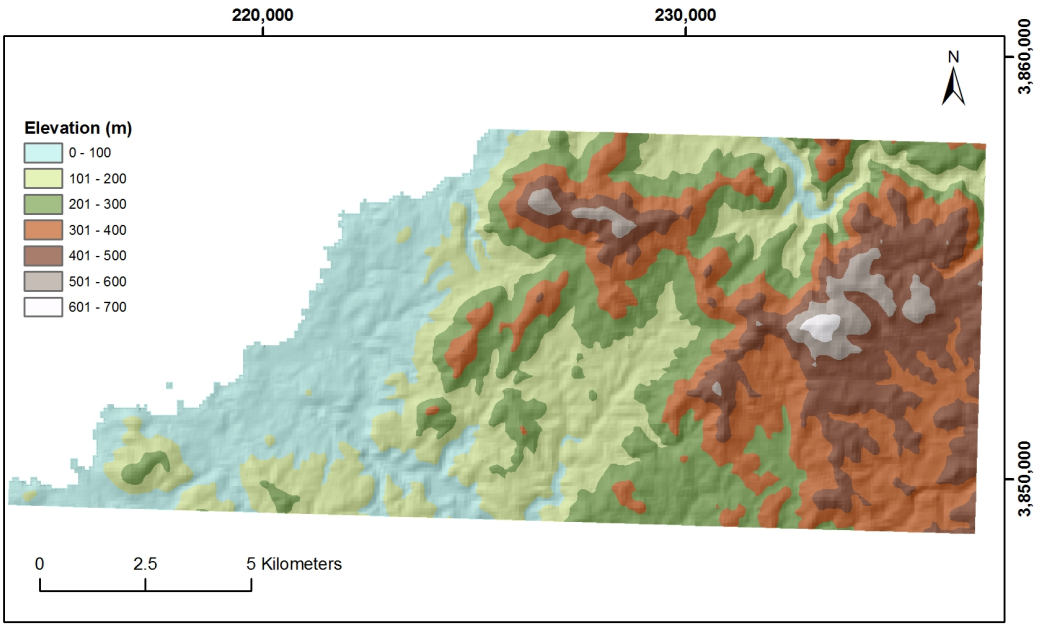

Figure 7.1: Elevation based on 90 m DEM draped over a hillshade model. Source Jarvis et al. (2008)

7.3.1.1 Slope gradient

Slope gradient map was produced from the elevation dataset described in the previous section (90-m resolution DEM) using ArcGIS slope function; that is based on the method by Zevenbergen and Thorne, as described by Burrough and MacDonnel (1998).

The resulting 0 to 65 degrees slope gradient raster was rounded to integer values (Fig. 7.2). Then it was normalized to susceptibility values; flat slopes, zero susceptibility; steepest slope, susceptibility one.

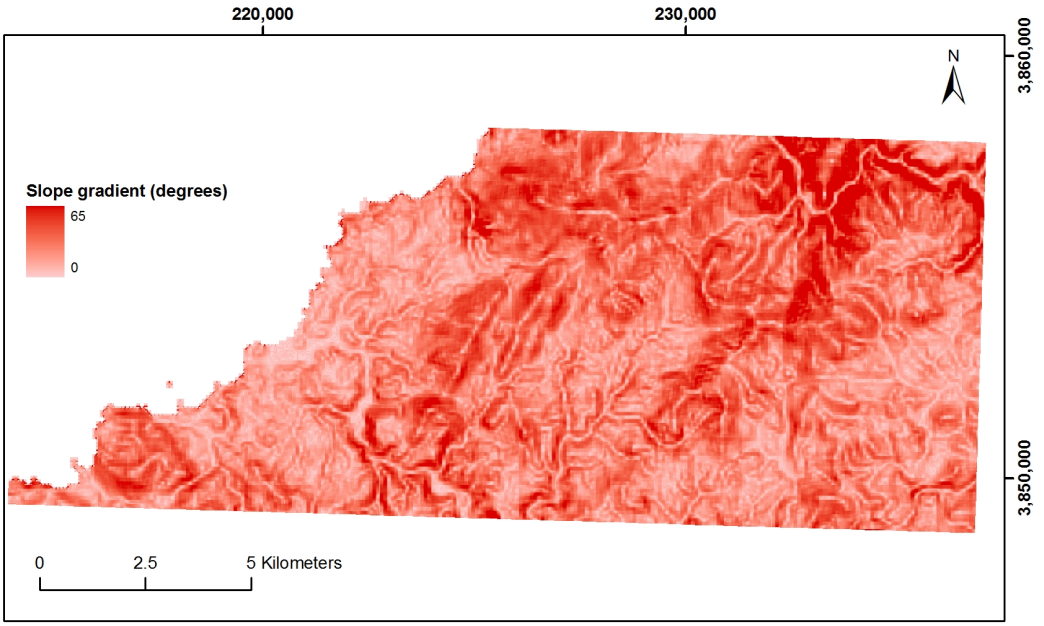

Figure 7.2: Slope gradient derived from 90 m DEM

7.3.1.2 Drainage density

Drainage density map was produced from elevation data (90-m resolution DEM) using Arc Hydro Tools, a free ArcGIS extension for hydrological modelling (ESRI, 2009).

Processing the DEM generated a polygon catchment layer and a line drainage layer. Drainage density for catchment areas was calculated from length and area attributes of the vector layers, and converted to raster.

Density values extremely high (>3 km/km^2) were reclassified to 3 km/km^2; in catchment areas with unnatural shape and very small area. Areas with no-data values cells, along the target area boundaries, were replaced with the lowest drainage density (0.035 km/km^2). Then the raster was normalized to express

susceptibility; lower density, higher susceptibility. Fig. 7.3 shows the drainage density map derived from 90 m DEM.

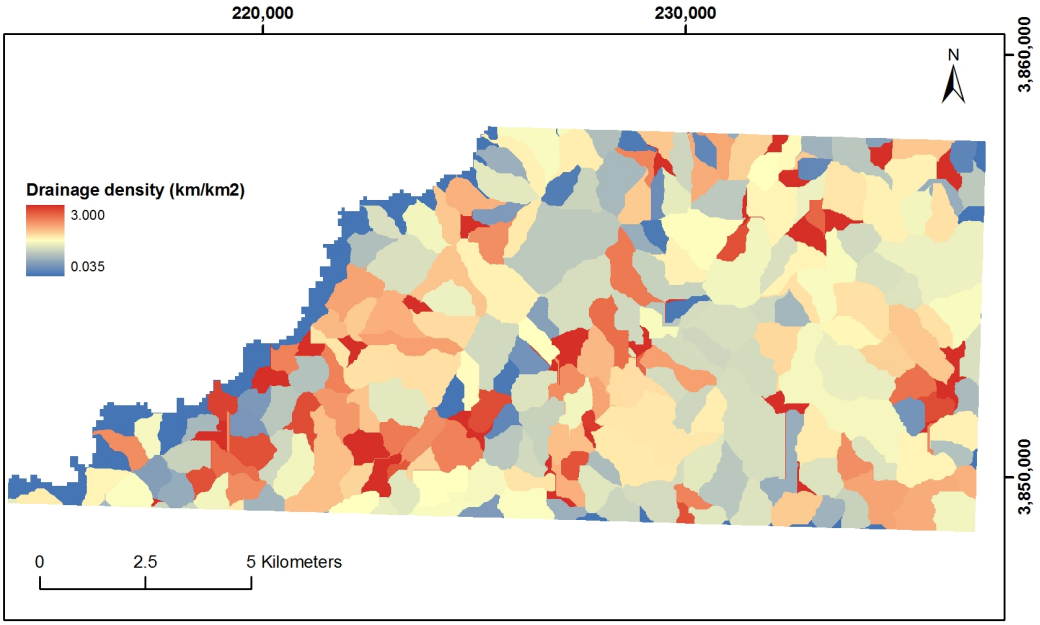

Figure 7.3: Drainage density derived from 90 m DEM

7.3.2 Land cover

For the target area, the dataset used is in MODIS Sinusoidal Tiling System (horizontal 28; vertical 5) and the projection is sphere-based Sinusoidal. The file name is MCD12Q1.A2005001.h28v05.005.2008310180817.hdf.

Processing consisted in selecting band 1 (corresponding to the IGBP land cover classification), and changing the image header format to BSQ; using MultiSpec, an image data analysis software (MultiSpec, 2009); then changing file extension to BSQ.

In ArcGIS, the image projection was defined as Sphere-based Sinusoidal; then it was projected to Tokyo UTM 53N Zone, and resampled to 10m grid and clipped to the target area.

Land cover susceptibility consists of 11 categories with landslide susceptibility between 0 and 1 (Hong et al., 2007). Table 7.1 shows MODIS values and sus-

ceptibility. However, only the land cover classes marked with (*) were present in the target area (Fig. 7.4). Reclassification of the land cover map produced the land cover susceptibility factor map.

Table 7.1: Land cover susceptibility according to Hong et al. (2007)

Category	Susceptibility	MODIS	Land cover
0	0.0	0, 15	Water bodies, permanent snow and ice
1	0.1	1, 2, 11	Evergreen forests, permanent wetland *
2	0.2	3, 4	Deciduous needle-leaf *, broad-leaf forests
3	0.3	5	Mixed forests *
4	0.4	6,7	Open, closed shrub lands
5	0.5	8, 9	Woody savannah *, savannah
6	0.6	10	Grass land
7	0.7	12	Croplands
8	0.8	14	Cropland/natural vegetation mosaic *
9	0.9	16	Barren or sparsely vegetated land *
10	1.0	13, 17	Developed land, road corridors, coastal area

7.3.3 Digital Soil Map of the World

The geographic data format is shapefile and attribute data are in spreadsheet files. Processing consisted in defining the spatial reference as WGS84 in ArcGIS; editing spreadsheet tables to produce database files for joining with geographic data; then clipping the target area and adding a susceptibility field according to susceptibility criteria.

In the target area, the soil map of the study consisted of two classes only, Table 7.2 and Fig. 7.5. Susceptibility was defined based on dominant soil texture classes (coarse, medium and fine). Susceptibility for soil type with fine texture was lower than for medium texture soil type.

Soil type map was rasterized based on soil type, then projected to Tokyo UTM 53N Zone and resampled to 10 m grid, and reclassified soil type to susceptibility values of Table 7.2.

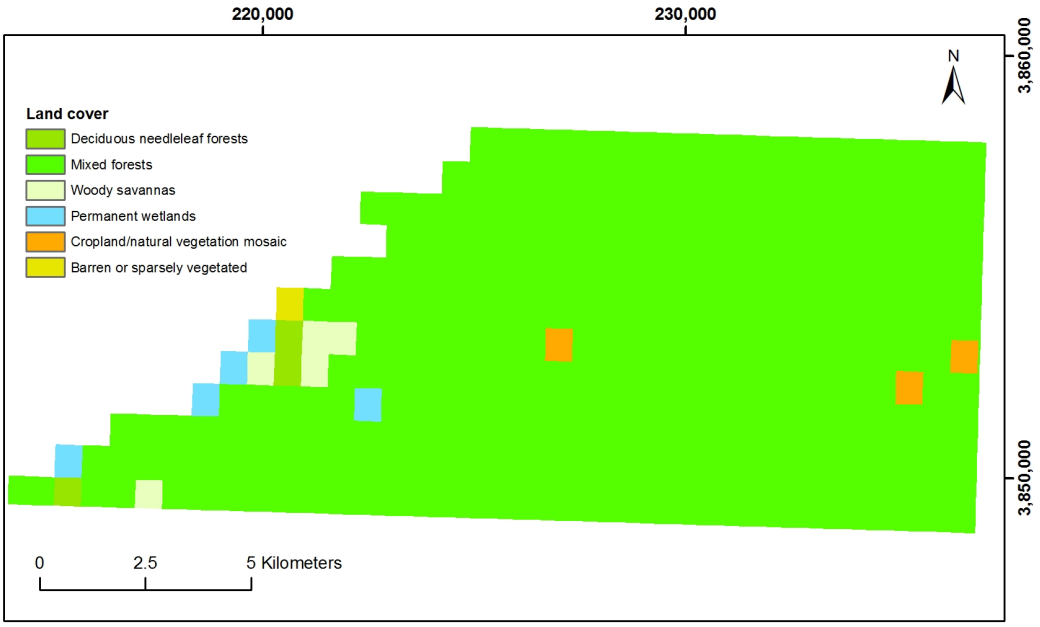

Figure 7.4: Land cover from MODIS

Table 7.2: Soil type and susceptibility according to Hong et al. (2007).

Value	Soil type	Texture	Susceptibility
1	Ao86-3b	Fine	0.33
2	Be88-2/3b	Medium/fine	0.66

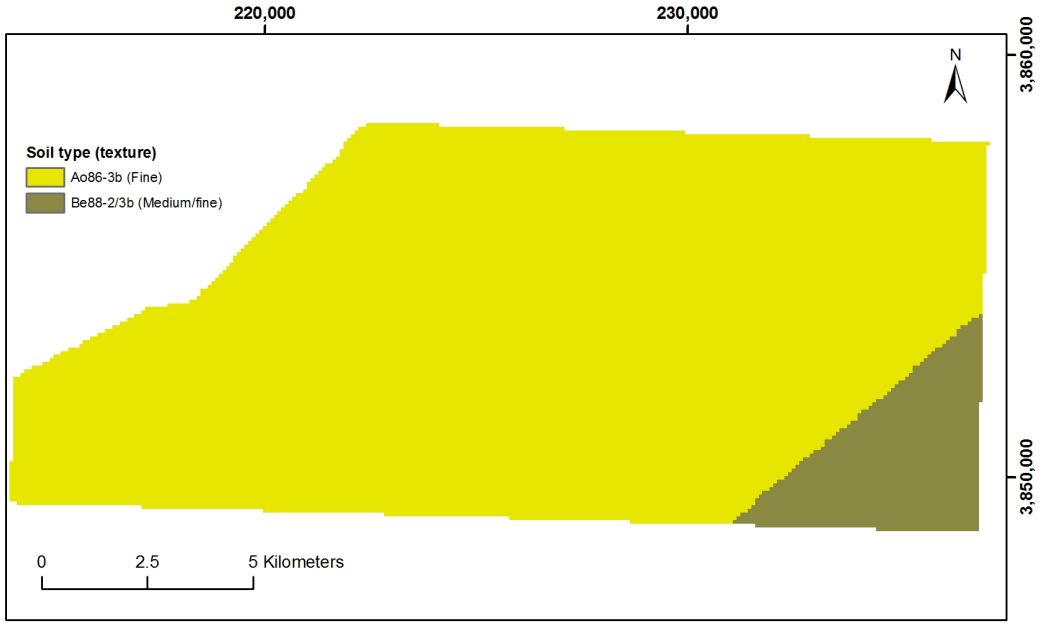

Figure 7.5: Soil type. Source FAO (2009)

7.3.4 Soil texture dataset

Dataset soil texture classes and raster values are: 0 water; 1 sand; 2 loamy sand; 3 sandy loam; 4 loam; 5 silt loam; 6 silt; 7 sandy clay loam; 8 clay loam; 9 silty clay loam; 10 sandy clay; 11 silty clay; 12 clay; and 13 permanent ice. These correspond to three texture categories, coarse (1 - 3 classes) with susceptibility 1; medium (4 - 8) with susceptibility of 0.66, and fine (8 - 12) with susceptibility of 0.33.

The ASCII file was converted to 1-degree raster map and the spatial reference defined as WGS84. The target area was inside one cell with texture value of 8, which corresponds to clay loam. Instead of projecting and clipping the target area, a constant raster was created.

The raster, with the same cell size and extent as the landslide distribution map 10 m grid and the spatial reference as Tokyo UTM 53N Zone, had a constant value for the susceptibility of clay loam class. In the susceptibility scale, fine texture (loam clay) corresponds to low susceptibility class, and it was assigned value of 0.33.

7.4 Map combination

After producing the landslide related parameter maps and the weights for their classes, this section describes the combination of the maps to estimate landslide susceptibility in the target area using the WLC approach. Besides the parameter class weight (primary weight), each parameter has a weight to express the contribution to the occurrence of landslides (secondary weight). Parameters and their secondary weights are as shown in Table 7.3.

Table 7.3: WLC parameter weights (or secondary weights) according to Hong et al. (2007).

Parameter	Weight
Slope gradient	0.3
Soil type	0.2
Soil texture	0.2
Elevation	0.1
Land cover	0.1
Drainage density	0.1

Total susceptibility value in each cell was the sum of the rasters (corresponding to primary weights) multiplied by their secondary weights, as expressed in Eq. 3.1. Hong et al. (2007) normalized total susceptibility from zero to one and reclassified the index in six categories based on abrupt changes in the histogram. Classes are 0-permanent snow or ice; 1-water bodies, 1-very low; 2-low; 3-moderate; 4-high; and 5-very high susceptibility.

Here, the resulting susceptibility map had values between 0.173 and 0.581, the mean was 0.334, and the standard deviation was 0.0515. Susceptibility was reclassified in ranks based on mean and standard deviation intervals to express relative susceptibility to the occurrence of shallow landslides (Fig. 7.6).

Susceptibility classes were very low, low, moderate, high, and very high. The moderate class is centred around the mean value. Table 7.4 shows the classification.

The categorized susceptibility map was compared to the landslide inventory map, to calculate the area of landslides in each class. Table 7.4 shows area of

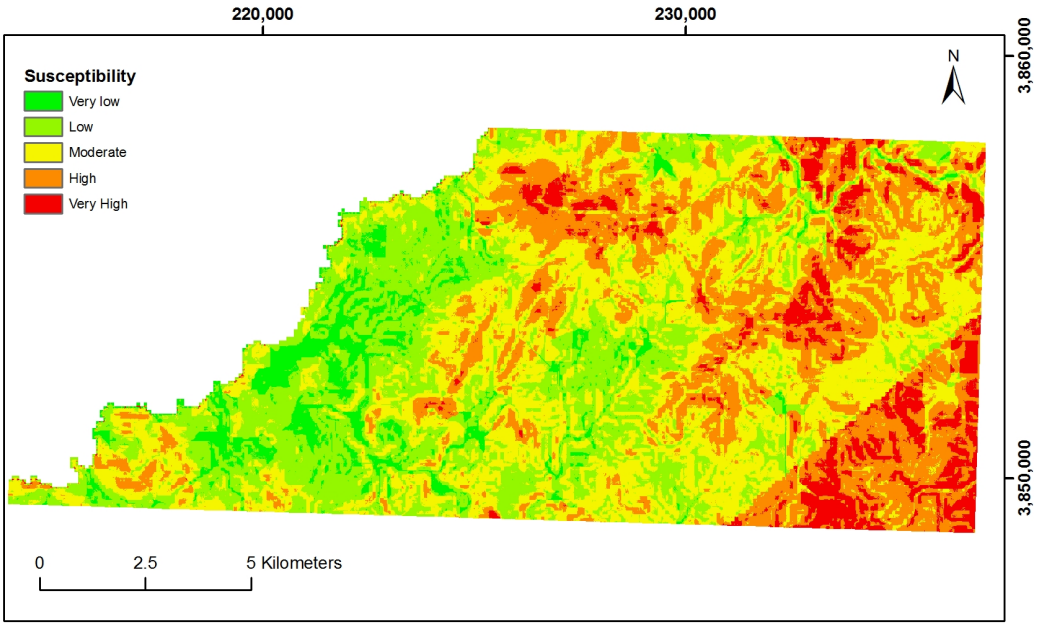

Figure 7.6: Shallow landslide susceptibility based on WLC

Table 7.4: Susceptibility classes

Class	Susceptibility	Category	Class area (km^2)	Landslide area (km^2)
1	0.173-0.257	Very low	9.43	0.14
2	0.257-0.308	Low	41.23	1.02
3	0.308-0.360	Moderate	54.69	1.37
4	0.360-0.411	High	39.35	0.80
5	0.411-0.581	Very high	11.49	0.11

failure based on number of cells.

Comparing the WLC model (Fig. 7.6) with the distribution of slope failures in the target area (Fig. 6.3), high and very high susceptibility classes are mainly in the eastern part of the target area. However, in the landslide inventory map, the highest failure concentration occurred in the western part of the target area.

7.5 Model validation

Validation consisted in ranking susceptibility, combining susceptibility classes with the landslide map and calculation of ROC plot and AUC. The susceptibility map was reclassified, based on the mean the standard deviation intervals, in 25 classes, from higher to lower susceptibility values. The tables resulting from the combination operation had the following information: area of susceptibility class and area of landslides within each susceptibility class.

Tables were joined and imported into a spreadsheet. Cumulative landslide area for each successive susceptibility class was estimated to calculate specificity using Eq. 5.3 and sensitivity using Eq. 5.2 and plot the ROC. AUC was calculated using Eq. 5.4.

The AUC of the model was 0.47. The result means that for the target area, the efficiency or prediction power of the model proposed by Hong et al. (2007) is lower than classification by chance. This indicates that it is not a good model to predict the occurrence of shallow landslides in the target area.

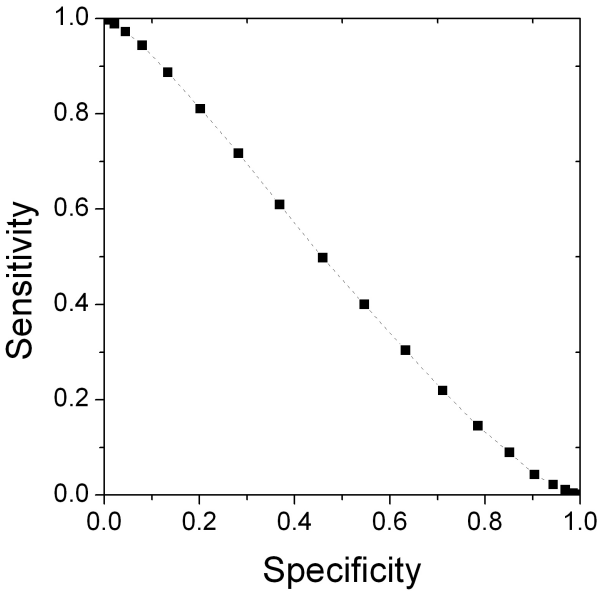

Figure 7.7: ROC plot of WLC model. AUC 0.47

Chapter 8

Weights of Evidence

This chapter is about applying the Weights of Evidence approach for the estimation of landslide susceptibility in the target area, using the same parameters used in the WLC model. The aim is to improve the accuracy of the WLC model by two approaches, using parameter weights calculated from practical data and replacing the elevation dataset with a higher resolution DEM.

Model validation requires the partition of the landslide database for the modelling and validation stages.

The general steps of the WOE approach were calculation of weights for parameter classes using the modelling landslide dataset, reclassification of parameters based on weights if necessary, combination of parameter weights and estimation of susceptibility, ranking susceptibility, comparison of susceptibility map with validation landslide dataset, and assessment of the model accuracy.

8.1 Landslide database partition

Based on the random partition approach, the landslide distribution database (Pimiento and Yokota, 2006) was divided in two sets. One for modelling and the other for validation. Furthermore, based on the data type, two different partitions were carried out; one based on landslide area (pixels) and the other based on point data.

For the raster based approach, a random raster with the same extent and cell size as the landslide inventory map was created. Then it was overlaid to

the landslide map to assign a probability value to each pixel. The landslide map was reclassified in two categories, pixels with values equal or lower than 0.5 were selected as the modelling dataset, and pixels with values greater than 0.5 were defined as validation dataset. The total number of pixels was 34,337. Table 8.1 and Fig. 8.1 show the classification.

Table 8.1: Area-based landslide database partition

Probability	Class	Dataset	Pixel count
0.00001-0.5	1	Modelling	17,202
0.5-0.99999	2	Validation	17,135

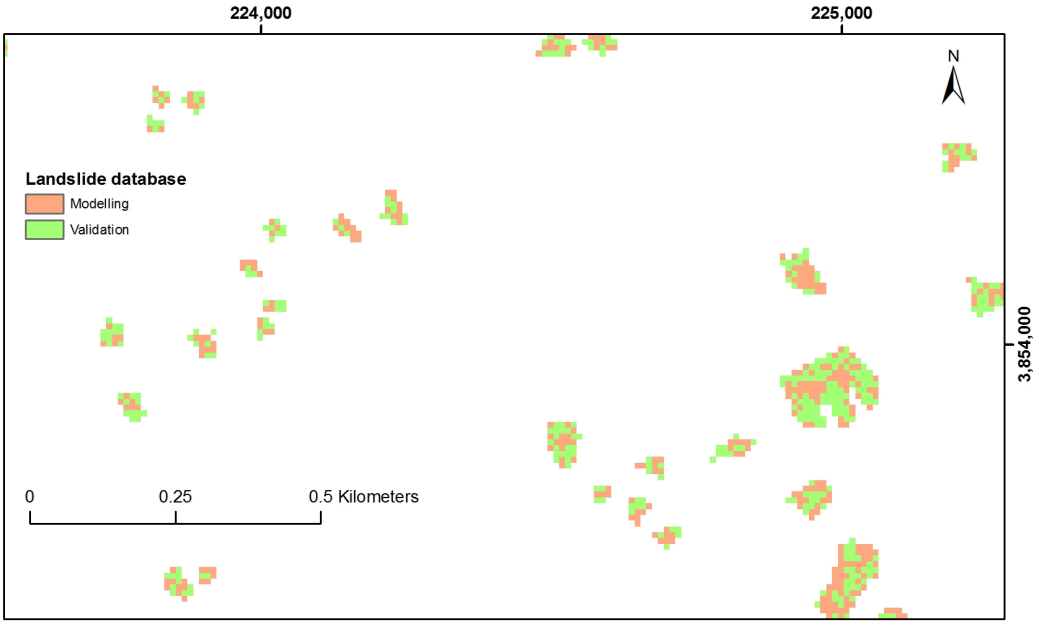

Figure 8.1: Area-based landslide database partition. Individual pixels in part of the target area

For the second partition, the landslide database was transformed to polygon data and it was randomly divided in a modelling (1,205 landslides) and validation (1,206 landslides) sets, using the Sampling Design Tool for ArcGIS (Center for Coastal Monitoring and Assessment, 2009).

Then, the polygon data were transformed to point data using ET Geowizards (ETGeoWizards, 2009), with the function label inside. The function calculates the centroid of the polygon, but if the centroid falls outside the polygon, the point is placed somewhere inside the polygon. A landslide was represented as a point of the centre of failure area, Fig. 8.2.

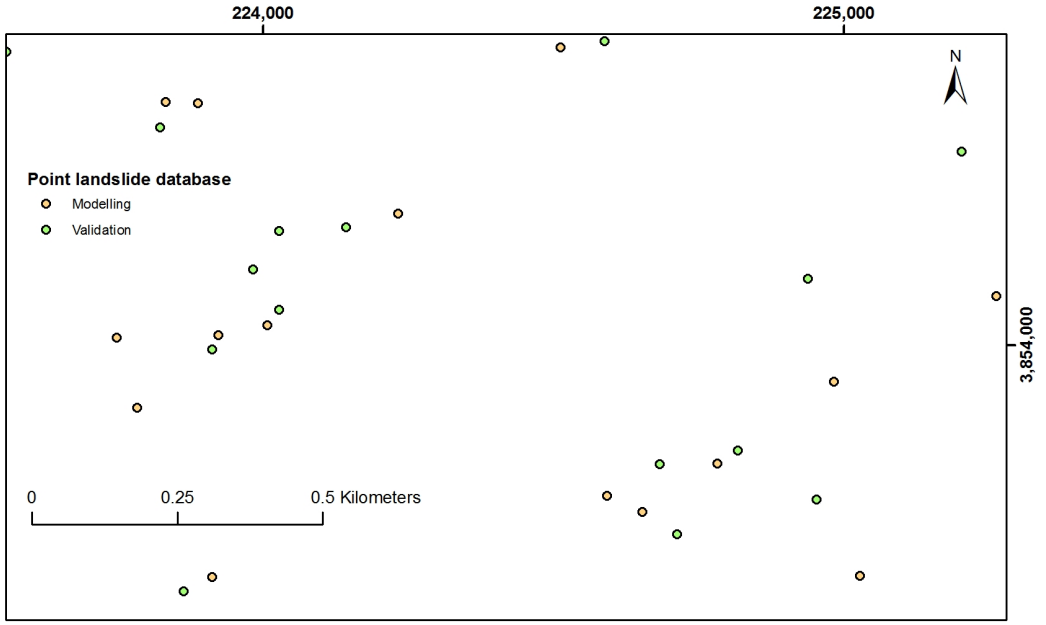

Figure 8.2: Point-based landslide database partition. Individual landslides in part of the target area

The area based partition datasets were mainly used for WOE modelling and validation. The point based partition datasets were used to compare results from using a smaller modelling dataset.

8.2 Parameter class weights 90 m DEM dataset

WOE requires the classification of continuous parameters. DEM was reclassified into 50-m interval classes, slope gradient data were categorized in 5-degree intervals and drainage density data were categorized in six classes based on mean and 1-standard deviation interval method.

Land cover and soil type were used with same classification (different from susceptibility) as in the WLC model. Soil texture map consisted of only one class, therefore it was not analysed. With a parameter with only one class, it is not possible to make a cross-area tabulation for the calculation of weights. However, the omission does not change the susceptibility, because susceptibility is relative.

Parameter maps were compared to the modelling landslide dataset (using the Sample function of ArcGIS). Results were tables of overlapping areas between landslide cells (landslide area) and parameter classes.

The total number of landslide cells was 34,337. Two cells fell outside the drainage density map and were classified as no-data; therefore only 34,335 cells were used in the analysis.

Table data were summarized to calculate the number of landslide cells within each class of the parameter maps.

Positive (W^+) and negative (W^-) weights for classes were calculated with Eq. 3.3 and Eq. 3.4. Using the weights and contrast, parameters were reclassified, and the weights were recalculated.

Slope gradient cells with values $> 35°$ were classified in one class. W^+ were negative for slope gradient classes $< 10°$ and $> 25°$. W^+ were positive for classes $> 10°$ and $< 25°$. This means that slope gradients $> 10°$ and $< 25°$ were favourable for the occurrence of landslides. Table 8.2 shows parameter categorization, positive and negative weights and contrast.

Table 8.2: 90 m DEM derived slope gradient classification and WOE weights

Slope gradient	W^+	W^-	Contrast
0 - 5	-0.358	0.055	-0.413
5 - 10	-0.034	0.013	-0.047
10 - 15	0.101	-0.034	0.135
15 - 20	0.180	-0.046	0.226
20 - 25	0.178	-0.018	0.196
25 - 30	-0.347	0.009	-0.356
30 - 35	-1.417	0.009	-1.426
35 - 65	-2.543	0.004	-2.547

For fine texture soil, W^+ was positive, while for medium/fine texture W^+ was negative. This indicates that medium/fine texture soil has lower susceptibility to landslides than fine texture soil, contrary to the criterion by Hong et al. (2007) (Table 8.3).

Table 8.3: Soil type classification and weights

Soil type	W^+	W^-	Contrast
Fine	0.0875	-1.730	1.817
Medium/fine	-1.7300	0.087	-1.817

Elevation classes between 600 and 700 m were grouped in one class. W^+ were negative for elevation classes lower than 100 and higher than 350 m. While elevation classes between 100 and 350 m had positive W^+, indicating higher landslide susceptibility (Table 8.4).

Table 8.4: 90 m DEM elevation classification and WOE weights

Elevation	W^+	W^-	Contrast
0-50	-0.707	0.054	-0.761
50-100	-0.143	0.016	-0.159
100-150	0.394	-0.055	0.449
150-200	0.241	-0.041	0.282
200-250	0.196	-0.024	0.219
250-300	0.325	-0.046	0.371
300-350	0.329	-0.039	0.368
350-400	-0.457	0.040	-0.497
400-450	-0.787	0.057	-0.844
450-500	-0.550	0.019	-0.570
500-550	-0.115	0.002	-0.117
550-600	-0.418	0.002	-0.420
600-700	-1.123	0.002	-1.124

Land cover classes, except for the Mixed forest, had negative W^+, indicating that they were not favourable for the occurrence of landslides. While Mixed forest

Table 8.5: Land cover classification and WOE weights

Land cover	W^+	W^-	Contrast
Permanent wetland	-1.098	0.008	-1.107
Deciduous needleleaf forests	-1.545	0.006	-1.551
Mixed forests	0.033	-0.949	0.982
Woody savannah	-0.737	0.008	-0.745
Cropland/natural vegetation mosaic	-0.478	0.003	-0.481
Barren or sparsely vegetated land	-1.353	0.007	-1.360

had a low positive W^+ and higher negative W^-, indicating higher susceptibility, Table 8.5.

Higher drainage density and low-density classes had negative W^+. Therefore, their relation to the occurrence of landslides was not clear, Table 8.6.

Table 8.6: 90 m DEM derived drainage density classification and WOE weights

Drainage density	W^+	W^-	Contrast
Very high	-0.218	0.015	-0.233
High	0.074	-0.015	0.089
Moderate	-0.015	0.016	-0.031
Low	0.064	-0.014	0.078
Very low	0.150	-0.007	0.157
Extremely low	-0.399	0.009	-0.408

8.2.1 Reclassification based on weights and contrast

As discussed in Section 3.4, the contrast helps to reclassify parameters in more significant classes. A positive contrast indicates the class is favourable for the occurrence of landslides, while a class with negative contrast is not.

Parameter maps were reclassified based on the WOE contrast value, except for soil type. Then, new weights were calculated using the modelling dataset, Table 8.7.

Table 8.7: Parameter reclassification and WOE weights

Slope gradient	W^+	W^-	Contrast
$> 10°$ and $< 25°$	0.1431	-0.1773	0.3205
$< 10°$ and $> 25°$	-0.1773	0.1431	-0.3205
Soil type			
Fine	0.0875	-1.730	1.817
Medium/fine	-1.7300	0.087	-1.817
Elevation	W^+	W^-	Contrast
100 - 350 m	0.2963	-0.4679	0.7642
< 100 and > 350 m	-0.4679	0.2963	-0.7642
Land cover			
Mixed forest	0.0332	-0.9492	0.9824
Other classes	-0.9492	0.0332	-0.9824
Drainage density			
Very high - moderate	0.0376	-0.0124	0.0500
Low - very low	-0.0124	0.0376	-0.0500

8.3 Parameter weights 50 m DEM dataset

In order to evaluate the significance of the elevation data used in the model by Hong et al. (2007), the 90 m DEM was replaced with a 50 m DEM from the Geographical Survey Institute of Japan (Geographical Survey Institute Japan, 1997).

8.3.1 Processing of 50 m DEM

Processing consisted in the transformation of elevation data to shapefile point data format using the software SMAPCNV (PASCO Co., 1997); inverse weighted distance interpolation; projection to UTM 53N Zone; and resampling using the same cell size as the landslide database.

Elevation values in target area were between 0 and 700 m. Elevation was re-classified in 50-m intervals classes. Only main land elevation data were considered for the analysis, small islands were not included (Fig. 8.3).

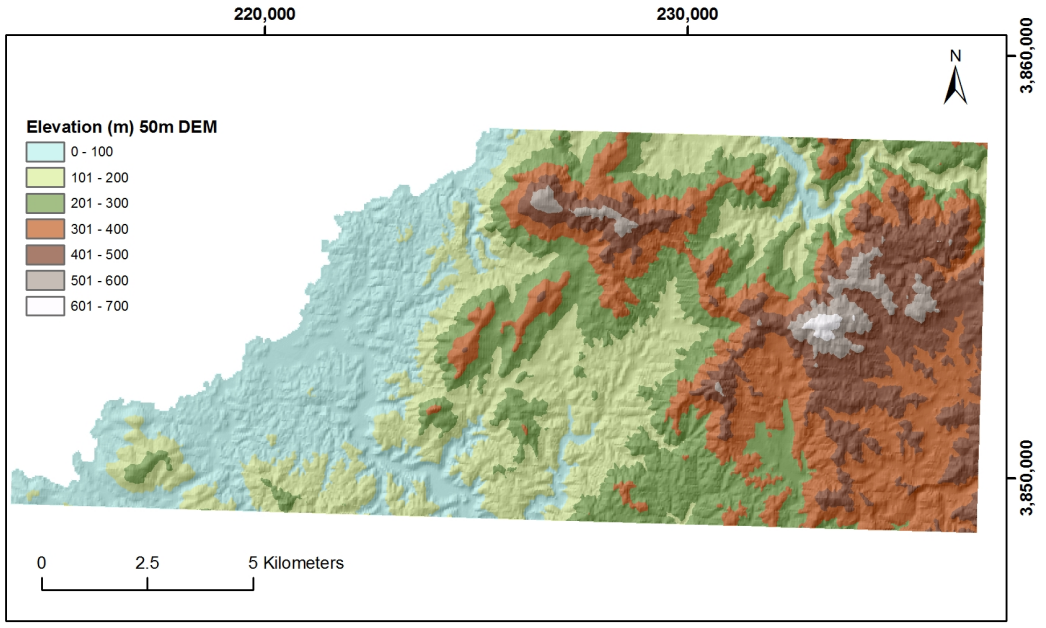

Figure 8.3: Elevation based on 50 m DEM. Source Geographical Survey Institute Japan (1997)

Slope gradient and drainage density were derived from the 50 m DEM, using the same approach used for the 90 m DEM data, Section 7.3.1.1 and Section 7.3.1.2. Resulting slope values, between 0° and 60°, were reclassified in 5-degree intervals (Fig 8.4).

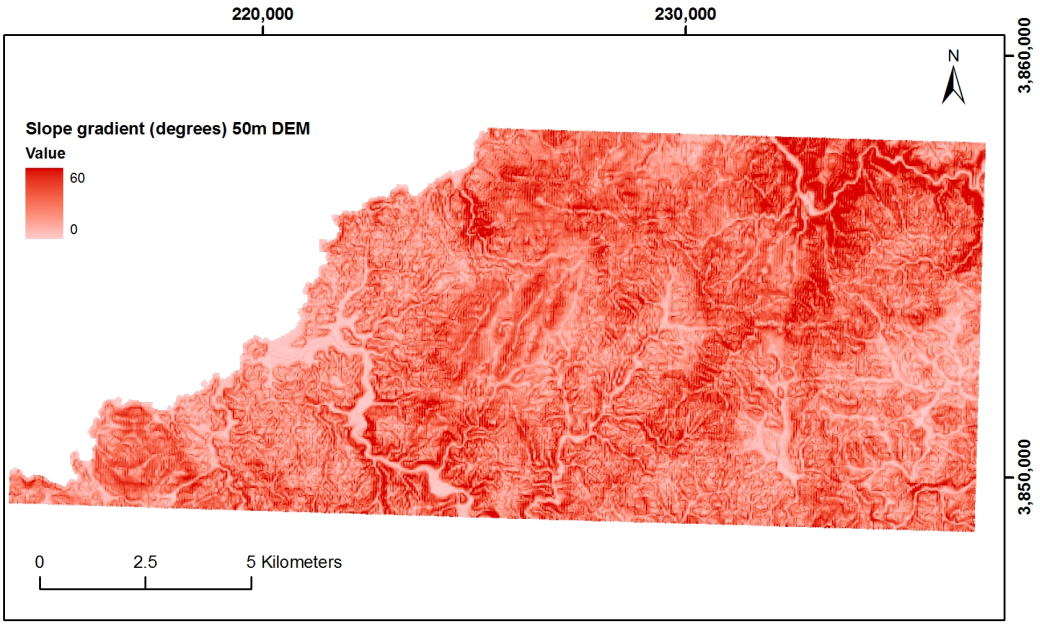

Figure 8.4: Slope gradient derived from 50 m DEM

Drainage density values were between 60 and 47,140 m/km^2. Values greater than 3,000 m/km^2 were grouped in one class. No data cells were assigned the minimum density (60.4 m/km^2). Finally, drainage density was classified in five categories, based on mean and 1 standard deviation values, as very high, high, moderate, low and very low (Fig 8.5).

The soil type and land cover data were clipped to the extension of the 50 m DEM.

8.3.2 Parameter weights

Weights were calculated for the initial categorization of the parameters derived from 50 m DEM. However, many slope and elevation classes did not have land-

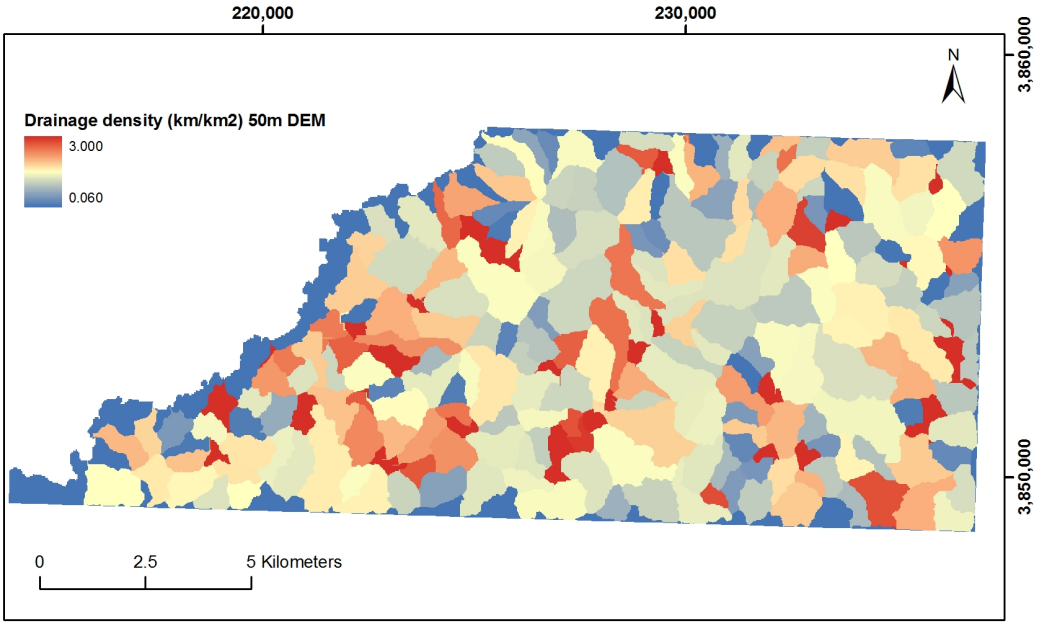

Figure 8.5: Drainage density derived from 50 m DEM

slides, therefore they were reclassified based on contrast values and new weights were calculated. Tables 8.8, 8.9, and 8.10 show the classifications and weights.

The relations between susceptibility and slope gradient and elevation were similar to those from the previous dataset. Slope gradients between 15° and 35°, and elevations between 100 and 350 m were favourable for the occurrence of landslides. Relation between drainage density and susceptibility was not clear.

Weights for soil type and land cover changed in relation to the 90 m DEM data scenarios, because the extent of the target area varied slightly, Tables 8.11 and 8.12.

8.4 Landslide susceptibility modelling

Modelling consisted in estimating susceptibility with weights calculated for three sets of parameters (called here as 90 m DEM dataset, reclassified 90 m DEM dataset, and 50 m DEM dataset) and 11 parameter combinations, to evaluate the significance of parameters, Table 8.13 shows parameter combinations. This

Table 8.8: 50 m DEM derived slope gradient classification and WOE weights

Slope gradient	W^+	W^-	Contrast
0 - 5	-1.1509	0.0695	-1.2204
5 - 10	-0.4968	0.0755	-0.5722
10 - 15	-0.1445	0.0350	-0.1795
15 - 20	0.1982	-0.0580	0.2562
20 - 25	0.3591	-0.0832	0.4423
25 - 30	0.3774	-0.0474	0.4248
30 - 35	0.2076	-0.0112	0.2188
35 - 40	-0.0858	0.0017	-0.0875
40 - 65	-0.6385	0.0047	-0.6432

Table 8.9: 50 m DEM derived elevation classification and WOE weights

Elevation	W^+	W^-	Contrast
0-50	-0.8161	0.0726	-0.8886
50-100	-0.0728	0.0085	-0.0814
100-150	0.3563	-0.0511	0.4074
150-200	0.3262	-0.0557	0.3819
200-250	0.2095	-0.0255	0.2351
250-300	0.3229	-0.0444	0.3673
300-350	0.2816	-0.0319	0.3136
350-400	-0.5432	0.0463	-0.5895
400-450	-0.7032	0.0480	-0.7512
450-500	-0.4389	0.0149	-0.4538
500-550	-0.2111	0.0034	-0.2145
550-600	-0.3455	0.0017	-0.3472
600-700	-1.2330	0.0017	-1.2347

Table 8.10: 50 m derived drainage density classification and WOE weights

Drainage density	W^+	W^-	Contrast
Very high	-0.4447	0.0429	-0.4876
High	-0.2281	0.0305	-0.2585
Moderate	0.2062	-0.2679	0.4741
Low	-0.1844	0.0375	-0.2219
Very low	-0.3255	0.0201	-0.3456

Table 8.11: Soil type WOE weights for 50 m DEM analysis

Soil type	W^+	W^-	Contrast
Fine	0.0861	-1.718	1.804
Medium/fine	-1.7175	0.086	-1.804

Table 8.12: Land cover WOE weights for 50 m DEM analysis

Land cover	W^+	W^-	Contrast
Permanent wetland	-1.4134	0.0132	-1.4265
Deciduous needleleaf forests	-1.6081	0.0061	-1.6142
Mixed forests	0.0408	-1.0774	1.1183
Woody Savannah	-0.7282	0.0081	-0.7362
Cropland/natural vegetation mosaic	-0.7315	0.0050	-0.7365
Barren or sparsely vegetated land	-1.3574	0.0073	-1.3646

resulted in 33 models.

Table 8.13: WOE parameter combinations

Combination	Parameters
A	Slope gradient, soil type, elevation, land cover, drainage density
B	Slope gradient, soil type, elevation, and land cover
C	Slope gradient, soil type, and land cover
D	Slope gradient, elevation, and land cover
E	Slope gradient, elevation, land cover and drainage density
F	Slope gradient, and elevation
G	Slope gradient, and land cover
H	Slope gradient and drainage density
I	Slope gradient and soil type
J	Slope gradient
K	Elevation

Estimation of landslide susceptibility consisted in the reclassification of the parameters maps with the W^+ weights and the calculation of the posterior probability, as described in Section 3.4.

This is the calculation of susceptibility for model WOEA1, the model with parameter combination A and the first dataset, 90 m DEM dataset. Susceptibility was calculated as follows:

Calculation of prior probability, $p(l)$, using Eq. 3.5, with areas in km^2,

$$p(l) = \frac{1.72}{157.06} = 0.011 \tag{8.1}$$

Estimation of $logit(l)$ with Eq. 3.6

$$logit(l) = \ln[\frac{0.011}{(1-0.011)}] = -4.4988 \tag{8.2}$$

For each cell in the target area, calculation of posterior logit, $logit(l|f)$, was the addition of rasters maps corresponding to the class weights for each parameter plus $logit(l)$; using Eq. 3.7.

$$logit(l|f) = -4.4988 + W_f^+ \tag{8.3}$$

Posterior logit $logit(l|f)$ was convert to posterior odds, $O(l|f)$ using Eq. 3.8

$$O(l|f) = \exp(logit(l|f)) \tag{8.4}$$

Finally, posterior probability $p(l|f)$ was calculated using Eq. 3.9

$$p(l|f) = \frac{O(l|f)}{(1 + O(l|f))} \tag{8.5}$$

The calculated posterior probability represents the landslide susceptibility based on the weights of the model parameters. Likelihood of occurrence of landslides is higher in cells with higher susceptibility. Similarly, low susceptibility cells indicate low probability to the occurrence of shallow landslides (Fig. 8.6).

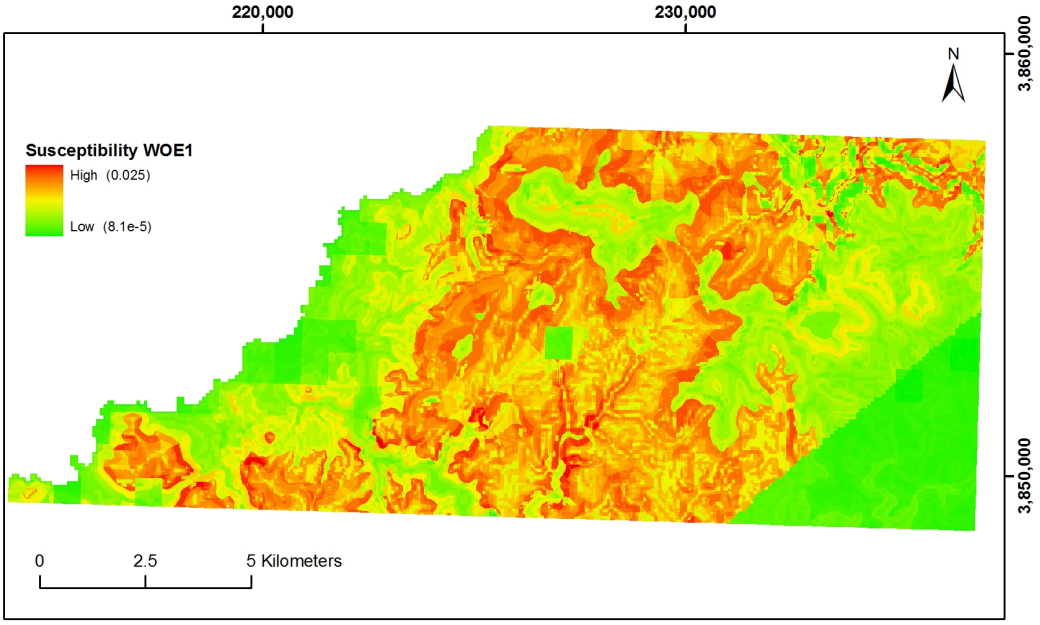

Figure 8.6: Landslide susceptibility model WOEA1. Posterior probability expresses landslide susceptibility from the model parameters

Susceptibility for all the parameter combinations and the three datasets was calculated as described above.

8.5 Model validation

The same procedure used to validate the WLC model was used to validate the WOE models. Susceptibility maps were classified in ranks based on mean value and 1/4 standard deviation intervals, and then compared to the landslide validation dataset. Reclassified maps did not have the same number of intervals.

Combination of susceptibility ranks and modelling dataset produced a table of landslide area within each rank. The ROC plot and AUC were calculated according to method described in Section 5.3 ROC plot and AUC.

Fig. 8.7 shows the ROC plot for the WOEA1 model, with the parameter combination A and the first dataset, 90 m DEM dataset. The AUC was 0.64.

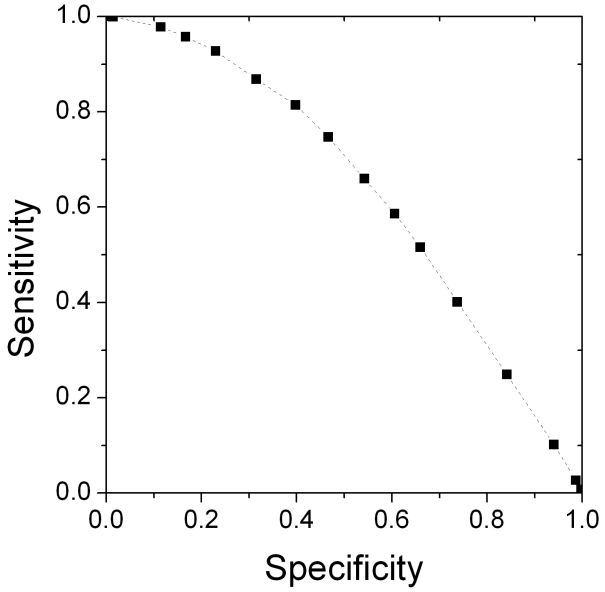

Figure 8.7: ROC plot of WOEA1 model. AUC 0.64

Table 8.14 shows the AUC calculated for the models, columns correspond to datasets and rows correspond to parameter combinations as describe earlier. The table permits to compare model accuracy for different parameter combinations using the three datasets.

Models using parameters derived from the 50 m DEM had better prediction power that models using data derived from 90 m DEM. Nevertheless, the dif-

ference was not big for some combinations. Models with reclassified parameters derived from the 90 m DEM had lower accuracies, contrary to what expected.

Table 8.14: AUC values for WOE models

Parameter combination	90-m-DEM	Reclassified 90-m-DEM	50-m-DEM
A	0.64	0.39	0.66
B	0.63	0.39	0.65
C	0.59	0.43	0.63
D	0.62	0.41	0.64
E	0.62	0.41	0.65
F	0.62	0.41	0.64
G	0.56	0.45	0.61
H	0.56	0.46	0.62
I	0.58	0.43	0.63
J	0.55	0.46	0.61
K	0.60	0.41	0.60

8.6 Landslide database partition assessment

Two approaches were used to assess the landslide database partition. The WOEA1 model was compared to the modelling dataset to evaluate the model performance instead of prediction performance. The ROC plot AUC calculated for the validation was 0.64, the same area as the AUC calculated from comparing to the validation dataset. This indicates that the random partition based on landslide area produced representative datasets.

The second assessment consisted in evaluating the effect of the size of the landslide modelling dataset in WOE modelling. Instead of using the raster-based validation dataset of 17,202 cells (or samples), the point-based validation dataset of 1,205 points was used.

The 50 m DEM data parameters were used for modelling. Parameter weights calculated using the validation dataset are in Tables 8.15, 8.16, 8.17, 8.18, and 8.19. Relations between susceptibility and parameter classes were similar to those estimated using the large raster-based dataset.

Slope gradient classes between 15° and 40° had positive W^+, while other classes had negative W^+. Fine texture soil type had positive W^+ and medium/fine texture soil type had negative W^+. Mixed forest land cover class had positive W^+, while other classes had negative W^+; however, elevation classes between 50 and 300 m and above 550 had positive W^+; while for the raster-based dataset estimation, all elevations classes above 300 m had negative W^+.

Here also, the relation between drainage density and susceptibility was not clear.

Table 8.15: Point-based partition slope gradient classification and WOE weights

Slope gradient	W^+	W^-	Contrast
0 - 5	-0.9288	0.0612	-0.9900
5 - 10	-0.2754	0.0469	-0.3223
10 - 15	-0.0528	0.0135	-0.0663
15 - 20	0.1644	-0.0471	0.2115
20 - 25	0.2271	-0.0485	0.2756
25 - 30	0.1882	-0.0212	0.2095
30 - 35	0.2298	-0.0126	0.2424
35 - 40	0.1571	-0.0036	0.1607
40 - 65	-0.3843	0.0031	-0.3875

Landslide susceptibility (model WOE45) was calculated with previous parameters and W^+ weights. Then it was validated using the validation dataset.

The ROC plot AUC for the model was 0.64; slightly lower than the AUC for the model using the same parameter combination (model WOEA3) and the large sampling dataset based on landslide area, that was 0.66. Nevertheless, the point based datasets represented well the characteristics of the landslides.

In general, the model produced using the smaller point-based modelling dataset produced acceptable accuracy, because the calculated weights expressed the parameter relations with the occurrence of shallow landslides.

Table 8.16: Point-based partition elevation classification and WOE weights

Elevation	W^+	W^-	Contrast
0-50	-0.4813	0.0499	-0.5312
50-100	0.0653	-0.0083	0.0736
100-150	0.4907	-0.0769	0.5676
150-200	0.3990	-0.0717	0.4707
200-250	0.0972	-0.0111	0.1084
250-300	0.2316	-0.0303	0.2619
300-350	-0.1687	0.0149	-0.1836
350-400	-0.6600	0.0530	-0.7130
400-450	-0.6590	0.0456	-0.7047
450-500	-0.4204	0.0144	-0.4347
500-550	-0.5754	0.0079	-0.5832
550-600	0.1330	-0.0008	0.1339
600-700	0.0469	-0.0001	0.0470

Table 8.17: Point-based partition drainage density classification and WOE weights

Drainage density	W^+	W^-	Contrast
Very high	-0.3285	0.0335	-0.3620
High	-0.3920	0.0479	-0.4399
Moderate	0.1986	-0.2568	0.4554
Low	-0.1167	0.0246	-0.1414
Very low	-0.3116	0.0193	-0.3309

Table 8.18: Point-based partition soil type classification and WOE weights

Soil type	W^+	W^-	Contrast
Fine	0.0745	-1.2456	1.3191
Medium/fine	-1.2446	0.0745	-1.3191

Table 8.19: Point-based partition land cover classification and WOE weights

Land cover	W^+	W^-	Contrast
Permanent wetland	-0.9444	0.0106	-0.9549
Deciduous needleleaf forests	-0.8131	0.0042	-0.8173
Mixed forests	0.0339	-0.7987	0.8326
Woody Savannah	-0.5185	0.0063	-0.5248
Cropland/natural vegetation mosaic	-0.3576	0.0029	-0.3605
Barren or sparsely vegetated land	-2.4527	0.0089	-2.4616

8.7 Overall conditional independence test

Conditional independence test was calculated for models with higher performance and different approaches, using Eq. 3.10. Table 8.20 shows the results. Model WOEA3 (parameter combination A and dataset 50 m DEM) presented the highest surplus. However, exceeding predicted events were below 15%, which means that conditional dependence was acceptable.

Table 8.20: Observed and predicted events

Model	Observed	Calculated	Exceeding (%)	AUC
WOEA1	17,133	17,959	4.8	0.64
WOEA3	17,133	19,077	11.3	0.66
WOE45	1,206	1,308	8.4	0.64

Chapter 9

Logistic Regression

This chapter is about producing a susceptibility model based on the characteristics of shallow landslides in the target area. Landslide susceptibility was estimated with the logistic regression approach using SDM, an extension for ArcGIS, Section 4.5.

The model parameters are generalized lithology, slope gradient, profile curvature, plan curvature, and elevation. Parameter selection was based on the mechanics of shallow landslides, as described earlier (Section 6.6).

The 10 m landslide inventory was used for modelling and validation. Parameters were derived from a 10 m DEM and 1:200,000 scale geological map.

9.1 Landslide and parameters databases

9.1.1 Landslide database

The landslide database is the 10 m grid landslide distribution map. However, the analysis requires training data as point data; therefore the point-based landslide database described in section 8.1 was used. Furthermore, SDM requires maximum 1,000 sample points for modelling. Therefore, the point database was reduced from 1,205 to 993 points, by random selection using Sampling Design Tool (Center for Coastal Monitoring and Assessment, 2009).

The validation database was produced in the same way. The point sample

dataset was reduced by randomly selecting 85% of the 1,206 points. The validation dataset consisted of 992 points.

9.1.2 10 m DEM

Elevation data was the 10 m DEM from the Geographical Survey Institute of Japan (GSI). DEM was produced produced from 1:25,000 topographic maps (Geographical Survey Institute Japan, 2009). For the target area, the elevation data is available for download in the Japanese version website since February 2009 at `http://fgd.gsi.go.jp/download/`.

The target area is in six 1:25,000 quadrangles; 5231-16; 5231-17; 5231-26; 5231-27; 5232-10, and 5232-20. The spatial reference is Japanese Geodetic Datum 2000 datum (JGD2000).

9.1.3 1:200,000 scale geological map of Japan

Geological map of Japan (scale 1:200,000) is an online geological database (GeoMapDB) from the Geological Survey of Japan (GSJ) (Geological Survey Japan, 2009). The target area is in two 1:200,000 quadrangles, Hamada and Mishima quadrangles. Quadrangles were downloaded from `http://iggis1.muse.aist.go.jp/ja/top.htm` as shapefiles. Spatial references of data are Japan Geodetic Datum 2000 (JGD2000) and UTM projection with JGD2000 datum.

The lithology distribution described in Section 6.2 was used for modelling.

9.2 Processing of parameters databases

Logistic regression analysis using SDM requires data projected in meters. Therefore the analyses were carried out with data projected in UTM 53N Zone with Tokyo datum.

Since the landslide map spatial reference was UTM, the elevation data and geological map data were processed. The general processing consisted in spatial reference transformation, clipping the study area and conversion to raster with same cell size as the landslide map, and categorization of continuous data.

9.2.1 10 m DEM

Processing consisted in transformation from XML format to BIL image using FGDDEMConv software tool (Akagi, 2009). The tool also permitted to make a mosaic of the files. Then, in ArcGIS (9.3) the image file was converted to integer raster. The spatial reference was defined as JGD2000, and the study area was clipped from the mosaic.

In the study area, elevations were between 0 and 712 m; the map was checked by comparing it with a paper topographic map of the area published by GSI, that was produced by photogrammetry. Elevation was reclassify in 50 m intervals, except for the class 600 to 712 m, Fig. 9.1.

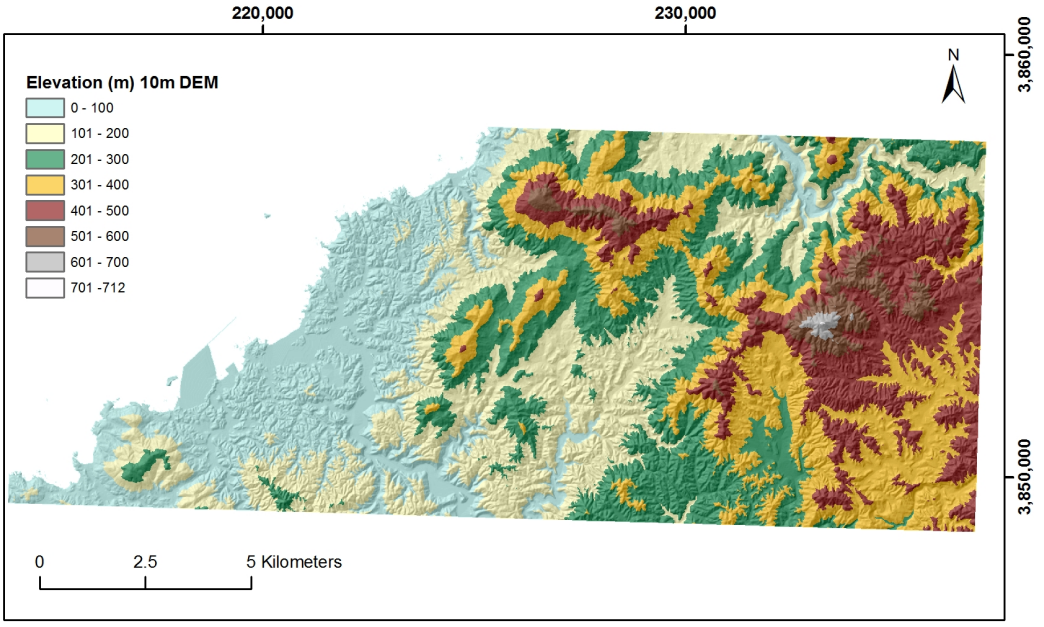

Figure 9.1: Elevation based on 10 m DEM. Source Geographical Survey Institute Japan (2009)

9.2.1.1 Slope gradient

Slope gradient, derived form the 10 m DEM as described previously (Section 7.3.1.1), was between 0° to 69° (Fig. 9.2). It was reclassified in 5 degree intervals, from 0° to 40° and the last except for gradients > 40°.

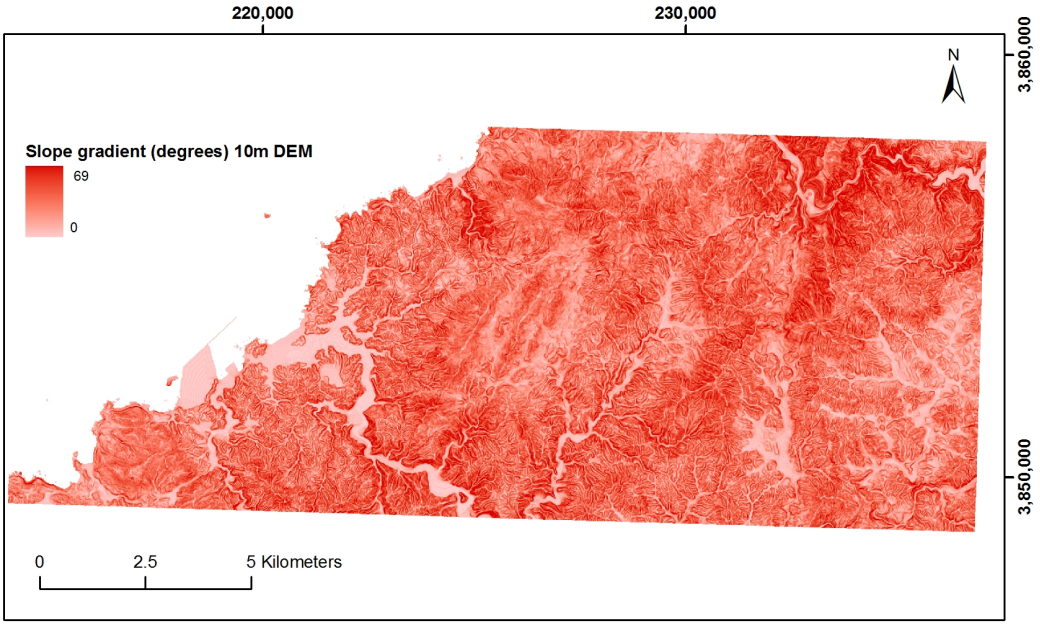

Figure 9.2: Slope gradient map derived from 10 m DEM

9.2.1.2 Profile and plan curvature

Curvature is a morphometric characteristic less known than slope or aspect. Therefore it is shortly described. Profile curvature is the terrain curvature in the slope direction. Plan curvature is the terrain curvature along a plane perpendicular to slope direction.

Profile and plan curvature were derived from the 10 m DEM using functions based on the methods by Zevenbergen and Thorne, as described by Burrough and MacDonnel (1998).

In ArcGIS, calculated negative profile curvature indicates an upwardly convex surface (hill). Positive profile curvature indicates an upwardly concave surface (depression). Positive plan curvature indicates a convex surface relative to elevation contour (nose). And negative plan curvature indicates a concave surface relative to contour (valley).

Profile curvature values varied between -22 and 26.7. Plan curvature values were between -24.77 and 20. Profile and plan curvature were reclassified by interactive visual inspection of data overlaid on elevation contour map, and based

on natural break values, in five qualitative classes. Classes were strongly concave, weakly concave, flat, weakly convex, and strongly convex. Table 9.1 and Fig. 9.3 show profile curvature classification and map. Table 9.2 and Fig. 9.4 show classification and map of plan curvature.

Table 9.1: Profile curvature classification

Value	Class	Gridcode
-22 - -2.5	Strongly convex	1
-2.5 - -0.6	Weakly convex	2
-0.6 - 1.1	Flat	3
1.1 - - 3.4	Weakly concave	4
3.4 - 26.7	Strongly concave	5

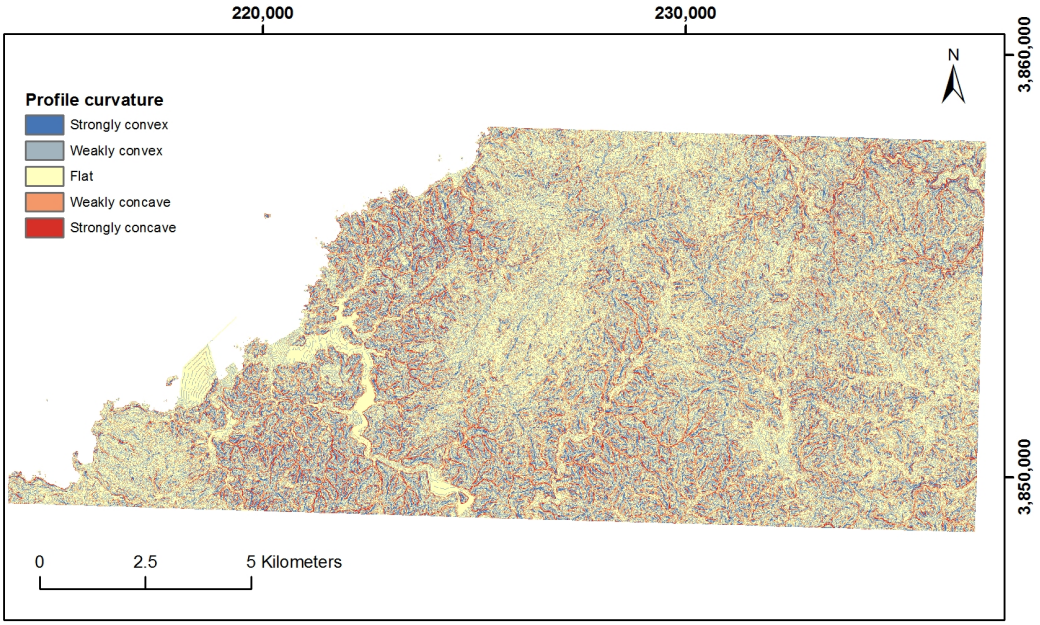

Figure 9.3: Profile curvature derived from 10 m DEM

9.2.2 Geological map of Japan

Downloaded data had shapefile format, and the spatial reference was JGD2000. Processing consisted in projecting to UTM 53N Zone with Tokyo datum, clip-

Table 9.2: Plan curvature classification

Value	Class	Gridcode
-24.8 - -3.5	Strongly concave	1
-3.5 - -1.1	Weakly concave	2
-1.1 - - 0.5	Flat	3
0.5 - - 2.4	Weakly convex	4
2.4 - - 20.0	Strongly convex	5

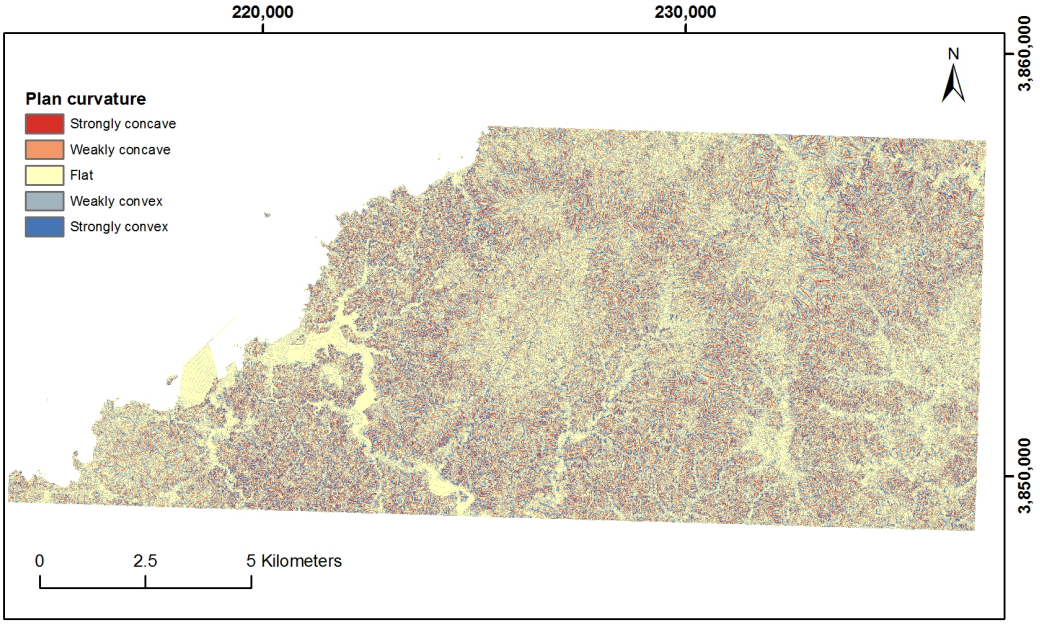

Figure 9.4: Plan curvature derived from 10 m DEM

ping the target area from each quadrangle, merging clipped areas, reclassifying lithologies, and dissolving units.

The geological map contained keys of lithology units based on the 1:200,000 scale geological map of Japan. A separate spreadsheet file contained the unit names in Japanese.

Lithology was reclassified according to lithofacies (rock type) without considering the age in: Quaternary deposits, volcanic and pyroclastic rocks, diorites and granitic rocks, and pelitic and psammitic schists. The lithology distribution of the target area is shown in Fig. 6.2. Table 9.3 shows classification and areas.

Generalized lithology map was converted to raster with same extent and cell size as elevation data, however lithology dataset had different extent; therefore, no data cells were assigned value -99 (a SDM requirement to identify missing information).

Table 9.3: Generalized lithology

Lithology	Area (km2)	Percentage
Quaternary deposits	1.43	0.91
Volcanic and pyroclastic rocks	50.43	32.04
Diorites and granitic rocks	43.37	27.55
Pelitic and psammitic schists	61.41	39.02
No-data	0.76	0.48
Total area	157.40	100.00

9.3 Landslide susceptibility modelling

After the definition of the model parameters, parameters were reclassified based on WOE approach using SDM, to maximize the association with shallow landslides. Then estimation of parameter weights, calculation of susceptibility for different parameter combinations, and validation of results.

9.3.1 Reclassification of parameters

Weights for parameter classes were calculated using the WOE tool from SDM with the point based modelling dataset. Calculation was based on categorical evidence type; descending or ascending type (used for proximity related evidence) were not appropriate. Weights and contrast indicate association with distribution of landslides, and are used as a guide for the categorization of parameters.

The SDM tool calculates the W^+ and W^- weights, contrast, their standard deviations, and the studentized contrast. Studentized contrast is the ratio of the contrast and the standard deviation of the contrast. A studentized contrast lower than 2.0 means an unacceptable contrast. It is a way to measure sampling confidence. Therefore, classes with a studentized contrast lower than 2.0 need to be reclassified.

Generalized lithology was reclassified by excluding Quaternary deposits. Volcanic and plutonic rocks had negative W^+ and contrast, while schists had positive W^+ and contrast, Table 9.4. These indicate that schists were more favourable for the occurrence of landslides than the other lithologies.

Table 9.4: Lithology reclassification and WOE weights

Generalized lithology	W^+	W^-	Contrast
Volcanic and pyroclastic rocks	-0.7142	0.2198	-0.9340
Diorites and granitic rocks	-0.1965	0.0669	-0.2634
Pelitic and psammitic schists	0.4353	-0.4415	0.8769

Slope gradient was reclassified in 5 degree intervals from 0° to 35° and the last class as >35°, Table 9.5. Similarly to the weights and contrast of lithology classes, classes of gradients below 20° were less favourable for the occurrence of landslides than classes with gradient >20°.

Profile curvature was reclassified by grouping the flat, weakly concave and strongly concave classes, Table 9.6. Strongly convex and weakly convex slopes had positive W^+; they were more favourable for the occurrence of landslides that flat to concave slopes.

Table 9.5: 10 m DEM derived slope gradient classification and WOE weights

Slope gradient	W^+	W^-	Contrast
0 - 5	-2.0084	0.0608	-2.0692
5 - 10	-1.0579	0.0625	-1.1204
10 - 15	-0.6406	0.0638	-0.7044
15 - 20	-0.2337	0.0347	-0.2684
20 - 25	0.1395	-0.0307	0.1702
25 - 30	0.3094	-0.0722	0.3816
30 - 35	0.5600	-0.1133	0.6733
35 - 69	0.2077	-0.0323	0.2400

Table 9.6: Profile curvature reclassification and WOE weights

Curvature	W^+	W^-	Contrast
Strongly convex	0.3987	-0.0409	0.4396
Weakly convex	0.1901	-0.0798	0.2699
Flat - concave	-0.1532	0.2399	-0.3931

Plan curvature was reclassified by grouping the flat to convex slopes, Table 9.7. Concave slopes had positive W^+; they were more favourable for the occurrence of landslides than flat to convex slopes.

Table 9.7: Plan curvature reclassification and WOE weights

Curvature	W^+	W^-	Contrast
Strongly concave	0.5158	-0.0274	0.5432
Weakly concave	0.2291	-0.0519	0.2810
Flat - convex	-0.0895	0.2902	-0.3797

Elevation was reclassified as shown in Table 9.8. The weights and contrast indicate that classes with elevation between 100 and 300 (with positive W^+) were favourable for the occurrence of landslides, while the other class were not.

Table 9.8: 10 m DEM derived elevation reclassification and WOE weights

Elevation	W^+	W^-	Contrast
0-100	-0.3156	0.0757	-0.3913
100-150	0.4863	-0.0750	0.5613
150-200	0.4437	-0.0829	0.5267
200-250	0.1954	-0.0239	0.2193
250-300	0.2019	-0.0260	0.2279
300-400	-0.3692	0.0703	-0.4394
400-450	-0.5665	0.0404	-0.6069
450-712	-0.3780	0.0226	-0.4005

9.3.2 Models

The logistic regression tool from SDM uses a unique condition table and it is limited to 6,000 unique conditions. The tool calculates the posterior probability and creates tables of the parameter coefficients.

Susceptibility was estimated for different parameter combinations. Table 9.9 shows the parameters combinations, lithology refers to generalized lithology. The training dataset consisted of point database. Fig. 9.5 shows landslide susceptibility calculated using LR1 with parameter classification and coefficients according to Table A.1. The other models were with parameter reclassification based on WOE, Tables A.2, A.3, A.4, A.5, and A.6.

Table 9.9: Logistic regression model parameter combinations

Model	Parameters
LR1	Lithology, slope gradient, profile curvature, plan curvature, elevation Parameters reclassified based on WOE
LR2	Lithology, slope gradient, profile curvature, plan curvature, elevation
LR3	Slope gradient, profile curvature, plan curvature, elevation
LR4	Lithology, slope gradient, elevation
LR5	Lithology, slope gradient, profile curvature, elevation
LR6	Lithology, slope gradient, plan curvature, elevation

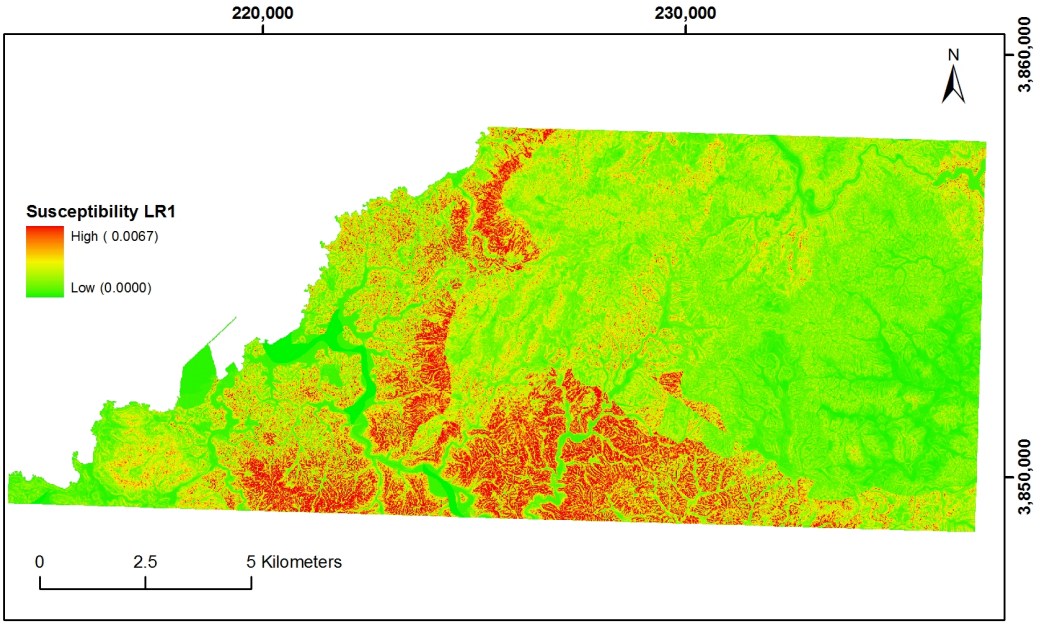

Figure 9.5: Landslide susceptibility calculated using LR1. The susceptibility pattern is similar to the slope failure distribution in Fig. 6.3

The coefficients express the relative importance of the parameter classes. A positive coefficient indicates that the class contributes to susceptibility, while a negative coefficient class reduces landslide susceptibility. Coefficients near zero indicate little relation to susceptibility and coefficients equal to zero indicate the class is not influencing the susceptibility. In Appendix A, Tables A.1, A.2, A.3, A.4, A.5, and A.6 show the coefficients calculated for the models.

For the LR1 model, lithology classes had positive and high values, except for Quaternary deposits where no landslides occurred, and schists the highest value. Therefore, schists had a high susceptibility for the occurrence of landslides. Other parameter classes contributing to susceptibility were slope gradients between 20° and 35°, convex profile curvature, concave plan curvature, and elevations between 100 and 300 m. However, classes with coefficients equal to zero correspond to classes with studentized contrast lower than two, and were not considered by the analysis.

Models LR2, LR3, LR4, LR5 and LR6 had the last parameter class with coefficient equal to zero. However, their studentized contrast were acceptable in the

weight calculation. Schists should be significant for the susceptibility estimation, while the other classes, slope gradient 35° - 69°, profile curvature flat - concave, plan curvature flat - convex, and elevation 450 - 712, are classes with little relation to the occurrence of landslides.

Bonham-Carter, who worked on the development of SDM, in a personal communication answered to a question regarding the coefficients with value of zero for the last classes. Bonham-Carter said that when the logistic regression tool of SDM is used with categorical data, each class is made into a new independent binary variable, and since the last class is determined by the state of the other classes (the presence of the other classes implies the absence of the last class), its coefficient turns out to be zero. The zero coefficients add nothing to posterior probabilities, and they can be ignored. In this way the logistic regression tool avoids conditional independence bias (Bonham-Carter, 2009).

Nevertheless, in general, model parameters had coefficients that expressed a direct relation to susceptibility; slope gradients between 25° and 35°; convex profile curvature; concave plan curvature; and elevations between 100 and 300 m. Volcanic and plutonic rocks had negative coefficients, which would mean lower susceptibilities than that for schists.

9.4 Model validation

All models were validated. The general steps were reclassification of susceptibility maps, combination with the validation dataset, and calculation of ROC plot AUC. Reclassification was based on dividing the susceptibility values in ten equal area zones. Ranking based on equal area zones and mean and standard deviation values produced the same results when calculating the ROC plot AUC.

The estimation of the ROC plot AUC was as described in Section 5.3. Fig. 9.6 shows the ROC plot of LR1 model. The AUC was 0.71. Table 9.10 shows ROC plot AUC for the models.

The model with the highest AUC was LR1, with all parameters before reclassification. The model with the same parameter combination (LR2), but with parameters reclassification based on weights had a slightly smaller AUC.

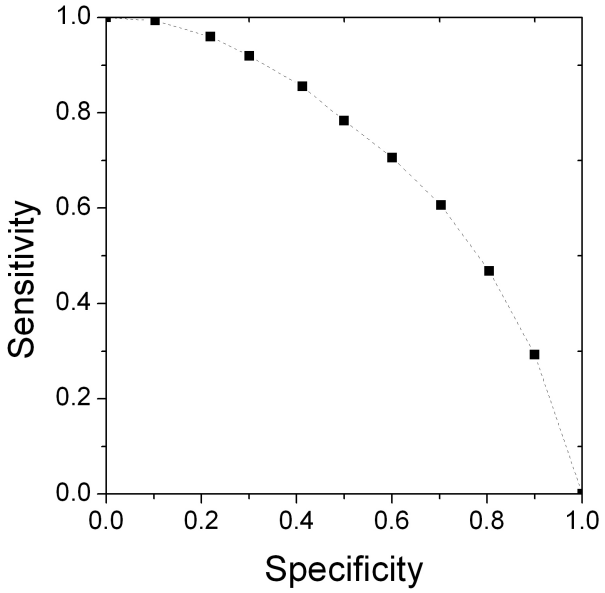

Figure 9.6: ROC plot of LR1 model. AUC 0.71

Table 9.10: AUC for logistic regression models

Model	ROC plot AUC
LR1	0.711
LR2	0.710
LR3	0.672
LR4	0.703
LR5	0.705
LR6	0.705

The model with the lowest prediction power is the one without including generalized lithology (LR3). This means that occurrence of shallow landslides in the target area depended on lithology. Replacing profile curvature with plan curvature produced the same accuracy (LR5 and LR6), and not including curvature produced a slightly lower accuracy (LR4).

Comparing the accuracy of all models, the LR models had higher accuracies than the WLC and WOE models.

Chapter 10

Discussion

Susceptibility to the occurrence of shallow landslides in the target area was estimated using WLC, WOE and LR approaches, and different parameter combinations. The models were validated and compared using the ROC plot AUC.

For validation, the database was randomly divided in modelling and validation datasets. However, using a time based partition should be ideal. Nevertheless, the random partition assessment indicated that the partition produced representative datasets. Furthermore, a point based database, consisting of the centroids of landslide source areas, represented adequately the landslide characteristics.

The accuracy of the WLC model proposed by Hong et al. (2007) was 0.47. The accuracy depends mainly on the model parameters, the definition of primary and secondary weights, and the spatial resolution of the datasets. Slope gradient is a fundamental parameter; however, the parameter weight does not express susceptibility for the occurrence of shallow landslides. Shallow landslides depend also on the availability of soil and colluvium, and in areas with slope gradient $> 35°$ soil and colluvium become less abundant (Crozier and Glade, 2005; Wieczorek, 1996). Therefore, very steep slopes do not necessarily have the highest susceptibility.

In the WLC model by Hong et al. (2007) the weight of soil characteristics is very high. Soil type is based on soil texture, but soil texture is also an other parameter. Therefore, texture is expressed by two parameters, and the combined secondary level weight (0.3) is the same as for slope gradient. However, the resolutions of soil type and texture dataset are very low compared to the landslides

in the target. The average size of the landslides in the target is 1,400 m^2 (Pimiento and Yokota, 2006). The low resolution of the datasets also produce abrupt changes in the estimated susceptibility. Like the soil type limit in the south-easter part of the target model, Fig. 7.6.

Drainage density was not a reliable model parameter. Estimated drainage density in the target area (Fig. 7.3) was very high in some catchment areas with unnatural shape. While other areas had nodata value cells. These conditions may also reduce the accuracy of the model.

Kirschbaum et al. (2009), assessing a system for global landslide monitoring based on the susceptibility model by Hong et al. (2007) and satellite-derived rainfall data, concluded that the system must be improved before its application. That the susceptibility model weighting is not correct, and contributions of slope gradient and soil conditions to landslide susceptibility should be considered regionally. Also, that the model requires higher resolution data and regional landslides inventories for calibration and validation.

Data driven weights express better the relationships between parameters and the occurrence landslides than the weights defined by Hong et al. (2007). This is because the occurrence of landslides depends on local conditions (van Westen et al., 2003).

The WLC weights and contrast showed that schists had higher susceptibility than plutonic and volcanic rocks; indicating that occurrence of landslides depends on lithology, as reported by Research Group of San-in Heavy Rainfall Disaster (1984) and Wada et al. (1984).

Slope gradients between 20° and 35° had also high susceptibilities; Research Group of San-in Heavy Rainfall Disaster (1984) and Okuda and Okimura (1984) reported high susceptibilities for 30° to 40° and 15° to 25° slopes respectively. In other shallow landslide analyses, slope gradients $> 35°$ also have low susceptibilities (Ahmad and McCalpin, 1999; Dai and Lee, 2002), as mentioned above.

Convex profile curvature and concave plan curvature had higher susceptibilities than other classes. In profile convex slopes the presence of residual soil may be higher; in plan concave slopes colluvium is more abundant (Turner, 1996) and concentration of run-off increases groundwater level promoting the occurrence of shallow landslides (Ayalew et al., 2004).

Elevation classes (between 100 and 300) with higher susceptibility indicate that occurrence of landslides depends on geomorphology and topography, as reported by Research Group of San-in Heavy Rainfall Disaster (1984) and Wada et al. (1984).

Drainage density class weights showed little relation to landslide susceptibility, as mentioned before. High density and low density classes had negative W^+ (Table 8.6).

In general, and contrary to the opinion of Bonham-Carter (1994), model parameter reclassification based on weights and contrast reduced the accuracy of WLC and WOE models. In LR models the difference was not significant.

The LR models had higher accuracies than the WLC and WOE models mainly for two reasons, the model parameters were defined considering the mechanics of the shallow landslides in the target area, and the higher spatial resolution of the DEM used. LR is a suitable multivariate statistical approach for estimating susceptibility. SDM is also a suitable tool for LR based modelling. However, SDM presented some limitations. Landslides (training data) are represented as points (maximum 1,000) and for large landslides the scarp centroid might not be enough to represent the scarp conditions. LR is based on a unique condition table of maximum 6,000 conditions. This limits the number of input parameters. More complex models require using external software for statistical analysis.

Quantitative modelling in the target area was based on the principles from Section 3.1 and the following assumptions. The estimated landslide susceptibility is for the occurrence of rainfall triggered shallow landslides. The rainfall distribution of the event that produced the landslides of the inventory was uniform, and future rainfall events also will have a uniform rainfall distribution. The validation landslides occurred after the modelling landslides. Terrain conditions that control the susceptibility, such as topography, availability of soil and colluvium, mechanic properties of soil, and hydrology, are static. However, the occurrence of landslides and erosion processes modify these conditions between triggering events. And, the model datasets, like elevation, land cover, and soil characteristics, correspond to the terrain conditions before the occurrence of the landslides.

Higher accuracy susceptibility models are required. For that it is necessary to use higher resolution elevation data, such as LIDAR data. For modelling it is

desired to have pre-event elevation dataset and use post-event elevation datasets for validation. An important limitation for quantitative modelling is availability of landslide inventories; therefore, the development of automatic detection of landslides is fundamental. Higher accuracy models should be produced by considering other parameters and other modelling approaches. Parameters that express the availability of soil and colluvium should significantly increase the accuracy of shallow landslide susceptibility models.

Chapter 11

Conclusions

In the target area, the accuracy of the qualitative susceptibility model was low (0.47). The main reasons for the low accuracy of the WLC model are the selection of parameters, the parameter weighting, and the low spatial resolution of the datasets. Landslide susceptibility in the target area depends on parameters and weights different from those of the WLC model. Two approaches raised the WLC model accuracy. Using parameter weights estimated from practical data produced a model with accuracy of 0.64, and replacing the 90 m DEM with a 50 m DEM produced a model with accuracy of 0.66.

The WOE parameter weights, estimated from the landslide distribution, expressed the relation between the parameters and susceptibility to the occurrence of landslides. In the target area, schists had higher susceptibility than plutonic and volcanic rocks. Slope gradients between 20° and 35° had higher susceptibility than other slope gradient classes. Convex profile curvature and concave plan curvature had higher susceptibilities than other classes. The susceptibility of elevation classes depended on local geomorphology. On the other hand, drainage density classes showed little relation to the occurrence of landslides. WOE base model had higher accuracy than the WLC model.

A quantitative susceptibility model based on LR had the highest accuracy (0.71). The definition of parameters was based on the characteristics of the landslides in the target area and the parameter weights were estimated from practical data. Shallow landslides triggered by rainfall greatly depend on topography, and higher resolution elevation produce higher accuracy models.

For medium-scale landslide susceptibility mapping, quantitative models produced better results than a qualitative model developed for global landslide susceptibility.

The parameter reclassification based on the WOE weight and contrast reduced the accuracy of WLC and WOE models, contrary to what expected. However, for the LR based models the accuracy decrease was not significant.

Validation of susceptibility models was fundamental to compare approaches and parameter combinations. The ROC plot AUC was a simple and objective approach for model validation. And random based database partition produced representative modelling and validation datasets.

Nevertheless, it is necessary to develop higher accuracy models. Using other approaches, higher spatial elevation data and other model parameters may produce better results.

Appendix A

Logistic Regression Coefficients

Table A.1: Logistic regression coefficients for model LR1

Factor	Coefficient
Constant Value	-16.927
Lithology	
Quaternary deposits	0.000
Volcanic and pyroclastic rocks	8.247
Diorites and granitic rocks	8.688
Pelitic and psammitic schists	9.310
Slope gradient	
0 - 5	-1.949
5 - 10	-1.130
10 - 15	-0.756
15 - 20	-0.354
20 - 25	0.018
25 - 30	0.181
30 - 35	0.420
35 - 40	0.120
35 - 69	0.000
Profile curvature	
Strongly convex	0.702
Weakly convex	0.594

Flat	0.312
Weakly concave	0.000
Strongly concave	-0.557
Plan curvature	
Strongly concave	0.931
Weakly concave	0.567
Flat	0.250
Weakly convex	0.000
Strongly convex	0.000
Elevation	
0 - 50	-0.222
50 - 100	0.000
100 - 150	0.399
150 - 200	0.566
200 - 250	0.359
250 - 300	0.378
300 - 350	0.000
350 - 400	-0.322
400 - 450	0.067
450 - 500	0.042
500 - 550	0.000
550 - 600	0.000
600 - 712	0.000

Table A.2: Logistic regression coefficients for model LR2

Factor	Coefficient
Constant Value	-6.826
Lithology	
Volcanic and pyroclastic rocks	-1.253
Diorites and granitic rocks	-0.702
Pelitic and psammitic schists	0.000
Slope gradient	
0 - 5	-1.873
5 - 10	-1.151
10 - 15	-0.807
15 - 20	-0.427
20 - 25	-0.069
25 - 30	0.083
30 - 35	0.322
35 - 69	0.000
Profile curvature	
Strongly convex	0.450
Weakly convex	0.365
Flat - concave	0.000
Plan curvature	
Strongly concave	0.667
Weakly concave	0.365
Flat - convex	0.000
Elevation	
0 - 100	-0.655
100 - 150	0.032
150 - 200	0.237
200 - 250	0.042
250 - 300	0.066
300 - 400	-0.210
400 - 450	-0.143
450 - 712	0.000

Table A.3: Logistic regression coefficients for model LR3

Factor	Coefficient
Constant Value	-8.049
Slope gradient	
0 - 5	-1.964
5 - 10	-1.165
10 - 15	-0.792
15 - 20	-0.385
20 - 25	-0.002
25 - 30	0.170
30 - 35	0.409
35 - 69	0.000
Profile curvature	
Strongly convex	0.550
Weakly convex	0.409
Flat - concave	0.000
Plan curvature	
Strongly concave	0.747
Weakly concave	0.403
Flat - convex	0.000
Elevation	
0 - 100	0.371
100 - 150	0.967
150 - 200	0.989
200 - 250	0.708
250 - 300	0.707
300 - 400	0.157
400 - 450	-0.074
450 - 712	0.000

Table A.4: Logistic regression coefficients for model LR4

Factor	Coefficient
Constant Value	-6.511
Lithology	
Volcanic and pyroclastic rocks	-1.290
Diorites and granitic rocks	-0.755
Pelitic and psammitic schists	0.000
Slope gradient	
0 - 5	-1.967
5 - 10	-1.190
10 - 15	-0.822
15 - 20	-0.429
20 - 25	-0.069
25 - 30	0.083
30 - 35	0.318
35 - 69	0.000
Elevation	
0 - 100	-0.700
100 - 150	0.000
150 - 200	0.214
200 - 250	-0.001
250 - 300	0.032
300 - 400	-0.239
400 - 450	-0.160
450 - 712	0.000

Table A.5: Logistic regression coefficients for model LR5

Factor	Coefficient
Constant Value	-6.660
Lithology	
Volcanic and pyroclastic rocks	-1.271
Diorites and granitic rocks	-0.734
Pelitic and psammitic schists	0.000
Slope gradient	
0 - 5	-1.932
5 - 10	-1.160
10 - 15	-0.799
15 - 20	-0.410
20 - 25	-0.054
25 - 30	0.093
30 - 35	0.326
35 - 69	0.000
Profile curvature	
Strongly convex	0.313
Weakly convex	0.256
Flat - concave	0.000
Elevation	
0 - 100	-0.670
100 - 150	0.022
150 - 200	0.229
200 - 250	0.026
250 - 300	0.054
300 - 400	-0.221
400 - 450	-0.148
450 - 712	0.000

Table A.6: Logistic regression coefficients for model LR6

Factor	Coefficient
Constant Value	-6.580
Lithology	
Volcanic and pyroclastic rocks	-1.282
Diorites and granitic rocks	-0.737
Pelitic and psammitic schists	0.000
Slope gradient	
0 - 5	-1.936
5 - 10	-1.194
10 - 15	-0.837
15 - 20	-0.448
20 - 25	-0.086
25 - 30	0.072
30 - 35	0.312
35 - 69	0.000
Plan curvature	
Strongly concave	0.508
Weakly concave	0.238
Flat - convex	0.000
Elevation	
0 - 100	-0.698
100 - 150	0.001
150 - 200	0.216
200 - 250	0.004
250 - 300	0.035
300 - 400	-0.236
400 - 450	-0.159
450 - 712	0.000

References

Agterberg, F. P., Bonham-Carter, G. F., Chen, Q. and Wright, D. F.: 1993, Logistic regression for mineral potential mapping, *in* J. C. Davis and U. C. Herzfeld (eds), *In Computers in geology: 25 years of progress*, Oxford University Press, New York, p. 298.

Ahmad, R. and McCalpin, J. P.: 1999, Landslide susceptibility maps for the Kingston Metropolitan Area, Jamaica with their notes on their use, Online `http://www.oas.org/cdmp/document/kma/udspub5.htm`.

Akagi, M.: 2009, FGDDEMConv, Online `http://space.geocities.jp/bischofia_vb/`.

Arthur, J., Wood, H., Baker, A., Cichon, J. and Raines, G.: 2007, Development and implementation of a bayesian-based aquifer vulnerability assessment in Florida, *Natural Resources Research* **16**(2), 93–107. 10.1007/s11053-007-9038-5.

Ayalew, L., Yamagishi, H., Marui, H. and 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. doi: DOI: 10.1016/j.enggeo.2005.08.004.

Ayalew, L., Yamagishi, H. and Ugawa, N.: 2004, Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano river, Niigata prefecture, Japan, *Landslides* **1**(1), 73–81. 10.1007/s10346-003-0006-9.

Begueria, S.: 2006, Validation and evaluation of predictive models in hazard assessment and risk management, *Natural Hazards* **37**(3), 315–329. 10.1007/s11069-005-5182-6.

Blue Marble: 2002, World map, Online `http://www.bluemarblegeo.com/products/worldmapdata.php?op=download`.

Bonham-Carter, G. F.: 1994, *Geographic Information Systems for Geoscientists: modelling with GIS*, Vol. 13 of *Computers Methods in Geosciences*, Pergamon Press.

Bonham-Carter, G. F.: 2009, Logistic regression coefficients, Personal communication.

Brenning, A.: 2005, Spatial prediction models for landslide hazards: review, comparison and evaluation, *Nat. Hazards Earth Syst. Sci.* **5**(6), 853–862.

Burrough, P. A. and MacDonnel, R. A.: 1998, *Principles of Geographical Information Systems*, Oxford University Press, Oxford.

Carrara, A., Cardinali, M., Guzzetti, F. and Riechenbach, P.: 1995, GIS technology in mapping landslide hazard, *in* A. Carrara and F. Guzzetti (eds), *Geographical information systems in assessing natural hazards*, Kluwer Academic Publishers, pp. 135–175.

Center for Coastal Monitoring and Assessment: 2009, Sampling design tool for ArcGIS, Online `http://ccma.nos.noaa.gov/products/biogeography/sampling/welcome.html`.

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

Chung, C.-J. F. and Fabbri, A. G.: 2003, Validation of spatial prediction models for landslide hazard mapping, *Natural Hazards* **30**(3), 451–472. 10.1023/B:NHAZ.0000007172.62651.2b.

Chung, C.-J. F. and Fabbri, A. G.: 2005, Systematic procedures of landslide hazard mapping for risk assessment using spatial prediction models, *in* T. Glade, M. Anderson and M. J. Crozier (eds), *Landslide Hazard and Risk*, Wiley, pp. 139–174.

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

Committee on the Review of the National Landslide Hazards Mitigation Strategy: 2004, *Paternship for Reducing Landslide Risk: Assessment of the National Landslide Hazards Mitigation Strategy*, The National Academy Press, Washington, D.C.

Crozier, M. J. and Glade, T.: 2005, Landslide hazard and risk: Issues, concepts and approach, *in* T. Glade, M. Anderson and M. J. Crozier (eds), *Landslide Hazard and Risk*, Wiley, pp. 1–40.

Cruden, D.: 1991, A simple definition of a landslide, *Bulletin of Engineering Geology and the Environment* **43**(1), 27–29. 10.1007/BF02590167.

Cruden, D. and Varnes, D. J.: 1996, Landslide types and processes, *in* A. Turner and R. Shuster (eds), *Landslides: investigation and mitigation*, National Academy Press, Washington, D. C., pp. 36–75.

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

ESRI: 2009, Hydro data model, Online `http://support.esri.com/index.cfm?fa=downloads.dataModels.filteredGateway&dmid=15`.

ETGeoWizards: 2009, Etgeowizards, Online `http://www.ian-ko.com/ET_GeoWizards/gw_main.htm`.

Fabbri, A. and Chung, C.-J.: 2008, On blind tests and spatial prediction models, *Natural Resources Research* **17**(2), 107–118. 10.1007/s11053-008-9072-y.

FAO: 2009, Digital Soil Map of the World (DSMW), Online http://www.fao.org/geonetwork/srv/en/metadata.show.

Fawcett, T.: 2006, An introduction to ROC analysis, *Pattern Recognition Letters* **27**(8), 861–874. doi: DOI: 10.1016/j.patrec.2005.10.010.

Fernandez, T., Irigaray, C., El Hamdouni, R. and Chacon, J.: 2003, Methodology for landslide susceptibility mapping by means of a GIS application to the Contraviesa area (Granada, Spain), *Natural Hazards* **30**(3), 297–308. 10.1023/B:NHAZ.0000007092.51910.3f.

Geographical Survey Institute Japan: 1997, Digital map 50 m grid (elevation). Nippon-III, CD ROM.

Geographical Survey Institute Japan: 2009, 10 m DEM, Online http://fgd.gsi.go.jp/download/.

Geological Survey Japan: 2009, Integrated geological map database (GeoMapDB), Online http://iggis1.muse.aist.go.jp/ja/top.htm.

Glade, T. and Crozier, M. J.: 2005, A review of scale dependency in landslide hazard and risk analysis, *in* T. Glade, M. Anderson and M. J. Crozier (eds), *Landslide Hazard and Risk*, Wiley, pp. 75–138.

Global Soil Data Task: 2000, Global gridded surfaces of selected soil characteristics (IGBP-DIS). International Geosphere-Biosphere Programme - Data and Information Services, Online http://daac.ornl.gov/SOILS/guides/igbp-surfaces.html.

Guzzetti, F., Carrara, A., Cardinali, M. and 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.

Highland, L. M. and Bobrowsky, P.: 2008, The landslide handbook. A guide to understanding landslides, *Technical Report Circular 1325*, U. S. Geological Survey.

Hong, Y., Adler, R. and Huffman, G.: 2007, Use of satellite remote sensing data in the mapping of global landslide susceptibility, *Natural Hazards* **43**(2), 245–256. 10.1007/s11069-006-9104-z.

Iida, T.: 1999, A stochastic hydro-geomorphological model for shallow landsliding due to rainstorm, *CATENA* **34**(3-4), 293–313. doi: DOI: 10.1016/S0341-8162(98)00093-9.

International Steering Committee for Global Mapping: 2009, Global mapping, Online `http://www.iscgm.org/cgi-bin/fswiki/wiki.cgi`.

Jager, S. and Wieczorek, G. F.: 2001, Landslide susceptibility in the Tully Valley area, Finger Lakes Region, New York, *Technical report*, USGS.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E.: 2008, Hole-filled SRTM for the globe version 4, Online `http://srtm.csi.cgiar.org`.

Kirschbaum, D., Adler, R., Hong, Y. and Lerner-Lam, A.: 2009, Evaluation of a preliminary satellite-based landslide hazard algorithm using global landslide inventories, *Nat. Hazards Earth Syst. Sci.* **9**(3), 673–686.

Land Processes Distributed Active Archive Center: 2009, MODIS overview, Online `https://lpdaac.usgs.gov/lpdaac/products/modis_overview`.

Laouafa, F. and Darve, F.: 2002, Modelling of slope failure by a material instability mechanism, *Computers and Geotechnics* **29**(4), 301–325. doi: DOI: 10.1016/S0266-352X(01)00030-1.

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

Lu, P. and Rosenbaum, M. S.: 2003, Artificial neural networks and grey systems for the prediction of slope stability, *Natural Hazards* **30**(3), 383–398. 10.1023/B:NHAZ.0000007168.00673.27.

Mark, R. K. and Ellen, S. D.: 1995, Statistical and simulation models for mapping debris-flow hazard, *in* A. Carrara and F. Guzzetti (eds), *Geographical information systems in assessing natural hazards*, Kluwer Academic Publishers, pp. 93–106.

Masetti, M., Poli, S. and Sterlacchini, S.: 2007, The use of the weights-of-evidence modeling technique to estimate the vulnerability of groundwater to nitrate contamination, *Natural Resources Research* **16**(2), 109–119. 10.1007/s11053-007-9045-6.

Ministry of Land, Infrastructure, Transport and Tourism: 2009, 1km-mesh orthophoto, Online `http://orthophoto.mlit.go.jp/`.

MultiSpec: 2009, A freeware multispectral image data analysis system, Online `http://cobweb.ecn.purdue.edu/~biehl/MultiSpec/`.

Nelson, E., Connors, K. and Suarez S, C.: 2007, GIS-based slope stability analysis, Chuquicamata open pit copper mine, Chile, *Natural Resources Research* **16**(2), 171–190. 10.1007/s11053-007-9044-7.

Nykanen, V. and Ojala, V.: 2007, Spatial analysis techniques as successful mineral-potential mapping tools for orogenic gold deposits in the Northern Fennoscandian Shield, Finland, *Natural Resources Research* **16**(2), 85–92. 10.1007/s11053-007-9046-5.

Okuda, S. and Okimura, T.: 1984, Characteristics of debris flows in Misumi town. research of the 1983 San'in heavy rainfall disaster, *Geological Reports of Shimane University* **3**, 33–47. In Japanese.

PASCO Co.: 1997, Smapcnv 4, Manual.

Pimiento, E. and Yokota, S.: 2006, Distribution of slope failures following the 1983 San'in heavy rainfall disaster in Misumi-Kitsuka area, western Shimane, southwest Japan, *Geological Reports of Shimane University* **25**, 25–30.

Poli, S. and Sterlacchini, S.: 2007, Landslide representation strategies in susceptibility studies using weights-of-evidence modeling technique, *Natural Resources Research* **16**(2), 121–134. 10.1007/s11053-007-9043-8.

Raines, G., Connors, K. and Chorlton, L.: 2007, Porphyry copper deposit tract definition a global analysis comparing geologic map scales, *Natural Resources Research* **16**(2), 191–198. 10.1007/s11053-007-9042-9.

Remondo, J., Gonzalez-Diez, A., De Teran, J. R. D. and Cendrero, A.: 2003, Landslide susceptibility models utilising spatial data analysis techniques. a case study from the Lower Deba Valley, Guipuzcoa (Spain), *Natural Hazards* **30**(3), 267–279. 10.1023/B:NHAZ.0000007202.12543.3a.

Research Group of San-in Heavy Rainfall Disaster: 1984, Geological character of slope failure by heavy rain of july, 1983, San'in district, *Geological Reports of Shimane University* **3**, 3–20. In Japanese.

Robinson, G. and Larkins, P.: 2007, Probabilistic prediction models for aggregate quarry siting, *Natural Resources Research* **16**(2), 135–146. 10.1007/s11053-007-9039-4.

Saito, H., Nakayama, D. and Matsuyama, H.: 2009, Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi mountains, Japan, *Geomorphology* **109**(3-4), 108–121. doi: DOI: 10.1016/j.geomorph.2009.02.026.

Sawatzky, D. L., Raines, G. L., Bonham-Carter, G. F. and Looney, C. G.: 2009, Spatial Data Modeller (SDM): ArcMAP 9.3 geoprocessing tools for spatial data modelling using weights of evidence, logistic regression, fuzzy logic and neural networks, Online `http://www.ige.unicamp.br/sdm/default_e.htm`.

Shuster, R. and Highland, L. M.: 2001, Socio-economic and environmental impacts of landslides in the western hemisphere, Online `http://pubs.usgs.gov/of/2001/ofr-01-0276/`.

Soeters, R. and van Westen, C.: 1996, Slope stability: recognition, analysis and zonation, *in* A. Turner and R. Shuster (eds), *Landslides: investigation and mitigation*, National Academy Press, Washington, D. C., pp. 129–177.

Suzen, M. and Doyuran, V.: 2004, A comparison of the gis based landslide susceptibility assessment methods: multivariate versus bivariate, *Environmental Geology* **45**(5), 665–679. 10.1007/s00254-003-0917-8.

Turner, A.: 1996, Colluvium and talus, *in* A. Turner and R. Shuster (eds), *Landslides: investigation and mitigation*, National Academy Press, Washington, D. C., pp. 525–554.

U S Geological Survey: 2004, Landslide types and processes, Online `http://pubs.usgs.gov/fs/2004/3072/fs-2004-3072.html`.

van Westen, C. J., Rengers, N. and Soeters, R.: 2003, Use of geomorphological information in indirect landslide susceptibility assessment, *Natural Hazards* **30**(3), 399–419. 10.1023/B:NHAZ.0000007097.42735.9e.

Varnes, D. J.: 1984, Landslide hazard zonation : A review of principles and practice, *Technical report*, UNESCO.

Wada, M., Inamoto, A. and Nagata, J.: 1984, Slope failure caused by heavy rainfall of july, 1983, in the west region of San'in district, *Geological Reports of Shimane University* **3**, 41–50. In Japanese with English abastract.

Wieczorek, G. F.: 1996, Landslide triggering mechanisms, *in* A. Turner and R. Shuster (eds), *Landslides: investigation and mitigation*, National Academy Press, Washington, D. C., pp. 76–90.

Wu, T. H., Tang, W. H. and Einstein, H. H.: 1996, Landslide hazard and risk assessment, *in* A. Turner and R. Shuster (eds), *Landslides: investigation and mitigation*, National Academy Press, Washington, D. C., pp. 106–120.

Yamagishi, H. and Iwahashi, J.: 2007, Comparison between the two triggered landslides in Mid-Niigata, Japan by july 13 heavy rainfall and october 23 intensive earthquakes in 2004, *Landslides* **4**(4), 389–397. 10.1007/s10346-007-0093-0.

Series from Lund University's Geographical Department

Master Thesis in Geographical Information Science (LUMA-GIS)

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland. (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research. (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing. (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia. (2009).
5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010)
8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010)
9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010)
10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011)
11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011)