Landslide Mapping Using A Semi-Automatic Bayesian Approach

Justine Douglas[1], Shou-Hao Chiang[2]

[1]International Environmental Sustainable Development Program (IESD),
No. 300 Jhongda Rd., Jhongli District, Taoyuan City 32001, Taiwan, 104350604@cc.ncu.edu.tw
[2]Center for Space and Remote Sensing Research (CSRSR),
No. 300 Jhongda Rd., Jhongli District, Taoyuan City 32001, Taiwan, gilbert@csrsr.ncu.edu.tw

KEY WORDS: Landslide Modeling, Taiwan, Bayesian Theory, Integrated Analysis

ABSTRACT: Landslide hazards are common in Taiwan due to its mountainous topography and high number of earthquakes and typhoons experienced yearly. As a result it is essential to develop a method of landslide detection that is capable of landslide providing results with a reasonable level of accuracy, and may also be integrated into an early warning or monitoring system. Probabilistic methods such as Bayesian analysis have gained interest in recent times over more traditional deterministic approaches due to their greater flexibility and the more informative nature of any obtained results. In this study, we will use a stepwise semi-automatic Bayesian analysis approach for mapping landslides in Huaguoshan, Taiwan. Prior probability of landsliding will be obtained using an integrated analysis and used to detect landslides from a post-typhoon satellite image. This information will be used to detect and map rainfall induced shallow landslides in a post typhoon environment.

1. INTRODUCTION

Though the term conjures a very specific image in the minds of a layperson, a landslide refers to any type of movement of rock, soil and debris down a slope. Further confusion about the concept of a landslide arises largely due to Landslides have always been a significant part of the human landscape, and are well known for their ability to create significant damage to infrastructure as well as many losses of human lives. In recent times, increased levels of urbanization and climate change altering the natural environment. (Aleotti and Chowdhury, 1999) For these reasons, many attempts have been made over time to predict when a landslide will occur. However, the mechanics behind landslide triggers are not very well understood and are also not easily transferrable from one area to another due to the nature of the Earth's topography.

Landslide hazard mapping (also known as Landslide Hazard Zonation or LHZ) aims to process and rank the odds of a landslide occurring in a specific area. Landslide hazard mapping is a fairly broad area, with many approaches typically falling under one of five major subheadings; inventory based, heuristic approach, probabilistic assessment, deterministic approach and statistical analysis. (Guzzetti et al., 1997) Each method has its own strengths and weaknesses, but current research trends toward a statistical analysis due to its data driven nature and the vast improvements that have been made in Geographical Information Systems (GIS). This makes it easier to achieve effective manipulation of spatial data while drastically simplifying the data analysis process. Statistical methods general fall under two major categories: bivariate statistical analysis e.g. Weights of Evidence model, Weighted Overlay model or multivariate statistical analysis e.g. Logistic Regression, Conditional Analysis methods.

1.1 Bayesian Theory

Bayesian theory is a mathematical framework for reasoning based on conditional probability. Using this theorem, we are able to calculate the probability of an event A happening given that a condition B is true. This can be expressed using the formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

Where P(A) is the prior probability of the hypothesis A, P(B) is the prior probability of the data B, P(A|B) is the posterior probability of A given B and P(B|A) is the posterior probability of B given A. In this paper, the Bayesian

equation to find the probability of a landslide occurring given that a critical rainfall threshold is met is calculated using the equation

$$p(S_1|v) = \frac{p(v|S_2)p(S_1)}{p(v)}; (O \leq p \leq 1) \quad (2)$$

Where S_1 represents the probability of the pixel containing a feature f (landslide free area, landslide area, source, runout), S_2 represents a landslide related feature f present before the landslide event, v represents the landslide influencing factor (critical rainfall) and p represents the probability of landslide occurrence. (Mondini et al., 2012)

1.2 Integrated Model

To investigate the conditions needed for the creation of shallow rainfall triggered landslides, the critical rainfall model has seen significant usage since its creation. Because this model is a physically based one, it is theoretically easier to use the model in a wide variety of areas. (Montgomery and Dietrich., 1994) In reality, because these models require a high amount of specific topographic attributes for the area they are often difficult to use. This is due to attributes such as soil type and geomorphologic structure not only being scarce in nature due to lack of funding for collection, but also requiring high levels of technical expertise that many countries may not be able to afford. On the other hand, while statistical models such as logistic and multivariate regression give a better idea of the correlation between landslides and their respective instability factors, they do not account for temporal changes and their understanding of landslide processes are limited. (Ayalew and Yamagishi, 2005; Clerici et al., 2002)

The integrated model aims to remedy the deficiencies of both approaches by combining a deterministic and statistic method. Topographic, soil and rainfall attributes are combined to estimate the probability of a landslide occurring in a specific area. While the integrated model is not capable of explaining the mechanical effects of higher rainfall intensity on landslide potential, it can estimate the probability that a landslide will occur through the use of Rainfall Intensity Difference (RID) and rainfall duration. (Chang, 2008)

The critical rainfall model uses a logistic regression in the form

$$logit(y) = a + b_1x_1 + b_2x_2 + \cdots + e \quad (3)$$

Where a is a constant, b_i is the ith regression coefficient and e is an error term. This represents the natural logarithm of the odds

$$logit(y) = \ln\left(\frac{p}{1-p}\right) \quad (4)$$

Where p represents the probability of the occurrence y. Equation (4) is then rewritten as

$$p = \frac{\exp(a+b_1x_1+b_2x_2+\cdots)}{1+\exp(a+b_1x_1+b_2x_2+\cdots)} \quad (5)$$

2. METHODS AND MATERIAL

2.1 Study Area

The study area for this paper is located in southern Taiwan's Huagoshan watershed. The study covers a region of 116km^2 and has elevation that ranges from 257m to 1673m. Slope of the study area ranges from 0° to 78° with an average slope of 26°. Average annual precipitation in the study area averages 2200mm and occurs mainly in May and October. The geology of the area consists of rocks from the Miocene and Holocene Alluvium from the Quaternary period. Alluvium, clay and limestone soils are also present in the area.

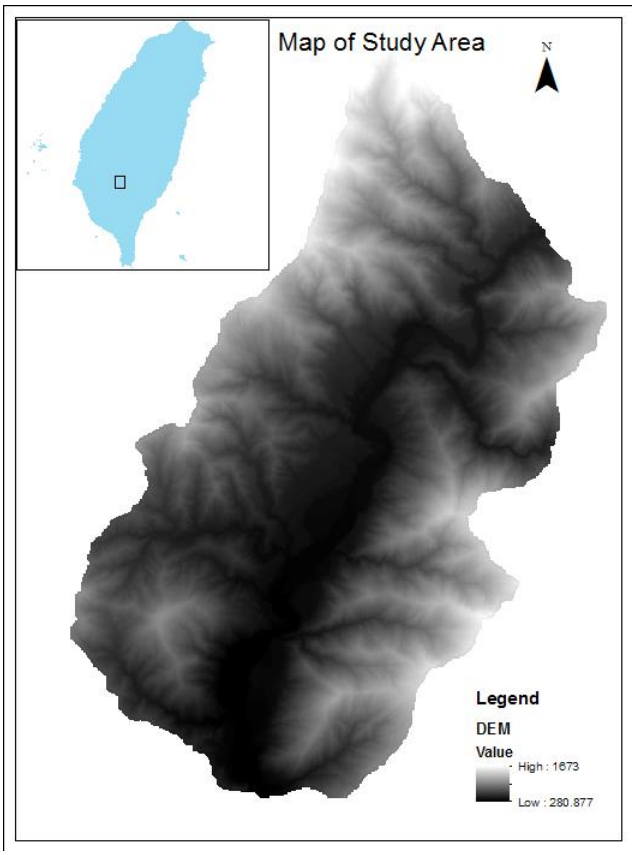

Figure 1. Location of the Huagoshan Watershed

2.2 Study Data

A 10m pre-event DEM for this study was obtained from Taiwan's Soil and Water Conservation Bureau and used to prepare maps showing terrain gradient and aspect. Shapefiles containing soil data e.g. soil conductivity, internal soil friction angle and soil bulk density were obtained and converted to raster format for use in the integrated model. These shapefiles were based on the 10m pre-event DEM for the study area. Daily rainfall data was obtained from the Central Weather Bureau (CWB) of Taiwan for 16 rain gauges located in the study area and the areas around it. All data from these rain gauges were standardized to contain only data for the month of August and used to calculate the daily and monthly rainfall total. A FORMOSAT-2 NDVI image of the surrounding area at a resolution of 8m was also used in the process.

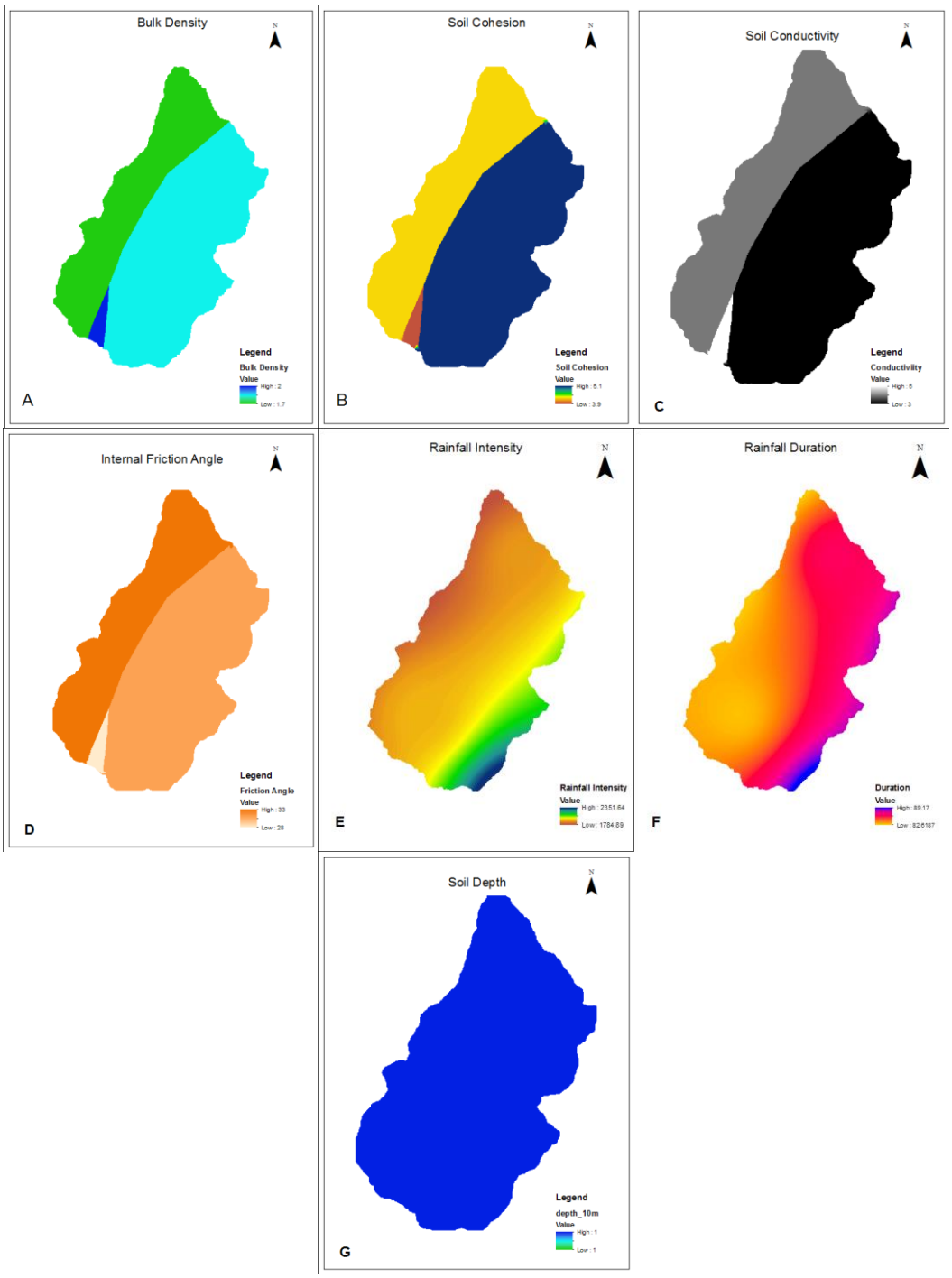

Figure 2. Data Used In the Study. A) Bulk Density, B) Soil Cohesion, C) Soil Conductivity, D) Internal Friction Angle, E) Rainfall Duration, F) Rainfall Intensity, G) Soil Depth

2.3 Method

2.3.1 Data Preprocessing

Before using them in their respective models, all data was standardized to the TWD97 projected co-ordinate system. Raster data was then resampled to a 10m resolution as necessary before being cropped to the study area extent. All shapefiles were then converted to raster files. Finally, any holes or no data areas in the rasters were filled using a 5x5 rectangular nearest neighbor function in raster calculator.

2.3.2 Integrated Model

The integrated model requires inputs of rainfall intensity, rainfall duration, soil conductivity, soil bulk density, soil cohesion, NDVI, soil internal friction angle and soil depth. Rainfall intensity and rainfall duration were calculated using Inverse Distance Weighted (IDW) interpolation of the rainfall totals obtained from 15 rain gauges in and around the study area. Due to unavailability of soil depth rasters for the study area, soil depth was assumed to be constant at 1m throughout the study area. Soil conductivity, bulk density, internal frictional angle and cohesion rasters were obtained from shapefiles which contained the relevant information.

2.3.3 Maximum Likelihood Classification

A Maximum Likelihood Classification was performed on a post typhoon event FORMOSAT-2 image. The classification was trained in 6 regions of interest (ROI) with four areas not affected by slope failures and two which represented the spectral characteristics of landslides in the study area. The ROIs were selected through visual inspection and after the classification was performed a majority filter was applied to the post classification image to reduce any stray pixels. Finally, all classes in the majority filtered image were combined except the ones representative of the landslide.

2.3.4 Normalization and Accuracy Assessment

Finally, all resulting rasters were normalized on a scale of 0 to 1 using the equation

$$\frac{\text{DEM}-\text{MIN(DEM)}}{MAX(DEM)-MIN(DEM)} \quad (6)$$

The accuracy of the integrated model performance was found using the modified success rate (MSR) (Huang and Kao, 2006). This uses the equation

$$MSR = 0.5(SR_{number}\ landslides) + 0.5(SR_{cell}\ stable\ areas) \quad (7)$$

Where SR_{number} represents the rate of successfully predicted landslides and SR_{cell} represents the proportion of successfully predicted stable areas. The equal weighting of each components allows for the consideration of the predictability of landslide sites and stable areas.

3. RESULTS & DISCUSSION

3.1 Integrated Model

Through experimentation, it was discovered that a critical rainfall co-efficient of 2 produced the best results. 5657 landslides were discovered by the model with an AUC of 0.718 and at a 0.5 significance level the model had a MSR of 86.5%.

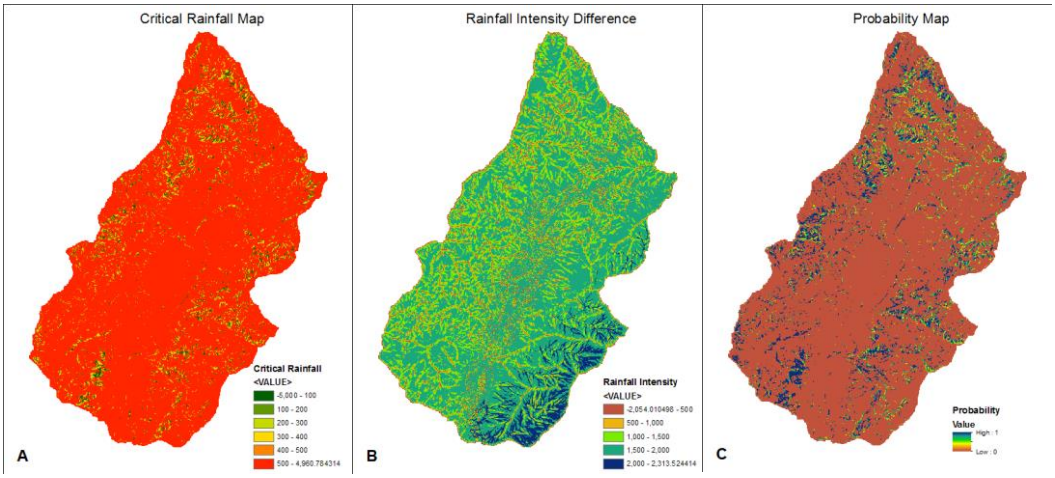

Figure 3. Integrated Model Results. A) Critical Rainfall Map, B) RID Map, C) Probability Map

The RID varies with rainfall intensity, with most stable values being located in the eastern sides of the map and decreasing eastward. The center of the map is completely stable, though some scattered areas of negative intensity exist in that region. On the probability map, a slice of high probability areas exists on the south western corner of the map.

3.2 Maximum Likelihood Classification

The results of the Maximum Likelihood Classification are shown in the figure below.

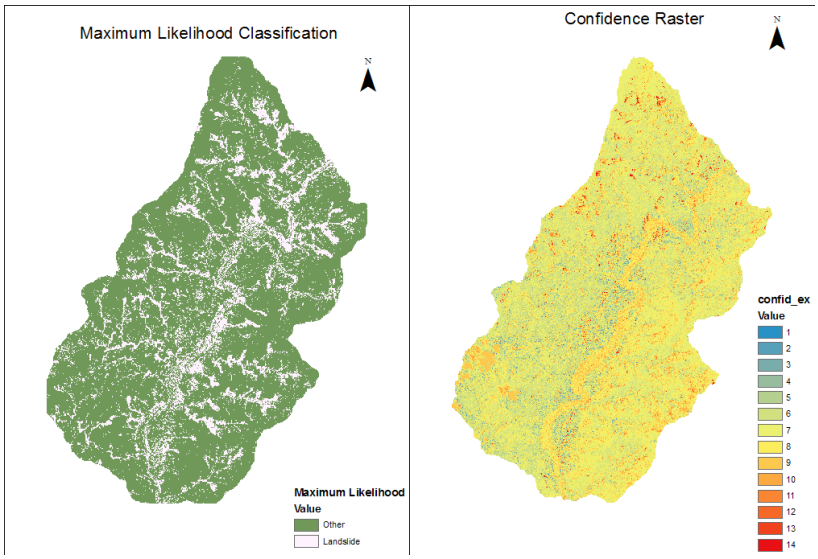

Figure 4. Maximum Likelihood Classification and Confidence Raster

A confidence raster was also produced. This raster gave the likelihood of a pixel being correctly predicted on a scale from a confidence level of or greater than 0.95 to a confidence level of 0. The majority of classified values fell in the confidence range of 0.5.

Table 1. Confidence Raster Results

Confidence Level	Count
0	4,839
0.005	1,874
0.01	6,481
0.025	12,162
0.05	26,810
0.1	113,307
0.25	251,915
0.5	455,541
0.75	299,987
0.9	58,177
0.95	21,633
0.975	10,280
0.99	2,714
0.995	2,359

Generally, the classification was able to determine the difference between the landslide source and runout fairly well. However, there tended to be misclassification of pixels representing areas such as the river channel and town areas with that of the landslide runout. This can be attributed to the high spectral similarity between these regions, making it difficult to near impossible to determine these areas unless done manually. Values that were identified as having a confidence level of 0 fell in areas that were either covered with clouds or were residential areas with roofs that possessed a similar spectral signature. Another potential issue is the relatively small and scattered size of landslides present in the area. Of the 6,440 landslides present in the Huagoshan basin, 6,149 possess a size that is smaller than

one hectare. This could result in potential false positive identification of landslides as these landslides may exist in areas with mixed pixel types.

3.3 Bayesian Model

After the results for both the Integrated Model and the Maximum Likelihood Classifications were obtained, they were then used as inputs in the Bayesian model. These inputs were then classified on from probabilities ranging from 0.1 – 1, where 0.1 represents 10% likelihood of landslide occurrence and 1 represents 100% likelihood of landslide occurrence.

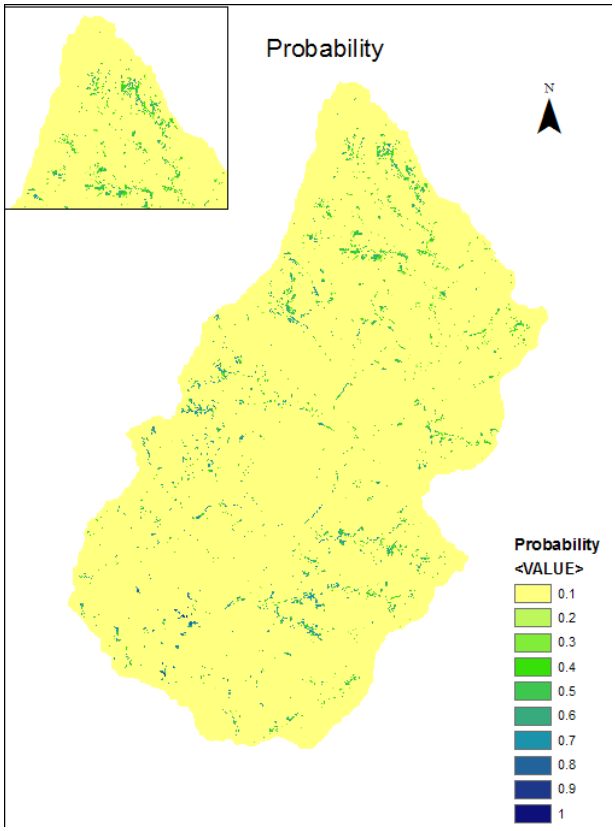

Figure 5. Bayesian Model Results

The histogram for this map indicated that the majority of values fell in the 0.1 (10%) likelihood range. This corresponds to the fact that these areas are typically located along ridgelines or other regions that are unconditionally stable in nature. The 0.6 (60%) likelihood range contained the second highest number of values, with the 0.5 (50%) likelihood of landslide occurrence coming in third. No area in the map contained a probability of 1 (100%), and the highest value for potential landslide occurrence was 0.8 (80%). These values were concentrated in a small triangular area in the southwestern part of the map. This corresponds to the area on the probability maps created by the Integrated Model with the highest values.

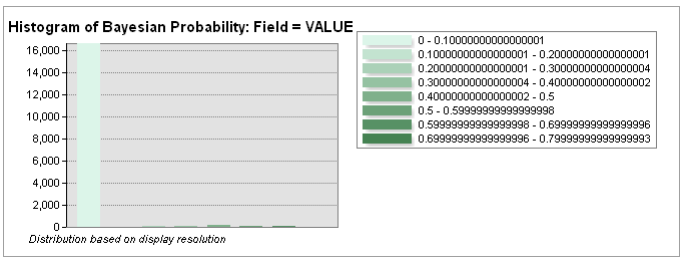

Figure 6. Histogram of Bayesian Output

Compared to existing landslide inventory maps for the area, the model underrepresented the total landslides present in the area by 11%. One possible reason for this is potentially the inconsistency between actual values for input parameters such as soil depth or soil permeability and the estimated value used in the Integrated Model. Soil depth in a landslide prone area typically varies with each landslide activity and the steepness of the slope. Slope was also a factor that this model did not take into account. Studies by other authors have indicated slope gradient as having a strong relationship with landslide occurrence, especially in combination with rock bed type. However, these researchers did indication elevation as being a reasonable approximation in the absence of slope data. (Ayalew and Yamagishi, 2005; Wang, 2014; Tsou, 2010). Interestingly, the model did not correctly classify landslide runout areas which coincided with stream or channel areas. This discrepancy is perhaps due to the functionally identical elevations and soil depths found in these areas. It did for the most part identify the source areas located along the sides of these runout areas.

5. CONCLUSION

The primary aim of this paper is to develop a landslide classification method which allows for the detection of shallow rainfall induced landslides while compensating for the weaknesses found in traditional deterministic and probabilistic techniques. It also aims to use statistical analysis to develop a classification technique that could be considered 'heuristic' in nature – mimicking the techniques that are often used by geomorphologists in map creation. (Mondini et al., 2013) This goal was partially achieved as the combined model was able to compensate for some of the weaknesses in the detection of landslides by Maximum Likelihood Classification – namely the identification of spectrally similar bare soil and landslide runout areas through its combination with the Integrated Model. However, this model still has some of the limitations present in traditional landslide models – namely the need for detailed geomorphological data such as soil conductivity and bulk density that may not be readily available in some regions. Despite this, is believed that with further refinement and some alteration this method will be one capable for use in the creation of landslide probability maps in other areas.

REFERENCE

Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bulletin Of Engineering Geology And The Environment 58 (1), pp. 21-44.

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

Chang, K. T., Chiang, S. H., 2009. An integrated model for predicting rainfall-induced landslides. Geomorphology 105 (3-4), pp. 366-373.

Chang, K. T., Chiang, S. H., Feng L., 2008. Analysing the relationship between typhoon‐triggered landslides and critical rainfall conditions. Earth Surf. Process. Landforms 33 (8), pp. 1261-1271.

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

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

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

Mondini, A. C., Marchesini, I., Rossi, M., Chang, K-T., Pasquariello, G., Guzzetti, F., 2013. Bayesian framework for mapping and classifying shallow landslides exploiting remote sensing and topographic data. Geomorphology (201), pp. 135-147.

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

Pardeshi, S. D., Autade, S., Pardeshi, S. S. 2013. Landslide hazard assessment: recent trends and techniques. Springerplus 2 (1), pp. 523.

Tsai, F., Hwang, J.-H., Chen, L.-C., Lin, T.-H., 2010. Post-disaster assessment of landslides in southern Taiwan after 2009 typhoon Morakot using remote sensing and spatial analysis. Nat. Hazards Earth Syst. Sci. 10 (10), pp. 2179-2190.

Tsou, C-Y., Feng, Z. Y., Chigira, M., 2011. Catastrophic landslide induced by typhoon Morakot, Shiaolin, Taiwan. Geomorphology 127 (3-4), pp. 166-178.

Wang, L. Jie., Guo, M., Sawada, K., Lin, J., Zhang, J., 2015. Landslide susceptibility mapping in mizunami city, japan: a comparison between logistic regression, bivariate statistical analysis and multivariate adaptive regression spline models. CATENA (135), pp. 271-282.